# Impure Altruism and Impure Selfishness

Kota SAITO[*]

California Institute of Technology

May 3, 2015

## Abstract

*Altruism* refers to a willingness to benefit others, even at one's own expense. In contrast, *selfishness* refers to prioritizing one's own interests with no consideration for others. However, even if an agent is selfish, he might nevertheless act as if he were altruistic out of selfish concerns triggered when his action is observed; that is, he might seek to feel *pride* in acting altruistically and to avoid the *shame* of acting selfishly. We call such behavior *impurely altruistic*. Alternatively, even if an agent is altruistic, he might nevertheless give in to the *temptation* to act selfishly. We call such behavior *impurely selfish*. This paper axiomatizes a model that distinguishes altruism from impure altruism and selfishness from impure selfishness. In the model, unique real numbers separately capture altruism and the other forces of pride, shame, and the

temptation to act selfishly. We show that the model can describe recent experiments on dictator games with an exit option. In addition, we describe an empirical puzzle that government spending only partially crowds out consumers' donations, contrary to the prediction based on standard consumer theory.

# 1   Introduction

*Altruism* refers to a willingness to benefit others, even at one's own expense. *Selfishness*, the opposite of altruism, refers to prioritizing one's own interests, with a concomitant lack of consideration for others. While these definitions may seem clear, it is difficult to detect from an agent's actions alone whether he is truly altruistic or selfish.

Even if an agent is indifferent to the welfare of others and therefore selfish by definition, he might nevertheless act as if he were truly altruistic (i.e., act to benefit others) out of selfish concerns triggered when his action is observed; that is, he might seek to feel *pride* in having acted altruistically or to avoid the *shame* of having acted selfishly.

In contrast to pure altruism, we say that an agent exhibits *impure altruism* if he chooses an action that benefits others in order to feel pride in acting altruistically and to avoid the shame of acting selfishly. (In this paper, the terms *altruism* and *pure altruism* will be used interchangeably, as will the terms *selfishness* and *pure selfishness*.)

On the other hand, even if an agent is willing to benefit others and is therefore by definition altruistic, he might nevertheless give in to the *temptation* to act selfishly. In contrast to pure selfishness, we say that an agent exhibits *impure selfishness* if the temptation to act selfishly motivates him to depart from his tendency to act altruistically.[1] An agent

---

[1]Of course, such an agent could also be said to exhibit *impure altruism*, since his altruistic tendencies could be seen as rendered "impure" whenever he gives in to the temptation to act selfishly. However, we feel

can exhibit impure selfishness, especially when an immediate payoff is at stake.[2]

Moreover, these three forces–pride, shame, and the temptation to act selfishly–could interact in conflicting ways in affecting an agent's actions.[3] For example, an agent could behave impurely altruistically even if he feels a temptation to act selfishly, when his desire to avoid the shame has overwhelmed the temptation. Conversely, an agent could behave impurely selfishly even if he knows that he will feel shame at doing so, when the temptation has overwhelmed such shame.

In the axiomatic literature on preferences over menus temptation and shame have been studied separately: Gul and Pesendorfer (2001) have proposed a model of temptation; Dillenberger and Sadowski (2012) have proposed a model of shame. In an independent project, Evren and Minardi (2014) study warm glow, which is a similar phenomenon to pride. Moreover, in the latter two works, the parameters capturing the psychological effects are not uniquely identified.

The purpose of this paper is to axiomatize a model in which unique real numbers separately capture altruism as well as pride in acting altruistically, shame of acting selfishly, and the temptation to act selfishly. The improved identification comes from a linear structure of our model. On the other hand, the linear structure makes our model restrictive. For example, our model satisfies the *independence of irrelevant alternatives* (IIA). That is, the decision maker will not change his choice even if an irrelevant element is removed from his set of alternatives. We discuss those restrictions at the end of Section 2.

Our model can capture the trade-off between the three forces and, thereby, distinguish

---

that the term *impure altruism* is more suitable for the selfish person who acts altruistically when motivated by his inherent selfishness. As a result, we use the term *impure selfishness* for the opposite situation, namely, a typically altruistic agent who sometimes gives in to the temptation to act selfishly.

[2]Indeed, Noor and Ren (2011) found that in their experiments the average donation from dictators to receivers has significantly decreased when the timing of the payment to the subjects is changed from one month later to immediately after the dictators' choices.

[3]Throughout this paper, when we use the terms, pride in acting altruistically, shame of acting selfishly, and the temptation to act selfishly, we are referring to individual acts of an agent, and *not* to an agent's personality or general tendencies. For example, if we say that an agent seeks to "feel pride in acting altruistically," that description will refer to the pride in one particular altruistic act, rather than to the agent's typical or habitual behavior.

3

altruism from impure altruism and selfishness from impure selfishness. For example, we can show that the agent could behave impurely altruistically if his index of shame is larger than that of the temptation; otherwise, an agent could behave impurely selfishly. Moreover, we provide axiomatizations of special cases which capture only pride, shame, or the temptation, respectively.

We adopt the same set up as Dillenberger and Sadowski (2012), which distinguishes the ex-ante private stage and the ex-post public stage. The setup is necessary to study pride, shame, and the temptation at the same time.

To conclude the introduction, we clarify (i) the definition of shame in comparison with a related emotion, guilt; (ii) the definition of pride; and (iii) the difference between temptation and two other psychological effects, pride and shame. There is a clear distinction between shame and guilt in Psychology. According to Hiebert (1985, p. 213), "An agent may suffer from guilt although no one else knows of his or her misdeed." This is the crucial difference between shame and guilt: an agent would not suffer from shame if no one else knows of his misdeed. In this paper, we do not consider guilt; we focus on shame alone. To study guilt, the two stage setup is not enough because the agent would feel guilt even at the ex-ante private stage.

There are various definitions of pride in Psychology. (See Lea and Webley (1997) for a survey on pride in Psychology and for a discussion on economic consequences of pride.) In Development Psychology, pride has often been studied in its comparison with shame. (See Scheff (1988) and Taylor (1985) for example.) Scheff (1988) claims that conformity to social expectations causes "the pleasure of pride and fellow feeling"; while disconformity causes "the punishment of embarrassment, shame, or humiliation". In this paper, we consider the particular social expectations that an agent should act altruistically. Hence, by pride of acting altruistically, we mean a feeling of pleasure derived from the social recognition of one's altruistic choice. Even without social recognition and even if an agent is not altruistic, he may derive a feeling of pleasure, such as warm glow, from one's altruistic choice. We do not study such feelings in this paper.

4

Finally, note that the temptation to act selfishly is different from shame and pride in the sense that the temptation is not caused by the publicity of choices but by the immediacy of the payoff at the ex-post stage. However, we think that the effect of temptation at the ex-post public stage is so significant that we should incorporate the effect into our model. Indeed, in experiments on dictator games, Noor and Ren (2011) found that the average donation from dictators to receivers is 65 percent of the total endowment when the payment to subjects is made one month later, but only 38 percent when the payment is made immediately after the dictators' choices.

The rest of the paper is organized as follows. In Section 2, we provide a preview of the results. In Section 3, we provide a literature review. In Section 4.1, we present the model and the axioms. In Section 4.2, we present a representation theorem and a sketch of the proof. In Section 4.3, we show the uniqueness of the representation. In Section 5, we provide axiomatizations of the special cases and discuss their relationship with Gul and Pesendorfer (2001), Dillenberger and Sadowski (2012), and Evren and Minardi (2014). In Section 6, we study characterizations of parameters and provide comparative statics. In Section 7, we study the agent's ex-post choices and provide characterizations of impure altruism and impure selfishness. In Section 8, we show that the model is consistent with existing experimental evidence. In Section 9, we provide an application. All proofs are in Appendix.

## 2　Preview of Results

### 2.1　Model

We investigate a decision maker who determines an allocation between himself and other passive agents. The decision maker's choice consists of two stages. In the first ex-ante stage, the decision maker chooses a set of allocations. We assume that the other agents do not know that the decision maker has such choices. His ex-ante choices are *private*.

In the ex-post stage, the decision maker chooses an allocation from the set that he chose ex-ante. We assume that other agents can observe which allocation the decision maker chooses from the set, even though they do not know that he chose the set ex-ante.

Ex-ante                             Ex-post

Choose a set $A$ of allocations privately      Choose an allocation from the set $A$ publicly

Figure 1: Two-Stage Decision Problem
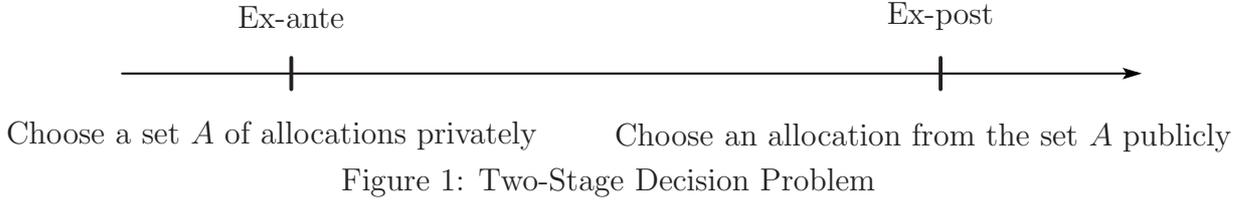
Because his choices will be *public* at the ex-post public stage, the decision maker could anticipate that he will feel pride in acting altruistically, or shame of acting selfishly at that stage. In addition, the decision maker could anticipate that he will suffer from the temptation to act selfishly ex-post. In light of these potential ex-post feelings, the decision maker chooses a set ex-ante. (We, henceforth, call these three feelings simply, pride, shame, and temptation, when there is no danger of confusion.) This set up is not new: Dillenberger and Sadowski (2012) have proposed the same model that distinguishes the ex-ante private stage and the ex-post public stage to study the effect of shame.

To introduce the representation, we first define some notation. The decision maker is denoted by 1 and the finite set of other agents is denoted by $S$. Define $I = \{1\} \cup S$ to be the set of all agents. A payoff profile $\boldsymbol{p} \equiv (p_i)_{i \in I}$ is called an *allocation*. An allocation $\boldsymbol{p}$ is also denoted by $(p_1, p_S)$, where $p_1$ is a payoff to the decision maker and $p_S \equiv (p_i)_{i \in S}$ is a payoff profile to the other agents.

When the decision maker chooses a set of allocations at the ex-ante private stage, he maximizes the following utility function. The utility of a set $A$ is :

$$V(A) = \max_{\boldsymbol{p} \in A} \Big[ \sum_{i \in I} \alpha_i u(p_i) + \beta_1 \max_{\boldsymbol{q} \in A} \alpha_1 (u(q_1) - u(p_1)) - \beta_S \max_{\boldsymbol{r} \in A} (\sum_{i \in S} \alpha_i u(r_i) - \sum_{i \in S} \alpha_i u(p_i)) \Big], \quad (1)$$

where $\alpha_1 > 0$, $\beta_1 < 1$, $\beta_S \geq 0$, and $\sum_{i \in S} \alpha_i = 1$. Moreover, $(\alpha_i)_{i \in I}$, $\beta_1$, and $\beta_S$ are unique and $u$ is unique up to positive affine transformation. The maximizer $\boldsymbol{p}$ over $A$ is the decision

maker's *ex-post public choice* from $A$.

The first term of the model, $\sum_{i \in I} \alpha_i u(p_i)$, captures the decision maker's *utilitarian* evaluation of his ex-post choice of $\boldsymbol{p}$. Hence, we call the model a *generalized utilitarian* (GU) model.

The second term, $\beta_1 \max_{\boldsymbol{q} \in A} \alpha_1(u(q_1) - u(p_1))$, captures the utility arising from *the pride of acting altruistically* if $\beta_1 \geq 0$ or the disutility arising from *the temptation to act selfishly* (i.e., the temptation to maximize his own allocation) if $\beta_1 \leq 0$. The utility and disutility are proportional to the difference between the maximum selfish utility $\max_{\boldsymbol{q} \in A} \alpha_1 u(q_1)$ and the actual utility $\alpha_1 u(p_1)$ attained by the decision maker's ex-post choice of $\boldsymbol{p}$. The difference captures how much the decision maker controls himself so as to keep himself from acting selfishly (i.e., from maximizing his selfish interest).

Similarly, the third term of the GU model, $\beta_S \max_{\boldsymbol{r} \in A}(\sum_{i \in S} \alpha_i u(r_i) - \sum_{i \in S} \alpha_i u(p_i))$, captures the disutility due to the *shame of acting selfishly*. The disutility is proportional to the difference between the maximum social utilities $\max_{\boldsymbol{r} \in A} \sum_{i \in S} \alpha_i u(r_i)$ (i.e., the maximum sum of the utilities of the other agents) and the actual social utilities $\sum_{i \in S} \alpha_i u(p_i)$ attained by the decision maker's ex-post choice of $\boldsymbol{p}$. Hence, the difference captures how much utility the other agents lose because of the decision maker's ex-post choice.

## 2.2 $\alpha$ captures Altruism and Selfishness; $\beta$ captures Impure Altruism and Impure Selfishness

In the GU model, the relative weight on the social utility with respect to the selfish utility captures *the level of pure altruism*. Since $\sum_{i \in S} \alpha_i = 1$, the relative weight is $(\sum_{i \in S} \alpha_i)/\alpha_1 = 1/\alpha_1$. In other words, $\alpha_1$ captures the level of pure selfishness. To see this interpretation of $\alpha_1$, note that whether or not the decision maker is purely altruistic is determined through his *ex-ante private* choices. This is because in private choices the decision maker would feel neither pride nor shame. Moreover, in ex-ante choices he would not feel temptation.

In the GU model, ex-ante private choices between allocations $\boldsymbol{p}$ and $\boldsymbol{q}$ are formalized as

choices between the singleton sets, such as $\{\boldsymbol{p}\}$ and $\{\boldsymbol{q}\}$. Hence, the utility function that captures only pure altruism is as follows:

$$V(\{\boldsymbol{p}\}) = \sum_{i \in I} \alpha_i u(p_i) = \alpha_1 u(p_1) + \sum_{i \in S} \alpha_i u(p_i). \tag{2}$$

Therefore, the smaller $\alpha_1$ is, the more the decision maker is willing to sacrifice his own allocation $p_1$ to improve the others' allocations $p_S$.
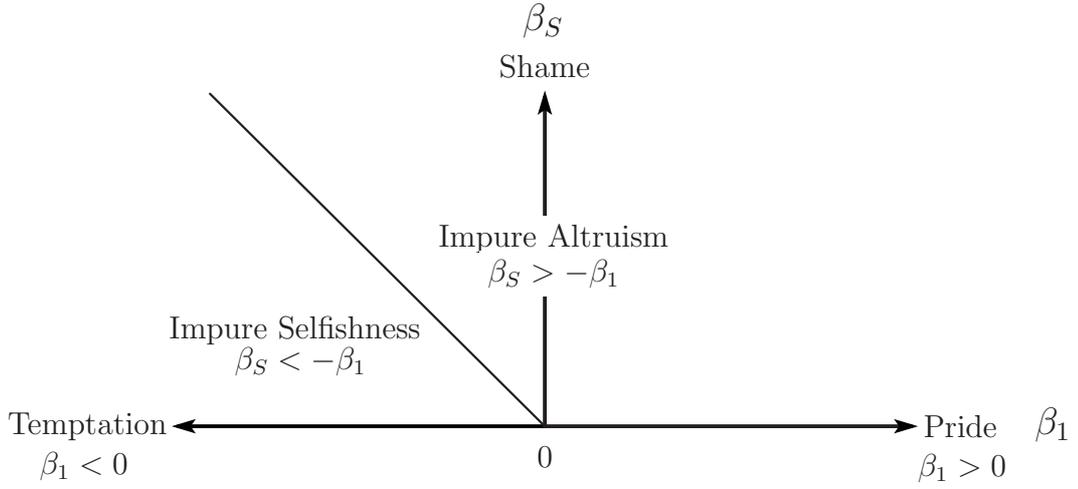


Figure 2: Pride, Shame, and Temptation Cause Impure Altruism and Impure Selfishness

In contrast to the ex-ante private choices, ex-post public choices can be affected by pride, shame, and the temptation to act selfishly. To see this effect, note that (1) can be expressed as $V(A) = \max_{\boldsymbol{p} \in A}[(1-\beta_1)\alpha_1 u(p_1) + (1+\beta_S)\sum_{i \in S}\alpha_i u(p_i)] + \beta_1 \max_{\boldsymbol{q} \in A}\alpha_1 u(q_1) - \beta_S \max_{\boldsymbol{r} \in A}\sum_{i \in S}\alpha_i u(r_i)$ for any set $A$ of allocations. This representation shows that the ex-post public choice $\boldsymbol{p}$ (i.e., the maximizer over $A$) maximizes

$$U(\boldsymbol{p}) \equiv (1-\beta_1)\alpha_1 u(p_1) + (1+\beta_S)\sum_{i \in S}\alpha_i u(p_i). \tag{3}$$

Note that since $\beta_1$ is always smaller than one, the function $U$ is monotonic in the decision maker's utility $u(p_1)$ and the social utility $\sum_{i \in S}\alpha_i u(p_i)$. Hence, no matter what the parameters are, the decision maker will *never* derive utility from decreasing his own utility (or

social utility) unless it increases social utility (or his own utility, respectively).

A comparison between (3) and (2) shows that the decision maker's choices become more altruistic at the ex-post public stage (i.e., in (3)) than they are at the ex-ante private stage (i.e., in (2)) if and only if $\beta_S > -\beta_1$. To see this note that the *relative* weight on the social utility $\sum_{i \in S} \alpha_i u$ at the ex-post public stage is $(1 + \beta_S)/((1 - \beta_1)\alpha_1)$. The relative weight at the ex-ante private stage is $1/\alpha_1$. Hence, the relative weight at the ex-post public stage is higher than the relative weight at the ex-ante private stage if and only if $1 + \beta_S > 1 - \beta_1$.[4]

In this way, $\beta_S > -\beta_1$ *captures impure altruism* caused by pride in acting altruistically or shame of acting selfishly. In the same way, $\beta_S < -\beta_1$ *captures impure selfishness* motivated by the temptation to act selfishly.

There are some limitations of the GU model. As it can be seen in (3), the ex-post choice satisfies IIA. That is, the decision maker will not change his ex-post choice even if an unchosen element is removed from his set of alternatives. In addition, the GU model does not incorporate inequality aversion since the GU model implies that the decision maker is a utilitarian. Both limitations result from the linearity of the GU model, which at the same time make it possible for us to identify the parameters (i.e., $\alpha$ and $\beta$) uniquely.

Second, since the model is based on observable preferences at the ex-ante (private) stage, an outside observer must know the decision maker's choices at the stage. Hence, the GU model does not incorporate pride and shame resulting from the outside observer. This limitation might not be so serious because, by definition, the outside observer is not included in the set $S$ of the other agents. Therefore, the outside observer's utility is not determined by the decision maker's choices. Consequently, the decision maker would feel little pride and shame resulting from the outside observer.[5]

Third, in the GU model, the decision maker evaluates the other agents' allocations $(p_i)_{i \in S}$ by using his own utility function $u$. This limitation can be fixed easily by using additional

---

[4]We provide behavioral characterizations of impure altruism and impure selfishness in Section 7.2.

[5]Indeed, in a well-designed experiment, subjects should not consider experimenters' intentions or their welfare. In such an experiment, the experimenters should have no effect on whether the decision maker feels shame or pride.

primitive preferences $(\succsim_i)_{i \in S}$ of the other agents. With these additional primitives, we could immediately obtain an extension of the GU model in which each agent's allocation $p_i$ is evaluated by the agent's utility function $u_i$ representing $\succsim_i$.

Finally, in the GU model, the temptation to act selfishly and pride of acting altruistically do not coexist. The decision maker can exhibit *either* pride in acting altruistically (i.e., $\beta_1 > 0$) *or* the temptation to act selfishly (i.e., $\beta_1 < 0$) *but not both*. This is because the decision maker feels pride in acting altruistically *only when* he does not yield to such a temptation. In other words, what can be captured in the GU model by the index $\beta_1$ is the *net* effect of the temptation and the pride. Since our main interest in the paper is to identify the effects of those feelings on choices, this limitation may not be so serious because at the end only the net effects matter on choices.

To conclude the section, we mention experimental evidence and an application. We show that the GU model can describe recent experimental evidence on dictator games with an exit option such as Dana et al. (2006), Broberg et al. (2007), and Lazear et al. (2012). The experimental design fits with our model. The GU model can be consistent not only with the choice of exit but also with the finding that *both low-level and high-level donors* play the dictator games, while the medium-level donors exit.

In addition, we show that the GU model is consistent with a classical empirical puzzle involving charitable donations. Standard consumer theory, which assumes that consumers' preferences are solely defined on allocations, predicts that government spending for charity should *completely crowd out* their donation. However, empirical evidence suggests that the crowding out is far from complete, and is at most about 50 percent. (See Andreoni (2006) for a survey of the evidence.) By applying the GU model, we describe the *partial crowding out* under the condition that consumers' pride dominates their shame (i.e., $\beta_1 > \beta_S$).

# 3 Related Literature

In this section, we overview the related literature. In Section 5, we explicitly compare our models and axioms with the models and axioms studied in Gul and Pesendorfer (2001), Dillenberger and Sadowski (2012), and Evren and Minardi (2014).

In the axiomatic literature on preferences over menus, since Dekel, Lipman, and Rustichini (2001) proposed a general frame work, many models have been proposed in specific contexts. Gul and Pesendorfer (2001) propose a general model of temptation. However, they do not study social decision making in particular. In their model, moreover, the self-control problem is captured by the difference between two von Neumann-Morgenstern utility functions. In contrast, we study a specific temptation in a social context: the temptation to act selfishly. Moreover, the GU model captures the self-control problem by the unique nonnegative number $\beta_1$, which facilitates distinguishing selfishness from impure selfishness.

Dillenberger and Sadowski (2012) have proposed the first model on social decision making within the literature on preferences over menus.[6] The purpose of their paper is to capture shame of acting selfishly; the other forces, namely pride and temptation, are outside of their scope. In their model, shame is captured by two nonunique functions. Because of this generality, their model can allow violations of IIA and inequality aversion.

Independent of our paper, several recent papers address related phenomena. Noor and Ren (2011) study guilt and temptation to avoid guilt by extending the model of Gul and Pesendorfer (2001) into preferences over menus of menus of allocations. Guilt is an importance issue but, as mentioned, there is a clear distinction between shame and guilt. In this paper, we focus on shame.

Evren and Minardi (2014) study warm glow. Their model is based on an interesting idea of relating warm glow to the freedom to be selfish. In Evren and Minardi's (2014) main result, their primitive preferences are different from ours: they focus a preference over sets of Pareto undominated allocations. Their model is general enough to allow for violations of IIA.

---

[6]Neilson (2009) also proposed a model of shame. Neilson's (2009) model is not axiomatic.

On the other hand, due to this generality, their representation lacks uniqueness properties.

Feddersen and Sandroni (2009) also provide an axiomatic model of warm glow. Their approach is different from the other axiomatic papers including ours: They study a decision maker's choice function and an *aspiration function* that captures the decision maker's normative concerns. One advantage of their approach is that they dispense with ex-ante choices of sets. However, they need the aspiration functions instead.

Yagasaki (2013) provides an axiomatic model of pride and shame. As primitives of his model, Yagasaki (2013) uses a function, which captures reference behavior, in addition to preference relations over sets. In his model, the decision maker may feel shame and pride by comparing his behavior and the reference behavior.

Finally, we mention two nonaxiomatic papers. Andreoni (1989, 1990) has proposed the celebrated model of warm glow. By using the model, Andreoni (1989, 1990) has obtained conditions that capture the partial crowding out, although the conditions are not easy to interpret. In Andreoni's (1989, 1990) model of warm glow, the decision maker can obtain positive utility by his donation even if his donation is private to the other agents and does not improve the welfare of the other agents. Hence, donations captured by Andreoni's (1989, 1990) model are essentially different from donations captured by the GU model.

Dufwenberg, Heidhues, Kirchsteiger, Riedel, and Sobel (2011) have studied menu-dependent other-regarding preferences. Their purpose is to analyze competitive market outcomes in economies where agents have such preferences. They proposed a general utility function that depends on the budge sets (opportunities, in their terminology), although they do not provide an axiomatization.

# 4   Model

Remember that the decision maker is denoted by 1 and the finite set of other agents is denoted by $S$. Hence, $I \equiv \{1\} \cup S$ is the set of all agents. Let $Z$ be a finite set. A *lottery* is a probability distribution over $Z$. We denote the set of lotteries by $\Delta(Z)$. For simplicity,

we assume that a payoff for each agent is a lottery. Hence, the set of allocations is $(\Delta(Z))^I$. Note that outcomes of lotteries are not correlated across agents. We denote the set of all nonempty closed subsets of $(\Delta(Z))^I$ by $\mathscr{A}$. An element of $\mathscr{A}$ is called a *set*.

The primitive of our model is a binary relation $\succsim$ on $\mathscr{A}$ that describes the decision maker's ex-ante private preference. We denote the asymmetric and symmetric parts of $\succsim$ by $\succ$ and $\sim$, respectively. We endow $\mathscr{A}$ with the topology generated by the Hausdorff metric.[7]

We use the following notation. Sets are denoted by $A, B$, and $C \in \mathscr{A}$. Allocations are denoted by $\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{l}$, and $\boldsymbol{r} \in (\Delta(Z))^I$. Lotteries are denoted by $p, q, l$, and $r \in \Delta(Z)$. In particular, lotteries for agent $i \in I$ are denoted by $p_i, q_i, l_i$, and $r_i \in \Delta(Z)$. We define convex combinations of two sets, two allocations, and two lotteries in the usual manner.[8] For a singleton, we write $\boldsymbol{p}$ instead of $\{\boldsymbol{p}\}$.

## 4.1 Axioms

**Axiom** (Standard): $\succsim$ is a nondegenerate, complete, transitive, and continuous binary relation.[9]

**Axiom** (Independence): For all $\alpha \in [0, 1]$ and $A, B, C \in \mathscr{A}$, $A \succsim B$ if and only if $\alpha A + (1 - \alpha)C \succsim \alpha B + (1 - \alpha)C$.

As mentioned, the GU model does not capture inequality aversion. This is mainly because of the independence axiom. (To incorporate inequality aversion, in Appendix C, we propose an extension of the GU model by weakening the independence axiom.)

We define the decision maker's risk preference $\succsim_1$ on $\Delta(Z)$ and social preference $\succsim_S$ on $(\Delta(Z))^S$ as follows:

---

[7]$d_h(A, B) = \max\{\max_{\boldsymbol{p} \in A} \min_{\boldsymbol{q} \in B} d(\boldsymbol{p}, \boldsymbol{q}), \max_{\boldsymbol{p} \in B} \min_{\boldsymbol{q} \in A} d(\boldsymbol{p}, \boldsymbol{q})\}$, where $d$ is the Euclidean metric. (Remember that $\boldsymbol{p}$ and $\boldsymbol{q}$ are $|Z| \cdot |I|$ dimensional real-valued vectors. Hence, $d$ is the Euclidean metric on $\mathbb{R}^{|Z| \cdot |I|}$.)

[8]For all $p_i, q_i \in \Delta(Z)$ and $\alpha \in [0, 1]$, $\alpha p_i + (1 - \alpha)q_i$ is a lottery such that $(\alpha p_i + (1 - \alpha)q_i)(z) = \alpha p_i(z) + (1 - \alpha)q_i(z)$ for each $z \in Z$. For all $\boldsymbol{p}, \boldsymbol{q} \in (\Delta(Z))^I$ and $\alpha \in [0, 1]$, $\alpha \boldsymbol{p} + (1 - \alpha)\boldsymbol{q}$ is an allocation such that $(\alpha \boldsymbol{p} + (1 - \alpha)\boldsymbol{q}) = (\alpha p_i + (1 - \alpha)q_i)_{i \in I}$. For all $A, B \in \mathscr{A}$ and $\alpha \in [0, 1]$, $\alpha A + (1 - \alpha)B = \{\alpha \boldsymbol{p} + (1 - \alpha)\boldsymbol{q} | \boldsymbol{p} \in A$ and $\boldsymbol{q} \in B\}$.

[9]Formally, the continuity is defined as follows: the sets $\{B \in \mathscr{A} | B \succsim A\}$ and $\{B \in \mathscr{A} | A \succsim B\}$ are closed in the Hausdorff metric topology. $\succsim$ is nondegenerate if $A \succ B$ for some $A, B \in \mathscr{A}$.

13

**Definition**: (i) For all $p_1, q_1 \in \Delta(Z)$, $p_1 \succsim_1 q_1$ if $(p_1, l_S) \succsim (q_1, l_S)$ for some $l_S \in (\Delta(Z))^S$;
(ii) for all $p_S, q_S \in (\Delta(Z))^S$, $p_S \succsim_S q_S$ if $(l_1, p_S) \succsim (l_1, q_S)$ for some $l_1 \in \Delta(Z)$.

The binary relation $\succsim_1$ captures the decision maker's *individual risk preference*, and $\succsim_S$ captures his *social preference* (i.e., his preference on other agents' allocations), which reflects the decision maker's belief about the other agents' preferences.[10] To see these interpretations, note that in part (i) of the definition, the two allocations differ *only* with respect to the decision maker's allocations, so that his choice between the two allocations does *not* affect other agents' allocations. Hence, the decision maker would allow himself to choose one allocation based *solely* on his individual preference.

Similarly, in part (ii) of the definition, the two allocations differ *only* with respect to other agents' allocations, so that the decision maker's choice between the two allocations does *not* affect his own allocation. Hence, the decision maker should choose one allocation to maximize the anticipated welfare of other agents. It can be shown that the preferences $\succsim_1$ and $\succsim_S$ are well defined under the independence axiom.

We assume the following consistency condition between the two preferences.

**Axiom** (Consistency): For all $p_S, q_S \in (\Delta(Z))^S$, if $p_i \succsim_1 q_i$ for all $i \in S$, then $p_S \succsim_S q_S$. Moreover, if $p_i \succsim_1 q_i$ for all $i \in S$ and $p_i \succ_1 q_i$ for some $i \in S$, then $p_S \succ_S q_S$.

The consistency condition means that if every other agents' allocation is better in $p_S$ than in $q_S$ according to the decision maker's individual preference $\succsim_1$, then $p_S$ is better than $q_S$ in his social preference $\succsim_S$. The decision maker uses his individual preference to evaluate the other agents' allocations because the decision maker here is *not* a social planner and might not know the other agents' preferences. This axiom implies that the decision maker evaluates the other agents' allocations $(p_i)_{i \in S}$ by using his own utility function $u$.

The next axiom requires that $\succsim$ satisfy the standard Pareto condition.

**Axiom** (Pareto): For any $\boldsymbol{p}, \boldsymbol{q} \in (\Delta(Z))^I$, if $p_1 \succsim_1 q_1$ and $p_S \succsim_S q_S$, then $\{\boldsymbol{p}\} \succsim \{\boldsymbol{q}\}$. Moreover if either $p_1 \succ_1 q_1$ or $p_S \succ_S q_S$, then $\{\boldsymbol{p}\} \succ \{\boldsymbol{q}\}$.

---

[10]For each $i \in \{1, S\}$, we write the asymmetric and symmetric parts of $\succsim_i$ by $\succ_i$ and $\sim_i$, respectively.

The next axiom, *Dominance*, says that the decision maker is indifferent to adding an allocation that is dominated both in individual preference and in social preference.

**Axiom** (Dominance): For any set $A \in \mathscr{A}$ and $\boldsymbol{p} \in (\Delta(Z))^I$, if there exists $\boldsymbol{q} \in A$ such that $q_1 \succsim_1 p_1$ and $q_S \succsim_S p_S$, then $A \sim A \cup \boldsymbol{p}$.

This axiom means that if an allocation $\boldsymbol{p}$ is dominated both in the individual preference and in his social preference, then the decision maker should not choose $\boldsymbol{p}$ from $A \cup \boldsymbol{p}$. Moreover, having the option to choose $\boldsymbol{p}$ does not cause the decision maker's feeling of shame, pride, and the temptation. Hence, the decision maker is indifferent between $A$ and $A \cup \boldsymbol{p}$.

The next axiom is a weaker version of the axiom of *Strategic Rationality*. Kreps (1963) defines Strategic Rationality as follows: $A \succsim B \Rightarrow A \sim A \cup B$. Strategic Rationality means that the decision maker has neither a preference for commitment nor a preference for flexibility. Because of the shame of acting selfishly and the temptation of acting selfishly, the decision maker can have a preference for commitment. On the other hand because of the feeling of pride of acting altruistically he can have a preference for flexibility.

To introduce the weaker version of Strategic Rationality, we need the following definitions. A similar definition appears in Kopylov (2012). Given a set $A$ and any element $\boldsymbol{p} \in A$, say that $\boldsymbol{p}$ is *individually optimal* in $A$ if $p_1 \succsim_1 q_1$ for all $\boldsymbol{q} \in A$ and that $\boldsymbol{p}$ is *socially optimal* in $A$ if $p_S \succsim_S q_S$ for all $\boldsymbol{q} \in A$. For any allocation $\boldsymbol{p} \in (\Delta(Z))^I$, define

$$
\begin{aligned}
\mathscr{A}_{\boldsymbol{p}}^1 &= \{A \in \mathscr{A} | \boldsymbol{p} \text{ is individually optimal in } A\}, \\
\mathscr{A}_{\boldsymbol{p}}^S &= \{A \in \mathscr{A} | \boldsymbol{p} \text{ is socially optimal in } A\}.
\end{aligned}
$$

For example, if $A, B \in \mathscr{A}_{\boldsymbol{p}}^1$ (or $\mathscr{A}_{\boldsymbol{p}}^S$), then $A$ and $B$ share the same optimal element in $\succsim_1$ (or in $\succsim_S$ respectively).

**Axiom** (Weak Strategic Rationality): For any $\boldsymbol{p}, \boldsymbol{q} \in (\Delta(Z))^I$ and $A, B \in \mathscr{A}_{\boldsymbol{p}}^1 \cap \mathscr{A}_{\boldsymbol{q}}^S$,

$$
A \succsim B \Rightarrow A \sim A \cup B.
$$

This axiom assumes three things: (i) the decision maker feels shame of acting selfishly by not choosing the socially optimal allocation; (ii) he feels temptation to choose the individually optimal allocation; and (iii) he can feel the pride of acting altruistically by overcoming such temptation. Consider two sets $A$ and $B$ that share the same socially optimal allocation and the same individually optimal allocation. Then $A$ and $A \cup B$ also share the same socially optimal allocation and the same individually optimal allocation. Hence, under the assumption, the decision maker's utility or disutility caused by the shame, the temptation, and the pride are the same in $A$, $B$, and $A \cup B$. Therefore, the only one reason that the decision maker could prefer $A$ to $B$ is that $A$ contains the optimal allocation in his ex-post preference. Consequently, the decision maker should be indifferent between $A$ and $A \cup B$ because he can select the same ex-post optimal allocation of $A$ from both $A$ and $A \cup B$.

The last axiom captures *shame of acting selfishly*: the decision maker feels shame when his ex-post choice is judged inferior by the social preference $\succsim_S$. Hence, the decision maker might have a preference for commitment in order to exclude the socially superior allocation from his choice set as follows:

**Axiom** (Shame of Acting Selfishly): For any $\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r} \in (\Delta(Z))^I$,

$$\left\{ \begin{array}{ll} \text{(i)} & q_S \succ_S p_S \\ \text{(ii)} & r_1 \succ_1 p_1 \\ \text{(iii)} & \{\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r}\} \succ \{\boldsymbol{q}, \boldsymbol{r}\} \end{array} \right\} \Rightarrow \{\boldsymbol{p}\} \succsim \{\boldsymbol{p}, \boldsymbol{q}\}.$$

By (i), $\boldsymbol{q}$ is better than $\boldsymbol{p}$ in the social preference (i.e., $q_S \succ_S p_S$). On the other hand, (ii) and (iii) imply that $\boldsymbol{p}$ is better than $\boldsymbol{q}$ in the ex-post preference. To see this, note that by (ii), $\boldsymbol{p}$ is not individually optimal (i.e., $r_1 \succ_1 p_1$). Since even a purely selfish agent, who maximizes $\succsim_1$, would not choose $\boldsymbol{p}$, the act of not choosing $\boldsymbol{p}$ cannot reveal the decision maker's altruism. Thus, the decision maker will not experience any feeling of pride from having (but not exercising) the option to choose $\boldsymbol{p}$. Therefore, the only one reason that the decision maker could prefer $\{\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r}\}$ to $\{\boldsymbol{q}, \boldsymbol{r}\}$ as in (iii), is that he prefers to choose $\boldsymbol{p}$

ex-post.

Hence, the decision maker will choose socially inferior allocation $\boldsymbol{p}$ from $\{\boldsymbol{p}, \boldsymbol{q}\}$ at the ex-post public stage, which may cause shame of acting selfishly. Consequently, the decision maker would prefer not to have the option to choose $\boldsymbol{q}$ (i.e., $\{\boldsymbol{p}\} \succsim \{\boldsymbol{p}, \boldsymbol{q}\}$) at the ex-ante private stage.

## 4.2 Representation Theorem

**Theorem**: *The following statements are equivalent:*

*(a) $\succsim$ satisfies Standard, Independence, Consistency, Pareto, Dominance, Weak Strategic Rationality, and Shame of Acting Selfishly.*

*(b) There exist a nonconstant von Neumann-Morgenstern function u on $\Delta(Z)$ and real numbers $(\boldsymbol{\alpha}, \beta_1, \beta_S)$ such that $\succsim$ is represented by*

$$V(A) = \max_{p \in A} \Big[ \sum_{i \in I} \alpha_i u(p_i) + \beta_1 \max_{q \in A} \alpha_1(u(q_1) - u(p_1)) - \beta_S \max_{r \in A} (\sum_{i \in S} \alpha_i u(r_i) - \sum_{i \in S} \alpha_i u(p_i)) \Big], \quad (4)$$

*where $\alpha_1 > 0$, $\sum_{i \in S} \alpha_i = 1$, $\beta_1 < 1$, and $\beta_S \geq 0$.*

By the definitions of $\succsim_1$ and $\succsim_S$, the theorem trivially implies that $\succsim_1$ and $\succsim_S$ are represented by $u$ and $\sum_{i \in S} \alpha_i u$, respectively.

In the rest of the section, we provide an intuition for the proof of sufficiency. By Independence, we show that there exist a von Neumann-Morgenstern utility function $u$ on $\Delta(Z)$ and positive numbers $\{\alpha_i\}_{i \in S}$ such that (i) $u$ represents $\succsim_1$ on $\Delta(Z)$ and (ii) $\sum_{i \in S} \alpha_i u$ represents $\succsim_S$ and $\sum_{i \in S} \alpha_i = 1$. For any set $A$ of allocations, we consider a set of utilities of allocations in $A$. Formally, for all $A \in \mathscr{A}$, define $\boldsymbol{u}(A) = \{(u(p_1), \sum_{i \in S} \alpha_i u(p_i)) \in \mathbb{R}^2 | \boldsymbol{p} \in A\}$ and $\mathscr{A}^* = \{\boldsymbol{u}(A) | A \in \mathscr{A}\}$. Define

$$A^* \succsim^* B^* \text{ if } A \succsim B,$$

where $\boldsymbol{u}(A) = A^*$ and $\boldsymbol{u}(B) = B^*$. By mainly using Dominance, we can show that $\succsim^*$ is well defined (i.e., if $A^* = B^*$, then $A \sim B$). It is easy to translate the axioms on $\succsim$ to axioms on $\succsim^*$.

Then, we apply Kopylov (2009, Theorem 2.1) to obtain a linear representation for $\succsim^*$. Because of the linearity of the model, we can obtain a function $\mu : \mathbb{R} \to \mathbb{R}$ such that $V^*(A^*) = \sum_{\lambda \in \mathbb{R}} \left( \max_{\boldsymbol{u} \in A^*} \lambda u_1 + (1 - \lambda) u_S \right) \mu(\lambda)$ represents $\succsim^*$. Moreover, $\text{supp}(\mu) \equiv \{\lambda \in \mathbb{R} | \mu(\lambda) \neq 0\}$ is finite.

To obtain the GU model, we need to show that (i) $\mu(1) \neq 0$, (ii) $\mu(0) < 0$, (iii) $\mu(\lambda^*) > 0$ for some $\lambda^* \in (0, 1)$, and (iv) $\mu(\lambda)$ is zero for all of the other $\lambda \in \mathbb{R}$. In other words, to obtain the GU model, unlike the other additive models such as Gul and Pesendorfer (2001) and Kopylov (2012), it is not enough to know the number of the elements of $\{\lambda \in \mathbb{R} | \mu(\lambda) > 0\}$ and $\{\lambda \in \mathbb{R} | \mu(\lambda) < 0\}$. To obtain the GU model, it is crucial to identify the elements of $\{\lambda \in \mathbb{R} | \mu(\lambda) > 0\}$ and $\{\lambda \in \mathbb{R} | \mu(\lambda) < 0\}$.

For this purpose, fix $\varepsilon < \frac{1}{2}$ and for all $\lambda \in \mathbb{R}$, define

$$u_1^*(\lambda) = \frac{1}{2} + \frac{\varepsilon \lambda}{\|(\lambda, 1 - \lambda)\|}, \quad u_S^*(\lambda) = \frac{1}{2} + \frac{\varepsilon(1 - \lambda)}{\|(\lambda, 1 - \lambda)\|}, \quad \text{and } \boldsymbol{u}^*(\lambda) = (u_1^*(\lambda), u_S^*(\lambda)).$$

Then $\boldsymbol{u}^*(\lambda)$ is on the boundary of $B_\varepsilon(\frac{1}{2}, \frac{1}{2})$ (i.e., the closed ball with radius $\varepsilon$ centered at $(\frac{1}{2}, \frac{1}{2})$). We can show that for all $\lambda, \lambda' \in \mathbb{R}$ such that $\lambda \neq \lambda'$, $\lambda u_1^*(\lambda) + (1 - \lambda) u_S^*(\lambda) > \lambda u_1^*(\lambda') + (1 - \lambda) u_S^*(\lambda')$. To put it simply, for any $\lambda \in \mathbb{R}$, $\boldsymbol{u}^*(\lambda) \equiv (u_1(\lambda), u_S(\lambda))$ is unique maximizer of $\lambda u_1 + (1 - \lambda) u_S$ among the sets $\{\boldsymbol{u}^*(\lambda') | \lambda' \in \mathbb{R}\}$ of utility pairs.

By using this property of $\boldsymbol{u}^*$, we can construct menus of utilities by which we can identify the sign of $\mu$ on any $\lambda \in \mathbb{R}$. To see this fix $\lambda \in \mathbb{R}$. Define $A^* = \{\boldsymbol{u}^*(\lambda') | \lambda' \in \text{supp}(\mu)\}$. Then, it must hold that $A^* \succ^* A^* \setminus \{\boldsymbol{u}^*(\lambda)\}$ if and only if $\mu(\lambda) > 0$.[11] Based on this idea, by using Dominance and Weak Strategic Rationality, we can show the aforementioned properties (i.e., (i)–(iv)). Consequently, there exists $\lambda^* \in (0, 1)$ such that $\mu(\lambda^*) > 0$ and $V^*(A^*) = \mu(1) \max_{\boldsymbol{u} \in A^*} u_1 + \mu(\lambda^*) \max_{\boldsymbol{u} \in A^*} (\lambda^* u_1 + (1 - \lambda^*) u_S) + \mu(0) \max_{\boldsymbol{u} \in A^*} u_S$. By using

---

[11]Because $\text{supp}(\mu)$ is finite, it can be shown that $A^* \setminus \{u^*(\lambda')\} \in \mathscr{A}^*$.

Shame of Acting Selfishly, we can show that $\mu(0) \leq 0$.

By defining $\alpha_1, \beta_1$, and $\beta_S$ based on $\mu$, we obtain $V^*(A^*) = \max_{\boldsymbol{u} \in A^*} \big((1 - \beta_1)\alpha_1 u_1 + (1 + \beta_S)u_S\big) + \beta_1 \alpha_1 \max_{\boldsymbol{u} \in A^*} u_1 - \beta_S \max_{\boldsymbol{u} \in A^*} u_S$.[12] Finally, for all $A \in \mathscr{A}$, we define $V(A) = V^*(A^*)$. Then, $A \succsim B \Leftrightarrow A^* \succsim^* B^* \Leftrightarrow V^*(A^*) \geq V^*(B^*) \Leftrightarrow V(A) \geq V(B)$. Hence, $V$ represents $\succsim$. By arranging the terms and substituting $u_1 = u(p_1)$ and $u_S = \sum_{i \in S} \alpha_i u(p_i)$, we obtain the GU model.

## 4.3 Uniqueness of Parameters

In the GU model, the parameters $\alpha$ and $\beta$ are unique.

**Proposition 1** *If two GU models with $(u, \boldsymbol{\alpha}, \boldsymbol{\beta})$ and $(u', \boldsymbol{\alpha}', \boldsymbol{\beta}')$ represent the same $\succsim$, then $\beta_1 = \beta_1'$, $\beta_S = \beta_S'$, $\alpha_i = \alpha_i'$ for all $i \in I$, and there exist real numbers $a, b$ such that $a > 0$ and $u = au' + b$.*

On the other hand, in the models of Gul and Pesendorfer (2001), Dillenberger and Sadowski (2012), and Evren and Minardi (2014), parameters capturing the temptation, shame, and pride are not uniquely identified. As mentioned, the improved identification in the GU model comes from its linear structure. At the same time, the linear structure makes the GU model restrictive. In Section 5.2, we discuss this issue in detail.

# 5 Axiomatizations of Special Cases

In Section 5.1, we provide axiomatizations of the special cases of GU models with $\beta_1 = 0$ and $\beta_S = 0$. In Section 5.2, we discuss the relationship between the special cases and Gul and Pesendorfer (2001), Dillenberger and Sadowski (2012), and Evren and Minardi (2014).

---

[12] $\alpha_1 = \mu(\lambda^*)\lambda^* + \mu(1)$, $\beta_1 = \frac{\mu(1)}{\mu(\lambda^*)\lambda^* + \mu(1)}$, and $\beta_S = -\mu(0)$.

## 5.1 Special Cases with $\beta_1 = 0$ and $\beta_S = 0$

In this section, we provide axiomatizations of two special cases of GU models. Each of them is a special cases of either Gul and Pesendorfer (2001), Dillenberger and Sadowski (2012), or Evren and Minardi (2014). Both models are characterized by stronger versions of Weak Strategic Rationality. Remember that Weak Strategic Rationality is defined as follows: for any $\boldsymbol{p}, \boldsymbol{q} \in (\Delta(Z))^I$ and $A, B \in \mathscr{A}_{\boldsymbol{p}}^1 \cap \mathscr{A}_{\boldsymbol{q}}^S$, if $A \succsim B$, then $A \sim A \cup B$. Remember that if $A, B \in \mathscr{A}_{\boldsymbol{p}}^S$ (or $\mathscr{A}_{\boldsymbol{p}}^1$), then $A$ and $B$ share the same optimal element in $\succsim_S$ (or in $\succsim_1$ respectively).

To obtain the special case of the GU model in which $\beta_1 = 0$, we strengthen Weak Strategic Rationality as follows:

**Axiom** (Weak Strategic Rationality I): For any $\boldsymbol{p}, \boldsymbol{q} \in (\Delta(Z))^I$ and $A, B \in \mathscr{A}_{\boldsymbol{p}}^S$,

$$A \succsim B \Rightarrow A \sim A \cup B.$$

This axiom assumes that the decision maker feels shame of acting selfishly by not choosing the socially optimal allocation. Consider two sets $A$ and $B$ that share the same socially optimal allocation. Then, $A$ and $A \cup B$ also share the same socially optimal allocation. Hence, under the assumption, the decision maker's disutility caused by shame are the same in $A$, $B$, and $A \cup B$. Therefore, the only one reason that the decision maker could prefer $A$ to $B$ is that $A$ contains the optimal allocation in his ex-post preference. Consequently, the decision maker should be indifferent between $A$ and $A \cup B$ because he can select the same ex-post optimal element of $A$ from both $A$ and $A \cup B$.

**Corollary 1** *The following statements (a) and (b) are equivalent:*

*(a) $\succsim$ satisfies Standard, Independence, Consistency, Pareto, Dominance, and Weak Strategic Rationality I, and Shame of Acting Selfishly.*

*(b) There exist a GU model $(u, \boldsymbol{\alpha}, (\beta_1, \beta_S))$ with $\beta_1 = 0$ that represents $\succsim$.*

*Moreover, if two GU models with $(u, \boldsymbol{\alpha}, (0, \beta_S))$ and $(u', \boldsymbol{\alpha}', (0, \beta_S'))$ represent the same $\succsim$,*

then $\beta_S = \beta'_S$, $\alpha_i = \alpha'_i$ for all $i \in I$, and there exist real numbers $a, b$ such that $a > 0$ and $u = au' + b$.

In contrast to Theorem, in Corollary 1 we impose Weak Strategic Rationality I, instead of Weak Strategic Rationality.

In the following, we provide a characterization of the other special case of GU model in which $\beta_S = 0$. For this purpose, we need to strengthen Weak Strategic Rationality as follows:

**Axiom** (Weak Strategic Rationality II): For any $\boldsymbol{p}, \boldsymbol{q} \in (\Delta(Z))^I$ and $A, B \in \mathscr{A}_{\boldsymbol{p}}^1$,

$$A \succsim B \Rightarrow A \sim A \cup B.$$

Weak Strategic Rationality II is different from Weak Strategic Rationality I only in one point. In Weak Strategic Rationality II, we require $A, B \in \mathscr{A}_{\boldsymbol{p}}^1$, not $A, B \in \mathscr{A}_{\boldsymbol{p}}^S$. This is because in the special case of the GU model in which $\beta_S = 0$, what matters to the decision maker is the individually optimal allocations, not the socially optimal allocations.

Weak Strategic Rationality II assumes that (i) the decision maker feels temptation to choose the individually optimal allocation; and (ii) he can feel the pride of acting altruistically by overcoming such temptation.

Instead of Shame of Acting Selfishly, we need a stronger version of the axiom:

**Axiom** (No Shame of Acting Selfishly): For any $\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r} \in (\Delta(Z))^I$,

$$\left\{ \begin{array}{ll} \text{(i)} & q_S \succ_S p_S \\ \text{(ii)} & r_1 \succ_1 p_1 \\ \text{(iii)} & \{\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r}\} \succ \{\boldsymbol{q}, \boldsymbol{r}\} \end{array} \right\} \Rightarrow \{\boldsymbol{p}\} \sim \{\boldsymbol{p}, \boldsymbol{q}\}.$$

As explained, the condition (i), (ii), and (iii) ensure that the decision maker would choose $\boldsymbol{p}$ ex-post but $\boldsymbol{p}$ is socially inferior to $\boldsymbol{q}$. Hence, if the decision maker anticipates the feeling of shame by not choosing $\boldsymbol{q}$, then he would prefer not to have the option to choose $\boldsymbol{q}$

(i.e., $\{\boldsymbol{p}\} \succ \{\boldsymbol{p}, \boldsymbol{q}\}$) at the ex-ante private stage. If the decision maker does not anticipate any feeling of shame, then he should be indifferent to have the option to choose $\boldsymbol{q}$ (i.e., $\{\boldsymbol{p}\} \sim \{\boldsymbol{p}, \boldsymbol{q}\}$).

**Corollary 2** *The following statements (a) and (b) are equivalent:*

*(a) $\succsim$ satisfies Standard, Independence, Consistency, Pareto, Dominance, Weak Strategic Rationality II, and No Shame of Acting Selfishly.*

*(b) There exist a GU model $(u, \boldsymbol{\alpha}, (\beta_1, \beta_S))$ with $\beta_S = 0$ that represents $\succsim$.*

*Moreover, if two GU models with $(u, \boldsymbol{\alpha}, (\beta_1, 0))$ and $(u', \boldsymbol{\alpha}', (\beta_1', 0))$ represent the same $\succsim$, then $\beta_1 = \beta_1'$, $\alpha_i = \alpha_i'$ for all $i \in I$, and there exist real numbers $a, b$ such that $a > 0$ and $u = au' + b$.*

## 5.2 Comparison with Dillenberger and Sadowski (2012), Evren and Minardi (2014), and Gul and Pesendorfer (2001)

In this section, we compare the two special cases of the GU model and Dillenberger and Sadowski (2012), Evren and Minardi (2014), and Gul and Pesendorfer (2001).[13] The GU model with $\beta_1 = 0$ is

$$V(A) = \max_{\boldsymbol{p} \in A} \left[ \sum_{i \in I} \alpha_i u(p_i) - \beta_S \max_{\boldsymbol{r} \in A} (\sum_{i \in S} \alpha_i u(r_i) - \sum_{i \in S} \alpha_i u(p_i)) \right]. \tag{5}$$

In the model of Dillenberger and Sadowski (2012), the primitive preferences are defined on the sets of allocations of monetary payoffs.

Dillenberger and Sadowski (2012) derive fairness preferences $\succ_f$ from their primitive preferences $\succsim$ as follows: $\boldsymbol{x} \succ_f \boldsymbol{y}$ if there exists a set $A$ such that $\boldsymbol{y} \in A$ and $A \succ A \cup \boldsymbol{x}$. Then, they impose axioms on $\succsim$ as well as on $\succ_f$ to axiomatize the following model:

$$V_{DS}(A) = \max_{\boldsymbol{x} \in A} \left[ v(\boldsymbol{x}) - g\big(\phi(\boldsymbol{x}), \max_{\boldsymbol{y} \in A} \phi(\boldsymbol{y})\big) \right], \tag{6}$$

---

[13]I really appreciate an associate editor who gave us insightful comments on the comparison.

where $\phi$ represents $\succ_f$ and $g$ captures the disutility of shame. The value of $g$ depends on the social utility $\phi(\boldsymbol{x})$ of the agent's choice $\boldsymbol{x}$ and the optimal social utility $\max_{\boldsymbol{y} \in A} \phi(\boldsymbol{y})$.

To compare (5) and (6), assume that outcomes are monetary payoffs in (5). Then, (5) is a special case of (6) in which $v = \sum_{i \in I} \alpha_i u$, $g(a,b) = \beta_S(a-b)$, and $\phi = \sum_{i \in S} \alpha_i u$. The linearity of $g$ results from Independence. In (6), $g$ and $\phi$ are not uniquely identified. While in (5), $\alpha$ and $\beta_S$ are uniquely identified. This stronger uniqueness property comes from the linearity of $g$, which, at the same time, makes (5) incompatible with violations of IIA.

The additional structure that $\phi = \sum_{i \in S} \alpha_i u$ comes mainly from Weak Strategic Rationality I and Shame of Acting Selfishly. Weak Strategic Rationality I implies that if $\boldsymbol{x} \succ_f \boldsymbol{y}$, then $x_S \succ_S y_S$. To see this suppose that not $x_S \succ_S y_S$ but $\boldsymbol{x} \succ_f \boldsymbol{y}$. Then there exists $A \in \mathscr{A}$ such that $\boldsymbol{y} \in A$ and $A \succ A \cup \boldsymbol{x}$. Since not $x_S \succ_S y_S$, we have $A, A \cup \boldsymbol{x} \in \mathscr{A}_{\boldsymbol{z}}^S$ for some $\boldsymbol{z} \in A$. Then by Weak Strategic Rationality I, we obtain $A \sim A \cup \boldsymbol{x}$, which is a contradiction. By using Shame of Acting Selfishly and other axioms, it can be shown that if $x_S \succ_S y_S$, then $\boldsymbol{x} \succ_f \boldsymbol{y}$. Hence, we obtain $x_S \succ_S y_S$ if and only if $\boldsymbol{x} \succ_f \boldsymbol{y}$. Since $\succ_S$ is represented by $\sum_{i \in S} \alpha_i u$ and $\succ_f$ is represented by $\phi$, we obtain $\phi = \sum_{i \in S} \alpha_i u$ under a normalization.

Next we discuss the other special of the GU model with $\beta_S = 0$:

$$V(A) = \max_{\boldsymbol{p} \in A} \left[ \sum_{i \in I} \alpha_i u(p_i) + \beta_1 \max_{\boldsymbol{q} \in A} \alpha_1 (u(q_1) - u(p_1)) \right]. \tag{7}$$

When $\beta_1 > 0$, the second term captures pride and when $\beta_1 < 0$, the second term captures the temptation to act selfishly. In the following, we discuss the relationship between (7) and Evren and Minardi (2014) as well as Gul and Pesendorfer (2001).

In the model of Evren and Minardi (2014), the primitive preferences are defined on the sets of allocations of monetary payoffs. Evren and Minardi (2014) have axiomatized the following model:

$$V_{EM}(A) = \max_{\boldsymbol{x} \in A} U\big(\boldsymbol{x}, \max_{\boldsymbol{y} \in A} y_1 - x_1\big), \tag{8}$$

where the first argument captures the evaluation of the allocation $x$ and the second argument captures the difference between the most selfish allocation and the actual choice $x_1$.

To compare (7) and (8), assume that outcomes are monetary payoffs in (7). If $\beta_1 > 0$ and $u$ is linear (i.e., $u(x_i) = x_i$), (7) is a special case of (8) in which $U(\boldsymbol{x}, d) = \alpha_1 x_1 + \alpha_2 x_2 + \beta_1 d$. This linearity comes from again Independence. In (8), $U$ is not uniquely identified. While in (5), $\alpha$ and $\beta_1$ are uniquely identified due to the the linearity of $U$, which makes (7) incompatible with violations of IIA.

It can be shown that Weak Strategic Rationality II and the completeness of the preferences imply one of the key axioms in Evren and Minardi (2014), Weak Instrumentalism. That axiom is defined as follows: Let $A, B, C \in \mathscr{A}$ and suppose that $C = A \cup B$. If there exists a $\boldsymbol{y} \in A \cap B$ such that $y_1 \geq x_1$ for every $\boldsymbol{x} \in C$, then $C \sim A$ or $C \sim B$.[14]

Gul and Pesendorfer's (2001) primitive preferences are defined on the sets of lotteries, not allocations of lotteries. In the following, in order to compare Gul and Pesendorfer's (2001) model and (7), we discuss the extended model of Gul and Pesendorfer (2001), in which preferences are defined on the set of allocations of lotteries. By using several axioms, including Independence, Gul and Pesendorfer (2001) obtain the following model:

$$V_{GP}(A) = \max_{\boldsymbol{p} \in A} \left[ w(\boldsymbol{p}) + v(\boldsymbol{p}) - \max_{\boldsymbol{q} \in A} v(\boldsymbol{q}) \right]. \tag{9}$$

If $\beta_1 = -1$, (7) is a special case of (9) in which $w(\boldsymbol{p}) = \sum_{i \in S} \alpha_i u(p_i)$ and $v(\boldsymbol{p}) = \alpha_1 u(p_1)$. In Gul and Pesendorfer (2001), $w$ and $v$ are unique up to positive affine transformation. In the GU model, $u$ is unique up to positive affine transformation and $\alpha$ and $\beta_1$ are unique. The additional structure that $w(\boldsymbol{p}) = \sum_{i \in S} \alpha_i u(p_i)$ and $v(\boldsymbol{p}) = \alpha_1 u(p_1)$ mainly comes from Weak Strategic Rationality II because the axiom implies that the temptation utility is the selfish utility (i.e., $v = \alpha_1 u$). This additional structure together with the richer domain in turn makes it possible for us to uniquely identify the parameters (i.e., $\alpha, \beta_1$) in the GU model.

---

[14]To see this note that if $\boldsymbol{y} \in A \cap B$ such that $y_1 \geq x_1$ for every $\boldsymbol{x} \in C$, then we obtain $A, B \in \mathscr{A}_{\boldsymbol{y}}^1$. Therefore, by Weak Strategic Rationality II, if $A \succsim B$ then $A \sim C$; and if $B \succsim A$ then $B \sim C$. Since $\succsim$ is complete, $A \succsim B$ or $B \succsim A$ holds, hence $C \sim A$ or $C \sim B$ holds.

# 6 Characterization of Parameters

Remember that in GU model, $\beta_1$ can be positive or negative. If $\beta_1$ is positive, then the GU model captures the pride of acting altruistically; if $\beta_1$ is negative, then the GU model captures the temptation to act selfishly. In Section 6.1, we provide two axioms that identify $\beta_1 \leq 0$ and $\beta_1 \geq 0$, respectively. In Section 6.2, we provide comparative statics on $\alpha_1$, $\beta_1$, and $\beta_S$.

## 6.1 Pride and Temptation

The next axiom captures a decision maker's *temptation to act selfishly* (i.e., the temptation to maximize his own allocation). To minimize the cost of self-control, the decision maker might prefer to commit a smaller set that does not contain an allocation that he will not choose ex-post but that is superior (i.e., tempting) in his individual preference. We formalize the axiom as follows:

**Axiom** (Temptation to Act Selfishly): For any $\boldsymbol{p}, \boldsymbol{q} \in (\Delta(Z))^I$,

$$\left\{ \begin{array}{ll} \text{(i)} & q_1 \succ_1 p_1 \\ \text{(ii)} & \{\boldsymbol{p}, \boldsymbol{q}\} \succ \{\boldsymbol{q}\} \\ \text{(iii)} & p_S \succ_S q_S \end{array} \right\} \Rightarrow \{\boldsymbol{p}\} \succsim \{\boldsymbol{p}, \boldsymbol{q}\}.$$

By (i), $\boldsymbol{q}$ is better than $\boldsymbol{p}$ in the individual preference (i.e., $q_1 \succ_1 p_1$). Moreover, (i) and (ii) imply that $\boldsymbol{p}$ is better than $\boldsymbol{q}$ in the ex-post preference. To see this, note that even a purely selfish agent, who maximizes $\succsim_1$, would not choose $\boldsymbol{p}$. Hence, the act of not choosing $\boldsymbol{p}$ cannot reveal the decision maker's altruism. Thus, the decision maker will not experience any feeling of pride from having (but not exercising) the option to choose $\boldsymbol{p}$. Therefore, the only one reason that the decision maker could prefer $\{\boldsymbol{p}, \boldsymbol{q}\}$ to $\{\boldsymbol{q}\}$ as in (ii), is that he prefers to choose $\boldsymbol{p}$ ex-post.

Moreover, by (iii), $\boldsymbol{p}$ is better than $\boldsymbol{q}$ in the social preference (i.e., $p_S \succ_S q_S$). Thus the

decision maker does not feel the shame of acting selfishly when he chooses $\boldsymbol{p}$ from $\{\boldsymbol{p}, \boldsymbol{q}\}$. Hence, the decision maker prefers $\{\boldsymbol{p}\}$ to $\{\boldsymbol{p}, \boldsymbol{q}\}$ only because of the cost of self-control of not choosing the individually optimal allocation (i.e., $\boldsymbol{q}$) from $\{\boldsymbol{p}, \boldsymbol{q}\}$.

In contrast, by not choosing such selfish allocations, a decision maker might feel *pride in acting altruistically*. Hence, such a decision maker might prefer to have (but not exercise) the option to choose the selfish allocations. This observation can be formalized as follows:

**Axiom** (Pride in Acting Altruistically): For any $\boldsymbol{p}, \boldsymbol{q} \in (\Delta(Z))^I$,

$$\left\{ \begin{array}{ll} \text{(i)} & q_1 \succ_1 p_1 \\ \text{(ii)} & \{\boldsymbol{p}, \boldsymbol{q}\} \succ \{\boldsymbol{q}\} \\ \text{(iii)} & p_S \succ_S q_S \end{array} \right\} \Rightarrow \{\boldsymbol{p}, \boldsymbol{q}\} \succsim \{\boldsymbol{p}\}.$$

**Proposition 2** *Suppose that $\succsim$ is represented by the GU model with $(u, \boldsymbol{\alpha}, \boldsymbol{\beta})$.*
*(i) $\succsim$ exhibits Temptation to Act Selfishly if and only if $\beta_1 \leq 0$,*
*(ii) $\succsim$ exhibits Pride in Acting Altruistically if and only if $\beta_1 \geq 0$.*

## 6.2 Comparative Statics

By exploiting the strong uniqueness of the GU model, we can show that the value of each parameter captures the degree of each phenomenon: $\alpha_1$ captures the level of selfishness (i.e., $1/\alpha_1$ captures the level of altruism). In contrast, $\beta_1$ and $\beta_S$ capture the levels of pride seeking (temptation aversion) and shame aversion, respectively.

**Definition**: For any preferences $\succsim^X$ and $\succsim^Y$ on $\mathscr{A}$ such that $\succsim_i^X = \succsim_i^Y$ for each $i \in \{1, S\}$, $\succsim^X$ is *more altruistic* than $\succsim^Y$ if the following is true: for any $\boldsymbol{p}, \boldsymbol{q} \in (\Delta(Z))^I$ such that $p_S \succ_S^i q_S$ for each $i \in \{X, Y\}$,

$$\boldsymbol{p} \succsim^Y \boldsymbol{q} \Rightarrow \boldsymbol{p} \succsim^X \boldsymbol{q}.$$

Note that in the definition above $\boldsymbol{p}$ is superior to $\boldsymbol{q}$ in the social preference (i.e., $p_S \succ_S^i q_S$). Hence, if an agent (i.e., agent $Y$) prefers $\boldsymbol{p}$ to $\boldsymbol{q}$, then a more altruistic agent (i.e., agent $X$)

also should prefer $\boldsymbol{p}$ to $\boldsymbol{q}$.

**Proposition 3** *Suppose $\succsim^X$ and $\succsim^Y$ are represented respectively by two GU models with $(u, \boldsymbol{\alpha}^X, \boldsymbol{\beta}^X)$ and $(u, \boldsymbol{\alpha}^Y, \boldsymbol{\beta}^Y)$ such that $\alpha_S^X = \alpha_S^Y$.[15] Then, $X$ is more altruistic than $Y$ if and only if $\alpha_1^X \leq \alpha_1^Y$.*

Next, we investigate comparative statics on $\beta_1$.

**Definition**: For preferences $\succsim^X$ and $\succsim^Y$ on $\mathscr{A}$ such that $\succsim^X$ and $\succsim^Y$ coincide on $(\Delta(Z))^I$. $\succsim^X$ is *more pride-seeking (less temptation-averse)* than $\succsim^Y$, if the following is true: for any $\boldsymbol{p}, \boldsymbol{q} \in (\Delta(Z))^I$ and $r \in \Delta(Z)$ such that (i) $q_1 \succ_1^j p_1$, (ii) $\{\boldsymbol{p}, \boldsymbol{q}\} \succ^j \{\boldsymbol{q}\}$, and (iii) $p_S \succ_S^j q_S$ for each $j \in \{X, Y\}$,

$$\{\boldsymbol{p}, \boldsymbol{q}\} \succsim^Y \{(r)_{i \in I}\} \Rightarrow \{\boldsymbol{p}, \boldsymbol{q}\} \succsim^X \{(r)_{i \in I}\}.$$

As in the axiom of Pride of Acting Altruistically, the conditions (i), (ii), and (iii) mean that $\boldsymbol{p}$ is ex-post superior to $\boldsymbol{q}$; however, $\boldsymbol{q}$ is superior in the individual preference (i.e., $q_1 \succ_1^j p_1$). Hence, by not choosing the selfish allocation $\boldsymbol{q}$, an agent might feel pride in acting altruistically. On the other hand, a singleton set $\{(r)_{i \in I}\}$ would not cause any psychological effects. Therefore, if an agent (i.e., agent $Y$) prefers $\{\boldsymbol{p}, \boldsymbol{q}\}$ to a singleton set $\{(r)_{i \in I}\}$, then a more pride seeking (or a less temptation-averse) agent (i.e., agent $X$) also should prefer $\{\boldsymbol{p}, \boldsymbol{q}\}$ to $\{(r)_{i \in I}\}$.

Next, we investigate comparative statics on $\beta_S$.

**Definition**: For preferences $\succsim^X$ and $\succsim^Y$ on $\mathscr{A}$ such that $\succsim^X$ and $\succsim^Y$ coincide on $(\Delta(Z))^I$. $\succsim^X$ is *more shame-averse* than $\succsim^Y$, if the following is true: for any $\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r} \in (\Delta(Z))^I$ and $l \in \Delta(Z)$ such that (i) $q_S \succ_S^j p_S$, (ii) $r_1 \succ_1^j p_1$, and (iii) $\{\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r}\} \succ^j \{\boldsymbol{q}, \boldsymbol{r}\}$ for each $j \in \{X, Y\}$,

$$\{(l)_{i \in I}\} \succsim^Y \{\boldsymbol{p}, \boldsymbol{q}\} \Rightarrow \{(l)_{i \in I}\} \succsim^X \{\boldsymbol{p}, \boldsymbol{q}\}.$$

As in the axiom of Shame of Acting Selfishly, the conditions (i), (ii), and (iii) mean that $\boldsymbol{p}$ is ex-post superior to $\boldsymbol{q}$; however, $\boldsymbol{q}$ is socially superior to $\boldsymbol{p}$ (i.e., $q_S \succ_S^j p_S$). Hence, the

---

[15]Since $\succsim_i^X = \succsim_i^Y$ for all $i \in \{1, S\}$, there exist two GU models that share the same $u$ and $\alpha_S$.

agent would feel shame at choosing the socially inferior allocation $\boldsymbol{p}$ from $\{\boldsymbol{p}, \boldsymbol{q}\}$. Therefore, if an agent (i.e., agent $Y$) prefers a singleton set $\{(l)_{i \in I}\}$ to $\{\boldsymbol{p}, \boldsymbol{q}\}$, then a more shame averse agent (i.e., agent $X$) also should prefer $\{(l)_{i \in I}\}$ to $\{\boldsymbol{p}, \boldsymbol{q}\}$.

**Proposition 4** *Suppose $\succsim^X$ and $\succsim^Y$ are represented respectively by two GU models with $(u, \boldsymbol{\alpha}^X, \boldsymbol{\beta}^X)$ and $(u, \boldsymbol{\alpha}^Y, \boldsymbol{\beta}^Y)$ such that $\boldsymbol{\alpha}^X = \boldsymbol{\alpha}^Y$.[16] Then, the following statements hold:*
*(i) $X$ is more pride-seeking (less temptation-averse) than $Y$ if and only if $\beta_1^X \geq \beta_1^Y$.*
*(ii) $X$ is more shame-averse than $Y$ if and only if $\beta_S^X \geq \beta_S^Y$.*

# 7　Ex-Post Choice

To model ex-post choice, we consider an additional primitive: for any set $A \in \mathscr{A}$, let $C(A) \subset A$ be the non-empty set of all alternatives in $A$ that the decision maker will choose at the ex-post stage. Consider the following conditions. The first condition is the standard condition proposed by Arrow (1959).

**Axiom** (Weak Axiom of Revealed Preference (WARP)): For all $A, B \in \mathscr{A}$ and $\boldsymbol{p}, \boldsymbol{q} \in (\Delta(Z))^I$, if $\boldsymbol{p} \in C(A)$, $\boldsymbol{q} \in A$, $\boldsymbol{q} \in C(B)$, and $\boldsymbol{p} \in B$, then $\boldsymbol{p} \in C(B)$.

The second condition is also a standard continuity condition:

**Axiom** (Closed Graph): The set $\{(A, \boldsymbol{p}) | A \in \mathscr{A}, \boldsymbol{p} \in C(A)\}$ is closed in $\mathscr{A} \times (\Delta(Z))^I$.

The last condition is new, although a similar condition appears in Kopylov (2012).

**Axiom** (Consistency): For all $\boldsymbol{p} \in (\Delta(Z))^I$ and $A \in \mathscr{A}$,

$$A \succ A \setminus \{\boldsymbol{p}\} \Rightarrow C(A) = \{\boldsymbol{p}\} \text{ or } A \in \mathscr{A}_{\boldsymbol{p}}^1.$$

This condition requires that the decision maker does not prefer to remove an element $\boldsymbol{p}$ from a set $A$ only if he plans to choose $\boldsymbol{p}$ ex-post or if $\boldsymbol{p}$ is individually optimal in $A$. In the latter

---

[16]Since $\succsim^X$ and $\succsim^Y$ coincide on $(\Delta(Z))^I$, there exist two GU models that share the same $u$ and $\boldsymbol{\alpha}$.

case, he could desire to have (but not exercise) the option to choose $\boldsymbol{p}$ in order to feel pride of acting altruistically.

## 7.1 Representation

**Proposition 5** *Let $\succsim$ be represented by the GU model with $(u, \boldsymbol{\alpha}, \boldsymbol{\beta})$. Then a choice correspondence $C(\cdot)$ satisfies WARP, Closed Graph, and Consistency if and only if for all $A \in \mathscr{A}$*

$$C(A) = \arg \max_{\boldsymbol{p} \in A} (1 - \beta_1)\alpha_1 u(p_1) + (1 + \beta_S) \sum_{i \in S} \alpha_i u(p_i).$$

If $C(\cdot)$ satisfies WARP, Closed Graph, and Consistency, then we call $C(\cdot)$ an *ex-post choice correspondence associated with* $\succsim$. We define $U(\boldsymbol{p}) \equiv (1 - \beta)\alpha_1 u(p_1) + (1 + \beta_S) \sum_{i \in S} \alpha_i u(p_i)$ for all $\boldsymbol{p} \in (\Delta(Z))^I$. We call $U(\cdot)$ the *ex-post utility function.*

Note that since $\beta_1 < 1$, the function $U$ is monotonic in the decision maker's utility $u(p_1)$ and the social utility $\sum_{i \in S} \alpha_i u(p_i)$. Hence, the decision maker will never derive utility from decreasing his own utility (or social utility) unless it increases social utility (or his own utility, respectively).

## 7.2 Impure Altruism and Impure Selfishness

We say that a decision maker exhibits *impure altruism* if he acts to benefit others' allocations because of pride in acting altruistically and shame of acting selfishly. Suppose that such a decision maker, at the ex-ante private stage, is indifferent between two allocations and that one of the allocations is ranked superior to the other allocation by his social preference $\succsim_S$. Then at the ex-post public stage, the decision maker should *strictly* prefer the socially superior allocation because the pride and the shame are triggered by the publicness of the choice.

**Axiom** (Impure Altruism): For all $\boldsymbol{p}, \boldsymbol{q} \in (\Delta(Z))^I$,

$$\{\boldsymbol{p}\} \sim \{\boldsymbol{q}\} \text{ and } p_S \succ_S q_S \Rightarrow C(\{\boldsymbol{p}, \boldsymbol{q}\}) = \{\boldsymbol{p}\}.$$

Note that $C(\{\boldsymbol{p}, \boldsymbol{q}\}) = \{\boldsymbol{p}\}$ implies that $\boldsymbol{q} \notin C(\{\boldsymbol{p}, \boldsymbol{q}\})$. Hence, the decision maker strictly prefers the socially superior allocation $\boldsymbol{p}$ at the ex-post public stage, even though he is indifferent between $\boldsymbol{p}$ and $\boldsymbol{q}$ at the ex-ante private stage.

In contrast, we say that a decision maker exhibits *impure selfishness* if the temptation to act selfishly motivates him to diverge from his ex-ante choices. Suppose that such a decision maker, at the ex-ante stage, is indifferent between two allocations and that one of the allocations is ranked superior to the other allocation by his individual preference $\succsim_1$. Then, the decision maker should *strictly* prefer the selfish allocation at the ex-post stage.

**Axiom** (Impure Selfishness): For all $\boldsymbol{p}, \boldsymbol{q} \in (\Delta(Z))^I$,

$$\{\boldsymbol{p}\} \sim \{\boldsymbol{q}\} \text{ and } p_1 \succ_1 q_1 \Rightarrow C(\{\boldsymbol{p}, \boldsymbol{q}\}) = \{\boldsymbol{p}\}.$$

**Proposition 6** *Suppose that $\succsim$ is represented by the GU model with $(u, \boldsymbol{\alpha}, \boldsymbol{\beta})$ and $C(\cdot)$ is the ex-post choice correspondence associated with $\succsim$. Then, the following statements hold:*
*(i) $\succsim$ exhibits Impure Altruism if and only if $\beta_S > -\beta_1$.*
*(ii) $\succsim$ exhibits Impure Selfishness if and only if $\beta_S < -\beta_1$.*

To understand the above result intuitively, remember that in the GU model, the ex-ante choices between allocations, say, between $\boldsymbol{p}$ and $\boldsymbol{q}$, are formalized as choices between the singleton sets, such as $\{\boldsymbol{p}\}$ and $\{\boldsymbol{q}\}$. So at the ex-ante stage, the ex-ante choices maximize $V(\{\boldsymbol{p}\}) = \alpha_1 u(p_1) + \sum_{i \in S} \alpha_i u(p_i)$. On the other hand, as shown in Proposition 5, the ex-post choices maximize $U(\boldsymbol{p}) \equiv (1 - \beta_1)\alpha_1 u(p_1) + (1 + \beta_S) \sum_{i \in S} \alpha_i u(p_i)$. By comparing the two functions, we can see that the decision maker's choices become more altruistic (i.e., the relative weight on $u_S$ becomes higher) at the ex-post stage than they are at the ex-ante stage if and only if $\beta_S > -\beta_1$.

30

# 8   Consistency with Experiments

In this section, we show that the special case of the GU model with $\beta_1 = 0$ is consistent with recent experimental evidence on dictator games with an exit option. Before describing the evidence, we clarify the meaning of *public* in the experiments. We call a dictator's choice public if there is common knowledge among the subjects that they are playing a dictator game, even though a receiver does not know the identity of his paired dictator. Given the common knowledge, the dictator may consider the receiver's wish that the dictator behaves altruistically.[17] Hence, the dictator could feel pride in acting altruistically by living up to the receiver's wish or shame from acting selfishly by denying that wish. (Indeed, in Appendix D we observe that anonymity does not change major tendencies of choices in some experiments.)

In the experiments conducted by Dana et al. (2006), Broberg et al. (2007), and Lazear et al. (2012), the exit option could be costly (in other words, playing the dictator game is subsidized) but the option ensured that receivers never knew the choice of exit. So, by using the exit option, dictators could consume the whole endowment (minus the cost of exit, if any) and leave nothing to receivers–*without* feeling shame of acting selfishly.[18] In these experiments, about one-third of subjects used the exit option privately, but when the same subjects played a standard dictator game without the exit option, they donated a positive amount to the receivers.

Moreover, the most recent experiments conducted by Lazear et al. (2012) found that *both* lower-level and higher-level donors decide to play the game, while the medium-level donors exit. Hence, screening is important to implement higher-level donations.[19] Other experiments such as Dana et al. (2006) found consistent evidence for the tendency. (See Figure 3 for this tendency and Appendix D for details.)

---

[17]Given the common knowledge, the dictator knows that the receiver knows that a dictator determines the receiver's allocation. Hence, the dictator would consider the receiver's wish.

[18]An experimenter observed the choice of exit. This is consistent with our model because the experimenter is an outside observer (i.e., the subjects' choices should not affect the experimenter's welfare).

[19]Notice that the low level donations minus the cost of subsidy can be negative. So the participation of low-level donors can be costly.
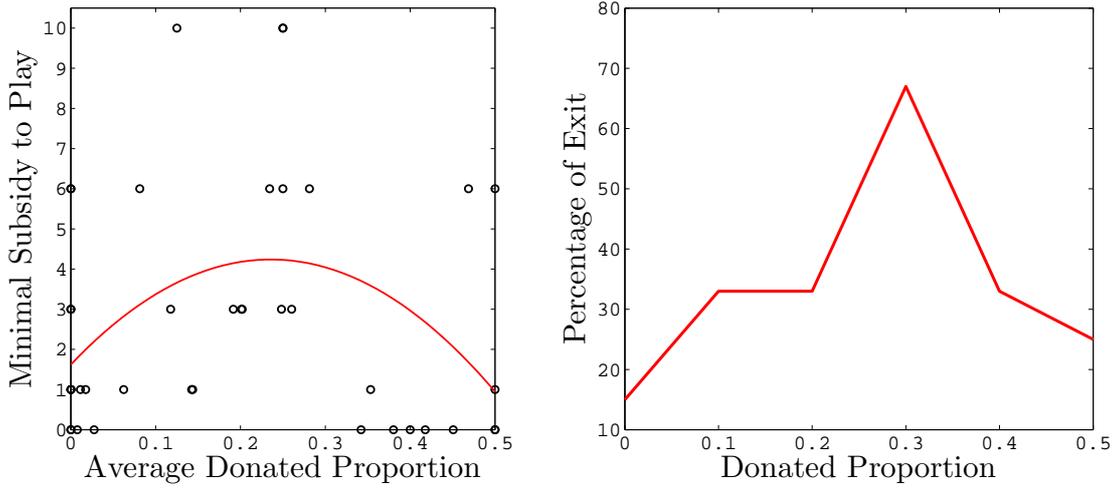
Figure 3: The relationship between donations and the choice of exit in Lazear et al. (2012) (left) and Dana et al. (2006) (right). (In the left figure, each point shows the choice of each subject. The curve is a smooth approximation of the choices. The right figure shows the percentage of subjects who existed, depending on the proportions of their donations.)

The theory of inequality aversion such as Fehr and Schmidt (1999) and Andreoni's (1989, 1990) model of warm glow are inconsistent with this robust tendency to exit, not to mention the above mentioned finding. In both theories, whether the dictator's choice is private or public does not make any difference in his utility. Moreover, by playing the dictator game, the dictator can allocate the whole endowment and the subsidy arbitrarily between himself and the receiver. Dillenberger and Sadowski (2012) can provide a sufficient condition for the choice of exit: if the value of shame (i.e., the value of the function $g$) is large enough, a subject should exit the game. In the following, by using the GU model, we will obtain a necessary and sufficient condition for the choice of exit by using the unique parameters (i.e., $\alpha$ and $\beta_S$). Moreover, we show that our model can be consistent with the finding that subjects whose donation is low or high tend to play the game, while subjects whose donation is medium tend to exit the game.

To see that the GU model is consistent with the experimental evidence, note that the singleton $\{(w,0)\}$ corresponds to exiting with endowment $w$; the set $\{(c,d) \in \mathbb{R}_+^2 | c + d \leq w + \tau\}$ corresponds to playing the dictator game publicly with total endowment $w + \tau$, where

$\tau > 0$ is the subsidy for playing the game (i.e., the cost of exiting the game). Hence, subjects exit if and only if $V(\{(w, 0)\}) \geq V(\{(c, d) \in \mathbb{R}^2_+ | c + d \leq w + \tau\})$.[20] With the GU model, we can characterize the choice of exit completely as follows:

**Proposition 7** *Suppose that $u' > 0$, $u'' < 0$, $u(0) = 0$, $u'(0) = +\infty$, $\alpha_1 > 0$, $\beta_1 = 0$, $\beta_S \geq 0$, and $\tau > 0$. Let $\underline{\beta_S} \equiv u(\tau)/(u(w + \tau) - u(\tau))$. There exist real-valued functions $\underline{\alpha_1}, \overline{\alpha_1}$ on $[\underline{\beta_S}, +\infty)$ such that*

*(i) given $\beta_S < \underline{\beta_S}$, then the decision maker plays the game;*

*(ii) given $\beta_S \geq \underline{\beta_S}$, then the decision maker exits the game if and only if $\alpha_1 \in [\underline{\alpha_1}(\beta_S), \overline{\alpha_1}(\beta_S)]$; the decision maker plays the game if and only if $\alpha_1 < \underline{\alpha_1}(\beta_S)$ or $\overline{\alpha_1}(\beta_S) < \alpha_1$.*

*Moreover, $\underline{\alpha_1}(\underline{\beta_S}) = \overline{\alpha_1}(\underline{\beta_S})$, $\underline{\alpha_1}$ is a strictly decreasing, and $\overline{\alpha_1}$ is a strictly increasing. Hence, $\underline{\alpha_1}(\beta_S) < \overline{\alpha_1}(\beta_S)$ for all $\beta_S > \underline{\beta_S}$.*
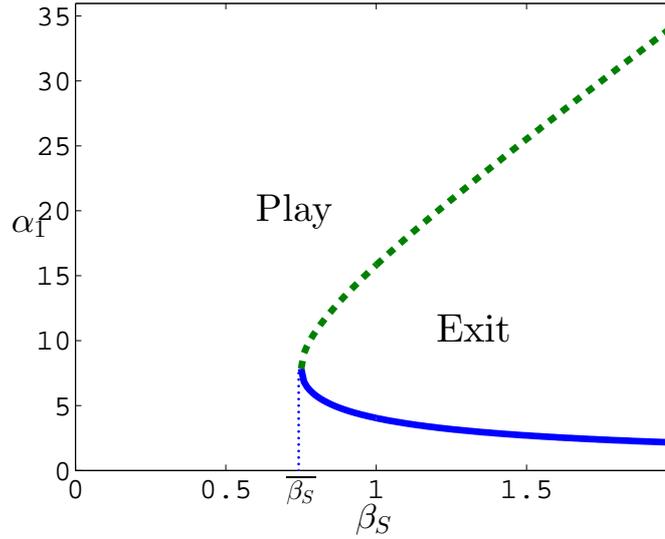


Figure 4: The optimal choice (i.e., exit or play) depending on the parameters (i.e., $\alpha_1$ and $\beta_S$). The lower solid curve corresponds to the graph of $\underline{\alpha_1}$. The upper dotted curve corresponds to the graph of $\overline{\alpha_1}$. Donation increases in the south-east direction (i.e., as $\alpha_1$ decreases and $\beta_S$ increases).

The proposition says that if $\beta_S$ is lower than the bound $\underline{\beta_S} \equiv u(\tau)/(u(w + \tau) - u(\tau))$,

---

[20]For simplicity, we assume that the decision maker exits if he is indifferent.

then any subjects play the game because the shame which they anticipate in the game is small.[21] When $\beta_S$ is larger than the bound $\underline{\beta_S}$, (ii) implies that subjects whose $\alpha_1$ is low or high tend to play the game, while subjects whose $\alpha_1$ is medium tend to exit the game. See Figure 4 for an example.[22]

In the following, we discuss how this proposition captures the finding that subjects whose donation is low or high tend to play the game, while subjects whose donation is medium tend to exit the game. First, note that, given $\beta_S$, donations increase as the index $\alpha_1$ of *selfishness* decreases. Hence, given $\beta_S$, the higher-level donors correspond to subjects whose $\alpha_1$ is low; the medium-level donors correspond to subjects whose $\alpha_1$ is medium; and the lower-level donors correspond to subjects whose $\alpha_1$ is high. So if $\beta_S$ and $\alpha_1$ are independently distributed, the proposition captures the finding that both lower-level and higher-level donors decide to play the game, while medium-level donors exit.

Given our purpose of separately capturing pure altruism and shame, there is no a prior reason to assume a specific correlation. So, independence may be a sensible assumption. Moreover, if we assume a specific correlation that $\beta_S$ is high when $\alpha_1$ is medium level, then it is easy to describe the observed finding. One may think that $\alpha_1$ and $\beta_S$ would be negatively correlated. That is, as a subject is more selfish (i.e., $\alpha_1$ is larger), he tends to be less shame-averse (i.e., $\beta_S$ is smaller). However, such a negative correlation implies that lower donors tend to play the game, which is incompatible with the observed finding. After all, how $\alpha_1$ and $\beta_S$ are distributed is an empirical question for future studies.

# 9  Application: Partial Crowding Out of Donations

In this section, we describe an empirical puzzle involving charitable donations: government spending only *partially* crowds out consumers' donations, even though standard consumer theory predicts that the crowding out should be *complete*.

---

[21]Note that the since $u' > 0$ and $u'' < 0$, the bound $\underline{\beta_S}$ increases as $\tau$ increases. This means that given $\beta_S$, as $\tau$ increases, subjects more likely to play the game.

[22]In the figure, we assume that $u(x) = -\exp(-x) + 1$, $\beta_1 = 0$, $w = 2$, and $\tau = .5$.

In his well-known paper, Andreoni (1989, 1990) has proposed a model of warm glow and obtained conditions that capture the puzzle. However, Andreoni's (1989, 1990) model is not axiomatic. Moreover, it is not easy to interpret his conditions that capture the puzzle: the conditions are imposed on derivatives of the first-order conditions of the decision maker's maximization problem. In the following, we obtain simpler conditions on $\alpha$ and $\beta$ that capture the puzzle.

We consider the following two-period decision problem. Period 1 consists of ex-ante private and ex-post public stages. At the ex-ante private stage, the decision maker divides his income $e$ between the savings $s$ for Period 2 and the budget $w$ for the ex-post public stage. At the ex-post public stage, the decision maker divides his budget $w$ between his donation $d$ and his consumption $c$. At Period 2, the decision maker consumes the savings $s$ privately. The two-period framework is adopted here for simplicity; the crucial assumption here is that the decision maker can consume some of his income privately.
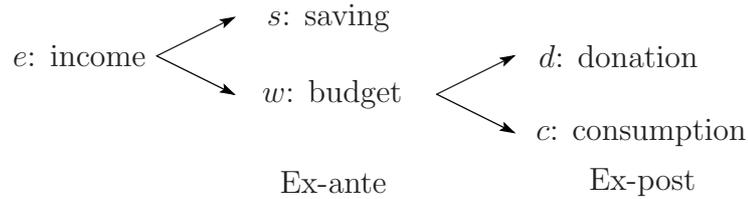
Figure 5: Decision Making in Period 1

We assume that the decision maker's preferences are separable across the two periods and that his utility at each period is represented by the GU model. Then, for any time-discount factor $\delta \in (0,1)$ and government spending $g$ and tax $\tau$, the decision maker's problem at the ex-ante stage of Period 1 is:

$$\max_{s,w} V(\{(c,d+g)|c+d \le w-\tau\}) + \delta V(\{(s,0)\}), \tag{10}$$

subject to $s + w \le e$.[23] Given the optimal budget $w^*(g,\tau)$, the decision maker's problem at

---

[23]Our results hold for any $\delta \in (0,1)$. It is easy to axiomatize this extended representation (10) by considering extended preferences $\hat{\succsim}$ over $\mathscr{A} \times (\Delta(Z))^I$. We could axiomatize this extended representation by

the ex-post stage of Period 1 is:

$$\max_{d,c} U(c, d+g), \tag{11}$$

subject to $d + c \leq w^*(g, \tau) - \tau$. We denote the *optimal donation* (i.e., the solution $d$ to (11)) by $d^*(g, \tau)$. We say that the decision maker's donation is *crowded out completely* if $d^*(g, \tau) + g$ is constant in $g$ and that his donation is *crowded out partially* if $d^*(g, \tau) + g$ is strictly increasing in $g$.

**Proposition 8** *Suppose that (a) government spending is financed by tax (i.e., $g = \tau$) and (b) $u' > 0$, $u'' < 0$, $u''' \geq 0$, $u'(0) = +\infty$, and $\alpha_1 \geq 1$. Then, the following holds:*

*(i) if the decision maker's index (i.e., $\beta_1$) for pride is larger than his index (i.e., $\beta_S$) of shame, then his donation is crowded out partially.*

*(ii) if he does not exhibit pride, shame, and temptation (i.e., $\beta_1 = 0 = \beta_S$), then his donation is crowded out completely.*

To see the result intuitively, note that $V(\{(c, d+g)|c+d \leq w - \tau\}) = (1 - \beta_1)\alpha_1 u(c^*) + (1 + \beta_S)u(d^*) + \beta_1\alpha_1(u(w - \tau) - u(c^*)) - \beta_S(u(w) - u(d^*))$, where $c^*$ and $d^*$ are the optimal consumption and the optimal donation respectively.[24] Under the assumption that $\alpha_1 \geq 1$, if $\beta_1 > \beta_S$, then $\beta_1\alpha_1 > \beta_S$. Hence, the net effect $\beta_1\alpha_1(u(w - \tau) - u(c^*)) - \beta_S(u(w) - u(d^*))$ caused by pride and shame is increasing in $w$.[25] Therefore, when $\tau$ increases, the decision maker would increase $w$ to compensate this increase of $\tau$. This increase of $w$ maintains the level of donations, which implies that the crowding out will be only partial.

---

assuming that the conditional preference of $\hat{\succsim}$ on $\mathscr{A}$ satisfies our axioms in the theorem as well as standard conditions, including the Independence axiom on $\mathscr{A} \times (\Delta(Z))^I$.

[24]In the equation, the disutility of the shame is $-\beta_S(u(w + g - \tau) - u(d^*))$, which is equal to $-\beta_S(u(w) - u(d^*))$ because $g = \tau$.

[25]By the envelope theorem, the derivative of the difference is $\beta_1\alpha_1 u'(w - \tau) - \beta_S u'(w)$. By the concavity, $u'(w - \tau) > u'(w)$. Hence, the derivative of the difference is positive because $\beta_1\alpha_1 > \beta_S$.

# A  Proof of Theorem

In this section, we prove the theorem. First, we show the sufficiency part of the theorem. Fix $\succsim$ that satisfies the axioms in the theorem. The next lemma provides representations for $\succsim_1$ and $\succsim_S$.

**Lemma 1** *There exist a von Neumann-Morgenstern utility function $u_1$ on $\Delta(Z)$ and positive numbers $\{\alpha_i\}_{i \in S}$ such that (i) $u_1$ represents $\succsim_1$ on $\Delta(Z)$, (ii) there exist $\overline{z}, \underline{z} \in Z$ such that $u_1(\overline{z}) = 1 \geq u_1(p) \geq 0 = u_1(\underline{z})$ for all $p \in \Delta(Z)$, and (iii) $u_S \equiv \sum_{i \in S} \alpha_i u_1$ represents $\succsim_S$ and $\sum_{i \in S} \alpha_i = 1$.*

**Proof of Lemma** 1: First, we show that $\succsim_1$ is well defined (i.e., if $(l_1, p_S) \succsim (r_1, p_S)$ for some $p_S \in (\Delta(Z))^S$, then $(l_1, q_S) \succsim (r_1, q_S)$ for all $q_S \in (\Delta(Z))^S$). Suppose by way of contradiction that $(l_1, p_S) \succsim (r_1, p_S)$ and $(l_1, q_S) \prec (r_1, q_S)$. By Independence, $(l_1, \frac{1}{2}p_S + \frac{1}{2}q_S) \succsim (\frac{1}{2}r_1 + \frac{1}{2}l_1, \frac{1}{2}p_S + \frac{1}{2}q_S)$ and $(l_1, \frac{1}{2}q_S + \frac{1}{2}p_S) \prec (\frac{1}{2}r_1 + \frac{1}{2}l_1, \frac{1}{2}q_S + \frac{1}{2}p_S)$. This is a contradiction. By the same way, we can show that $\succsim_S$ is well defined.

To show $\succsim_1$ satisfies Independence, fix $p_1, q_1, l_1 \in \Delta(Z)$ and $\alpha \in [0,1]$. Then, for any $p_S, q_S \in \Delta(Z)$, $p_1 \succsim_1 q_1 \Leftrightarrow (p_1, p_S) \succsim (q_1, p_S) \Leftrightarrow \alpha(p_1, p_S) + (1-\alpha)(l_1, q_S) \succsim \alpha(q_1, p_S) + (1-\alpha)(l_1, q_S) \Leftrightarrow \alpha p_1 + (1-\alpha)l_1 \succsim_1 \alpha q_1 + (1-\alpha)l_1$. By the same way, we can show that $\succsim_S$ satisfies Independence. Then, by the standard argument with von Neumann-Morgenstern's Theorem, there exist mixture-linear utility functions $u_1$ and $\hat{u}_S$, which are unique up to positive affine transformation. Then, (ii) follows from the finiteness of $Z$ and a normalization. Finally, by Consistency, it follows from Fishburn (1984) that there exist a positive numbers $\{\alpha_i\}_{i \in S}$ and a real number $\gamma$ such that $\hat{u}_S(p_S) = \sum_{i \in S} \alpha_i u_1(p_i) + \gamma$ and $\sum_{i \in S} \alpha_i = 1$. By defining $u_S = \hat{u}_S - \gamma$, we obtain (iii). ∎

The next lemma shows a stronger version of Dominance.

**Lemma 2** *For any set $A, B \in \mathscr{A}$, if for any $\boldsymbol{p} \in B$, there exists $\boldsymbol{q} \in A$ such that $q_1 \succsim_1 p_1$ and $q_S \succsim_S p_S$, then $A \sim A \cup B$.*

**Proof of Lemma** 2: Fix $A, B \in \mathscr{A}$. Since $\Delta(Z)$ is separable, there exists a countable set $B' = \{\boldsymbol{p}^1, \boldsymbol{p}^2, \boldsymbol{p}^3, \dots\}$ such that $\mathrm{cl}.B' = B$. Let $B_0 = \emptyset$ and $B_n = \{\boldsymbol{p}^1, \dots, \boldsymbol{p}^n\}$, so that $B_n \subset B_{n+1}$ for all $n$ and $B' = \cup_{n=1}^\infty B_n$. Then, $B_n \to \mathrm{cl}.B' = B$ in the Hausdorff metric topology.[26] Since for any $\boldsymbol{p}^n$, there exists $\boldsymbol{q} \in A$ such that $q_1 \succsim_1 p_1^n$ and $q_S \succsim_S p_S^n$. Then, by Dominance, $A \cup B_n \equiv (A \cup B_{n-1}) \cup \boldsymbol{p}^n \sim A \cup B_{n-1}$. By transitivity, $A \cup B_n \sim A$ for all $n \in \mathbb{N}$. By continuity, hence, $A \cup B \sim A$. $\blacksquare$

For all $A \in \mathscr{A}$, define $\boldsymbol{u}(A) = \{(u_1(p_1), \sum_{i \in S} \alpha_i u(p_i)) \in \mathbb{R}^2 | \boldsymbol{p} \equiv (p_1, p_S) \in A\}$ and $\mathscr{A}^* = \{\boldsymbol{u}(A) | A \in \mathscr{A}\}$. Since $A \in \mathscr{A}$ is closed, $\boldsymbol{u}(A)$ is also closed by the continuity of $u_1$ and $u_S$. Since $u_1(\Delta(Z)) = [0, 1]$ and $u_S((\Delta(Z))^S) = [0, 1]$, $\mathscr{A}^*$ is a set of compact subsets of $[0, 1]^2$. We endow $\mathscr{A}^*$ with the Hausdorff metric $d_h(A^*, B^*) = \max\{\max_{\boldsymbol{u} \in A^*} \min_{\boldsymbol{v} \in B^*} d(\boldsymbol{u}, \boldsymbol{v}), \max_{\boldsymbol{u} \in B^*} \min_{\boldsymbol{v} \in A^*} d(\boldsymbol{u}, \boldsymbol{v})\}$, where $d$ is the Euclidean metric. We define a mixture on $\mathscr{A}^*$ as follows: for all $A^*, B^* \in \mathscr{A}^*$ and $\alpha \in [0, 1]$, $\alpha A^* + (1 - \alpha)B^* = \{\boldsymbol{u} \in \mathbb{R}^2 | \boldsymbol{u} = \alpha \boldsymbol{v} + (1 - \alpha)\boldsymbol{w} \text{ for some } \boldsymbol{v} \in A^*, \boldsymbol{w} \in B^*\}$. Define $\succsim^*$ on $\mathscr{A}^*$ as follows:

$$A^* \succsim^* B^* \text{ if } A \succsim B,$$

where $\boldsymbol{u}(A) = A^*$ and $\boldsymbol{u}(B) = B^*$. Define $\succ^*$ and $\sim^*$ as the asymmetric and symmetric parts of $\succsim^*$, respectively.

Lemma 2 shows that $\succsim^*$ is well defined in the following sense:

**Lemma 3** *If $A^* = B^*$, then $A \sim B$.*

**Proof of Lemma** 3: If $\boldsymbol{u}(A) = \boldsymbol{u}(B)$, then for any $\boldsymbol{p} \in A$ there exists $\boldsymbol{q} \in B$ such that $q_1 \sim_1 p_1$ and $q_S \sim_S p_S$. Therefore, Lemma 2 shows that $A \sim A \cup B$. In the same way, we can show $B \sim A \cup B$. Hence, $A \sim B$. $\blacksquare$

Next lemma shows that $\succsim^*$ satisfies the following axioms.

**Axiom** (Independence*): For all $\alpha \in [0, 1]$ and $A^*, B^*, C^* \in \mathscr{A}^*$, $A^* \succsim^* B^*$ if and only if $\alpha A^* + (1 - \alpha)C^* \succsim^* \alpha B^* + (1 - \alpha)C^*$.

---

[26]$\mathrm{cl}.A$ denotes the closure of $A$ for all $A \in \mathscr{A}$.

**Axiom:** (Pareto*) For any $\boldsymbol{u}, \boldsymbol{v} \in [0,1]^2$, if $u_1 \geq v_1$ and $u_S \geq v_S$, then $\boldsymbol{u} \succsim^* \boldsymbol{v}$. Moreover if $u_1 > v_1$ or $u_S > v_S$, then $\boldsymbol{u} \succ^* \boldsymbol{v}$.

**Axiom:** (Dominance*) For any $A \in \mathscr{A}$ and $\boldsymbol{u} \in [0,1]^2$, if there exists $\boldsymbol{v} \in A^*$ such that $v_1 \geq u_1$ and $v_S \geq u_S$, then $A^* \sim^* A^* \cup \{\boldsymbol{u}\}$.

**Axiom:** (Weak Strategic Rationality*) If there exist $\boldsymbol{u}, \boldsymbol{v} \in A^*$ and $\boldsymbol{u}', \boldsymbol{v}' \in B^*$ such that for all $\boldsymbol{w} \in A^* \cup B^*$, $u_1 = u_1' \geq w_1$ and $v_S = v_S' \geq w_S$, then $A^* \succsim^* B^* \Rightarrow A^* \cup B^* \sim^* A^*$.

**Axiom:** (Shame of Acting Selfishly*) If (i) $v_S > u_S$, (ii) $w_1 > u_1$, and (iii) $\{\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}\} \succ^* \{\boldsymbol{v}, \boldsymbol{w}\}$, then $\{\boldsymbol{u}\} \succsim^* \{\boldsymbol{u}, \boldsymbol{v}\}$.

**Lemma 4** $\succsim^*$ *is a complete, transitive, and continuous binary relation that satisfies Independence\*, Pareto\*, Dominance\*, Weak Strategic Rationality\*, and Shame of Acting Selfishly\*.*

**Proof of Lemma** 4: Remember that $\succsim$ is a complete and transitive binary relation that satisfies Independence, Pareto, Dominance, Weak Strategic Rationality, and Shame of Acting Selfishly. Hence, by the definition of $\succsim^*$, $\succsim^*$ also satisfies those properties. In the following, we show that $\succsim^*$ is continuous. Choose any $A^* \in \mathscr{A}^*$ to show $\{B^* \in \mathscr{A}^* | B^* \succsim^* A^*\}$ and $\{B^* \in \mathscr{A}^* | A^* \succsim B^*\}$ are closed.

Let $\{B_n^*\}$ be a sequence such that $B_n^* \succsim^* A^*$ and $B_n^* \to B^*$ to show $B^* \succsim^* A^*$. By definition, there exists a sequence $\{B_n\}$ such that $\boldsymbol{u}(B_n) = B_n^*$ and $B_n \succsim A$. Since $B_n \in \mathscr{A}$ and $\mathscr{A}$ is compact, there exists a convergent subsequence $\{B_k'\}$ such that $B_k' \to B'$. Since $B_k' \succsim A$ for all $k$, then the continuity of $\succsim$ shows that $B' \succsim A$. Since $\boldsymbol{u}$ is continuous, then $\boldsymbol{u}(B_k') \to \boldsymbol{u}(B')$. Since $\{\boldsymbol{u}(B_k')\}$ is a subsequence of $\{B_n^*\}$ and $B_n^* \to B^*$, it must hold that $\boldsymbol{u}(B') = B^*$. Since $B' \succsim A$, we obtain $B^* \succsim^* A^*$. In the same way, we can show that $\{B^* \in \mathscr{A}^* | A^* \succsim^* B^*\}$ is closed. ∎

Now, we show a general representation for $\succsim^*$.

**Lemma 5** *There exists a function $\mu : \mathbb{R} \to \mathbb{R}$ such that $V^*(A^*) = \sum_{\lambda \in \mathbb{R}} \left( \max_{\boldsymbol{u} \in A^*} \lambda u_1 + (1 - \lambda) u_S \right) \mu(\lambda)$ represents $\succsim^*$. Moreover, $supp(\mu) \equiv \{\lambda \in \mathbb{R} | \mu(\lambda) \neq 0\}$ is finite.*

39

**Proof of Lemma** 5:

**Step** 1: (i) There exist finite sets $K$ and $L$ of functions on $[0,1]^2$ such that $\succsim^*$ is represented by $V^*(A^*) = \sum_{U \in K} \max_{\boldsymbol{u} \in A^*} U(\boldsymbol{u}) - \sum_{U \in L} \max_{\boldsymbol{u} \in A^*} U(\boldsymbol{u})$, (ii) For any $U \in K \cup L$, $\lambda \in [0,1]$, and $\boldsymbol{u}, \boldsymbol{v} \in [0,1]^2$, $U(\lambda \boldsymbol{u} + (1-\lambda)\boldsymbol{v}) = \lambda U(\boldsymbol{u}) + (1-\lambda)U(\boldsymbol{v})$, (iii) For any $U \in K \cup L$, there exist no $U' \in K \cup L \setminus \{U\}$, $a > 0$, and $b \in \mathbb{R}$ such that $U = aU' + b$.

**Proof of Step** 1: By Lemma 4, $\succsim^*$ is a continuous, complete, and transitive relation that satisfies Independence*. To show Step 1, by Kopylov (2009, Theorem 2.1), it suffices to show that $\succsim^*$ satisfies his Finiteness axiom, that is, for any sequence $\{A_n^*\}$ of $\mathscr{A}^*$ there exists a positive integer $N$ such that $\bigcup_{n=1}^{N} A_n^* \sim^* \bigcup_{n=1}^{N+1} A_n^*$.[27] To show that $\succsim^*$ satisfies this axiom, choose any $A_1^*, A_2^*, A_3^*, A_4^* \in \mathscr{A}^*$. Let $\boldsymbol{u}^* = \arg\max_{\boldsymbol{u} \in A_1^* \cup A_2^* \cup A_3^* \cup A_4^*} u_1$ and $\boldsymbol{v}^* = \arg\max_{\boldsymbol{v} \in A_1^* \cup A_2^* \cup A_3^* \cup A_4^*} v_S$. (Such $\boldsymbol{u}^*$ and $\boldsymbol{v}^*$ exist because each $A_i^*$ is compact.) Without loss of generality, assume $\boldsymbol{u}^* \in A_1^*$, $\boldsymbol{v}^* \in A_2^*$, and $A_1^* \cup A_2^* \cup A_3^* \succsim^* A_1^* \cup A_2^* \cup A_4^*$. Since $\boldsymbol{u}^* \in A_1^*$ and $\boldsymbol{v}^* \in A_2^*$, Weak Strategic Rationality* shows that $A_1^* \cup A_2^* \cup A_3^* \cup A_4^* \sim^* A_1^* \cup A_2^* \cup A_3^*$. So Finiteness is satisfied with $N = 3$. $\qquad\square$

Normalize each $U \in K \cup L$ by adding a constant number so as to obtain $U(0,0) = 0$.

**Step** 2: For all $U \in K \cup L$, there exists $(a_1(U), a_S(U)) \in \mathbb{R}^2$ such that $U(\boldsymbol{u}) = a_1(U)u_1 + a_S(U)u_S$. Moreover, for any $U, U' \in K \cup L$ such that $U \neq U'$, $\frac{a_1(U)}{a_1(U)+a_S(U)} \neq \frac{a_1(U')}{a_1(U')+a_S(U')}$.

**Proof of Step** 2: For all $U \in K \cup L$, define $a_1(U) = U(1,0)$ and $a_S(U) = U(0,1)$. Fix $\boldsymbol{u} \equiv (u_1, u_S) \in [0,1]^2$. Consider the case where $u_1 + u_S \geq 1$. Then, $\frac{1}{u_1+u_S}U(u_1, u_S) + (1 - \frac{1}{u_1+u_S})U(0,0) = U\left(\frac{u_1}{u_1+u_S}, \frac{u_S}{u_1+u_S}\right) = \frac{u_1}{u_1+u_S}U(1,0) + \frac{u_S}{u_1+u_S}U(0,1) = \frac{1}{u_1+u_S}(a_1(U)u_1 + a_S(U)u_S)$. Since $U(0,0) = 0$, then $U(\boldsymbol{u}) = a_1(U)u_1 + a_S(U)u_S$. The other case where $u_1 + u_S \leq 1$ can be proved in the same way. Step 1 (iii) shows that for any $U, U' \in K \cup L$ such that $U \neq U'$, $\frac{a_1(U)}{a_1(U)+a_S(U)} \neq \frac{a_1(U')}{a_1(U')+a_S(U')}$. $\qquad\square$

---

[27]Kopylov (2009, Theorem 2.1) shows that Step 1 holds if and only if $\succsim^*$ is a continuous, complete, and transitive relation that satisfies Independence and Finiteness.

For all $\lambda \in \mathbb{R}$ define

$$
\mu(\lambda) = \begin{cases}
a_1 + a_S & \text{if } \lambda = \frac{a_1(U)}{a_1(U)+a_S(U)} \text{ for some } U \in K, \\
-(a_1 + a_S) & \text{if } \lambda = \frac{a_1(U)}{a_1(U)+a_S(U)} \text{ for some } U \in L, \\
0 & \text{otherwise.}
\end{cases}
$$

Note that $\mu$ is well defined because for any $U, U' \in K \cup L$ such that $U \neq U'$, $\frac{a_1(U)}{a_1(U)+a_S(U)} \neq \frac{a_1(U')}{a_1(U')+a_S(U')}$. Therefore, by Step 1 and Step 2, we establish Lemma 5. Moreover, since $K$ and $L$ are finite, $\mathrm{supp}(\mu)$ is finite. ∎

Fix $\varepsilon < \frac{1}{2}$. For all $\lambda \in \mathbb{R}$, define

$$
u_1^*(\lambda) = \frac{1}{2} + \frac{\varepsilon \lambda}{\|(\lambda, 1-\lambda)\|}, \quad u_S^*(\lambda) = \frac{1}{2} + \frac{\varepsilon(1-\lambda)}{\|(\lambda, 1-\lambda)\|}, \quad \text{and } \boldsymbol{u}^*(\lambda) = (u_1^*(\lambda), u_S^*(\lambda)).
$$

Then $\boldsymbol{u}^*(\lambda)$ is on the boundary of $B_\varepsilon(\frac{1}{2}, \frac{1}{2})$ (i.e., the closed ball with radius $\varepsilon$ centered at $(\frac{1}{2}, \frac{1}{2})$). The next lemma is useful to characterize $V$.

**Lemma 6** *(i) For all $\lambda, \lambda' \in \mathbb{R}$ such that $\lambda \neq \lambda'$, $\lambda u_1^*(\lambda) + (1 - \lambda)u_S^*(\lambda) > \lambda u_1^*(\lambda') + (1 - \lambda)u_S^*(\lambda')$. (ii) For all $\lambda \in [0, 1]$, $u_1^*(1) \geq u_1^*(\lambda)$ and $u_S^*(0) \geq u_S^*(\lambda)$. (iii) If $\lambda \notin [0, 1]$, $u_1^*(\frac{|\lambda|}{|\lambda|+|1-\lambda|}) \geq u_1^*(\lambda)$ and $u_S^*(\frac{|\lambda|}{|\lambda|+|1-\lambda|}) \geq u_S^*(\lambda)$.*

**Proof of Lemma** 6: To show (i), choose any $\boldsymbol{u} \in B_\varepsilon(\frac{1}{2}, \frac{1}{2})$ and any $\lambda \in \mathbb{R}$. Then, $\lambda u_1^*(\lambda) + (1 - \lambda)u_S^*(\lambda) = \frac{1}{2} + \varepsilon \|(\lambda, 1-\lambda)\| \geq \frac{1}{2} + \|\boldsymbol{u} - (\frac{1}{2}, \frac{1}{2})\| \|(\lambda, 1-\lambda)\| \geq \frac{1}{2} + |(\boldsymbol{u} - (\frac{1}{2}, \frac{1}{2})) \cdot (\lambda, 1-\lambda)| = |\lambda u_1 + (1-\lambda)u_S| \geq \lambda u_1 + (1-\lambda)u_S$, where the first inequality holds because $\boldsymbol{u} \in B_\varepsilon(\frac{1}{2}, \frac{1}{2})$ and the second inequality holds by Cauchy-Scharz's inequality. The two inequalities hold with equalities if and only if $\boldsymbol{u} - (\frac{1}{2}, \frac{1}{2}) = \frac{\varepsilon}{\|(\lambda, 1-\lambda)\|}(\lambda, 1-\lambda)$, or $\boldsymbol{u} = \boldsymbol{u}^*(\lambda)$. Since $u^*(\lambda) = u^*(\mu)$ if and only if $\lambda = \mu$, (i) holds. (ii) and (iii) follow from direct calculations.[28] ∎

---

[28] (ii) holds because $u_1^*(1) = \frac{1}{2} + \varepsilon \geq u_1^*(\lambda)$ and $u_S^*(0) = \frac{1}{2} + \varepsilon \geq u_S^*(\lambda)$ for all $\lambda \in \mathbb{R}$. To see (iii) holds, consider the case where $\lambda < 0$. Then $\frac{|\lambda|}{|\lambda|+|1-\lambda|} \in [0, 1]$. Hence, $u_1(\frac{|\lambda|}{|\lambda|+|1-\lambda|}) \geq \frac{1}{2} > u_1(\lambda)$. Moreover, by a direct calculation, $u_S(\frac{|\lambda|}{|\lambda|+|1-\lambda|}) = \frac{1}{2} + \frac{\varepsilon|1-\lambda|}{\|(\lambda, 1-\lambda)\|} = u_S(\lambda)$ because $1 - \lambda > 0$. The other case where $\lambda > 1$ can be proved in the same way.

**Lemma 7** *There exists $\lambda^* \in (0,1)$ such that $\mu(\lambda^*) > 0$ and $V^*(A^*) = \mu(1) \max_{\boldsymbol{u} \in A^*} u_1 + \mu(\lambda^*) \max_{\boldsymbol{u} \in A^*}(\lambda^* u_1 + (1 - \lambda^*)u_S) + \mu(0) \max_{\boldsymbol{u} \in A^*} u_S.$*

**Proof of Lemma** 7: Given Lemma 5, it suffices to show the following: (i) For any $\lambda \notin [0,1]$, then $\mu(\lambda) = 0$; (ii) For any $\lambda \notin \{0,1\}$, $\mu(\lambda) \geq 0$; (iii) There exists unique $\lambda^* \in (0,1)$ such that $\mu(\lambda^*) > 0$.

To show (i), suppose by way of contradiction that there exists $\lambda' \notin [0,1]$ such that $\mu(\lambda') \neq 0$. Define $A^* = \{\boldsymbol{u}^*(\lambda) \in \mathbb{R}^2 | \lambda \in \{0, 1, \frac{|\lambda'|}{|\lambda'| + |1 - \lambda'|}\} \cup \mathrm{supp}(\mu) \setminus \{\lambda'\}\}$. Since $\mathrm{supp}(\mu)$ is finite, both $A^*$ and $A^* \cup \boldsymbol{u}^*(\lambda')$ are closed. Therefore, $V^*(A^*)$ and $V^*(A^* \cup \boldsymbol{u}^*(\lambda'))$ are well defined. Since $\boldsymbol{u}^*(\frac{|\lambda'|}{|\lambda'| + |1 - \lambda'|}) \in A^*$, Lemma 6 (iii) and Dominance* show that $A^* \sim^* A^* \cup \boldsymbol{u}^*(\lambda')$, so that $V^*(A^*) = V^*(A^* \cup \boldsymbol{u}^*(\lambda'))$.

However, for all $\lambda \in \{0, 1, \frac{|\lambda'|}{|\lambda'| + |1 - \lambda'|}\} \cup \mathrm{supp}(\mu) \setminus \{\lambda'\}$, Lemma 6 (i) shows that $\max_{\boldsymbol{u} \in A^* \cup \boldsymbol{u}^*(\lambda')} \lambda u_1 + (1 - \lambda)u_S = \lambda u_1^*(\lambda) + (1 - \lambda)u_S^*(\lambda) = \max_{\boldsymbol{u} \in A^*} \lambda u_1 + (1 - \lambda)u_S$. Moreover, by (i) again, $\max_{\boldsymbol{u} \in A^* \cup \boldsymbol{u}^*(\lambda')} \lambda' u_1 + (1 - \lambda')u_S = \lambda' u_1^*(\lambda') + (1 - \lambda')u_S^*(\lambda') > \max_{\boldsymbol{u} \in A^*} \lambda_1' u_1 + (1 - \lambda')u_S$. Therefore, $V^*(A^* \cup \boldsymbol{u}^*(\lambda')) - V^*(A^*) = (\lambda' u_1^*(\lambda') + (1 - \lambda')u_S^*(\lambda') - \max_{\boldsymbol{u} \in A^*} \lambda_1' u_1 + (1 - \lambda')u_S)\mu(\lambda') \neq 0$ because $\mu(\lambda') \neq 0$. This is a contradiction. Hence, (i) holds.

To show (ii), suppose by way of contradiction that there exists $\xi \notin \{0,1\}$ such that $\mu(\xi) < 0$. Define $A^* = \{\boldsymbol{u}^*(\lambda) | \lambda \in \{0,1\} \cup \mathrm{supp}(\mu) \setminus \{\xi\}\}$. Since $\mathrm{supp}(\mu)$ is finite, $A^*$ is closed. Hence, $A^* \in \mathscr{A}^*$. By Lemma 6 (i), $V(A^* \cup \boldsymbol{u}^*(\xi)) - V(A^*) = (\max_{\boldsymbol{u} \in A^* \cup \boldsymbol{u}^*(\xi)}(\xi u_1 + (1 - \xi)u_S) - \max_{\boldsymbol{u} \in A^*}(\xi u_1 + (1 - \xi)u_S))\mu(\xi) < 0$. Hence, $A^* \succ^* A^* \cup \boldsymbol{u}^*(\xi)$. Since $\boldsymbol{u}^*(1), \boldsymbol{u}^*(0) \in A^*$, it follows from Lemma 6 (ii) that $u_1^*(1) \geq w_1$ and $u_S^*(0) \geq w_S$ for all $\boldsymbol{w} \in A^* \cup \boldsymbol{u}^*(\xi)$. By Weak Strategic Rationality*, we obtain $A^* \sim^* A^* \cup A^* \cup \boldsymbol{u}^*(\xi) = A^* \cup \boldsymbol{u}^*(\xi)$, which is a contradiction.

To show (iii), first we show that there exists at most one $\lambda \notin \{0,1\}$ such that $\mu(\lambda) > 0$. Suppose by way of contradiction that there exist distinct elements $\xi, \eta \notin \{0,1\}$ such that $\mu(\xi) > 0$ and $\mu(\eta) > 0$. Define $A^* = \{\boldsymbol{u}^*(\lambda) | \lambda \in \{0,1\} \cup \mathrm{supp}(\mu) \setminus \{\xi\}\}$ and $B^* = \{\boldsymbol{u}^*(\lambda) | \lambda \in \{0,1\} \cup \mathrm{supp}(\mu) \setminus \{\eta\}\}$. Since $\mathrm{supp}(\mu)$ is finite, $A^*$ and $B^*$ are closed. Hence, $A^*, B^* \in \mathscr{A}^*$. Note that $\boldsymbol{u}^*(1) \in A^* \cap B^*$ and $\boldsymbol{u}^*(0) \in A^* \cap B^*$. Hence, Lemma 6 (ii) and Weak Strategic

Rationality* imply that if $A^* \succsim^* B^*$, then $A^* \cup B^* \sim^* A^*$; and that if $B^* \succsim^* A^*$, then $A^* \cup B^* \sim^* B^*$. Since $\succsim^*$ is a weak order, either $A^* \cup B^* \sim^* A^*$ or $A^* \cup B^* \sim^* B^*$ holds. However, by Lemma 6 (i), $V(A^* \cup B^*) - V(B^*) = \left(\max_{\boldsymbol{u} \in A^* \cup B^*} \eta u_1 + (1 - \eta)u_S - \max_{\boldsymbol{u} \in A^*}(\eta u_1 + (1 - \eta)u_S)\right)\mu(\eta) > 0$ and $V(A^* \cup B^*) - V(A^*) = \left(\max_{\boldsymbol{u} \in A^* \cup B^*} \xi u_1 + (1 - \xi)u_S - \max_{\boldsymbol{u} \in B^*}(\xi u_1 + (1 - \xi)u_S)\right)\mu(\xi) > 0$. This is a contradiction.

Next we show that there exists at least one $\lambda \notin \{0, 1\}$ such that $\mu(\lambda) > 0$. Suppose not. Then, by (i) and (ii), $\mu(\lambda) = 0$ for all $\lambda \notin \{0, 1\}$. Therefore, $V^*(\boldsymbol{u}) = \mu(1)u_1 + \mu(0)u_S$ for all $u \in [0, 1]^2$. Since $\succsim$ satisfies Pareto*, it must hold that $\mu(1) > 0$ and $\mu(0) > 0$. However, this representation contradicts with Shame of Acting Selfishly*. To see this note that if (i) $v_S > u_S$, (ii) $w_1 > u_1$, and (iii) $\{\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}\} \succ^* \{\boldsymbol{v}, \boldsymbol{w}\}$, then $V^*(\{\boldsymbol{u}\}) = \mu(1)u_1 + \mu(0)u_S < \mu(1)\max\{u_1, v_1\} + \mu(0)v_S = V^*(\{\boldsymbol{u}, \boldsymbol{v}\})$. ∎

**Lemma 8** $\mu(0) \leq 0$.

**Proof of Lemma** 8: Choose $\boldsymbol{v}, \boldsymbol{w} \in [0, 1]^2$ such that $w_1 > v_1$ and $v_S > w_S$. Let $\boldsymbol{u} \in [0, 1]^2$ be such that $w_1 > u_1 > v_1$ and $v_S > u_S > w_S$. By making $u_1$ and $u_S$ close enough to $w_1$ and $v_S$ respectively, we can obtain $\lambda^* u_1 + (1 - \lambda^*)u_S > \max\{\lambda^* v_1 + (1 - \lambda^*)v_S, \lambda^* w_1 + (1 - \lambda^*)w_S\}$ because $\lambda^* \in (0, 1)$. Since $\mu(\lambda^*) > 0$, it follows that $V^*(\{\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}\}) > V^*(\{\boldsymbol{v}, \boldsymbol{w}\})$. Therefore, Shame of Acting Selfishly* requires that $V^*(\{\boldsymbol{u}\}) \geq V^*(\{\boldsymbol{u}, \boldsymbol{v}\})$. Hence, $0 \leq V^*(\{\boldsymbol{u}\}) - V^*(\{\boldsymbol{u}, \boldsymbol{v}\}) = \mu(0)[u_S - v_S]$. Since $v_S > u_S$, we obtain $\mu(0) \leq 0$. ∎

By Lemma 7, for all $\boldsymbol{u} \in [0, 1]^2$, $V^*(\boldsymbol{u}) = (\mu(\lambda^*)\lambda^* + \mu(1))u_1 + (\mu(\lambda^*)(1 - \lambda^*) + \mu(0))u_S$. Normalize $\mu$ so as to hold $\mu(\lambda^*)(1 - \lambda^*) + \mu(0) = 1$. By Pareto*, $\mu(\lambda^*)\lambda^* + \mu(1) > 0$. Define

$$\alpha_1 = \mu(\lambda^*)\lambda^* + \mu(1), \ \beta_1 = \frac{\mu(1)}{\mu(\lambda^*)\lambda^* + \mu(1)}, \ \text{and} \ \beta_S = -\mu(0).$$

Then, $\alpha_1 > 0$. Since $\lambda^* \in (0, 1)$ and $\mu(\lambda^*) > 0$, we obtain $\beta_1 < 1$. Since $\mu(0) \leq 0$, we obtain, $\beta_S \geq 0$. By the normalization and the definitions, $\mu(1) = \beta_1 \alpha_1$, $\mu(\lambda^*)\lambda^* = (\mu(\lambda^*)\lambda^* + \mu(1))\left(1 - \frac{\mu(1)}{\mu(\lambda^*)\lambda^* + \mu(1)}\right) = \alpha_1(1 - \beta_1)$, and $\mu(\lambda^*)(1 - \lambda^*) = 1 - \mu(0) = 1 + \beta_S$. Hence, by Lemma 7, we obtain $V^*(A^*) = \max_{\boldsymbol{u} \in A^*}\left((1 - \beta_1)\alpha_1 u_1 + (1 + \beta_S)u_S\right) + \beta_1 \alpha_1 \max_{\boldsymbol{u} \in A^*} u_1 - \beta_S \max_{\boldsymbol{u} \in A^*} u_S$.

43

For all $A \in \mathscr{A}$, define $V(A) = V^*(A^*)$. Then, $A \succsim B \Leftrightarrow A^* \succsim^* B^* \Leftrightarrow V^*(A^*) \geq V^*(B^*) \Leftrightarrow V(A) \geq V(B)$. Hence, $V$ represents $\succsim$. By arranging the terms and substituting $u_1 = u(p_1)$ and $u_S = \sum_{i \in S} \alpha_i u(p_i)$, we obtain the GU model.

Finally, we show the necessity. We show that the representation satisfies Weak Strategic Rationality and Shame of Acting Selfishly. It is easy to see the necessity of the other axioms.

To show Weak Strategic Rationality, consider $\boldsymbol{p}, \boldsymbol{q} \in (\Delta(Z))^I$ and $A, B \in \mathscr{A}_{\boldsymbol{p}}^1 \cap \mathscr{A}_{\boldsymbol{q}}^S$ such that $A \succsim B$. Then, $0 \leq V(A) - V(B) = \max_{\boldsymbol{p}' \in A} U(\boldsymbol{p}') - \max_{\boldsymbol{p}' \in B} U(\boldsymbol{p}')$. Hence, $\max_{\boldsymbol{p}' \in A} U(\boldsymbol{p}') \geq \max_{\boldsymbol{p}' \in B} U(\boldsymbol{p}')$, so that $\max_{\boldsymbol{p}' \in A \cup B} U(\boldsymbol{p}') = \max_{\boldsymbol{p}' \in A} U(\boldsymbol{p}')$. It follows that $V(A \cup B) - V(A) = \max_{\boldsymbol{p}' \in A \cup B} U(\boldsymbol{p}') - \max_{\boldsymbol{p}' \in A} U(\boldsymbol{p}') = 0$.

To show Shame of Acting Selfishly, suppose $q_S \succ_S p_S$, $r_1 \succ_1 p_1$, and $\{\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r}\} \succ \{\boldsymbol{q}, \boldsymbol{r}\}$. Then, $0 < V(\{\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r}\}) - V(\{\boldsymbol{q}, \boldsymbol{r}\}) = \max_{\boldsymbol{p}' \in \{\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r}\}} U(\boldsymbol{p}') - \max_{\boldsymbol{p}' \in \{\boldsymbol{q}, \boldsymbol{r}\}} U(\boldsymbol{p}')$. It follows that $U(\boldsymbol{p}) > U(\boldsymbol{q})$. Since $q_S \succ_S p_S$, we have $p_1 \succ_1 q_1$. Hence, $V(\{\boldsymbol{p}\}) - V(\{\boldsymbol{p}, \boldsymbol{q}\}) = \beta_S[\sum_{i \in S} \alpha_i u(q_i) - \sum_{i \in S} \alpha_i u(p_i)] \geq 0$, so that $\{\boldsymbol{p}\} \succsim \{\boldsymbol{p}, \boldsymbol{q}\}$.

# B  Proof of Corollaries and Propositions

In the following, for any $p \in \Delta(Z)$, we denote by $(p)_{i \in S}$ an allocation over the other agents (i.e., over the set $S$) that yields $p$ for each $i \in S$.

**Proof of Corollary** 1: Note that Weak Strategic Rationality I implies Weak Strategic Rationality. So all of the axioms in Theorem are satisfied. Hence, we can show Lemma 1–8 in the proof of Theorem. So, there exists $\lambda^* \in (0, 1)$ such that $\mu(\lambda^*) > 0$ and $V^*(A^*) = \mu(1) \max_{\boldsymbol{u} \in A^*} u_1 + \mu(\lambda^*) \max_{\boldsymbol{u} \in A^*} (\lambda^* u_1 + (1 - \lambda^*) u_S) + \mu(0) \max_{\boldsymbol{u} \in A^*} u_S$ and $\mu(0) \leq 0$. Finally, we show that $\mu(1) = 0$. To see this note that choose $\boldsymbol{v}, \boldsymbol{w} \in [0, 1]^2$ such that $w_1 > v_1$ and $v_S > w_S$. Let $\boldsymbol{u} \in [0, 1]^2$ be such that $w_1 > u_1 > v_1$ and $v_S > u_S > w_S$. By making $u_1$ and $u_S$ close enough to $w_1$ and $v_S$ respectively, we can obtain $\lambda^* u_1 + (1 - \lambda^*) u_S > \max\{\lambda^* v_1 + (1 - \lambda^*) v_S, \lambda^* w_1 + (1 - \lambda^*) w_S\}$ because $\lambda^* \in (0, 1)$. Then, $V^*(\{\boldsymbol{u}, \boldsymbol{v}\}) - V^*(\{\boldsymbol{v}, \boldsymbol{w}\}) = \mu(1) u_1 + \mu(\lambda^*)(\lambda^* u_1 + (1 - \lambda^*) u_S) + \mu(0) v_S - \Big[\mu(1) w_1 + \mu(\lambda^*) \max\{\lambda^* v_1 + (1 - \lambda^*) v_S, \lambda^* w_1 + (1 - \lambda^*) w_S\} + \mu(0) v_S\Big] = \mu(1)(u_1 - w_1) + \mu(\lambda^*)\Big(\lambda^* u_1 + (1 - \lambda^*) u_S - \max\{\lambda^* v_1 + (1 -$

$\lambda^*)v_S, \lambda^*w_1 + (1-\lambda^*)w_S\}\Big)$. When $u_1$ and $u_S$ are close enough to $w_1$ and $v_S$ respectively, we have $V^*(\{\boldsymbol{u}, \boldsymbol{v}\}) - V^*(\{\boldsymbol{v}, \boldsymbol{w}\}) > 0$, so that $\{\boldsymbol{u}, \boldsymbol{v}\} \succ \{\boldsymbol{v}, \boldsymbol{w}\}$. Since $\{\boldsymbol{u}, \boldsymbol{v}\}, \{\boldsymbol{v}, \boldsymbol{w}\} \in \mathscr{A}_{\boldsymbol{v}}^S$, Weak Strategic Rationality I shows that $\{\boldsymbol{u}, \boldsymbol{v}\} \sim \{\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}\}$, so that $V^*(\{\boldsymbol{u}, \boldsymbol{v}\}) = V^*(\{\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}\})$. Since $w_1 > v_1$, $v_S > w_S$, and $\lambda^*u_1 + (1-\lambda^*)u_S > \max\{\lambda^*v_1 + (1-\lambda^*)v_S, \lambda^*w_1 + (1-\lambda^*)w_S\}$, we obtain $V^*(\{\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}\}) - V^*(\{\boldsymbol{u}, \boldsymbol{v}\}) = \mu(1)(w_1 - v_1)$, so that $\mu(1)(w_1 - v_1) = 0$. Since $w_1 - v_1 > 0$, it follows that $\mu(1) = 0$.

**Proof of Corollary** 2: Note that Weak Strategic Rationality II implies Weak Strategic Rationality. Hence, by using No Shame of Acting Selfishly, we can show Lemma 1–7 hold. Therefore, there exists $\lambda^* \in (0,1)$ such that $\mu(\lambda^*) > 0$ and $V^*(A^*) = \mu(1)\max_{\boldsymbol{u} \in A^*} u_1 + \mu(\lambda^*)\max_{\boldsymbol{u} \in A^*}(\lambda^*u_1 + (1-\lambda^*)u_S) + \mu(0)\max_{\boldsymbol{u} \in A^*} u_S$. Finally, we show that $\mu(0) = 0$. To see this note that choose $\boldsymbol{v}, \boldsymbol{w} \in [0,1]^2$ such that $w_1 > v_1$ and $v_S > w_S$. Let $\boldsymbol{u} \in [0,1]^2$ be such that $w_1 > u_1 > v_1$ and $v_S > u_S > w_S$. By making $u_1$ and $u_S$ close enough to $w_1$ and $v_S$ respectively, we can obtain $\lambda^*u_1 + (1-\lambda^*)u_S > \max\{\lambda^*v_1 + (1-\lambda^*)v_S, \lambda^*w_1 + (1-\lambda^*)w_S\}$ because $\lambda^* \in (0,1)$. Then, $V^*(\{\boldsymbol{u}, \boldsymbol{w}\}) - V^*(\{\boldsymbol{v}, \boldsymbol{w}\}) = \mu(1)w_1 + \mu(\lambda^*)(\lambda^*u_1 + (1-\lambda^*)u_S) + \mu(0)u_S - \Big[\mu(1)w_1 + \mu(\lambda^*)\max\{\lambda^*v_1 + (1-\lambda^*)v_S, \lambda^*w_1 + (1-\lambda^*)w_S\} + \mu(0)v_S\Big] = \mu(\lambda^*)\Big(\lambda^*u_1 + (1-\lambda^*)u_S - \max\{\lambda^*v_1 + (1-\lambda^*)v_S, \lambda^*w_1 + (1-\lambda^*)w_S\}\Big) - \mu(0)(v_S - u_S)$. When $u_1$ and $u_S$ are close enough to $w_1$ and $v_S$ respectively, we have $V^*(\{\boldsymbol{u}, \boldsymbol{w}\}) - V^*(\{\boldsymbol{v}, \boldsymbol{w}\}) > 0$, so that $\{\boldsymbol{u}, \boldsymbol{w}\} \succ \{\boldsymbol{v}, \boldsymbol{w}\}$. Since $\{\boldsymbol{u}, \boldsymbol{w}\}, \{\boldsymbol{v}, \boldsymbol{w}\} \in \mathscr{A}_{\boldsymbol{w}}^1$, Weak Strategic Rationality II shows that $\{\boldsymbol{u}, \boldsymbol{w}\} \sim \{\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}\}$, so that $V^*(\{\boldsymbol{u}, \boldsymbol{w}\}) = V^*(\{\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}\})$. Since $w_1 > v_1$, $v_S > w_S$, and $\lambda^*u_1 + (1-\lambda^*)u_S > \max\{\lambda^*v_1 + (1-\lambda^*)v_S, \lambda^*w_1 + (1-\lambda^*)w_S\}$, we obtain $V^*(\{\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}\}) - V^*(\{\boldsymbol{u}, \boldsymbol{w}\}) = \mu(0)(v_S - u_S)$, so that $\mu(0)(v_S - u_S) = 0$. Since $v_S - u_S > 0$, it follows that $\mu(0) = 0$.

**Proof of Proposition** 1: By the standard uniqueness result on $(\Delta(Z))^I$ and the normalization (i.e., $\sum_{i \in S} \alpha_i = 1$), $\alpha_i = \alpha_i'$ for all $i \in I$ and $u = au' + b$ for some real numbers $a > 0$ and $b$. By the nondegeneracy, $\overline{z} \succ \underline{z}$ for some $\overline{z}, \underline{z} \in Z$. Normalize $u$ such that $u(\overline{z}) = 1$ and $u(\underline{z}) = -1$. For all $x \in [0,1]$ define $p(x) = x\delta_{\overline{z}} + (1-x)\left(\frac{1}{2}\delta_{\overline{z}} + \frac{1}{2}\delta_{\underline{z}}\right) \in \Delta(Z)$, $p_S(x) = (p(x))_{i \in S}$, and $\boldsymbol{p}(x) = (p(x), p_S(x))$. Then, $u(p(x)) = x$ and $\sum_{i \in S} \alpha_i u(p(x)) = x$.

Now, we show $\beta_1 = \beta_1'$. Since $1 - \beta_1 > 0$, $1 + \beta_S > 0$, $\alpha_1 > 0$, and $\alpha_i > 0$ for some $i \in S$, then for all $a \in [0,1]$, there exists $\varepsilon_a \in [0,a)$ such that $(1 - \beta_1)\alpha_1\varepsilon_a + (1 + \beta_S)\varepsilon_a = (1 - \beta_1)\alpha_1 a$. Since $u(p(x)) = x = u_S(p_S(x))$ for all $x \in [0,1]$, then $(1 - \beta_1)\alpha_1 u\big(p(\varepsilon_a)\big) + (1 + \beta_S)u_S\big(p_S(\varepsilon_a)\big) = (1 - \beta_1)\alpha_1 u(p(a)) + (1 + \beta_S)\sum_{i \in S}\alpha_i u(p(0))$.

For all $a \in [0,1]$, define $g(a) = V\big(\{\boldsymbol{p}(\varepsilon_a),\ \big(p(a), p_S(0)\big)\}\big)$. Since $\sum_{i \in S}\alpha_i = 1$, we have $\sum_{i \in S}\alpha_i u(p(a)) = u(p(a))$. Hence, $g(a) = (1 - \beta_1)\alpha_1 u(p(\varepsilon_a)) + (1 + \beta_S)\sum_{i \in S}\alpha_i u(p(\varepsilon_a)) + \beta_1\alpha_1 u(p(a)) - \beta_S\sum_{i \in S}\alpha_i u(p(\varepsilon_a)) = (\alpha_1 + 1)u(p(\varepsilon_a)) + \beta_1\alpha_1(u(p(a)) - u(p(\varepsilon_a)))$.

Since $\varepsilon_0 = 0$, then $g(0) = V(p(0))$. Since $g$ is continuous and $V(\overline{z}) > g(0) > V(\underline{z})$, there exists a positive number $a$ such that $V(\overline{z}) > g(a) > V(\underline{z})$, so that $\overline{z} \succ \big\{\boldsymbol{p}(\varepsilon_a),\ \big(p(a), p_S(0)\big)\big\} \succ \underline{z}$. By the continuity, there exists $\eta \in [0,1]$ such that $\eta\delta_{\overline{z}} + (1 - \eta)\delta_{\underline{z}} \sim \big\{\boldsymbol{p}(\varepsilon_a),\ \big(p(a), p_S(0)\big)\big\}$. To make notation simple, let $r = \eta\delta_{\overline{z}} + (1 - \eta)\delta_{\underline{z}}$.

Since $\alpha_1 u(r) + \sum_{i \in S}\alpha_i u(r) = (\alpha_1 + 1)u(r)$, it follows that $(\alpha_1 + 1)u(p(\varepsilon_a)) + \beta_1\alpha_1(u(p(a)) - u(p(\varepsilon_a))) = g(a) = (\alpha_1 + 1)u(r)$. Since $u(p(a)) - u(p(\varepsilon_a)) \neq 0$, we obtain $\beta_1 = \frac{(\alpha_1 + 1)(u(r) - u(p(\varepsilon_a)))}{\alpha_1(u(p(a)) - u(p(\varepsilon_a)))} = \frac{(\alpha_1' + 1)(u'(r) - u'(p(\varepsilon_a)))}{\alpha_1'(u'(p(a)) - u'(p(\varepsilon_a)))} = \beta_1'$, where the equality holds because $\alpha_i = \alpha_i'$ and $u = au' + b$ for some $a > 0$ and $b \in \mathbb{R}$.

Next, we show $\beta_S = \beta_S'$. Since $1 - \beta_1 > 0$, $1 + \beta_S > 0$, and $\alpha_i > 0$ for some $i \in I$, then for all $a \in [0,1]$, there exists $\xi_a \in [0,a)$ such that $(1 - \beta_1)\alpha_1 u(p(\xi_a)) + (1 + \beta_S)\sum_{i \in S}\alpha_i u(p(\xi_a))) = (1 - \beta_1)\alpha_1 u(p(0)) + (1 + \beta_S)\sum_{i \in S}\alpha_i u(p(a))$. Given this $\xi$, we can show $\beta_S = \beta_S'$ by considering $\big\{(\boldsymbol{p}(\xi_a),\ \big(p(0), p_S(a)\big)\big\}$, instead of $\big\{\boldsymbol{p}(\varepsilon_a),\ \big(p(a), p_S(0)\big)\big\}$, in the same way of proving $\beta_1 = \beta_1'$.

**Proof of Proposition** 2: Suppose that $q_1 \succ_1 p_1$, $\{\boldsymbol{p}, \boldsymbol{q}\} \succ \{\boldsymbol{q}\}$, and $p_S \succ_S q_S$. Then, $0 < V(\{\boldsymbol{p}, \boldsymbol{q}\}) - V(\boldsymbol{q}) = (\max\{U(\boldsymbol{p}), U(\boldsymbol{q})\} + \beta_1\alpha_1 u(q_1) - \beta_S\sum_{i \in S}\alpha_i u(p_i)) - (U(\boldsymbol{q}) + \beta_1\alpha_1 u(q_1) - \beta_S\sum_{i \in S}\alpha_i u(q_i)) = (\max\{U(\boldsymbol{p}), U(\boldsymbol{q})\} - U(\boldsymbol{q})) + \beta_S(\sum_{i \in S}\alpha_i u(q_i) - \sum_{i \in S}\alpha_i u(p_i))$. Hence, $0 \leq \beta_S(\sum_{i \in S}\alpha_i u(p_i) - \sum_{i \in S}\alpha_i u(q_i)) < \max\{U(\boldsymbol{p}), U(\boldsymbol{q})\} - U(\boldsymbol{q})$ because $\beta_S \geq 0$, so that $U(\boldsymbol{p}) > U(\boldsymbol{q})$. Hence, $V(\{\boldsymbol{p}, \boldsymbol{q}\}) - V(\boldsymbol{p}) = \beta_1\alpha_1(u(q_1) - u(p_1))$. Since $q_1 \succ_1 p_1$, we have $u(q_1) - u(p_1) > 0$. Therefore, $\{\boldsymbol{p}, \boldsymbol{q}\} \succsim \{\boldsymbol{p}\}$ if and only if $V(\{\boldsymbol{p}, \boldsymbol{q}\}) - V(\boldsymbol{p}) \geq 0$ if and only if $\beta_1\alpha_1(u(q_1) - u(p_1)) \geq 0$ if and only if $\beta_1 \geq 0$. Hence, Proposition 2 holds.

**Proof of Proposition** 3: Let $\alpha_1^X \leq \alpha_1^Y$ to show that $X$ is more altruistic than $Y$. Fix $\boldsymbol{p}, \boldsymbol{q}$ such that $p_S \succ_S^j q_S$ for all $j \in \{X, Y\}$ and $\boldsymbol{p} \succsim^Y \boldsymbol{q}$. Since $p_S \succ_S^j q_S$ for all $j \in \{X, Y\}$,

then $\sum_{i \in S} \alpha_i u(p_i) > \sum_{i \in S} \alpha_i u(q_i)$. First consider the case where $u(p_1) \geq u(q_1)$. Since $\alpha_1^X > 0$, we obtain $\alpha_1^X u(p_1) + \sum_{i \in S} \alpha_i u(p_i) \geq \alpha_1^X u(q_1) + \sum_{i \in S} \alpha_i u(q_i)$. Hence, $\boldsymbol{p} \succsim^X \boldsymbol{q}$, as desired. Next consider the case where $u(p_1) < u(q_1)$. Since $\boldsymbol{p} \succsim^Y \boldsymbol{q}$, then $\alpha_1^Y(u(p_1) - u(q_1)) \geq \sum_{i \in S} \alpha_i u(q_i) - \sum_{i \in S} \alpha_i u(p_i)$. Since $u(p_1) - u(q_1) < 0$ and $\alpha_1^X \leq \alpha_1^Y$, then $\alpha_1^X(u(p_1) - u(q_1)) \geq \sum_{i \in S} \alpha_i u(q_i) - \sum_{i \in S} \alpha_i u(p_i)$. Therefore, $\alpha_1^X u(p_1) + \sum_{i \in S} \alpha_i u(p_i) \geq \alpha_1^X u(q_1) + \sum_{i \in S} \alpha_i u(q_i)$. Hence, $\boldsymbol{p} \succsim^X \boldsymbol{q}$.

To show the converse, choose any $\boldsymbol{p}, \boldsymbol{q} \in (\Delta(Z))^I$ such that $p_S \succ_S^j q_S$, $q_1 \succ_1^j p_1$ for all $j \in \{X, Y\}$, and $\boldsymbol{p} \sim^Y \boldsymbol{q}$. (Such $\boldsymbol{p}$ and $\boldsymbol{q}$ exist because $\alpha_1 > 0$ and $\alpha_i > 0$ for some $i \in S$.) Hence, $\alpha_1^Y u(p_1) + \sum_{i \in S} \alpha_i u(p_i) = \alpha_1^Y u(q_1) + \sum_{i \in S} \alpha_i u(q_i)$, so that $\alpha_1^Y = \frac{\sum_{i \in S} \alpha_i u(p_i) - \sum_{i \in S} \alpha_i u(q_i)}{u(q_1) - u(p_1)}$. On the other hand, since $X$ is more altruistic than $Y$, then $\boldsymbol{p} \succsim^X \boldsymbol{q}$, so that $\alpha_1^X u(p_1) + \sum_{i \in S} \alpha_i u(p_i) \geq \alpha_1^X u(q_1) + \sum_{i \in S} \alpha_i u(q_i)$, so that $\frac{\sum_{i \in S} \alpha_i u(p_i) - \sum_{i \in S} \alpha_i u(q_i)}{u(q_1) - u(p_1)} \geq \alpha_1^X$. Therefore, $\alpha_1^Y \geq \alpha_1^X$.

**Proof of Proposition** 4: To show the proposition, for all $j \in \{X, Y\}$ and $\boldsymbol{p} \in (\Delta(Z))^I$, define $U^j(\boldsymbol{p}) = (1 - \beta_1^j)\alpha_1 u(p_1) + (1 + \beta_S^j)\sum_{i \in S} \alpha_i u(p_i)$: for all $j \in \{X, Y\}$ and for all $A \in \mathscr{A}$, define $V^j(A) = \max_{\boldsymbol{p} \in A} U^j(\boldsymbol{p}) + \beta_1^j \max_{\boldsymbol{q} \in A} \alpha_1 u(q_1) - \beta_S^j \max_{\boldsymbol{r} \in A} \sum_{i \in S} \alpha_i u(r_i)$

First, we show (i). Choose $\boldsymbol{p}, \boldsymbol{q} \in (\Delta(Z))^I$ and $r \in \Delta(Z)$ such that (i) $q_1 \succ_1^j p_1$, (ii) $\{\boldsymbol{p}, \boldsymbol{q}\} \succ^j \{\boldsymbol{q}\}$, (iii) $p_S \succ_S^j q_S$ for each $j \in \{X, Y\}$, and $\{\boldsymbol{p}, \boldsymbol{q}\} \sim^Y \{(r)_{i \in I}\}$. Then, by the argument in Proposition 2, $U^j(\boldsymbol{p}) > U^j(\boldsymbol{q})$ for each $j \in \{X, Y\}$. Since $\succsim^X$ is more pride-seeking (less temptation-averse) than $\succsim^Y$, we have $\{\boldsymbol{p}, \boldsymbol{q}\} \succsim^X \{(r)_{i \in I}\}$. Since $\{\boldsymbol{p}, \boldsymbol{q}\} \sim^Y \{(r)_{i \in I}\}$, we have $(1 - \beta_1^Y)\alpha_1 u(p_1) + \sum_{i \in S} \alpha_i u(p_i) + \beta_1^Y \alpha_1 u(q_1) = V^Y(\{\boldsymbol{p}, \boldsymbol{q}\}) = u(r)$. Since $\{\boldsymbol{p}, \boldsymbol{q}\} \succsim^X \{(r)_{i \in I}\}$, we have $u(r) \leq V^X(\{\boldsymbol{p}, \boldsymbol{q}\}) = (1 - \beta_1^X)\alpha_1 u(p_1) + \sum_{i \in S} \alpha_i u(p_i) + \beta_1^X \alpha_1 u(q_1)$. Hence, $(1 - \beta_1^Y)\alpha_1 u(p_1) + \sum_{i \in S} \alpha_i u(p_i) + \beta_1^Y \alpha_1 u(q_1) \leq (1 - \beta_1^X)\alpha_1 u(p_1) + \sum_{i \in S} \alpha_i u(p_i) + \beta_1^X \alpha_1 u(q_1)$, so that $\beta_1^Y \alpha_1(u(q_1) - u(p_1)) \leq \beta_1^X \alpha_1(u(q_1) - u(p_1))$. Hence, $\beta_1^Y \leq \beta_1^X$ because $u(q_1) > u(p_1)$.

To show the converse assume $\beta_1^Y \leq \beta_1^X$. Choose $\boldsymbol{p}, \boldsymbol{q} \in (\Delta(Z))^I$ and $r \in \Delta(Z)$ such that (i) $q_1 \succ_1^j p_1$, (ii) $\{\boldsymbol{p}, \boldsymbol{q}\} \succ^j \{\boldsymbol{q}\}$, (iii) $p_S \succ_S^j q_S$ for each $j \in \{X, Y\}$, and $\{\boldsymbol{p}, \boldsymbol{q}\} \succsim^Y \{(r)_{i \in I}\}$. Then, $u(r) \leq V^Y(\{\boldsymbol{p}, \boldsymbol{q}\}) \equiv (1 - \beta_1^Y)\alpha_1 u(p_1) + \sum_{i \in S} \alpha_i u(p_i) + \beta_1^Y \alpha_1 u(q_1) \leq (1 - \beta_1^X)\alpha_1 u(p_1) + \sum_{i \in S} \alpha_i u(p_i) + \beta_1^X \alpha_1 u(q_1) \equiv V^X(\{\boldsymbol{p}, \boldsymbol{q}\})$. So, we obtain $u(r) \leq V^X(\{\boldsymbol{p}, \boldsymbol{q}\})$. Hence, the converse holds.

47

Next we show (ii). Choose $\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r} \in (\Delta(Z))^I$ and $l \in \Delta(Z)$ such that (i) $q_S \succ_S^j p_S$, (ii) $r_1 \succ_1^j p_1$, (iii) $\{\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r}\} \succ^j \{\boldsymbol{q}, \boldsymbol{r}\}$ for each $j \in \{X, Y\}$, and $\{(l)_{i \in I}\} \sim^Y \{\boldsymbol{p}, \boldsymbol{q}\}$. Since $\succsim^X$ is more shame averse than $\succsim^Y$, we have $\{(l)_{i \in I}\} \succsim^X \{\boldsymbol{p}, \boldsymbol{q}\}$. Now we will show that $U^j(\boldsymbol{p}) > U^j(\boldsymbol{q})$ for all $j \in \{X, Y\}$. To see this note that $0 < V^j(\{\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r}\}) - V^j(\{\boldsymbol{q}, \boldsymbol{r}\}) = \left[\max\{U^j(\boldsymbol{p}), U^j(\boldsymbol{q}), U^j(\boldsymbol{r})\} + \beta_1^j \alpha_1 \max\{u(q_1), u(r_1)\} - \beta_S^j \max\{\sum_{i \in S} \alpha_i u(q_i), \sum_{i \in S} \alpha_i u(r_i)\}\right] - \left[\max\{U^j(\boldsymbol{q}), U^j(\boldsymbol{r})\} + \beta_1^j \alpha_1 \max\{u(q_1), u(r_1)\} - \beta_S^j \max\{\sum_{i \in S} \alpha_i u(q_i), \sum_{i \in S} \alpha_i u(r_i)\}\right] = \max\{U^j(\boldsymbol{p}), U^j(\boldsymbol{q}), U^j(\boldsymbol{r})\} - \max\{U^j(\boldsymbol{q}), U^j(\boldsymbol{r})\}$, so that $U^j(\boldsymbol{p}) > \max\{U^j(\boldsymbol{q}), U^j(\boldsymbol{r})\}$ for each $j \in \{X, Y\}$. Since $q_S \succ_S^j p_S$, we have $p_1 \succ_1^j q_1$ for each $j \in \{X, Y\}$. Moreover, we have $V^j(\{\boldsymbol{p}, \boldsymbol{q}\}) = U^j(\boldsymbol{p}) + \beta_1^j \alpha_1 u(p_1) - \beta_S^j \sum_{i \in S} \alpha_i u(q_i) = \alpha_1 u(p_1) + (1 + \beta_S^j) \sum_{i \in S} \alpha_i u(p_i) - \beta_S^j \sum_{i \in S} \alpha_i u(q_i)$ for each $j \in \{X, Y\}$.

Since $\{(l)_{i \in I}\} \sim^Y \{\boldsymbol{p}, \boldsymbol{q}\}$, we have $\alpha_1 u(p_1) + (1 + \beta_S^Y) \sum_{i \in S} \alpha_i u(p_i) - \beta_S^Y \sum_{i \in S} \alpha_i u(q_i) = V^Y(\{\boldsymbol{p}, \boldsymbol{q}\}) = u(l)$. Since $\{(l)_{i \in I}\} \succsim^X \{\boldsymbol{p}, \boldsymbol{q}\}$, we have $u(l) \geq V^X(\{\boldsymbol{p}, \boldsymbol{q}\}) = \alpha_1 u(p_1) + (1 + \beta_S^X) \sum_{i \in S} \alpha_i u(p_i) - \beta_S^X \sum_{i \in S} \alpha_i u(q_i)$. Hence, $\alpha_1 u(p_1) + (1 + \beta_S^X) \sum_{i \in S} \alpha_i u(p_i) - \beta_S^X \sum_{i \in S} \alpha_i u(q_i) \leq \alpha_1 u(p_1) + (1 + \beta_S^Y) \sum_{i \in S} \alpha_i u(p_i) - \beta_S^Y \sum_{i \in S} \alpha_i u(q_i)$, so that $\beta_S^Y (\sum_{i \in S} \alpha_i u(q_i) - \sum_{i \in S} \alpha_i u(p_i)) \leq \beta_S^X (\sum_{i \in S} \alpha_i u(q_i) - \sum_{i \in S} \alpha_i u(p_i))$. Hence, $\beta_S^Y \leq \beta_S^X$.

To show the converse assume $\beta_S^Y \leq \beta_S^X$. Choose $\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r} \in (\Delta(Z))^I$ and $l \in \Delta(Z)$ such that (i) $q_S \succ_S^j p_S$, (ii) $r_1 \succ_1^j p_1$, (iii) $\{\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r}\} \succ^j \{\boldsymbol{q}, \boldsymbol{r}\}$ for each $j \in \{X, Y\}$, and $\{(l)_{i \in I}\} \succsim^Y \{\boldsymbol{p}, \boldsymbol{q}\}$. Then, $u(l) \geq V^Y(\{\boldsymbol{p}, \boldsymbol{q}\}) \equiv \alpha_1 u(p_1) + (1 + \beta_S^Y) \sum_{i \in S} \alpha_i u(p_i) - \beta_S^Y \sum_{i \in S} \alpha_i u(q_i) \geq \alpha_1 u(p_1) + (1 + \beta_S^X) \sum_{i \in S} \alpha_i u(p_i) - \beta_S^X \sum_{i \in S} \alpha_i u(q_i) \equiv V^X(\{\boldsymbol{p}, \boldsymbol{q}\})$. So, we obtain $u(l) \geq V^X(\{\boldsymbol{p}, \boldsymbol{q}\})$. Hence, the converse holds.

**Proof of Proposition** 5: Let $\succsim$ be represented by the GU model with $(u, \boldsymbol{\beta}, \boldsymbol{\alpha})$. By the nondegeneracy, $\bar{z} \succ \underline{z}$ for some $\bar{z}, \underline{z} \in Z$. Normalize $u$ such that $u(\bar{z}) = 1$ and $u(\underline{z}) = 0$. For all $x \in [0, 1]$ define $p(x) = x \delta_{\bar{z}} + (1 - x) \delta_{\underline{z}} \in \Delta(Z)$ and $p_S(x) = (p(x))_{i \in S}$. There exists $\varepsilon \in (0, 1)$ such that $[(1 - \beta_1)\alpha_1 + (1 + \beta_S)](1 - \varepsilon) > \max\{(1 - \beta_1)\alpha_1, 1 + \beta_S\}$. Define $\boldsymbol{p}^U = (p(1 - \varepsilon), p_S(1 - \varepsilon))$, $\boldsymbol{p}^1 = (p(1), p_S(0))$, and $\boldsymbol{p}^S = (p(0), p_S(1))$. Then, $U(\boldsymbol{p}^U) > \max\{U(\boldsymbol{p}^1), U(\boldsymbol{p}^S)\}$, $u(p_1^1) > \max\{u(p_1^U), u(p_1^S)\}$, and $u_S(p_S^S) > \max\{u_S(p_S^1), u_S(p_S^U)\}$. Take any set $A \in \mathscr{A}$. Choose any $\boldsymbol{p}^* \in \arg\max_{\boldsymbol{p} \in A} U(\boldsymbol{p})$ and $\boldsymbol{q}_* \in A$ such that $U(\boldsymbol{q}_*) < U(\boldsymbol{p}^*)$. It suffices to show $\boldsymbol{p}^* \in C(A)$ and $\boldsymbol{q}^* \notin C(A)$.

Fix $\alpha \in (0,1)$. Let $\boldsymbol{q}^* = \alpha \boldsymbol{p}^U + (1-\alpha)\boldsymbol{p}^*$ and $B = \{\boldsymbol{q}^*\} \cup (\alpha\{\boldsymbol{p}^1, \boldsymbol{p}^S\} + (1-\alpha)A)$. Then $U(\boldsymbol{q}^*) > U(\boldsymbol{r})$ for all $\boldsymbol{r} \in B$ such that $\boldsymbol{r} \neq \boldsymbol{q}^*$. Moreover, $u(q_1^*) < u(\alpha p_1^1 + (1-\alpha)p_1^*)$ and $u_S(q_S^*) < u_S(\alpha p_S^S + (1-\alpha)p_S^*)$. Then $B \notin \mathscr{A}_{\boldsymbol{q}^*}^1$. We obtain $V(B) - V(B \setminus \{\boldsymbol{q}^*\}) = U(\boldsymbol{q}^*) - \max_{\boldsymbol{r} \in B \setminus \{\boldsymbol{q}^*\}} U(\boldsymbol{r}) > 0$, so that $B \succ B \setminus \{\boldsymbol{q}^*\}$. By Consistency, $C(B) = \{\boldsymbol{q}^*\}$. By choosing $\alpha$ arbitrarily small and using Closed Graph, we obtain $\boldsymbol{p}^* \in C(A)$.

To show $\boldsymbol{q}^* \notin C(A)$, let $D = \{\boldsymbol{q} \in (\Delta(Z))^I | U(\boldsymbol{q}_*) < U(\boldsymbol{q}) < U(\boldsymbol{p}^*)\}$. There exist $\bar{l}, \underline{l} \in \Delta(Z)$ such that $U(\boldsymbol{q}_*) < u(\underline{l}) < u(\bar{l}) < U(\boldsymbol{p}^*)$. By using $\underline{l}$ and $\bar{l}$, we can define $\boldsymbol{q}^U, \boldsymbol{q}^1, \boldsymbol{q}^S \in D$ such that $U(\boldsymbol{q}^U) > \max\{U(\boldsymbol{q}^1), U(\boldsymbol{q}^S)\}$, $u(q_1^1) > \max\{u(q_1^U), u(q_1^S)\}$, and $u_S(q_S^S) > \max\{u_S(q_S^1), u_S(q_S^U)\}$, as we defined $\boldsymbol{p}^U, \boldsymbol{p}^1, \boldsymbol{p}^S$ by using $\bar{z}$ and $\underline{z}$. Let $B = \{\boldsymbol{q}^U, \boldsymbol{q}^1, \boldsymbol{q}^S, \boldsymbol{q}_*\}$. Hence, $V(B) - V(B \setminus \{\boldsymbol{q}^U\}) = U(\boldsymbol{q}^U) - \max\{U(\boldsymbol{q}^1), U(\boldsymbol{q}^S), U(\boldsymbol{q}_*)\}$, so that $B \succ B \setminus \{\boldsymbol{q}^U\}$. However, $\boldsymbol{q}^U$ is not individually optimal in $B$. Hence, by Consistency, $C(B) = \{\boldsymbol{q}^U\}$.

Define $A' = A \cup \{\boldsymbol{q}^U\}$. Since $\boldsymbol{q}^U \in D$, we obtain $U(\boldsymbol{q}^U) < U(\boldsymbol{p}^*)$, so that $\boldsymbol{p}^* \in \arg\max_{\boldsymbol{p} \in A'} U(\boldsymbol{p})$. By the first part of the proof, $\boldsymbol{p}^* \in C(A')$. Suppose that $\boldsymbol{q}_* \in C(A)$ to show a contradiction. By WARP, $\boldsymbol{q}_* \in C(A')$ because $\boldsymbol{p}^* \in C(A') \cap A$. Moreover, $\boldsymbol{q}_* \in C(B)$ because $\boldsymbol{q}^U \in C(B) \cap A'$. This contradicts with $C(B) = \{\boldsymbol{q}^U\}$.

**Proof of Proposition** 6: In the following, we show (i). (ii) can be proved in the same way. Fix $\bar{z}, \underline{z} \in Z$ such that $\bar{z} \succ \underline{z}$. Normalize $u$ by $u(\bar{z}) = 1$ and $u(\underline{z}) = 0$. Consider the case where $\alpha_1 \geq 1$. Define $q_1 = \frac{1}{\alpha_1}\bar{z} + \frac{\alpha_1 - 1}{\alpha_1}\underline{z}$, $\boldsymbol{p} = (\underline{z}, (\bar{z})_{i \in S})$, and $\boldsymbol{q} = (q_1, (\underline{z})_{i \in S})$. By a direct calculation, we obtain $\boldsymbol{p} \sim \boldsymbol{q}$ and $\boldsymbol{p} \succ_S \boldsymbol{q}$.[29] Hence, Impure Altruism implies $C(\{\boldsymbol{p}, \boldsymbol{q}\}) = \{\boldsymbol{p}\}$. Hence, by Proposition 5, $0 < U(\boldsymbol{p}) - U(\boldsymbol{q}) = \sum_{i \in I} \alpha_i(u(p_i) - u(q_i)) + \beta_S(\sum_{i \in S} \alpha_i u(p_i) - \sum_{i \in S} \alpha_i u(q_i)) + \beta_1 \alpha_1(u(q_1) - u(p_1)) = \beta_S + \beta_1$, so that $\beta_S > -\beta_1$. In the case where $\alpha_1 \leq 1$, we can show the result in the same way by defining $q_1 = \alpha_1 \bar{z} + (1-\alpha_1)\underline{z}$.

Next, we show the converse (i.e., $\beta_S > -\beta_1$ implies Impure Altruism). Choose any $\boldsymbol{p}, \boldsymbol{q} \in \Delta(Z)$ such that $\boldsymbol{p} \sim \boldsymbol{q}$ and $p_S \succ_S q_S$ to show $C(\{\boldsymbol{p}, \boldsymbol{q}\}) = \{\boldsymbol{p}\}$. It suffices to show that $U(p) > U(q)$. By assumption, $\beta_1 > -\beta_S$. Hence, $U(\boldsymbol{p}) - U(\boldsymbol{q}) > \sum_{i \in I} \alpha_i(u(p_i) - u(q_i)) +$

---

[29]Then, $u(\underline{z}) = 0$, $u_S(\bar{z}) = 1$, $u(q_1) = 1/\alpha_1$, and $\sum_{i \in S} \alpha_i u(q_i) = 0$. Hence, $\sum_{i \in I} \alpha_i u(p_i) = 1 = \sum_{i \in I} \alpha_i u(q_i)$.

$\beta_S(\sum_{i \in S} \alpha_i u(p_i) - \sum_{i \in S} \alpha_i u(q_i)) - \beta_S \alpha_1(u(q_1) - u(p_1)) = (1 + \beta_S) \sum_{i \in I} \alpha_i(u(p_i) - u(q_i)) \geq 0$,

where the equality holds because $\boldsymbol{q} \sim \boldsymbol{p}$. Therefore, $U(\boldsymbol{p}) > U(\boldsymbol{q})$ so that $C(\{\boldsymbol{p}, \boldsymbol{q}\}) = \{\boldsymbol{p}\}$.

**Proof of Proposition** 7: First note that since $u' > 0$, $u'(0) = +\infty$, $\alpha_1 > 0$, and $\beta_S \geq 0$, the budget constraint is binding (i.e., $c + d = w + \tau$). Hence, $V(\{(c, d) \in \mathbb{R}_+^2 | c + d \leq w + \tau\}) = V(\{(c, w + \tau - c) | c \in [0, w + \tau]\}) = \alpha_1 u(c^*) + (1 + \beta_S) u(w + \tau - c^*) - \beta_S u(w + \tau)$, where $c^* \in \arg\max_{c \in [0, w+\tau]} \alpha_1 u(c) + (1 + \beta_S) u(w + \tau - c)$. Moreover, $c^*(\alpha_1, \beta_S, \tau)$ is an interior solution determined by the first order condition: $0 = \alpha_1 u'(c^*) + (1 + \beta_S) u'(w + \tau - c^*)$. For any $\beta_S$, $c^* \to 0$ as $\alpha_1 \to 0$ and $c^* \to w + \tau$ as $\alpha_1 \to \infty$. Therefore, by the intermediate value theorem, for any $\beta_S \geq 0$, there exists $\alpha_1(\beta_S)$ such that $c^*(\alpha_1(\beta_S), \beta_S, \tau) = w$.

Define $f(\alpha_1, \beta_S, \tau) = V(\{(w, 0)\}) - V(\{(c, d) \in \mathbb{R}_+^2 | c + d \leq w + \tau\})$. Since $u(0) = 0$, we have $f(\alpha_1, \beta_S, \tau) = \alpha_1 u(w) - [\alpha_1 u(c^*) + (1 + \beta_S) u(w + \tau - c^*) - \beta_S u(w + \tau)] = \alpha_1(u(w) - u(c^*)) - u(w + \tau - c^*) - \beta_S(u(w + \tau - c^*) - u(w + \tau))$. Hence, $f(\alpha_1(\beta_S), \beta_S, \tau) = -u(\tau) - \beta_S(u(\tau) - u(w + \tau)) = \beta_S(u(w + \tau) - u(\tau)) - u(\tau)$. Define $\underline{\beta_S} = \frac{u(\tau)}{u(w+\tau) - u(\tau)}$. It follows that $f(\alpha_1(\beta_S), \beta_S, \tau) > 0$ if and only if $\beta_S > \underline{\beta_S}$.

Therefore, for any $\beta_S > \underline{\beta_S}$, $f$ tends to strictly positive as $\alpha_1 \to \alpha(\beta_S)$. Moreover, $f$ tends to strictly negative as $\alpha_1 \to \infty$.[30] Therefore, by the intermediate value theorem, for any $\beta_S > \underline{\beta_S}$, there exists $\overline{\alpha_1}(\beta_S)$ such that $f(\overline{\alpha_1}(\beta_S), \beta_S, \tau) = 0$ and $\infty > \overline{\alpha_1}(\beta_S) > \alpha_1(\beta_S)$. Moreover, $f$ tends to strictly negative as $\alpha_1 \to 0$.[31] Therefore, by the intermediate value theorem, for any $\beta_S > \underline{\beta_S}$, there exists $\underline{\alpha_1}(\beta_S)$ such that $f(\underline{\alpha_1}(\beta_S), \beta_S, \tau) = 0$ and $0 < \underline{\alpha_1}(\beta_S) < \alpha(\beta_S)$. By this way, we have defined $\underline{\alpha_1}, \overline{\alpha_1}$ on $(\underline{\beta_S}, +\infty)$. Finally, define $\underline{\alpha_1}(\underline{\beta_S}) = \alpha_1(\underline{\beta_S})$ and $\overline{\alpha_1}(\underline{\beta_S}) = \alpha_1(\underline{\beta_S})$. Then, $f(\overline{\alpha_1}(\underline{\beta_S}), \underline{\beta_S}, \tau) = 0 = f(\underline{\alpha_1}(\underline{\beta_S}), \underline{\beta_S}, \tau)$.

By the implicit function theorem, it can be shown that $\frac{\partial c^*}{\partial \alpha_1}(\alpha_1, \beta_S, \tau) > 0$. By the envelop theorem, we have $\frac{\partial f}{\partial \alpha_1}(\alpha_1, \beta_S, \tau) = u(w) - u(c^*)$, hence $\frac{\partial f}{\partial \alpha_1}(\alpha_1, \beta_S, \tau) < 0$ if and only if $w < c^*(\alpha_1, \beta_S, \tau)$ if and only if $\alpha_1 > \alpha_1(\beta_S)$. In addition, $\frac{\partial f}{\partial \beta_S}(\alpha_1, \beta_S, \tau) = u(w + \tau) - u(w + \tau - c^*) > 0$.

---

[30] To see this note that, $c^* \to w + \tau$ as $\alpha_1 \to \infty$. Hence, $f(\alpha_1, \beta_S, \tau) = \alpha_1(u(w) - u(c^*)) - u(w + \tau - c^*) - \beta_S(u(w + \tau - c^*) - u(w + \tau)) \to -\infty$.

[31] To see this note that, $c^* \to 0$ as $\alpha_1 \to 0$. Hence, $f(\alpha_1, \beta_S, \tau) = \alpha_1(u(w) - u(c^*)) - u(w + \tau - c^*) - \beta_S(u(w + \tau - c^*) - u(w + \tau)) \to -u(w + \tau)$.

Remember that $\overline{\alpha_1}(\beta_S) > \alpha_1(\beta_S)$. Hence, $\frac{\partial f}{\partial \alpha_1}(\overline{\alpha_1}(\beta_S), \beta_S, \tau) < 0$ and $\frac{\partial f}{\partial \beta_S}(\overline{\alpha_1}(\beta_S), \beta_S, \tau) > 0$. Since $f(\overline{\alpha_1}(\beta_S), \beta_S, \tau) = 0$, it follows that $\overline{\alpha_1}(\beta_S)$ is strictly increasing.[32]

Similarly, remember also that $\underline{\alpha_1}(\beta_S) < \alpha_1(\beta_S)$. Hence, $\frac{\partial f}{\partial \alpha_1}(\underline{\alpha_1}(\beta_S), \beta_S, \tau) > 0$ and $\frac{\partial f}{\partial \beta_S}(\underline{\alpha_1}(\beta_S), \beta_S, \tau) > 0$. Since $f(\underline{\alpha_1}(\beta_S), \beta_S, \tau) = 0$, it follows that $\underline{\alpha_1}(\beta_S)$ is strictly decreasing.

Moreover, for any $\beta_S$, $f(\alpha_1, \beta_S, \tau) \geq 0$ if and only if $\alpha_1 \in [\underline{\alpha_1}(\beta_S), \overline{\alpha_1}(\beta_S)]$ and $f(\alpha_1, \beta_S, \tau) < 0$ if and only if $\alpha_1 < \underline{\alpha_1}(\beta_S)$ or $\alpha_1 > \overline{\alpha_1}(\beta_S)$.

**Proof of Proposition** 8: Define $\hat{d} \equiv d^* + \tau = d^* + g$ for $\tau = g$, where $d^*$ is the optimal donation. Since $u' > 0$, the budget constraints must be binding (i.e., $s + w = e$ and $c + d = w - \tau$). Hence, the decision maker's problem is as follows: $\max_w V(\{(w - \tau - d^*, d^* + g) | d^* \leq w - \tau\}) + \delta V(\{(e - w, 0)\}) = \max_{\hat{d}, w} h(w, \hat{d}, \tau)$, where $h(w, \hat{d}, \tau) = (1 - \beta_1)\alpha_1 u(w - \hat{d}) + (1 + \beta_S)u(\hat{d}) + \beta_1 \alpha_1 u(w - \tau) - \beta_S u(w) + \delta \alpha_1 u(e - w)$.

If $\beta_1 = 0 = \beta_S$, then $h$ does not depend on $\tau$ (i.e., $g$), so that the solution $\hat{d}$ is constant in $g$. Hence, (ii) holds. To show (i), assume $\beta_1 > \beta_S$. Define $w^*(\hat{d}, \tau) = \arg\max_w h(w, \hat{d}, \tau)$, $f(\hat{d}, \tau) = h(w^*(\hat{d}, \tau), \hat{d}, \tau)$, and $\hat{d}(\tau) = \arg\max_{\hat{d}} f(\hat{d}, \tau)$. Since $u' > 0$, $u'' < 0$, and $u'(0) = +\infty$, both $w^*$ and $\hat{d}$ are interior solutions.

We show that $\hat{d}(\tau)$ is strictly increasing. Given $\beta_1 \alpha_1 > \beta_S \geq 0$ and $u''' \geq 0$, direct calculation shows $\frac{\partial^2 h(w, \hat{d}, \tau)}{\partial w^2} < 0$.[33] Moreover, $\frac{\partial^2 h(w, \hat{d}, \tau)}{\partial w \partial \tau} = -\beta_1 \alpha_1 u''(w - \tau) > 0$, for $\beta_1 \alpha_1 > \beta_S \geq 0$. Therefore, by the standard result on monotone comparative statics, $\frac{\partial w^*(\hat{d}, \tau)}{\partial \tau} > 0$. Hence, by the envelope theorem, $\frac{\partial^2 f(\hat{d}, \tau)}{\partial \hat{d} \partial \tau} = -(1 - \beta_1)\alpha_1 u''(w^*(\hat{d}, \tau) - \hat{d})\frac{\partial w^*(\hat{d}, \tau)}{\partial \tau} > 0$.

By the implicit function theorem, a direct calculation shows that $\frac{\partial w^*(\hat{d}, \tau)}{\partial \hat{d}} < 1$.[34] Hence,

---

[32]To see this suppose by way of contradiction that $\beta_S < \beta_S'$ and $\overline{\alpha_1}(\beta_S) \geq \overline{\alpha_1}(\beta_S')$. Then, $0 = f(\overline{\alpha_1}(\beta_S), \beta_S, \tau) < f(\overline{\alpha_1}(\beta_S), \beta_S', \tau) \leq f(\overline{\alpha_1}(\beta_S'), \beta_S', \tau) = 0$, which is a contradiction. The first inequality holds because $\frac{\partial f}{\partial \beta_S}(\overline{\alpha_1}(\beta_S), \beta_S, \tau) > 0$. The second inequality holds because $\overline{\alpha_1}(\beta_S') > \alpha_1(\beta_S')$ and $\frac{\partial f}{\partial \alpha_1}(\alpha_1, \beta_S', \tau) < 0$ if and only if $\alpha_1 > \alpha_1(\beta_S')$.

[33]$\frac{\partial^2 h(w, \hat{d}, \tau)}{\partial w^2} = (1 - \beta_1)\alpha_1 u''(w - \hat{d}) + \beta_1 \alpha_1 u''(w - \tau) - \beta_S u''(w) + \delta \alpha_1 u''(e - w)$. Since $u''' \geq 0$, then $u''(w - \tau) \leq u''(w) < 0$. Hence, $\beta_1 \alpha_1 > \beta_S$ implies $\frac{\partial^2 h(w, \hat{d}, \tau)}{\partial w^2} < 0$.

[34]To see this, note that $w^*(\hat{d}, \tau)$ is characterized by the first order condition: $0 = \frac{\partial h(w, \hat{d}, \tau)}{\partial w} = (1 - \beta_1)\alpha_1 u'(w - \hat{d}) + \beta_1 \alpha_1 u'(w - \tau) - \beta_S u'(w) - \delta \alpha_1 u'(e - w) \equiv k(w, \hat{d}, \tau)$. By the implicit function theorem, $\frac{\partial w^*(\hat{d}, \tau)}{\partial \hat{d}} = -\frac{\partial k/\partial \hat{d}}{\partial k/\partial w} = \left(1 + \frac{\beta_1 \alpha_1 u''(w - \tau) - \beta_S u''(w) + \delta \alpha_1 u''(e - w)}{(1 - \beta_1)\alpha_1 u''(w - \hat{d})}\right)^{-1} < 1$ because $\beta_1 \alpha_1 u''(w - \tau) - \beta_S u''(w) < 0$.

$\frac{\partial^2 f(\hat{d}, \tau)}{\partial \hat{d}^2} = -(1 - \beta_1)\alpha_1 u''(w^*(\hat{d}, \tau) - \hat{d})(\frac{\partial w^*(\hat{d}, \tau)}{\partial \hat{d}} - 1) + (1 + \beta_S)u''(\hat{d}) < 0$. Therefore, by the monotone comparative statics again, $\hat{d}$ is strictly increasing in $\tau$ (i.e., $g$).

# References

[1] Andreoni, J., Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence, Journal of Political Economy 97 (1989) 1447–1458.

[2] Andreoni, J., Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving, Economic Journal 100 (1990) 464–477.

[3] Andreoni, J., Philanthropy, in Handbook on the Economics of Giving, Altruism and Reciprocity ed. by S. Kolm and J. M. Ythier. Elsevier, (2006) 1201–1269.

[4] Arrow, K., Rational Choice Functions and Orderings, Econometrica 26 (1959) 121–127.

[5] Broberg, T., Ellingsen, T., and M. Johannesson, Is Generosity Involuntary? Economics Letters 94 (2007) 32–37.

[6] Dana, J., Cain, D. M., and R. M. Dawes, What You Don't Know Won't Hurt Me, Organizational Behavior and Human Decision Processes 100 (2006) 193–201.

[7] Dekel, E., B. L. Lipman, and A. Rustichini, Representing Preferences with a Unique Subjective State Space, Econometrica 69 (2001) 891–934.

[8] Dufwenberg, M., Heidhues, P., Kirchsteiger, G., Riedel, F., and J. Sobel, Other-Regarding Preferences in General Equilibrium, Review of Economic Studies 78 (2001) 613–639.

[9] Dillenberger, D. and P. Sadowski, Ashamed to be Selfish, Theoretical Economics 7 (2012) 99–124.

[10] Evren, Ö. and S. Minardi, Warm-Glow Giving and Freedom to be Selfish, working paper, 2014.

[11] Feddersen, T. and A. Sandroni, The Foundation of Warm-Glow Theory, working paper, 2009.

[12] Fehr, E. and K. M. Schmidt, A Theory of Fairness, Competition, and Cooperation, Quarterly Journal of Economics 114 (1999) 817–868.

[13] Fishburn, P. C., On Harsanyi's Utilitarian Cardinal Welfare Theorem, Theory and Decision, 17 (1984) 21–28.

[14] Gilboa, I. and D. Schmeidler, Maxmin Expected Utility with a Non-unique Prior, Journal of Mathematical Economics 18 (1989) 141–153.

[15] Gul, F. and W. Pesendorfer, Temptation and Self-Control, Econometrica, 69 (2001) 1403–1435.

[16] Herstein, I. N. and J. Milnor, An Axiomatic Approach to Measurable Utility, Econometrica 21 (1953), 291–297.

[17] Kopylov, I., Finite Additive Utility Representations for Preferences over Menus, Journal of Economic Theory 144 (2009) 354–374.

[18] Kopylov, I., Perfectionism and Choice, Econometrica 80 (2012) 1819–1843.

[19] Lea, S. E., and Webley, P., Pride in Economic Psychology, Journal of Economic Psychology 18 (1997) 323–340.

[20] Lazear, E. P., Malmendier, U., and R. A. Weber, Sorting in Experiments with Application to Social Preferences, American Economic Journal: Applied Economics 4 (2012) 136–163.

[21] Neilson, W. S., A Theory of Kindness, Reluctance, and Shame for Social Preferences, Games and Economic Behavior 66 (2009), 394–403.

[22] Noor, J. and L. Ren, Temptation and Social Preference, working paper, 2011.

[23] Hiebert, P. G., Anthropological insights for missionaries. Grand Rapids: Baker Book House, 1985.

[24] Sheff, T. J., Shame and Conformity: The Deference-Emotion System, American sociological review, (1988) 395–406.

[25] Taylor, G., Pride, shame, and guilt: Emotions of self-assessment. Oxford University Press, 1985.

[26] Yagasaki, M. Pride and Shame, working paper, 2013.

# C    Extension (Not for Publication)

In this section, to incorporate inequality aversion, we axiomatize an extended GU model, in which $u_S$ is a maxmin utility function. We consider a decision maker who is inequality-averse among other agents' allocations.

It is well known that the independence axiom may fail in social context because mixtures among allocations can offset inequality in the mixed allocation. However, mixing with constant allocations does not offset inequality. Hence, we keep the following weaker version of the independence axiom:

**Definition**: A set $C \in \mathscr{A}$ is called *constant over $S$* if $p_i = p_j$ for any $i, j \in S$ and $p \in C$.

**Axiom** (Weak Independence): Let $\alpha \in [0, 1]$ and $A, B, C \in \mathscr{A}$. Suppose that $C$ is constant over $S$. Then $A \succsim B$ if and only if $\alpha A + (1 - \alpha)C \succsim \alpha B + (1 - \alpha)C$.

We need an additional axiom to make sure that $\succsim_1$ and $\succsim_S$ well-defined.[35]

**Axiom** (Separability): For all $p_1, q_1, l_1, r_1 \in \Delta(Z)$ and $p_S, q_S, l_S, r_S \in (\Delta(Z))^S$, (i) $(p_1, l_S) \succsim (q_1, l_S)$ if and only if $(p_1, r_S) \succsim (q_1, r_S)$; (ii) $(l_1, p_S) \succsim (l_1, q_S)$ if and only if $(r_1, p_S) \succsim (r_1, q_S)$.

The next axiom captures inequality aversion among other agents' allocations.

**Axiom** (Quasi-Concavity) For any $p_S, q_S \in (\Delta(Z))^S$, if $p_S \sim_S q_S$, then $\frac{1}{2}p_S + \frac{1}{2}q_S \succsim_S p_S$.

**Corollary**: *The following statements are equivalent:*

*(a) $\succsim$ satisfies Quasi-Concavity, Weak Independence, and Separability as well as the axioms*

---

[35]I appreciate a referee who points out the necessity of this axiom.

*in the theorem except Independence.*

*(b) There exists an extended GU model in which $\sum_{i \in S} \alpha_i u(p_i) = \min_{\alpha_S \in C} \sum_{i \in S} \alpha_i u(p_i)$ for some $C \subset \Delta(S)$.*

**Proof:** It is easy to see the necessity of the axioms. To show the sufficiency it suffices to show the following two lemmas. First, instead of Lemma 1, we prove the next lemma by using the standard argument with the von Neumann-Morgenstern's theorem and Gilboa and Schmeidler's (1989) theorem.

**Lemma 9** *There exist a mixture linear function $u_1$ on $\Delta(Z)$ and a closed subset $C$ of $\Delta(S)$ such that (i) $u_1$ represents $\succsim_1$ on $\Delta(Z)$, (ii) there exist $\overline{z}, \underline{z} \in Z$ such that $u_1(\overline{z}) = 1 \geq u_1(p) \geq 0 = u_1(\underline{z})$ for all $p \in \Delta(Z)$, and (iii) $\sum_{i \in S} \alpha_i u(p_i) \equiv \min_{\alpha \in C} \sum_{i \in S} \alpha_i u_1(p_i)$ represents $\succsim_S$.*

Given the above $u_1$ and $u_S$, we define $\succsim^*$ in the same way as in the proof of theorem. Weak Independence of $\succsim$ on $\mathscr{A}$ implies Independence of $\succsim^*$ on $\mathscr{A}^*$.

**Lemma 10** $\succsim^*$ *satisfies Independence\*.*

**Proof of Lemma** 10: Fix $C^* \in \mathscr{A}^*$. For all $x \in [0,1]$, define $p(x) = x\delta_{\overline{z}} + (1-x)\delta_{\underline{z}}$ and $p_S(x) = (p(x))_{i \in S}$. Then, for all $\boldsymbol{u} \equiv (u_1, u_S) \in [0,1]^2$, $u_1(p(u_1)) = u_1$ and $\sum_{i \in S} \alpha_i u(p_i(u_S)) = u_S$. Define $C = \{(p(u_1), p_S(u_S)) | \boldsymbol{u} \equiv (u_1, u_S) \in C^*\}$. Then, $C$ is constant over $S$ and $\boldsymbol{u}(C) = C^*$. Therefore, by Weak Independence, $A^* \succsim^* B^* \Leftrightarrow A \succsim B \Leftrightarrow \alpha A + (1-\alpha)C \succsim \alpha B + (1-\alpha)C \Leftrightarrow \alpha A^* + (1-\alpha)C^* \succsim^* \alpha B^* + (1-\alpha)C^*$. $\square$

Since $\succsim^*$ satisfies the same properties as in the proof of the theorem, Lemma 4–8 hold in the same way. Hence, the sufficiency of the axioms holds with $u_1 = u_1(p_1)$ and $u_S = \min_{\alpha \in C} \sum_{i \in S} \alpha_i u_1(p_i)$. Therefore, Corollary holds. ∎

# D    Discussion on Experiments (Not for Publication)

In the experiments conducted by Lazear et al. (2012), we could observe that medium-level donors exit more often than low-level and high-level donors, when playing the dictator game is subsidized. In the experiments, 96 subjects (48 dictators) participated in five sequential sessions of dictator games with an exit option. Lazear et al. (2012) provided dictators with $10 as baseline endowment and, on top of that, added subsidies of $0, $1, $3, $6, and $10 to the baseline endowment in order. For each subsidy value, dictators decided whether to play the dictator game or exit. Then, the dictators decided the donation amount publicly if they did not exit. For each dictator, the left figure in Figure 3 (p. 14) of Section 8 shows the minimal subsidy needed to play the dictator game and the dictator's average donated proportion.[36] Clearly, the figure shows tendency (i).[37]

We found consistent evidence for the tendency in the earliest experiments on dictator games with an exit option conducted by Dana et al. (2006). Dana et al. (2006) provided dictators with $10 as an endowment and asked dictators the donation amount before the dictators knew that they could exit privately. When the dictators exited, they obtained $9 privately and receivers obtained $0 without knowing that this is a consequence of the dictators' choice. The right figure in Figure 3 shows the percentage of dictators who exited and their (intended) donated proportion, which clearly exhibits the tendency.[38]

In the experiment conducted by Dana et al. (2006), dictators were anonymous, while in the treatment conducted by Lazear et al. (2012), receivers could identify dictators. Hence, the consistency between these two experiments, as captured by Figure 3, would support our hypothesis: as long as playing dictator games is common knowledge among subjects, the dictator would consider the receiver's wish that the dictator should act altruistically. Hence,

---

[36]We regressed donated proportion on subsidy size. The estimated coefficient on the subsidy size is $-1.6 \cdot 10^{-4}$ ($p = 0.887$), which is not significantly different from zero. Hence, the donated proportion is statistically constant across the treatments.

[37]We made the left figure of Figure 3 based on the no-anonymity treatment in Experiment 2 in Lazear et al. (2012).

[38]We made the right figure of Figure 3 based on Figure 1 (p.197) in Dana et al. (2006).

the dictator could feel pride in acting altruistically by living up to the receiver's wish and ashamed of acting selfishly by denying their wish, even though the receiver could not identify the dictator.