LEVELS OF OTHER-REGARDING PREFERENCES AND THE STRUCTURE OF THE INTERACTION[1]

Olena Orlova*
Université Paris 1 Panthéon-Sorbonne, Universität Bielefeld

Abstract: *The paper contributes to the literature on other-regarding preferences challenging the narrow self-interest assumption. Experimental evidence confirms that the same individuals might express different other-regarding preferences in different situations or contexts. The structure of their interaction, their relative positions in it might trigger different behavioral patterns. In this paper we propose a model of multi-level other-regarding preferences assuming that different levels are actualized depending on the context in which an individual has to take her decision. We analyze the experimental trust game letting the players have multi-level preferences. Under certain parameterization and asymmetric information assumption, we show that the share given up by the leader of the game in favor of the follower is strictly monotonically increasing with altruism of the former. It is also demonstrated that utilitarian social welfare is increasing with the leader's altruism if the players are not extremely risk-averse. In the case when information for both players is incomplete, a separating equilibrium exists allowing to distinguish between leaders with different other-regarding preferences.*

Key words: *other-regarding preferences, trust game, social welfare.*

POZIOMY NIEEGOISTYCZNYCH PREFERENCJI A STRUKTURA INTERAKCJI

Streszczenie: *Praca wpisuje się w literaturę na temat nieegoistycznych preferencji, kwestionującą założenie o kierowaniu się przy podejmowaniu decyzji wyłącznie wąsko pojmowanym interesem własnym. Wyniki eksperymentów potwierdzają, że te same osoby mogą ujawniać różne nieegoistyczne preferencje*

[1] I am greatly indebted to Miguel A. Ballester for his valuable feedback and generous comments throughout the course of this work. I would also like to thank two anonymous referees for their valuable suggestions. The financial support of the European Commission is also greatly acknowledged.

* Olena Orlova, Université Paris 1 Panthéon-Sorbonne, Paris, France, Universität Bielefeld, Bielefeld, Germany, e-mail: olena.orlova@uni-bielefeld.de

DOI: 10.7206/DEC.1733-0092.93

w różnych sytuacjach czy kontekstach. Struktura interakcji między osobami i ich pozycje w niej mogą uruchamiać rozmaite wzorce postępowania. W pracy proponujemy model wielopoziomowych nieegoistycznych preferencji przy założeniu, że poszczególne poziomy są uruchamiane w zależności od kontekstu, w jakim osoba ma podjąć decyzję. Zajmujemy się eksperymentalną grą zaufania. Przy pewnej parametryzacji i założeniu wielopoziomowych preferencji graczy oraz asymetrycznej informacji pokazujemy, że kwota przekazywana przez gracza powierzającego jest ściśle rosnącą funkcją jego altruizmu. Pokazujemy także, że wzrost altruizmu gracza powierzającego podnosi dobrobyt społeczny, o ile tylko gracze nie są skrajnie niechętni ryzyku. W przypadku, gdy obaj gracze mają niekompletną informację, istnieje równowaga rozdzielająca, w której gracze powierzający z różnymi nieegoistycznymi preferencjami wybierają różne decyzje.

Słowa kluczowe: *nieegoistyczne preferencje, gra zaufania, dobrobyt społeczny.*

1. Introduction

For more than two decades economists have been challenging the classical assumption of selfish preferences. Numerous experiments and empirical observations have confirmed that other-regarding considerations play an important role in decision-making, providing scientific support for the intuitive view that concern about others constitutes a natural feature of human beings. Meant to benefit other individuals like altruism, or to harm them like spite or revenge; unconditional or conditional on the actions of others like positive or negative reciprocity; relative to others' payoffs like fairness or status-seeking – however different all these phenomena are they share an important feature: incompatibility with the self-interest assumption embedded in the homo oeconomicus model.

Understanding decisions taken in multiple economic environments is virtually impossible without acknowledgement of other-regarding preferences of the agents involved. The very existence of charitable organizations, for example, or voluntary contributions to public goods can be taken as a clear indication of the presence of other-regarding concerns. Other examples include family transfers and inheritance, international aid, cooperation and reciprocity in labor relations, or punishing "free-riders" in collective tasks, to name but a few.[2]

[2] See the surveys of Laferrere and Wolff (2006), Andreoni (2006), Rotemberg (2006), and Kanbur (2006) in the *Handbook of the Economics of Giving, Altruism, and Reciprocity*, vol. 2, for applications in family transfers, philanthropy, labor economics, and international aid.

Although definite attempts to formalize and model the actual behavior of individuals by including other-regarding considerations in utility functions have been made, there is still no unified theory which would encompass and explain all the observed phenomena.[3] Some models assume that there are different "types" of agents in the economy, concerned about others in a different way, and analyze how these agents interact with each other in a given environment.[4] They show that for a specific environment the outcome of the interaction is determined by prevalence of some but not other "types" (or sometimes even mere presence of a particular "type" suffices to determine the outcome). While this, the interplay between different individuals in one situation, is undoubtedly a very interesting and important subject for analysis, it is not less interesting to examine how the same individuals behave in different situations or contexts, since there is a reason to believe that their behavioral patterns are not constant. So far this question has not received sufficient attention in the literature, yet its careful investigation could shed light on many curious phenomena, including variations in behavioral responses of individuals to changes in framing. Instead of assuming constant other-regarding "types" and trying to calibrate the distribution of these "types" in the population by means of laboratory experiments, we propose a model of "mobile" preferences, validity of which is to be tested by means of various within-subject experimental designs.

The very fact that the same individuals might reason differently depending on the context has been confirmed in many experiments. Among other features, the framing of the situation, the structure of the economic interaction or the individuals' relative roles may lead them to think in different ways. There is rich empirical evidence confirming that, in the workplace, different motives play a role in conditioning the actions that subordinates take towards their supervisors and vice versa. While altruism can often explain behavior of the latter, reciprocity usually drives decisions of the former.[5] As well as the structural position, the order in a decision-making chain might trigger one's way of reasoning if decisions are taken sequentially by individuals. There is a wide range of economic situations in which a first-mover makes some choice concerning a second-mover, and then the latter in her turn takes a decision that influences the first-mover. Here the framing of being the first induces the use of unconditional other-regarding preferences, like altruism, while the second-mover is motivated by the framing to use preferences conditional on the treatment she has received.

[3] See, for example, Fehr and Schmidt (2006) for a good survey of theoretical models of other-regarding preferences.

[4] Thus Fehr and Schmidt (1999) examine interactions of individuals with heterogeneous other-regarding preferences in various economic environments, given by experimental economic games such as market or cooperation games.

[5] An excellent survey on other-regarding preferences in the workplace and influence of hierarchical positions is that of Rotemberg (2006). On other context effects see also Camerer and Thaler (1995).

Understanding how different other-regarding preferences are triggered would provide key insights for many economic problems. Given a set of individuals with different other-regarding preferences, how to allocate the roles or positions between them so that to maximize overall happiness? How to allocate them so that to minimize inequality of the outcome, or to get some tradeoff between overall utility surplus and inequality? Should one place more altruistic individuals as first-movers and more reciprocal as followers? In the workplace, whom to appoint as a manager so that to maximize the firm's profits? In the social choice framework, whom within a given set of candidates to give authority to take decisions and interact with a heterogeneous population so that to maximize expected overall output or overall happiness? In this paper we are trying to give partial answers to these questions.

We start by proposing a certain multi-level classification of outcome-based other-regarding preferences, and assume that individuals reason and decide using different levels of preferences in different situations or framings.[6] In particular, we focus our attention on altruistic preferences (henceforth also referred to as level 1), and reciprocal preferences conditional on the level 1 "types" of the individuals (henceforth also level 2). Following most of the literature, we define altruism as willingness to sacrifice one's own gains in favor of others' gains. As for reciprocity, we consider type-based reciprocity depending on the degree of altruism of the opponent.[7] That is, for a reciprocal person her conditional altruism is increasing both with her own unconditional altruism and that of her opponent, and in the limit case implies "paying back with the same". Thus, in our model the behavior of an individual is determined both by her constant other-regarding "type" (specific parameters of altruism and reciprocity), and by the level of preferences triggered by the structure or framing of a particular situation.

We proceed by examining the trust game, an experimental game which has a structure very appealing for our analysis.[8] The design of the trust game induces different levels of other-regarding preferences in players both because of different privileges implied by their roles and because of the order of their decisions. It seems quite reasonable to assume that the Proposer, the first-mover in the trust game and the one given discretion over the total surplus, would express altruistic (level 1) preferences, and the Responder, the second-mover influencing allocation of the surplus, would choose according to her reciprocal (level 2) preferences.

[6] In contrast to outcome-based preferences, one might think of preferences accommodating concerns for procedure, i.e. not only *what* the payoffs are, but also *how* they are generated. For the latter see Borah (2010). In this paper we consider other-regarding preferences over outcomes only.

[7] Type-based reciprocity depends on the "type" of the reference individual – altruistic, selfish, spiteful etc. – as opposed to intention-based reciprocity which implies reciprocating to the *intentions* of the opponent in a one-shot action. See, for example, Levine (1998) for a model of type-based reciprocity, and Rabin (1993) for a model of intention-based reciprocity.

[8] For description and discussion of the experiments see Berg et al. (1995).

In our model two risk-averse individuals with other-regarding preferences interact in the trust game and sequentially make their choices. Under certain parameter specification of the players' preferences and asymmetric information assumption, we show that the share given up by the Proposer in favor of the Responder at the first stage of the game is strictly monotonically increasing with the Proposer's altruism.[9] This result carries an immediate policy implication: if overall surplus is the variable to be maximized and it is determined by the first-mover's choice, then the most altruistic candidate should be assigned the first-mover's role whatever the probability distribution of the second-mover's "types" and whatever the degree of the individuals' risk-aversion is. We also examine social welfare implications of monotonicity of the optimal choice for the case of a utilitarian social welfare function. It turns out that altruistic candidates being placed as first-movers increase not only overall surplus but also overall happiness, at least when risk aversion is not extreme.

Later we relax the complete information assumption for the Responder and let both players be unaware of each other's "types". Although strict monotonicity of the Proposer's transfer suggests that the "type" of the Proposer might be revealed by her behavior, there is room for strategic considerations on the Proposer's part: different "types" might be willing to mimic each other's choices if this gives them higher utility in the end. With the introduction of a quite intuitive form of beliefs used by the Responder, we show that a separating equilibrium exists nevertheless. Thus, even without knowing the Proposer's "type" it is possible to clearly distinguish altruists from egoists by the share they send.

Moreover, as in a separating equilibrium Proposers of different "types" have no incentives to mimic each other, they send their optimal shares. This implies, in particular, that whenever a selfish Proposer gives up a non-zero share, it is done because she expects to be paid back by a high enough proportion of unconditional altruists and not because she trusts to be reciprocated as an altruist she is pretending to be.

As it is pointed by Cox (2004), the Berg et al. (1995) experimental design of the trust game "does not allow one to distinguish between transfers resulting from trust [or strategic considerations dependent on the beliefs of the Responder] and transfers resulting from altruistic other-regarding preferences. Similarly, their design does not provide data that distinguish between second-mover return transfers motivated by reciprocity and returns resulting from unconditional other-regarding preferences." Our model accounts for all the possibilities listed in the above quote. It gives combinations of parameters (distribution of other-regarding "types", specific shape of risk-aversion etc.) for which the Proposer's transfers are zero, or non-zero up to

[9] The only exception is the case when giving up everything is optimal for the Proposer regardless of her "type". Obviously, *strict* monotonicity does not hold then.

the whole endowment given up to the Responder. It also explains zero responses, as well as accounts for both reciprocal and altruistic motivation for non-zero ones.

The rest of the paper is organized as follows. In section 2 we describe the basic model as it applies to the analysis of the trust game. Section 3 presents the main results for the case of complete information for the Responder, in particular monotonicity property of the Proposer's choice. Section 4 outlines several directions for extending the model, among which continuous other-regarding parameters (subsection 4.1), incomplete information for both players (subsection 4.2) and a more general model of levels of other-regarding preferences (subsection 4.3). Finally, section 5 concludes the paper. The proofs are presented in the appendix.

2. The Model and Preview of the Analysis

2.1. Structure of the game, preferences and other-regarding "types"

In the following analysis we proceed by using a slightly modified version of the trust game introduced in Berg et al. (1995). The structure of the game is as follows. There are two players – the Proposer and the Responder – who make their choices sequentially. The Proposer receives some amount of money, normalized to 1, from the experimenter. At the first stage of the game she can voluntary give up some share $s \in [0, 1]$ of the amount to the Responder, and this share is multiplied by a constant c (so that the Responder gets cs, while the Proposer is left with $1 - s$). At the second stage the Responder is free to return any sum $k \in [0, cs]$ to the Proposer. Thus, the final monetary payoffs are $1 - s + k$ for the Proposer and $cs - k$ for the Responder.[10]

The Berg et al. (1995) design of the game differs from our in one detail: at the first stage of their game the Responder receives from the experimenter the same amount of money as the Proposer does. We eliminate this transfer, as otherwise an altruistic Proposer would only share because it increases the total surplus (for now we omit strategic considerations related to the presence of reciprocal Responders). If the amount she sends to the Responder was not multiplied, she would not share, as long as the players are concerned about own payoff not less than about their opponent's and utility of money is strictly concave, which will be assumed below.

In our model the level k utility function of individual i interacting with individual j takes the following form:

$$U_i^k(x_i, x_j) = \frac{1}{1+\beta_{ij}^k} u(x_i) + \frac{\beta_{ij}^k}{1+\beta_{ij}^k} u(x_j),$$

[10] The constant c is greater than 1, which means that the Proposer can increase the total surplus by sharing more. For the ease of exposition, we take c equal to 3, which is a standard value in experimental designs.

where $(x_i, x_j) \in \mathbb{R}^2$ is an outcome assigning material payoffs x_i and x_j to individuals i and j respectively, and $\beta_{ij}^k \in [0,1]$ is the relative weight of the other's payoff in i's utility function at level k.

In the models of other-regarding preferences different assumptions are made on whether other-regarding motives are applied to material payoffs or to utility payoffs.[11] For our theoretical study we adopt the second approach, interpreting it for the case of monetary payoffs in the trust game as concave utility of money. Thus, we assume strictly increasing, twice continuously differentiable and strictly concave function u.

Note also that, following most of the literature, we assume that concern about the other cannot exceed concern about self. In this paper we normalize the weights of the utility payoffs in order to allow for interpersonal comparisons and aggregation of other-regarding utilities, which is necessary for deriving policy implications such as allocating the roles between individuals so that to maximize overall welfare.[12]

For the purpose of our analysis we assume a specific parameterization, such that

$$\beta_{ij}^k = \begin{cases} a_i & if\ k = 1\ (altruistic\, preferences) \\ \frac{a_i + \mu_i a_j}{1+\mu_i} & if\ k = 2\ (reciprocal\, preferences) \end{cases}$$

Here k – the level of other-regarding preferences – can be either unconditional (level 1) in the form of altruism, or conditional (level 2) on the altruistic "type" of the opponent. At each level one new parameter is introduced: $a \in \{0, 1\}$ is the altruism parameter, and $\mu \in \{0, +\infty\}$ is the parameter of reciprocity.

The altruism parameter $a_i \in \{0, 1\}$ is the measure of altruism of individual i. We start by assuming that it can take only two values: if $a_i = 0$ then i's preferences are purely selfish, and if $a_i = 1$ then i is a pure altruist, that is, concerned about the other's payoff in the same way as about own.

The parameter $\mu_i \in \{0, +\infty\}$ is the measure of reciprocity of individual i. Again, we assume for now two values: $\mu_i = 0$ means that the individual is non-reciprocal and behaves according to her level 1 altruism parameter, while $\mu_i = +\infty$ indicates a pure reciprocator who disregards her own level 1 characteristic and acts according to that of her opponent.

The expression for reciprocal preferences is taken from Levine (1998), but we have changed the ranges of parameter values. In particular, we do not allow a to take negative values, as we are primarily interested in altruism and positive reciprocity

[11] While the first is more common for experimental economics literature, dealing with monetary payoffs given out in the experiments, the second is usual for the literature on applications.

[12] It should be mentioned, however, that there is no consensus on the correct formulation of social welfare. See, among others, Decerf and Van der Linden (2016) and Treibich (2014).

phenomena; exhibition of spitefulness and thus negative reciprocity is ruled out by the design of the game. The upper bound for admissible values of the reciprocity parameter μ is also modified. Levine sets it equal to 1, which corresponds to averaging own and the opponent's altruism. However, we find it more reasonable to regard $\mu = +\infty$ as pure reciprocity, implying "paying back with the same" – being as kind to the opponent as she is to others.

To sum up, there are four different "types" of individuals in our model:

- non-reciprocal selfish (with $a = 0, \mu = 0$),
- non-reciprocal altruistic (with $a = 1, \mu = 0$),
- reciprocal selfish (with $a = 0, \mu = +\infty$),
- reciprocal altruistic (with $a = 1, \mu = +\infty$).

If an individual belongs to the first "type", she behaves selfishly whatever level of other-regarding preferences is forced on her by the framing. For individuals of the second "type" choices at either level are altruistic, and the "type" of her opponent does not matter for her choices. For selfish reciprocal ones, the third "type", preferences at level 1 are selfish, while at level 2 they reflect the other individual's unconditional altruism ($\beta_{ij}^2 = a_j$). Finally, individuals of the last "type" behave selfishly only towards selfish ones, otherwise they display altruism.

As we have argued previously, the framing of being a first-mover and being given a greater discretion induces the Proposer to use level 1 altruistic preferences, while the framing of being a second-mover motivates the Responder to choose with level 2 reciprocal preferences. Hence, when the final monetary payoffs $x_p = 1 - s + k$ for the Proposer and $x_R = 3s - k$ for the Responder are realized, the utilities of the players are respectively:

$$U_P^1(x_P, x_R) = \frac{1}{1+a_P} u(x_P) + \frac{a_P}{1+a_P} u(x_R),$$

and

$$U_R^2(x_P, x_R) = \frac{1}{1+\beta_R^2} u(x_R) + \frac{\beta_R^2}{1+\beta_R^2} u(x_P)$$

where $\beta_R^2 = \frac{a_R + \mu_R a_P}{1+\mu_R}$.

Let us note that the Proposer, being a first-mover in the game, does not have a chance to reveal her reciprocity parameter; thereby we can only differentiate between selfish (first and third "types") and altruistic (second and fourth "types") Proposers. At the same time, Responders fall into three groups – non-reciprocal selfish (first "type"), non-reciprocal altruistic (second "type") and reciprocal (third and fourth "types").[13, 14]

2.2. Complete information

We assume that the structure of the game, the players' payoffs, and available actions are common knowledge. The players are also aware of the fact that the framing induces Proposers to behave according to their unconditional preferences, while it drives Responders, whenever they are reciprocal, to express reciprocity. The only information which can be held privately is the exact "type" of the player. Such a setting seems to be quite natural for many economic situations, when "the rules" are announced to all and possible consequences of interactions are understood intuitively, yet the actual preferences of the opponent might be unknown ex ante.

We start the analysis by assuming complete information for both players, implying no uncertainty about the "types" of the players. Since their positions in the game trigger different levels of preferences in them, the Proposer chooses an amount that maximizes her corresponding level 1 utility function:

$$U_{SP}(s) \equiv u\big(1 - s + k^*(s)\big)$$

for a selfish Proposer, and

$$U_{AP}(s) \equiv \frac{1}{2}u\big(1 - s + k^*(s)\big) + \frac{1}{2}u\big(3s - k^*(s)\big)$$

for an altruistic one, where $k^*(s)$ is the Responder's response function. The Responder acts according to her level 2 preferences, and hence the response functions $k^*(s)$ can be simply obtained by maximizing

$$U_{SR}(k) \equiv u(3s - k)$$

for a non-reciprocal selfish Responder,

$$U_{AR}(k) \equiv \tfrac{1}{2}u(3s - k) + \tfrac{1}{2}u(1 - s + k)$$[15]

for a non-reciprocal altruist, and

$$U_{RR}(k) = \tfrac{1}{1+a_P}u(3s - k) + \tfrac{a_P}{1+a_P}u(1 - s + k)$$

[13] With $\mu_R = +\infty$ the altruism parameter a_R of the Responder becomes irrelevant.

[14] Henceforth we will also refer to non-reciprocal selfish and non-reciprocal altruistic Responders as "unconditionally selfish" and "unconditionally altruistic" respectively.

[15] Note that the level 2 utility functions (à la Levine) for non-reciprocal ($\mu = 0$) players coincide with the level 1 utility functions for the same players.

for a reciprocal Responder. Note that for the latter, her concern about the Proposer depends on the Proposer's altruism, defined by a_P, which can take values 0 or 1.

Thus, we can observe either selfish behavior, implying that the player is concerned solely about her own payoff, or altruistic behavior, maximizing the equally weighted sum of both players' payoffs, or reciprocal behavior which can be selfish or altruistic conditional on the treatment received from the first-mover.

Under complete information assumption for both players, we can easily obtain optimal transfers, s^* and k^*, and get basic intuition about optimality of these choices for the players of different "types", which is done in subsection 3.1.

2.3. Incomplete information for the Proposer

In multiple economic situations the complete information assumption cannot be justified. It often happens that the other-regarding preferences of the opponent, sometimes even her identity, are unknown. There might be some information available about the distribution of possible preferences which comes, for example, from personal experience of previous interactions with heterogeneous individuals, and this information is the only grounds for the decision of the first-mover. The position of the second-mover is different, however, as the treatment she receives gives her an additional source for forming her belief about the first-mover's "type". If only first-movers can be differentiated by their behavior, this, and not the general distribution of other-regarding "types", is key for the follower's decision.

Yet we leave the case of incomplete information for both players for later consideration, and focus our attention on the asymmetric information case where the leader is unaware of the follower's actual preferences, but the latter knows exactly the other-regarding "type" of her opponent.[16] Such situations arise naturally if the leader is in a position of power and interacts with a heterogeneous population of the followers. Leaders are continuously in the public eye and their personalities are revealed more to members of the public, while the masses usually appear to leaders rather on statistical basis than on personal one. Take, for instance, a local authority representative using her power for providing additional assistance to local residents who might (or might not) stage a demonstration in support of her if needed. Alternatively, in the workplace, consider a manager giving support to her subordinates who might (or might not) pay back by raising their effort and thus increasing the manager's bonus payment.

From the technical side, if we keep the complete information assumption for the Responder, it allows to isolate the Proposer's intrinsic other-regarding considerations

[16] The assumption of complete information for the follower (the Responder) is relaxed in subsection 4.2.

given by her altruism parameter from her strategic considerations aimed at winning favor with reciprocal Responders. In this case reciprocal Responders would treat a selfish Proposer as selfish even if they received a generous transfer from her, because what matters for them is the actual "type" of their opponent.

Thus, we consider complete information for the Responder, but let the Proposer be unaware of the Responder's other-regarding "type". Assume the following probability distribution of these "types": let p be the proportion of non-reciprocal altruistic Responders, q – the proportion of reciprocal ones (selfish as well as altruistic reciprocators), and 1 – p – q – non-reciprocal selfish Responders. This distribution is common knowledge before the game. In all the other respects, both players share the same information: the structure of the game and payoffs are communicated to them beforehand, and the form of the utility functions except for the exact parameter values is also known to both.

Now, since the Proposer faces uncertainty about the Responder's "type", she maximizes her expected utility, based on the probability distribution of these "types":

$$\mathbb{E}U_{SP}(s) = pu\big(1 - s + k_{AR}^{*}(s)\big) + (1-p)u\big(1 - s + k_{SR}^{*}(s)\big)$$

for a selfish Proposer, and

$$\mathbb{E}U_{AP}(s) = (p+q)\left[\frac{1}{2}u\big(1 - s + k_{AR}^{*}(s)\big) + \frac{1}{2}u\big(3s - k_{AR}^{*}(s)\big)\right] +$$
$$(1-p-q)\left[\frac{1}{2}u\big(1 - s + k_{SR}^{*}(s)\big) + \frac{1}{2}u\big(3s - k_{SR}^{*}(s)\big)\right]$$

for an altruistic one.[17]

The utility functions and maximization problems for Responders of different "types" are the same as in subsection 2.2, since information for the Responder is complete. Optimal transfers s_S^* for a selfish Proposer and s_A^* for an altruistic one with incomplete information are given by lemmas 4 and 5 in subsection 3.2.[18]

2.4. Social welfare function

As it has been mentioned in the introduction, several social welfare implications arise immediately from our model. Among them there is the question of the optimal allocation of positions or roles between individuals. It might be that given a set of individuals with heterogeneous other-regarding preferences, a social planner has to

[17] Note that a reciprocal Responder, knowing the Proposer's "type", reacts as a non-reciprocal selfish Responder if the Proposer is selfish or as a non-reciprocal altruistic Responder if the Proposer is altruistic.

[18] Since the incomplete information case is the matter of further more detailed analysis, we use subscripts to differentiate between a selfish and an altruistic Proposer's optimal choices, what we do not do in the complete information case.

put them into a given structure so that to maximize aggregate welfare. As different roles activate different levels of other-regarding preferences, and thus induce different behavior in the same individuals occupying one position or another, social welfare varies with allocations, and the problem is to find the optimal one. Alternatively, we could have a set of candidates for the leadership role and the population of followers, and the problem is to choose the candidate who maximizes expected aggregate welfare from her interactions with the population. By way of illustration, consider the following problem: knowing the "types" of team members whom to select as a team leader so that to maximize overall satisfaction from work.

As we have defined the utility function form which allows for comparison of the utilities of different "types" of individuals (recall, for this we let the value of an egalitarian outcome (x, x) be the same for any individual), let us define a social welfare function W as a sum of equally weighted utilities of the individuals participating in the game. With incomplete information for the Proposer it becomes the sum of the Proposer's expected utility and the utility of the Responder:

$$W = \mathbb{E}U_P(s^*) + U_R\big(k^*(s^*)\big).$$

We assume that a social planner is empowered to choose the Proposer from a given set of candidates of different "types", although she has no control over Responders. Anyone, in accordance with the probability distribution of the Responder's "types", may become the Responder. One might think of repeated yet independent interactions of the same Proposer with different Responders randomly drawn from the population for every single interaction.[19] Hence, the (utilitarian) social welfare function takes the form:

$$W = \mathbb{E}U_P(s^*) + pU_{AR}\big(k^*_{AR}(s^*)\big) + qU_{RR}\big(k^*_{RR}(s^*)\big) + (1-p-q)U_{SR}\big(k^*_{SR}(s^*)\big).$$

After obtaining optimal solutions for different "types" of Proposers, we can compute and compare social welfare for the cases when altruists or selfish ones are leading the game, and make some policy recommendations concerning the best choice of the leader. This is done in subsection 3.2.4.

3. Analysis of the Model

3.1. Complete information

In this subsection we derive the optimal share s^* to be sent by the Proposer and the optimal return amount k^* to be paid back by the Responder under the assumption

[19] Again, interpretations might be various: appointing a manager interacting with subordinates in the workplace, choosing the local authority for communication and cooperation with residents, etc.

of complete information for both players. Let us proceed backwards, starting with response functions of different “types” of Responders.

From the problem of a non-reciprocal selfish Responder,

$$\max_{k\in[0,3s]} u(3s - k),$$

we get that the optimal amount to be returned is $k^* = 0$ whatever is the received amount s and the “type” of the Proposer.

A non-reciprocal altruistic Responder’s problem is

$$\max_{k\in[0,3s]} \frac{1}{2}u(3s - k) + \frac{1}{2}u(1 - s + k),$$

which gives $k^* = \frac{4s-1}{2}$ if $s > \frac{1}{4}$, or $k^* = 0$ otherwise.[20] Thus, an unconditional altruist would try to equalize the final payoffs for both players by sending back $k^* = \frac{4s-1}{2}$ whenever the amount she has after the first stage of the game is greater than the one left for the Proposer, that is, whenever $s > \frac{1}{4}$.

Finally, a reciprocal Responder solves one of the two maximization problems above, depending on the “type” of the Proposer she interacts with. She would behave in the same way as an unconditionally selfish Responder if she encounters a selfish Proposer, or in the same way as an unconditionally altruistic Responder if the Proposer is altruistic.

The above results are summarized in the following lemma.

LEMMA 1:

If information for the Responder is complete, the response functions are the following:

- *for an unconditionally selfish Responder* $k^*_{SR}(s) = 0$,
- *for an unconditionally altruistic Responder* $k^*_{AR}(s) = \begin{cases} \frac{4s-1}{2} & if\ s > \frac{1}{4} \\ 0 & otherwise \end{cases}$, *and*
- *for a reciprocal Responder* $k^*_{RR}(s) = \begin{cases} k^*_{SR}(s) & if\ the\ Proposer\ is\ selfish \\ k^*_{AR}(s) & if\ the\ Proposer\ is\ altruistic \end{cases}$.

From this simple analysis it becomes obvious that if $s \le \frac{1}{4}$, then regardless of the Proposer’s “type”, every Responder finds it optimal to send back nothing. If $s > \frac{1}{4}$, then a non-reciprocal altruistic Responder always responds with $k^* = \frac{4s-1}{2}$,

[20] Note that the response function of a non-reciprocal altruistic Responder is continuous, and in particular, it is continuous at $\frac{1}{4}$. The response functions of Responders of the other “types” are also continuous, which implies continuity of the Proposer’s utility functions on [0, 1], as well as continuity of the Proposer’s expected utility functions in the case of incomplete information (see the next subsection).

a reciprocal Responder responds in such a way only to an altruistic Proposer, while giving nothing to a selfish one, and a non-reciprocal selfish Responder never pays back.

Now, let us solve the Proposer's problem for different "types" of Proposers. Being aware of the Responder's "type", she anticipates the response she would get, and maximizes her utility without uncertainty.

A selfish Proposer's problem is

$$\max_{s\in[0,1]} u\big(1 - s + k^*(s)\big).$$

From an unconditionally selfish or from reciprocal Responder she would always receive back $k^* = 0$, which implies an optimal transfer $s^* = 0$. If the Responder is unconditionally altruistic, the response function is $k^*(s) = \begin{cases}\frac{4s-1}{2} & if\ s > \frac{1}{4}\\ 0 & otherwise\end{cases}$. Thus, a selfish Proposer would compare her utility from sending nothing, which is $u(1)$, with her utility from sending some $s > \frac{1}{4}$ and receiving back $\frac{4s-1}{2}$, which is $u\left(s + \frac{1}{2}\right)$. Note that $u\left(s + \frac{1}{2}\right)$ on $(\frac{1}{4}, 1]$ is maximized at $s = 1$, and $u(1) < u\left(\frac{3}{2}\right)$. Hence, the Proposer's optimal choice would be $s^* = 1$. The intuition for this result is straightforward: whatever the Proposer gives up at the first stage of the game is tripled, and the total sum is shared equally by an unconditionally altruistic Responder (if the share given up is higher than $\frac{1}{4}$); since a half of a maximal possible total sum of 3 is more valuable than the whole endowment of 1 left for herself, a selfish Proposer would transfer everything.

Lemma 2 summarizes the solution to a selfish Proposer's problem.

LEMMA 2:

In the case of complete information for both players, the optimal share sent by a selfish Proposer is $s^* = \begin{cases}0 & if\ the\ Responder\ is\ unconditionally\ selfish\ or\ reciprocal\\ 1 & if\ the\ Responder\ is\ unconditionally\ altruistic\end{cases}$.

In her turn, an altruistic Proposer solves the following problem:

$$\max_{s\in[0,1]} \frac{1}{2}u\big(1 - s + k^*(s)\big) + \frac{1}{2}u\big(3s - k^*(s)\big).$$

If her opponent is selfish, she should expect nothing to be returned, and thus maximizes $\frac{1}{2}u(1-s) + \frac{1}{2}u(3s)$. If we look at it over the real line, this function is strictly concave and attains its maximum either at some $s \in [0,1]$ or at $s > 1$.[21] Yet the Proposer's problem is constrained, and these two possibilities have to be treated separately. If the first derivative of the objective function is non-negative at 1, i.e.

[21] It cannot attain its maximum at some $s < 0$, as it is strictly concave and its first derivative at $s = 0$ is positive.

if $\frac{u'(3)}{u'(0)} \geq \frac{1}{3}$, then choosing $s^* = 1$ would be optimal. Otherwise the optimal share is given by the first-order condition: $\frac{u'(3s^*)}{u'(1-s^*)} = \frac{1}{3}$. Obviously, in this case the Proposer's choice is determined by the curvature of u: the higher risk aversion is, the lower the optimal transfer s^* would be. However, in any case it is higher than $\frac{1}{4}$, as every unit given up by the Proposer triples for the Responder, and thus at least up to $\frac{1}{4}$ (when the payoffs of both players become equal) marginal increase in utility from raising the Responder's payoff outweighs marginal decrease in utility from reducing at the same time the payoff of the Proposer (remember that an altruistic Proposer values equally own and her opponent's gains).

When an altruistic Proposer faces a reciprocal or an unconditionally altruistic Responder, she maximizes

$$U_{AP}(s) = \begin{cases} \frac{1}{2}u(1-s) + \frac{1}{2}u(3s) & if\ s \leq \frac{1}{4} \\ u\left(s + \frac{1}{2}\right) & if\ s > \frac{1}{4} \end{cases}.$$

It can be easily shown that $s = \frac{1}{4}$ maximizes the Proposer's utility on $[0, \frac{1}{4}]$, and $s = 1$ is optimal on $(\frac{1}{4}, 1]$. Comparing $U_{AP}\left(\frac{1}{4}\right) = u\left(\frac{3}{4}\right)$ with $U_{AP}(1) = u\left(\frac{3}{2}\right)$, an altruistic Proposer chooses $s^* = 1$ as the optimal share. The intuition for this is again straightforward. Firstly, choosing $\frac{1}{4}$ is better than any $s < \frac{1}{4}$, as it gives the most egalitarian allocation and the larger total surplus. Secondly, knowing for sure that the total surplus will be shared equally whenever she transfers more than $\frac{1}{4}$, the Proposer maximizes her utility by transferring everything she has, as then the total surplus is maximized.

The solution to the altruistic Proposer's problem is summarized below.

LEMMA 3:

In the case of complete information for both players, the optimal share sent by an altruistic Proposer is

$$s^* = \begin{cases} s\ \ given\ by\ \frac{u'(3s)}{u'(1-s)} = \frac{1}{3}\ if\ the\ Responder\ is\ unconditionally\ selfish\ and\ \frac{u'(3)}{u'(0)} < \frac{1}{3} \\ 1 \qquad if\ the\ Responder\ is\ unconditionally\ selfish\ and\ \frac{u'(3)}{u'(0)} \geq \frac{1}{3} \\ \qquad or\ if\ the\ Responder\ is\ unconditionally\ altruistic\ or\ reciprocal \end{cases}$$

3.2. Incomplete information for the Proposer

Having formed the basic intuition, we move to the less trivial incomplete information case. We assume that the Proposer knows only the probabilities of meeting different "types" of Responders: p, q and $1 - p - q$ for unconditionally altruistic, reciprocal and non-reciprocal selfish Responders respectively. Thus, she faces uncertainty and has to maximize her expected utility, as stated in subsection 2.3.

As for the Responder, before making her choice she is aware of whom she is interacting with – an altruist or a selfish person – whatever the actions of the Proposer have been. Thus, the response functions of the different "types" of Responders are the same as in the previous subsection, given by lemma 1.

3.2.1. Selfish Proposer. Let us turn to the Proposer's problem. If the Proposer is selfish, then substituting the response functions of the different "types" of Responders, we get the expected utility function to be maximized by a selfish Proposer:

$$\mathbb{E}U_{SP}(s) = \begin{cases} u(1-s) & if\ s < \frac{1}{4} \\ pu\left(s+\frac{1}{2}\right) + (1-p)u(1-s) & if\ s \geq \frac{1}{4} \end{cases}.$$

If the share sent to the Responder is below $\frac{1}{4}$, and thus after the first stage of the game the Responder's payoff is already lower than the Proposer's one, then no Responder would pay back, and the Proposer's utility is simply $u(1 - s)$. Obviously, in this case sending zero would be optimal, $s^* = 0$, and maximal possible utility which can be obtained in this interval is $u(1)$. However, if the share sent is above $\frac{1}{4}$, some Responders, namely altruistic ones, would pay back so that to make the final payoffs egalitarian and equal to $s + \frac{1}{2}$, while the others would keep the whole transfer for themselves, leaving the Proposer with $1 - s$. Thus, now a selfish Proposer choosing $s \geq \frac{1}{4}$ enters a lottery in which she could get the higher utility payoff $u\left(s+\frac{1}{2}\right)$ with probability p and the lower payoff $u(1 - s)$ with probability $1 - p$. Note that variance of the lottery grows with s.

In order to get the solution to her problem, the Proposer has to compare her utility from choosing $s^* = 0$ with the maximum of her expected utility if she sends $s \geq \frac{1}{4}$. The latter is determined by the first-order condition unless it gives the value outside the interval $[\frac{1}{4}, 1]$ which depends on the curvature of u.[22] The following three cases are possible:

[22] The first-order condition implies: $pu'\left(s+\frac{1}{2}\right) - (1-p)u'(1-s) = 0$, or equivalently $\frac{u'\left(s+\frac{1}{2}\right)}{u'(1-s)} = \frac{1-p}{p}$. The second-order condition: $pu''\left(s+\frac{1}{2}\right) + (1-p)u''(1-s) < 0$. Thus, the objective function is strictly concave and attains its maximum on $[\frac{1}{4}, 1]$.

(i) The first-order condition gives a value above or equal to 1. Then $s^* = 1$, as the expected utility function being maximized on $[\frac{1}{4}, 1]$ is concave.

(ii) The first-order condition gives a value $s \in (\frac{1}{4}, 1)$. Then it determines s^*.

(iii) The first-order condition gives a value below or equal to $\frac{1}{4}$. Then $s^* = \frac{1}{4}$, again because of concavity of the expected utility function.

As it has been already mentioned, lower risk aversion, that is, the less sharp curvature of u, implies the higher optimal share s^*. If risk aversion is that low that $\frac{u'\left(\frac{3}{2}\right)}{u'(0)} \geq \frac{1-p}{p}$, meaning that the expected utility function is still not decreasing at 1, then we are in the case (i) and $s^* = 1$. If risk aversion is higher, and $\frac{u'\left(\frac{3}{2}\right)}{u'(0)} < \frac{1-p}{p}$, then we are either in the case (ii) where s^* is given by the first-order condition, or in the case (iii) where $s^* = \frac{1}{4}$.

Now, comparing utility at zero, which is $u(1)$, with expected utility at s^* optimal on $[\frac{1}{4}, 1]$, we get the solution to a selfish Proposer's problem – the optimal transfer s_S^*.

In order to simplify future reference we denote the following sets of conditions:

$$\frac{u'\left(\frac{3}{2}\right)}{u'(0)} \geq \frac{1-p}{p} \text{ and } pu\left(\frac{3}{2}\right) + (1-p)u(0) > u(1), \tag{1}$$

and

$$\frac{u'\left(\frac{3}{2}\right)}{u'(0)} < \frac{1-p}{p} \text{ and } pu\left(s+\frac{1}{2}\right) + (1-p)u(1-s) > u(1). \tag{2}$$

Note that if we are in the case (iii), conditions (2) do not hold, since $u(1)$ is greater than $\mathbb{E}U_{SP}\left(\frac{1}{4}\right) = u\left(\frac{3}{4}\right)$, and thus the optimal share is $s_S^* = 0$.

Let us sum up the above in lemma 4.

LEMMA 4:

In the case of complete information for the Responder and incomplete for the Proposer, when the latter knows only a probability distribution of the possible "types" of her opponent, the optimal share sent by a selfish Proposer is

$$s_S^* = \begin{cases} 1 & if\ conditions\ (1)\ hold \\ s\ \ given\ by\ \frac{u'\left(s+\frac{1}{2}\right)}{u'(1-s)} = \frac{1-p}{p} & if\ conditions\ (2)\ hold \cdot \\ 0 & otherwise \end{cases}$$

We assume no ties, meaning that if two different shares s give the same, maximal on [0, 1] utility to a player, she chooses the lowest share. With this assumption and strict concavity of u, the solution to the Proposer's problem is unique.

By means of some simple analysis, we can specify further the solution. In particular, we can show that unless the proportion of unconditional altruists among possible Responders is sufficiently high (at least $\frac{2}{3}$) a selfish Proposer would not transfer anything. On the other hand, whenever she decides to transfer a non-zero share, this share is above one half. These two refinements of the result obtained in lemma 4 are summarized in the following remarks. The rigorous proofs can be found in the appendix.

Remark 1: *If the proportion of unconditionally altruistic Responders is not greater than two thirds ($p \leq \frac{2}{3}$), then the optimal share sent by a selfish Proposer is zero: $s_S^* = 0$.*

Remark 2: *If conditions (2) hold, and thus the optimal share s_S^* sent by a selfish Proposer is given by $\frac{u'\left(s+\frac{1}{2}\right)}{u'(1-s)} = \frac{1-p}{p}$, then it belongs to the interval $(\frac{1}{2}, 1)$.*

3.2.2. Altruistic Proposer. For an altruistic Proposer, concerned both about own payoff and the Responder's payoff, the expected utility to be maximized is the following:

$$\mathbb{E}U_{AP}(s) = \begin{cases} \frac{1}{2}u(1-s) + \frac{1}{2}u(3s) & if\ s \leq \frac{1}{4} \\ (p+q)u\left(s+\frac{1}{2}\right) + (1-p-q)\left[\frac{1}{2}u(1-s) + \frac{1}{2}u(3s)\right] & if\ s > \frac{1}{4} \end{cases}.$$

If she gives up a share below or equal to $\frac{1}{4}$, no Responder would pay back, thus the final monetary payoffs are 1 – *s* for the Proposer and 3*s* for the Responder, receiving the same weights in the Proposer's utility function. If the Proposer offers a more generous share, advantageous for the Responder, then she has a chance to get back some money from altruistic and reciprocal Responders, although selfish would still leave everything for themselves. The utility of an altruistic Proposer would be then $u\left(s+\frac{1}{2}\right)$ with probability $p + q$ or, as in the previous case, $\frac{1}{2}u(1-s) + \frac{1}{2}u(3s)$ with probability 1 – p – q. Because of strict concavity of u, the first outcome with egalitarian payoffs is always better. Moreover, it follows that the optimal value of the objective function on the second interval is always higher than that on the first interval; thus the solution to the Proposer's problem is determined by the optimum on $(\frac{1}{4}, 1]$.[23]

[23] Recall continuity of the expected utility function (footnote 20), implying that $(p+q)u\left(s+\frac{1}{2}\right) + (1-p-q)\left[\frac{1}{2}u(1-s) + \frac{1}{2}u(3s)\right] = \frac{1}{2}u(1-s) + \frac{1}{2}u(3s)$ at $\frac{1}{4}$. Then note that on both intervals, $[0, \frac{1}{4}]$ and $(\frac{1}{4}, 1]$, the objective function is strictly concave. Lastly, both right and left first derivatives are positive at $\frac{1}{4}$, thus the optimum on $[0, \frac{1}{4}]$ is attained at $\frac{1}{4}$, and on $(\frac{1}{4}, 1]$ the optimum gives even higher value of the objective function, because of the above remarks concerning its continuity and concavity.

The optimum on $(\frac{1}{4}, 1]$, in its turn, is determined by the first-order condition, unless it gives the value outside the interval.[24] This value is certainly above $\frac{1}{4}$, but it might reach 1 or even above if either

$$u'(0) \leq 3u'(3) \text{ or} \left(u'(0) > 3u'(3) \text{ and } \frac{u'\left(\frac{3}{2}\right)}{u'(0)-3u'(3)} \geq \frac{1-p-q}{2(p+q)} \right) \tag{3}$$

holds.[25]

The solution to an altruistic Proposer's problem is summarized in lemma 5, and a quite straightforward specification of the result is given in following remark.

LEMMA 5:

In the case of complete information for the Responder and incomplete for the Proposer, when the latter knows only a probability distribution of the possible "types" of her opponent, the optimal share sent by an altruistic Proposer is

$$s_A^* = \begin{cases} 1 & if\ conditions\ (3)\ hold \\ s\ \ given\ by\ \frac{u'\left(s+\frac{1}{2}\right)}{u'(1-s)-3u'(3s)} = \frac{1-p-q}{2(p+q)} & otherwise \end{cases}.$$

Again, we should note that strict concavity of u implies uniqueness of the solution to an altruistic Proposer's problem.

Remark 3: *If conditions (3) do not hold, and thus the optimal share s_A^* sent by an altruistic Proposer is given by $\frac{u'\left(s+\frac{1}{2}\right)}{u'(1-s)-3u'(3s)} = \frac{1-p-q}{2(p+q)}$, then it belongs to the interval $(\frac{1}{4}, 1)$.*

3.2.3. Monotonicity of the optimal choice. Under complete information for the Responder and incomplete for the Proposer we have formulated the optimal solutions to a selfish and an altruistic Proposer's problems. We assumed that ties, if any, are broken in favor of a lower value. Consequently, for every particular function u and a set of values of p and q, there exists a unique optimal share $s_S^* \in \{0\} \cup (\frac{1}{2}, 1]$ to be

[24] The first-order condition implies: $(p+q)u'\left(s+\frac{1}{2}\right) + \frac{1-p-q}{2}[-u'(1-s)+3u'(3s)] = 0$, or equivalently $\frac{u'\left(s+\frac{1}{2}\right)}{u'(1-s)-3u'(3s)} = \frac{1-p-q}{2(p+q)}$.

[25] From the first-order condition it follows that $u'(1-s) - 3u'(3s) > 0$, implying that the value it gives is above $\frac{1}{4}$. However, this value might be equal to 1 or above, if the first derivative of the objective function is non-negative at 1, that is, if $(p+q)u'\left(\frac{3}{2}\right) + \frac{1-p-q}{2}[-u'(0)+3u'(3)] \geq 0$, which splits into two cases – either $u'(0) \leq 3u'(3)$, or $u'(0) > 3u'(3)$ and $\frac{u'\left(\frac{3}{2}\right)}{u'(0)-3u'(3)} \geq \frac{1-p-q}{2(p+q)}$.

sent by a selfish Proposer and a unique optimal share $s_A^* \in (\frac{1}{4}, 1]$ to be sent by an altruistic Proposer.

Intuitively, we would not expect the share given up by a selfish Proposer to be greater than that given up by an altruistic Proposer. Recall that in the complete information case, a non-reciprocal selfish Responder receives nothing from a selfish Proposer, while an altruistic Proposer shares some positive amount even if she knows she will get nothing back. With a reciprocal Responder a selfish Proposer shares nothing as well, while an altruistic Proposer, quite the contrary, gives up everything. Finally, a non-reciprocal altruistic Responder receives the whole endowment from either selfish or altruistic Proposer, since she always pays back. Thus, in the complete information case the Proposer's optimal choice is increasing with her altruism. Now we want to prove the same for the case when the Proposer does not have complete information.

This non-trivial yet very important and intuitive result is stated in proposition 1. It tells that under no conditions a selfish Proposer would transfer more than an altruistic Proposer, that is, the share sent to the Responder is monotonically increasing with the Proposer's altruism. The proof of the proposition is presented in the appendix.

PROPOSITION 1 (Monotonicity Property):

In the case of complete information for the Responder and incomplete for the Proposer, when the latter knows only a probability distribution of the possible "types" of her opponent, the optimal share sent by a selfish Proposer cannot be greater than the optimal share sent by an altruistic Proposer: $s_S^* \leq s_A^*$.

Moreover, whenever $s_S^* \neq 1$, *the optimal share sent by a selfish Proposer is strictly lower than the optimal share sent by an altruistic Proposer:* $s_S^* < s_A^*$.

Thus, if for some function u and a set of values of p and q a selfish Proposer decides to give up everything ($s_S^* = 1$), then an altruistic Proposer definitely gives up everything as well ($s_A^* = 1$). If, however, the former leaves something for herself ($s_S^* < 1$), then the latter always leaves less, i.e. strict monotonicity takes place: $s_S^* < s_A^*$. The last statement is obvious if $s_S^* = 0$, yet it is also true if s_S^* is an interior solution, as shown in the appendix.

Together with remark 2, implying that whenever a selfish Proposer transfers a non-zero share this share is above $\frac{1}{2}$, monotonicity property gives the following corollary:

Corollary: *Whenever the optimal share sent by a selfish Proposer is non-zero, the optimal share sent by an altruistic Proposer is greater than one half:*

$$s_S^* > 0 \Longrightarrow s_A^* > \frac{1}{2}.$$

Later, as an extension, we will show that even if information for the Responder is incomplete as well, with an intuitive form of beliefs on the Proposer's "type", a separating equilibrium exists in which a selfish Proposer always chooses a transfer below the optimal choice of an altruistic one. Thus, monotonicity of the Proposer's choice can be sustained even with incomplete information for both players.

3.2.4. Social welfare implications. Let us turn to the implications of the above analysis. In subsection 2.4 we have defined a social welfare function measuring expected overall happiness of the players. It would be interesting to compare social welfare (as it is defined above) for the cases when the Proposer is selfish and when she is altruistic: $W_{/SP} = \mathbb{E}U_{SP}(s_S^*) + \mathbb{E}U_R\big(k^*(s_S^*)\big)$ versus $W_{/AP} = \mathbb{E}U_{AP}(s_A^*) + \mathbb{E}U_R\big(k^*(s_A^*)\big)$. Monotonicity property (proposition 1) implies that overall surplus increases with the Proposer's altruism, but does the same hold for overall happiness?

It appears that if the risk aversion of the players is not particularly high, to be precise, that $\frac{1}{2}u(1-s_S^*) + \frac{1}{2}u(3s_S^*) < \frac{1}{4}u(1-s_A^*) + \frac{3}{4}u(3s_A^*)$, then monotonicity property holds also for social welfare.[26] This result is stated in proposition 2 and the complete proof of it can be found in the appendix.

PROPOSITION 2:

Utilitarian social welfare increases with the leader's altruism, at least if individuals are not extremely risk averse.

If $\frac{1}{2}u(1-s_S^*) + \frac{1}{2}u(3s_S^*) \geq \frac{1}{4}u(1-s_A^*) + \frac{3}{4}u(3s_A^*)$, that is, risk aversion is high, conclusions are less obvious.[27] Rearranging the terms, we can get:

$$\frac{1}{2}\left(W_{/AP} - W_{/SP}\right) = p\left[u\left(s_A^* + \frac{1}{2}\right) - u\left(s_S^* + \frac{1}{2}\right)\right] + \frac{1-p}{2}\left[u(3s_A^*) - u(3s_S^*)\right] +$$

$$\frac{1-p-q}{4}\left[u(3s_A^*) - u(1-s_A^*)\right] + q\left[u\left(s_A^* + \frac{1}{2}\right) - \left[\frac{1}{4}u(1-s_A^*) + \frac{3}{4}u(3s_A^*)\right]\right] -$$

$\frac{1-p}{2}\left[u(1-s_S^*) - u(1-s_A^*)\right]$, where all the terms except for the last one are at least non-negative. Whether they altogether outweigh the last negative term depends mainly on the curvature of u.

Intuitively, if risk aversion is particularly high on [0, 1], and especially if at the same time the proportion of non-reciprocal selfish Responders is large, it is not impossible that $W_{/AP} < W_{/SP}$. This might happen, if the utility which a selfish Proposer enjoys when she encounters the Responder giving nothing back (utility $u(1-s_S^*)$ of the share she has left for herself) is much higher than the utility for an altruistic Proposer in the

[26] In this case the difference $u(1-s_S^*) - u(1-s_A^*)$ is not large compared to the difference $u(3s_A^*) - u(3s_S^*)$.

[27] In this case the difference $u(1-s_S^*) - u(1-s_A^*)$ is large enough compared to $u(3s_A^*) - u(3s_S^*)$.

same situation (which is $\frac{1}{2}u(1-s_A^*)+\frac{1}{2}u(3s_A^*)$), as the latter leaves less for herself. Due to computational difficulties, we leave for future analysis the question of deriving explicitly configurations of parameters (the distribution of the Responder's "types" and the shape of risk-aversion) leading to $W_{/AP} > W_{/SP}$ or otherwise.

4. Extensions

4.1. Continuous other-regarding parameters

It can be shown that the problem and the results extend naturally to the continuous case where the altruism and reciprocity parameters take any values between pure selfishness/non-reciprocity and pure altruism/reciprocity. Now different "types" of individuals $t \equiv (a, \mu)$ are given by various combinations of the altruism parameter $a \in [0, 1]$ and the reciprocity parameter $\mu \in [0, +\infty]$. For simplicity, assume that a and μ are independent random variables with cumulative distribution functions $\mathcal{F}_a(x)$ and $\mathcal{F}_\mu(y)$ respectively; thus, the probability distribution of other-regarding "types" is given by $\mathcal{F}_t(x,y) = \mathcal{F}_a(x) \times \mathcal{F}_\mu(y)$.

As in our basic analysis, the Proposer conditions her decision on the different probabilities of facing different "types" of Responders, since the reactions of different "types" to her own "type" are not the same. Given a_p, which determines the Proposer's "type" (remember that the reciprocity parameter is irrelevant for Proposers), the reaction of the Responder is determined by $\beta_R = \frac{a_R+\mu_R a_P}{1+\mu_R}$.

The Responder is solving the following problem:

$$\max_{k\in[0,3s]} \frac{1}{1+\beta_R}u(3s-k) + \frac{\beta_R}{1+\beta_R}u(1-s+k).$$

The first-order condition implies that $\frac{u'(3s-k)}{u'(1-s+k)} = \beta_R$. Let us note that the return transfer k, given by this condition, is monotonically increasing with the share s received from the Proposer.[28] But if the share s is too small, namely if $\frac{u'(3s)}{u'(1-s)} \geq \beta_R$, then it is optimal for the Responder to send back nothing, as she already feels that she has at least as much as (or even less than) she would like to have in the optimal allocation. Hence, the response function of the Responder can be formulated as

[28] It can be easily proved by means of the implicit function theorem: with the assumptions made on u, $\frac{dk}{ds} = -\frac{3u''(3s-k)u'(1-s+k)+u'(3s-k)u''(1-s+k)}{-u''(3s-k)u'(1-s+k)-u'(3s-k)u''(1-s+k)} > 0$.

$$k^*(s,\beta_R) = \begin{cases} 0 & if \ \frac{u'(3s)}{u'(1-s)} \geq \beta_R \\ k \ \ given\ by \frac{u'(3s-k)}{u'(1-s+k)} = \beta_R \ \ otherwise \end{cases},$$

or alternatively,

$$k^*(s,\beta_R) = \max\{0,k\} \ where \ \ k \ \ is\ given\ by \ \ \frac{u'(3s-k)}{u'(1-s+k)} = \beta_R.$$

Let us note that, for a given type of reaction β_R, the response function is continuous on $s \in [0, 1]$, and the optimal response k^* is monotonically (but not strictly) increasing with the share s received from the Proposer. Moreover, for a given share s, the response function is continuous on $\beta_R \in [0, 1]$, and the optimal k^* is monotonically increasing with β_R (from $k^* = 0$ for the Responders reacting in a purely selfish way to $k^* = \frac{4s-1}{2}$ for those reacting as pure altruists).[29]

The distribution of β_R (determining different reactions) conditional on a_P, denoted by $\mathcal{F}_{\beta_R/a_P=a_P}(z)$, can be obtained from the distribution of "types" $\mathcal{F}_t(x,y)$. A complete analysis of this model would then require solving the Proposer's maximization problem:

$$\max_{s\in[0,1]} \mathbb{E}U_P(s),$$

where

$$\mathbb{E}U_P(s) = \int_{z\in[0;1]}\left(\frac{1}{1+a_P}u\big(1-s+k^*(s,z)\big) + \frac{a_P}{1+a_P}u\big(3s-k^*(s,z)\big)\right)d\mathcal{F}_{\beta_R/a_P=a_P}(z).$$

4.2. Incomplete information for both players

In this subsection we analyze how to relax the assumption that the "type" of the Proposer is known to the Responder ex ante. Now the Responder has to elicit her opponent's "type" during the game, and the only way to do it is through the share s given up for her. In a separating equilibrium the shares chosen by Proposers of different "types" have to be different; otherwise the Responder is unable to identify the "type" of the opponent and respond accordingly.

As it has been shown in the previous section, it is not only altruistic Proposers who are willing to share. A selfish Proposer also might be willing to give up a positive amount if she has good reason to believe that she will face an unconditionally altruistic

[29] Let $\beta_R' < \beta_R''$. If $\frac{u'(3s)}{u'(1-s)} < \beta_R'$, then we compare interior solutions, and by the implicit function theorem: $\frac{dk^*}{d\beta_R} = -\frac{-1}{\frac{-u''(3s-k^*)u'(1-s+k^*)-u'(3s-k^*)u''(1-s+k^*)}{\left(u'(1-s+k^*)\right)^2}} > 0$. If $\beta_R' \leq \frac{u'(3s)}{u'(1-s)} < \beta_R''$, then $k^*(\beta_R') = 0$, while $k^*(\beta_R'') > 0$. Finally, if $\frac{u'(3s)}{u'(1-s)} \geq \beta_R''$, then $k^*(\beta_R') = k^*(\beta_R'') = 0$. Monotonicity holds in all the cases.

Responder. Thus, identification of the Proposer's "type" is not straightforward. Monotonicity of the optimal choice (proposition 1) raises hopes for possibility of successful identification, yet the basic difficulty related to incompleteness of the Responder's information needs to be resolved, namely, the possibility that the Proposer would mimic the behavior of the other "type". Hence, we need to introduce incentive compatibility constraints in order for a separating equilibrium to be sustained.

To characterize the equilibrium we proceed backwards: we assume that the shares sent by Proposers of different "types" do differ, and the Responder makes her choice of the amount to be returned according to the Proposer's "type" that she has elicited. The Proposer, when designing her strategy, takes into account the Responder's beliefs according to which the latter decides on the Proposer's "type" (one quite natural form of such beliefs, assuming a specific threshold which separates selfish behavior from altruistic, is introduced in more detail below). We consider pure symmetric strategies $s \in [0, 1]$, meaning that in equilibrium all Proposers of the same "type" choose the same optimal share with certainty. Next, we have to ensure that the optimal strategies satisfy the announced incentive compatibility constraints.

We are back to the binary case described and analyzed in sections 2 and 3. Let us define the Responder's belief function as a surjective function $b\colon [0; 1] \rightarrow \{selfish; altruistic\}$ which assigns the Proposer's "type" $b(s)$ to every observed share $s \in [0, 1]$. We consider a particular belief function of the following form:

$$b(s) = \begin{cases} altruistic & if\ s \geq \bar{s} \\ selfish & if\ s < \bar{s} \end{cases} \tag{4}$$

which implies existence of a threshold for s, separating altruistic givers from selfish ones in the Responder's mind. If the Proposer sends a share below this threshold, she is believed to be selfish, otherwise she is perceived as an altruist. The assumption of the Responder's beliefs of this form, henceforth referred to as threshold beliefs, sounds quite realistic and intuitive, and thus we adopt it for our analysis.[30]

Hence, a separating equilibrium is defined as the set of strategies $(\tilde{s}_S^*, \tilde{s}_A^*)$ for a selfish Proposer and an altruistic Proposer respectively and a belief function $b\colon [0; 1] \rightarrow \{selfish; altruistic\}$ assigning the Proposer's "type" to the observed share s such that:

(i) strategies $(\tilde{s}_S^*, \tilde{s}_A^*)$ are optimal for these beliefs;

(ii) beliefs are correct, i.e. $b(\tilde{s}_S^*) = selfish$ and $b(\tilde{s}_A^*) = altruistic$.[31]

[30] It is quite important to note that we do not assume any utility reduction in the case if the Responder's belief proves wrong after uncertainty is resolved. In fact, in this paper we do not consider such scenarios, as we are looking only for separating equilibria, in which beliefs are always correct.

[31] We use tildes above (s_S^*, s_A^*) in the definition of a separating equilibrium to distinguish between the optimal choices in the case of complete information for the Responder (given by lemmas 4 and 5) and those in the

Recall, the expected utility maximized by a selfish Proposer is

$$\mathbb{E}U_{SP}(s)=\begin{cases}u(1-s) & if\ s<\frac{1}{4}\\ pu\left(s+\frac{1}{2}\right)+(1-p)u(1-s) & if\ s\geq\frac{1}{4}\end{cases},$$

and that maximized by an altruistic Proposer is

$$\mathbb{E}U_{AP}(s)=\begin{cases}\frac{1}{2}u(1-s)+\frac{1}{2}u(3s) & if\ s\leq\frac{1}{4}\\ (p+q)u\left(s+\frac{1}{2}\right)+(1-p-q)\left[\frac{1}{2}u(1-s)+\frac{1}{2}u(3s)\right] & if\ s>\frac{1}{4}\end{cases}.$$

We denote by $\mathbb{E}U_{SP/AP}(s)$ and $\mathbb{E}U_{AP/SP}(s)$ the expected utilities of a selfish Proposer perceived by the Responder as altruistic and of an altruistic Proposer seen as selfish respectively. They are given by

$$\mathbb{E}U_{SP/AP}(s)=\begin{cases}u(1-s) & if\ s<\frac{1}{4}\\ (p+q)u\left(s+\frac{1}{2}\right)+(1-p-q)u(1-s) & if\ s\geq\frac{1}{4}\end{cases},$$

and

$$\mathbb{E}U_{AP/SP}(s)=\begin{cases}\frac{1}{2}u(1-s)+\frac{1}{2}u(3s) & if\ s\leq\frac{1}{4}\\ pu\left(s+\frac{1}{2}\right)+(1-p)\left[\frac{1}{2}u(1-s)+\frac{1}{2}u(3s)\right] & if\ s>\frac{1}{4}\end{cases}.$$

Now we turn to the analysis of the game assuming threshold beliefs. Let the Responder put the threshold value $\bar{s}$ equal to the optimal share sent by an altruistic Proposer in the case of complete information for the Responder, i.e. equal to s_A^* given by lemma 5. This ensures that an altruistic Proposer would always choose $\tilde{s}_A^*=\bar{s}$ and never deviate from it, and thus incentive compatibility constraints for an altruistic Proposer are irrelevant.[32]

Proposition 1 implies strict monotonicity of the optimal choices if the Responder's information is complete, i.e. $s_S^*<s_A^*$, whenever $s_S^*\neq 1$. Thus, if only s_S^* does not equal 1, it is strictly below the threshold $\bar{s}=s_A^*$, and consequently, it is an optimal choice for a selfish Proposer on $s\in[0,\bar{s})$, where a selfish Proposer is perceived as selfish. In order for it to be optimal on $[\bar{s},1]$ we have to impose the following incentive compatibility constraints:

case of incomplete information for both players, which might coincide (as shown in proposition 3) but not necessarily.

[32] For an altruistic Proposer s_A^* is optimal on $[\bar{s},1]$, as according to the belief function she is viewed as an altruist there, and s_A^* is the optimal solution for the case of complete information for the Responder (when altruists are perceived as altruists). She would also not deviate from s_A^* to any $s\in[0,\bar{s})$, as then she would be treated as selfish, which reduces her utility even more compared to the same deviation but without misperception (only unconditionally altruistic Responders would pay back now, while reciprocal would not any more): $\mathbb{E}U_{AP}(s_A^*)>\mathbb{E}U_{AP}(s)\geq\mathbb{E}U_{AP/SP}(s)\ \forall s\in[0,\bar{s})$.

$$\begin{cases} u(1) \geq (p+q)u\left(s_A^* + \frac{1}{2}\right) + (1-p-q)u(1-s_A^*) \; if \; s_S^* = 0 \\ pu\left(s_S^* + \frac{1}{2}\right) + (1-p)u(1-s_S^*) \geq (p+q)u\left(s_A^* + \frac{1}{2}\right) + (1-p-q)u(1-s_A^*) \; if \; s_S^* > 0 \end{cases} \quad (5)$$

Here the first constraint means that whenever a selfish Proposer sends zero to the Responder who knows her "type" with certainty, she would not send s_A^*, pretending to be an altruist, to the Responder who is unaware of her "type". The second constraint implies that mimicking an altruistic Proposer by sending s_A^* is neither profitable in the case when the optimal share for a selfish Proposer under complete information for her opponent is positive. As it is proved in the appendix, these two constraints are sufficient to guarantee that a selfish Proposer would neither deviate to any share above the threshold value $\bar{s} = s_A^*$.

Thus, we have shown that with proportions p (of non-reciprocal altruistic Responders) and q (of reciprocal Responders) and function u such that $s_S^* \neq 1$ and the incentive compatibility constraints for a selfish Proposer given by (5) hold, a separating equilibrium of the trust game with incomplete information exists, and it is given by the strategies $(\tilde{s}_S^*, \tilde{s}_A^*) = (s_S^*, s_A^*)$ and threshold beliefs with $\bar{s} = s_A^*$. The shares s_S^* and s_A^* are optimal with such beliefs, and the beliefs are correct, i.e. the threshold clearly separates selfish Proposers from altruistic ones. This result is summarized in proposition 3.

PROPOSITION 3 (existence of a separating equilibrium):

In the case of incomplete information for both players, when the Proposer's "type" is unknown to the Responder, and after observing the share s at the first stage of the game she forms her belief about it, if conditions (1) do not hold, and incentive compatibility constraints (5) for a selfish Proposer are satisfied, then the strategies (s_S^*, s_A^*) given by lemmas 4 and 5 and threshold beliefs of the form (4) with a threshold $\bar{s}$ equal to s_A^* constitute a separating equilibrium of the trust game.

In this equilibrium a selfish Proposer always gives up a share s_S^* and an altruistic Proposer – a share s_A^* which is strictly higher, and the Responder always elicits the Proposer's "type" correctly by the transfer she receives.

4.3. *k*-levels of other regarding preferences

In this subsection we broaden and generalize our concept of different levels of other-regarding preferences. Above we were discussing two levels of preferences – unconditional (altruism) and conditional on the opponent's preference profile of the previous level (type-based reciprocity). One might think of other kinds of unconditional preferences such as spitefulness, envy or inequality aversion, which can be put into this multi-level framework.[33] Then we could model such phenomena as negative reciprocity, for example, or conditional inequality aversion when an individual is concerned about equality only with some specific "types" of opponents. On the other hand, one might develop the model in the vertical direction, allowing for more than two levels of preferences. These would be of use when modeling situations in which treatment of the opponent is conditional on her reciprocity characteristics.

The main idea of our generalized model is that preferences at each subsequent level may depend on own and other reference individuals' preferences at the previous level:

$$U_i^k(x) = f_i^k\left(x, U_i^{k-1}(X), \left\{U_j^{k-1}(X)\right\}_{j\neq i}\right),$$

where $x \in X$ is some outcome from the set of possible outcomes, specifying payoffs to each of the individuals (now we allow for concern about multiple opponents simultaneously); f_i^k is an individual- and level-specific function determining other-regarding utility of the outcome x for the individual i at level k, conditional on her own preference profile at level k – 1 and on other individuals' preference profiles at level k – 1 (which gives partiality in treatment of different "types" defined by other-regarding preference profiles at the previous level).

As the basic level, referred to as level 0, it is intuitive to take classical selfish preferences implying that an individual is concerned solely about her own payoff. Formally, let X be a set of possible outcomes where each outcome $x = (x_1, x_2, ..., x_N)$ assigns a utility payoff to each of N individuals. The classical selfish preferences for an individual i can be represented by the utility function $U_i^0(x) = x_i$. Here the superscript indicates level 0 of other-regarding preferences – the selfish preferences level.

The next level, level 1, is qualitatively different as it introduces an additional concern of an individual – concern about the other individuals' payoffs. Many phenomena fall into this category – altruism or spitefulness, inequality aversion, status-seeking etc.[34] In general, treatment of the others does not have to be impartial; it might depend on some individual characteristics which appeal to a greater or lesser extent to the concerned individual. In many cases, among which inequality aversion is, relative payoffs are important too, while in the other cases they are not. However different all these phenomena appear at first sight, the common feature of all these types of preferences is that, unlike selfish preferences, they account for

[33] It might be argued that envy or inequality aversion are "conditional" other-regarding preferences, as they depend on the other's payoff, but we would rather call them "relative"; by "conditional" preferences we mean those conditional on the opponent's "type" given by her preferences (of a lower level), and not just conditional on her payoff.

[34] For simple, yet quite general models of such *unconditional* other-regarding preferences see, for example, Charness and Rabin (2002) or Erlei (2008). They also try to incorporate reciprocity, but do it in a very primitive form, by adding a binary indicator for misbehaving. Other examples include Andreoni and Miller (2002) for examination of altruistic preferences, Bolton (1991) for the model of envy, Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) for inequality aversion models.

other individuals. Formally, we would write the utility function of an individual i at level 1 of other-regarding preferences as $U_i^1(x) = f_i^1\left(x, U_i^0(X), \left\{U_j^0(X)\right\}_{j \neq i}\right)$, where f_i^1 gives the functional form of i's preferences (altruistic, competitive, etc.), and $U_i^0(X), \left\{U_j^0(X)\right\}_{j \neq i}$ are the individuals' preference profiles, indicating how they value different own payoffs (this allows partiality in treatment of the opponents based on their level 0 preferences, which might reflect tastes, speed of satiation, initial wealth effects etc.).

Level 2 of other-regarding preferences could be called reciprocal preferences. We refer to type-based reciprocity here (in contrast to intention-based reciprocity).[35] That is, now preferences of an individual not only include the payoffs that the others get, but also depend on unconditional other-regarding "types" of these others: for example, one might exhibit altruistic preferences only towards those who are altruistic, and be spiteful towards spiteful ones. This level of preferences, unlike the previous level, is conditional on other-regarding preference profiles at level 1 of all the other individuals, as well as on own level 1 preference profile. The latter, inter alia, can account for the so-called "Warm Glow" effect, which means getting satisfaction from the very fact of being altruistic.[36] The utility function at level 2 is the following: $U_i^2(x) = f_i^2\left(x, U_i^1(X), \left\{U_j^1(X)\right\}_{j \neq i}\right)$, where f_i^2 gives the functional form of i's reciprocal preferences, that is, how to treat different "types", and $U_i^1(X), \left\{U_j^1(X)\right\}_{j \neq i}$ are other-regarding level 1 preference profiles. Thus, now partiality in treatment of the others is also based on their level 1 preferences, i.e. is conditional on their other-regarding "types".

We might construct the next level of other-regarding preferences in the same manner: $U_i^3(x) = f_i^3\left(x, U_i^2(X), \left\{U_j^2(X)\right\}_{j \neq i}\right)$. At this level, preferences over the outcomes depend on the individuals' reciprocal (level 2) preference profiles. For example, one might treat better those who are reciprocal than those who are not. However, it seems that this and higher levels are less likely to happen in other-regarding reasoning.

5. Concluding Remarks

As it follows from the above discussion, there is a wide spectrum of implications of the multi-level preferences idea. Empirical evidence confirms that individuals do

[35] For specific models of type-based reciprocity see Levine (1998) or Rotemberg (2008).

[36] See Andreoni (1989).

behave differently in different environments, and the structure of their interaction plays an important role in determining behavioral patterns they follow.

The main result of the analysis of the trust game is monotonicity of the Proposer's optimal transfer. Although quite intuitive, it carries significant implications, including policy recommendations maximizing social welfare. Existence of a separating equilibrium under incomplete information for both sides is also an important result, implying among other things that the optimal under complete information strategies can be sustained even with less strict informational assumptions.

As immediate directions for future research we see investigating separating equilibria (existence and uniqueness) with different forms of beliefs from the Responder's side and with a more general parameterization of other-regarding preferences. Linearity of surplus-generating technology, namely, tripling of the share sent to the Responder in Berg et al. (1995), can be relaxed too. It also seems worthwhile to develop further the idea of *k*-levels of other-regarding preferences, as we believe it would open a number of novel and interesting questions. Finally, various social welfare implications mentioned above should be an important subject for further analysis.

References

Andreoni, J. (1989). Giving with Impure Altruism: Application to Charity and Ricardian Equivalence. *Journal of Political Economy*, *vol. 97*(6), 1447-1458.

Andreoni, J., Miller, J. (2002). Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism. *Econometrica*, *vol. 70*(2), 737-753.

Andreoni, J. (2006). Philanthropy. In: Kolm, S.-C., Mercier Ythier, J. (Eds.), *Handbook of the Economics of Giving, Altruism, and Reciprocity*, vol. 2. Elsevier.

Berg, J., Dickhaut, J., McCabe, K. (1995). Trust, Reciprocity, and Social History. *Games and Economic Behavior*, *vol. 10*(1), 122-142.

Bolton, G.E. (1991). A Comparative Model of Bargaining: Theory and Evidence. *The American Economic Review*, *vol. 81*(5), 1096-1136.

Bolton, G.E., Ockenfels, A. (2000). A Theory of Equity, Reciprocity and Competition. *The American Economic Review*, *vol. 90*(1), 166-193.

Borah, A. (2010). Other-Regarding Preferences and Consequentialism. *Publicly accessible Penn Dissertations*. Paper 132.

Camerer, C., Thaler, R.H. (1995). Anomalies: Ultimatums, Dictators and Manners. *The Journal of Economic Perspectives*, *vol. 9*(2), 209-219.

Charness, G., Rabin, M. (2002). Understanding Social Preferences with Simple Tests. *The Quarterly Journal of Economics*, *vol. 117*(3), 817-869.

Cox J.C. (2004). How to Identify Trust and Reciprocity. *Games and Economic Behavior*, *vol. 46*(2), 260-281.

Decerf, B., Van der Linden, M. (2016). Fair Social Orderings with Other-Regarding Preferences. *Social Choice and Welfare, vol. 46*, 655-694.

Erlei, M. (2008). Heterogeneous Social Preferences. *Journal of Economic Behavior and Organization, vol. 65*(3-4), 436-457.

Fehr, E., Schmidt, K.M. (1999). A Theory of Fairness, Competition and Cooperation. *The Quarterly Journal of Economics, vol. 114*(3), 817-868.

Fehr, E., Schmidt, K.M. (2006). The Economics of Fairness, Reciprocity, and Altruism – Experimental Evidence and New Theories. In: Kolm, S.-C., Mercier Ythier, J. (Eds.), *Handbook of the Economics of Giving, Altruism, and Reciprocity*, vol. 1. Elsevier.

Kanbur, R. (2006). The Economics of International Aid. In: Kolm, S.-C., Mercier Ythier, J. (Eds.), *Handbook of the Economics of Giving, Altruism, and Reciprocity*, vol. 2. Elsevier.

Laferrere, A., Wolff, F.-C. (2006). Microeconomic Models of Family Transfers. In: Kolm, S.-C., Mercier Ythier, J. (Eds.), *Handbook of the Economics of Giving, Altruism, and Reciprocity*, vol. 2. Elsevier.

Levine D.K. (1998). Modeling Altruism and Spitefulness in Experiments. *Review of Economic Dynamics*, vol. 1 (3), 593-622.

Rabin M. (1993). "Incorporating Fairness into Game Theory and Economics". *The American Economic Review, vol. 83*(5), 1281-1302.

Rotemberg J.J. (2006). Altruism, Reciprocity and Cooperation in the Workplace. In: Kolm, S.-C., Mercier Ythier, J. (Eds.), *Handbook of the Economics of Giving, Altruism, and Reciprocity*, vol. 2. Elsevier.

Rotemberg J.J. (2008). Minimally Acceptable Altruism and the Ultimatum Game. *Journal of Economic Behavior and Organization, vol. 66*(3-4), 457-476.

Treibich, R. (2014). Welfare Egalitarianism with Other-Regarding Preferences. Discussion Papers on Business and Economics no. 22/2014, University of Southern Denmark.

Appendix

Proof of remark 1:

The condition $pu\left(\frac{3}{2}\right)+(1-p)u(0)>u(1)$ is a necessary condition for $s_S^*=1$. Since u is strictly concave, its value at 1 has to be above the line connecting points $\left(0,u(0)\right)$ and $\left(\frac{3}{2},u\left(\frac{3}{2}\right)\right)$ in $\left(s,u(s)\right)$-space, that is, $u(1)>\frac{u\left(\frac{3}{2}\right)-u(0)}{\frac{3}{2}}\cdot 1+u(0)=\frac{2}{3}u\left(\frac{3}{2}\right)+\frac{1}{3}u(0)$. Combining these two inequalities, we conclude that $s_S^*=1$ holds only if $p>\frac{2}{3}$.

Similarly, for $s_S^*=s$, given by $\frac{u'\left(s+\frac{1}{2}\right)}{u'(1-s)}=\frac{1-p}{p}$, it has to hold that $pu\left(s+\frac{1}{2}\right)+(1-p)u(1-s)>u(1)$. By remark 2 (proved below), $s\in(\frac{1}{2},1)$, and thus

$1 - s < 1 < s + \frac{1}{2}$. Concavity of u implies

$u(1) > \frac{u(s+\frac{1}{2}) - u(1-s)}{2s-\frac{1}{2}} \cdot s + u(1-s) = \frac{s}{2s-\frac{1}{2}} u\left(s+\frac{1}{2}\right) + \frac{s-\frac{1}{2}}{2s-\frac{1}{2}} u(1-s)$.

It follows that in order for $s_S^* = s$ to hold, p has to be above $\frac{s}{2s-\frac{1}{2}}$, which in its turn is above $\frac{2}{3}$ for $s \in (\frac{1}{2}, 1)$.

Summarizing the above, s_S^* might be higher than zero only if $p > \frac{2}{3}$.

Q.E.D.

Proof of remark 2:

Since s, given by $\frac{u'\left(s+\frac{1}{2}\right)}{u'(1-s)} = \frac{1-p}{p}$, belongs to the interval $(\frac{1}{4}, 1)$, it holds that $s + \frac{1}{2} > 1 - s$, and thus, $pu\left(s+\frac{1}{2}\right) + (1-p)u(1-s) \leq u\left(s+\frac{1}{2}\right)$ $\forall p$. If the solution to the Proposer's problem $s_S^* = s$, then $pu\left(s+\frac{1}{2}\right) + (1-p)u(1-s) > u(1)$, implying $u\left(s+\frac{1}{2}\right) > u(1)$. It follows that then $s > \frac{1}{2}$, which allows to move the lower bound for s up to $\frac{1}{2}$.

Q.E.D.

Proof of proposition 1 (Monotonicity Property):

Since s_A^* is always positive, whenever $s_S^* = 0$ strict monotonicity takes place: $s_S^* < s_A^*$. Let us analyze the case when $s_S^* \neq 0$.

Accounting for remarks 2 and 3, we have $s_S^* \in (\frac{1}{2}, 1]$ and $s_A^* \in (\frac{1}{4}, 1]$. If the optimal share for an altruistic Proposer $s_A^* = 1$, then monotonicity holds: $s_S^* \leq s_A^*$, and it is strict unless $s_S^* = 1$. If the optimal share $s_A^* \in (\frac{1}{4}, 1)$, that is, we have an interior solution $s_A^* = s$, given by $\frac{u'\left(s+\frac{1}{2}\right)}{u'(1-s)-3u'(3s)} = \frac{1-p-q}{2(p+q)}$, strict monotonicity takes place. Let us prove it by contradiction. Suppose the opposite: $s_S^* \geq s_A^*$. Then the two conditions should be satisfied simultaneously: $\mathbb{E}U'_{AS}(s) = 0$ and $\mathbb{E}U'_{SS}(s) \geq 0$. They imply respectively $\frac{u'\left(s+\frac{1}{2}\right)}{u'(1-s)-3u'(3s)} = \frac{1-p-q}{2(p+q)}$ and $\frac{u'\left(s+\frac{1}{2}\right)}{u'(1-s)} \geq \frac{1-p}{p}$. As $\frac{u'\left(s+\frac{1}{2}\right)}{u'(1-s)-3u'(3s)} > \frac{u'\left(s+\frac{1}{2}\right)}{u'(1-s)}$, it follows that $\frac{1-p-q}{2(p+q)} > \frac{1-p}{p}$. But for any $p > 0$ and for any q it is true that $\frac{1-p-q}{2(p+q)} < \frac{1-p}{p}$, which contradicts what we have just obtained. Thus, $s_S^* < s_A^*$ whenever $s_A^* \in (\frac{1}{4}, 1)$

and $p > 0$. If $p = 0$, a selfish Proposer maximizes $u(s - 1)$, hence $s_S^* = 0$, and strict monotonicity holds as well.

To sum up, $s_S^* \leq s_A^*$ always holds, and when it is not the case that $s_S^* = s_A^* = 1$ it holds strictly.

Q.E.D.

Proof of proposition 2:

Substituting expressions for expected utilities together with response functions and optimal transfers into the social welfare functions for a selfish Proposer and an altruistic Proposer cases, we get respectively:

$$\frac{W_{/SP}}{2} = \begin{cases} p\left[\frac{1}{4}u(0) + \frac{3}{4}u(1)\right] + (1-p)\left[\frac{1}{2}u(0) + \frac{1}{2}u(1)\right] & if\ s_S^* = 0 \\ pu\left(s_S^* + \frac{1}{2}\right) + (1-p)\left[\frac{1}{2}u(1 - s_S^*) + \frac{1}{2}u(3s_S^*)\right] & if\ s_S^* \neq 0 \end{cases},$$

and

$$\frac{W_{/AP}}{2} = (p+q)u\left(s_A^* + \frac{1}{2}\right) + (1-p-q)\left[\frac{1}{4}u(1 - s_A^*) + \frac{3}{4}u(3s_A^*)\right].$$

We consider the expressions for social welfare functions divided by 2 to simplify the following calculations.

Let us consider first the case $s_S^* = 0$. Then $\frac{W_{/SP}}{2}$ can be interpreted as the expected value of the lottery giving a higher payoff $\frac{1}{4}u(0) + \frac{3}{4}u(1)$ with probability p and a lower payoff $\frac{1}{2}u(0) + \frac{1}{2}u(1)$ with probability $1 - p$. Similarly, we interpret $\frac{W_{/AP}}{2}$ as the expected value of the lottery giving $u\left(s_A^* + \frac{1}{2}\right)$ with probability $p + q$ and $\frac{1}{4}u(1 - s_A^*) + \frac{3}{4}u(3s_A^*)$ with probability $1 - p - q$. However, in this case it is not obvious which payoff is higher. Yet the better payoff in the first lottery is lower than either of the possible payoffs in the second lottery. This follows from strict concavity of u and the fact that $s_A^* > \frac{1}{4}$, implying that $\frac{1}{4}u(0) + \frac{3}{4}u(1) < u\left(\frac{3}{4}\right) < u\left(s_A^* + \frac{1}{2}\right)$ and $\frac{1}{4}u(0) + \frac{3}{4}u(1) < \frac{1}{4}u(1 - s_A^*) + \frac{3}{4}u(3s_A^*)$ respectively. Hence, $W_{/AP} > W_{/SP}$, meaning that in the case when a selfish Proposer keeps everything for herself, social welfare would be strictly higher if the Proposer was altruistic.

If $s_S^* \neq 0$, then, similarly to the previous case, we can think of $\frac{W_{/SP}}{2}$ and $\frac{W_{/AP}}{2}$ as expected values of the lotteries. Note that $\frac{1}{2}u(1 - s_S^*) + \frac{1}{2}u(3s_S^*) < u\left(s_S^* + \frac{1}{2}\right) \leq u\left(s_A^* + \frac{1}{2}\right)$, where the first inequality follows from strict concavity of u, and the last one – from proposition 1. If $\frac{1}{2}u(1-s_S^*)+\frac{1}{2}u(3s_S^*)<\frac{1}{4}u(1-s_A^*)+\frac{3}{4}u(3s_A^*)$, that is, the difference $u(1-s_S^*)-u(1-s_A^*)$ is not large compared to the difference $u(3s_A^*)-u(3s_S^*)$, it can be easily shown that $W_{/AP}>W_{/SP}$. There are two possibilities. If $u\left(s_A^*+\frac{1}{2}\right)\leq\frac{1}{4}u(1-s_A^*)+\frac{3}{4}u(3s_A^*)$, then the lowest possible payoff in the case of an altruistic Proposer is at least as high as the highest payoff in the case of a selfish Proposer. If $u\left(s_A^*+\frac{1}{2}\right)>\frac{1}{4}u(1-s_A^*)+\frac{3}{4}u(3s_A^*)$, then both possible payoffs are higher in the case of an altruistic Proposer, and the better payoff happens with even higher (at least not lower) probability than the better payoff in the case of a selfish Proposer.

Thus, placing an altruistic person as the Proposer ensures higher social welfare at least in the case when risk aversion is not very high.

Q.E.D.

Proof of proposition 3 (existence of a separating equilibrium):

To prove the proposition, we need to show that with threshold beliefs of the form

$$b(s)=\begin{cases} altruistic\ if\ s\geq s_A^* \\ selfish\quad if\ s<s_A^*, \end{cases}$$

and whenever $s_S^*\neq 1$ and the incentive compatibility constraints (5) for a selfish Proposer are satisfied,

(i) s_S^* and s_A^* constitute the optimal strategies for a selfish and an altruistic Proposers respectively;

(ii) beliefs are correct, i.e. $b(s_S^*)=selfish$ and $b(s_A^*)=altruistic$.

According to the monotonicity property (proposition 1), if $s_S^*\neq 1$ then $s_S^*<s_A^*$, that is, the strategies are different which is necessary for an equilibrium to be separating. Obviously, with the form of beliefs defined above, condition (ii) is satisfied.

It is left to verify whether the strategies s_S^* and s_A^* are optimal. The fact that s_A^* is optimal for an altruistic Proposer has been proved in footnote 32. To repeat briefly, s_A^* is optimal for her on $[0,1]$ if her “type” is perceived correctly, and if she is perceived as selfish her utility is even lower, since the probability of a positive response decreases from $p+q$ to p. As for a selfish Proposer, because of monotonicity, $s_S^*\in[0,\bar{s})$. She would not deviate within this set where she is perceived as selfish, as s_S^* is already an optimal solution there. She would neither deviate to $\bar{s}=s_A^*$, as it is ruled out by the incentive compatibility constraints (5). Finally, these constraints ensure also that a selfish Proposer would not deviate to any s above the threshold $\bar{s}=s_A^*$, because they imply $\mathbb{E}U_{SP/AP}(s_A^*)\geq\mathbb{E}U_{SP/AP}(s)\ \forall s>s_A^*$.

To prove the last inequality, recall that s_A^* is optimal on $[\bar{s}, 1]$ for an altruistic Proposer, that is,

$$\mathbb{E}U_{AP}(s_A^*) = (p+q)u\left(s_A^* + \tfrac{1}{2}\right) + (1-p-q)\left[\tfrac{1}{2}u(1-s_A^*) + \tfrac{1}{2}u(3s_A^*)\right] >$$
$$\mathbb{E}U_{AP}(s) = (p+q)u\left(s + \tfrac{1}{2}\right) + (1-p-q)\left[\tfrac{1}{2}u(1-s) + \tfrac{1}{2}u(3s)\right] \ \ \forall s > s_A^*,$$

which implies

$$(1-p-q)[u(1-s_A^*) - u(1-s)] - (1-p-q)[u(3s) - u(3s_A^*)] >$$
$$2(p+q)\left[u\left(s+\tfrac{1}{2}\right) - u\left(s_A^* + \tfrac{1}{2}\right)\right] \ \ \forall s > s_A^*$$

or equivalently,

$$(1-p-q)[u(1-s_A^*) - u(1-s)] - (p+q)\left[u\left(s+\tfrac{1}{2}\right) - u\left(s_A^*+\tfrac{1}{2}\right)\right] >$$
$$(p+q)\left[u\left(s+\tfrac{1}{2}\right) - u\left(s_A^*+\tfrac{1}{2}\right)\right] + (1-p-q)[u(3s) - u(3s_A^*)] \ \ \forall s > s_A^*.$$

Then

$$\mathbb{E}U_{SP/AP}(s_A^*) - \mathbb{E}U_{SP/AP}(s) = \left[(p+q)u\left(s_A^*+\tfrac{1}{2}\right) + (1-p-q)u(1-s_A^*)\right] -$$
$$\left[(p+q)u\left(s+\tfrac{1}{2}\right) + (1-p-q)u(1-s)\right] = (1-p-q)[u(1-s_A^*) -$$
$$u(1-s)] - (p+q)\left[u\left(s+\tfrac{1}{2}\right) - u\left(s_A^*+\tfrac{1}{2}\right)\right] > (p+q)\left[u\left(s+\tfrac{1}{2}\right) -\right.$$
$$\left.u\left(s_A^*+\tfrac{1}{2}\right)\right] + (1-p-q)[u(3s) - u(3s_A^*)] > 0 \ \ \forall s > s_A^*.$$

Thus, condition (i) is also satisfied.

Let us note that conditions (1) are necessary for (s_S^*, s_A^*) to be equilibrium strategies. If they do not hold, together with the monotonicity property it implies $s_S^* = s_A^* = 1$, and thus a separating equilibrium does not exist. To see this, note that an altruistic Proposer would always choose $\tilde{s}_A^* = s_A^* = 1$, as it has been reasoned above. At the same time a selfish Proposer would choose $\tilde{s}_S^* = s_S^* = 1$, since it is already optimal on [0, 1], and being perceived as altruist increases her expected utility even more. Thus, in this case no incentive compatibility constraint can prevent a selfish Proposer from choosing 1 and appearing as an altruist to the Responder. Consequently, the Responder is unable to differentiate between a selfish Proposer and an altruistic Proposer.

Q.E.D.