



SP-SP

Working Paper  
WP no 652  
September, 2006

University of Navarra

**HELPING THE MEANER, HURTING THE NICER:  
THE CONTRIBUTION *VERSUS* DISTRIBUTION GAME**

Gianandrea Staffiero

IESE Business School – University of Navarra

Avda. Pearson, 21 – 08034 Barcelona, Spain. Tel.: (+34) 93 253 42 00 Fax: (+34) 93 253 43 43

Camino del Cerro del Águila, 3 (Ctra. de Castilla, km 5,180) – 28023 Madrid, Spain. Tel.: (+34) 91 357 08 09 Fax: (+34) 91 357 29 13

Copyright © 2006 IESE Business School.

The Public-Private Center is a Research Center based at IESE Business School. Its mission is to develop research that analyses the relationships between the private and public sectors primarily the following areas: regulation and competition, innovation, regional economy and industrial politics and health economics.

Research results are disseminated through publications, conferences and colloquia. These activities are aimed to foster cooperation between the private sector and public administrations, as well as the exchange of ideas and initiatives.

The sponsors of the SP-SP Center are the following:

- Accenture
- Ajuntament de Barcelona
- Official Chamber of Commerce, Industry and Navigation of Barcelona
- BBVA
- Diputació de Barcelona
- Garrigues, Abogados y Asesores Tributarios
- Catalan Government (Generalitat de Catalunya)
- Sanofi-Aventis
- Telefonica
- T-Systems
- VidaCaixa

The content of this publication reflects the conclusions and findings of the individual authors, and not the opinions of the Center's sponsors.

# HELPING THE MEANER, HURTING THE NICER: THE CONTRIBUTION *VERSUS* DISTRIBUTION GAME

Gianandrea Staffiero\*

## Abstract

A wide experimental evidence shows that people, do care about their opponents' payoff during social interaction. Our design aims at shed light over the relative importance of different motives behind non selfish choices highlighted in the recent literature. After a standard public good game, one player is given the possibility to increase or decrease his opponent's payoff. While our baseline treatment replicates the tendency to hurt richer lower contributors and help poorer higher contributors. By adding exogenous assignments by the experimenter we find substantial willingness to hurt the richer even if he or she had contributed more and to help the poorer even if they had contributed less. These results show a greater focus on correcting inequality rather than on punishing or rewarding behavior. Moreover, we also find that subjects disregard efficiency, in terms of the overall "pie" to be shared. Overall, our data support inequality aversion as a more robust phenomenon than reciprocity and efficiency considerations.

\* Post-Doctoral Research Fellow, IESE

Classification JEL: C91, D63, H41

**Keywords:** Fairness, Cooperation, Inequality, Reciprocity.

# HELPING THE MEANER, HURTING THE NICER: THE CONTRIBUTION *VERSUS* DISTRIBUTION GAME<sup>1</sup>

## 1. Introduction

The importance of social preferences is emerging as a prominent feature of the recent literature in behavioral and experimental economics. In this study, we present data collected using an experimental design in which decision nodes are created, where the forces of inequality aversion and reciprocity push in opposite directions. More specifically, the payoff distribution derived from contribution levels in a previously played voluntary contribution game are changed via exogenous assignments by the experimenter. Choices in those nodes, which may be helpful, neutral or hurtful towards opponents, reveal whether players reward or punish higher or lower contributions or, rather, tend to reduce inequality in payoff distributions. The relative importance of contributions and distribution revealed by those choices, therefore, is used as a test to compare the robustness of reciprocity orientation and inequality aversion. Moreover, the data from the same game also shed light in the debate over the impact of equality and efficiency in affecting individual choices. The aim of this study is to show how inequality aversion emerges as a robust phenomenon, within the realm of “fairness” motives, when the other forces could potentially counterbalance its effects.

The growing interest in reciprocity and inequality aversion, i.e. on deviations from “pure” selfishness based on opponents’ intentions and on distributional preferences, is motivated by the data obtained in experiments which show that in symmetric public good games agents are willing to spend in order to punish free riders (see Ostrom et al., 1992, and Fehr and Gächter, 2000); other studies replicate these results and also present some, albeit less conclusive, evidence of the willingness to reward highly cooperative behavior (see Andreoni et al., 2003, Sefton et al., 2002). In particular, in Fehr and Gächter (2000) punishment provokes positive effects on cooperation rates across rounds, in the repeated version of the game, as free riders

---

<sup>1</sup> A previous version of this paper was part of my Ph.D. thesis at the I.D.E.A. program at the Universitat Autònoma de Barcelona. I would like to thank my supervisor, Jordi Brandts, for his continuous support, and Klaus Abbink, Joan Maria Esteban, Marco Casari, Cristina Blanco, Marcela Arqueros Wood, Pablo Guillén, Raúl López and Rosella Nicolini for useful suggestions. The usual disclaimer applies. Financial support from the Generalitat de Catalunya during my Ph.D. studies (Beca FI 2000 487) and from the SP-SP Research Center at IESE Business School-University of Navarra is gratefully acknowledged.

learn to contribute in order to avoid sanctions. This result stands in contrast to what is found in standard repeated public good and dilemma games, where cooperation tends to fade away. The term “punishment” in their title appeals to a reciprocity motivation: it is the action of hurting somebody whose previous actions have hurt the punisher. However, in other frequently cited papers, Fehr himself and Schmidt (1999) and Bolton and Ockenfels (2000) present models of preferences based on inequality aversion, which are consistent with a large set of experimental results including the ones in Fehr and Gaechter (2000). That is due to the fact that free riders achieve higher payoffs in the game, before punishment can occur. The “punishment technology”, according to which the expense for the punisher is lower than the loss for the punished, is such this action reduces (and possibly eliminates) the inequality caused by contribution differentials. A natural question, then, concerns the real motivation for decreasing free riders’ payoffs: is it negative reciprocity, i.e. the desire to “hurt who hurts you” as argued, for instance, in Rabin (1993), or rather pure inequality aversion? The same kind of question, in fact, can be asked when observing “rewarding” behavior, like, e.g., in gift-exchange (Fehr et al., 1993) and “trust” games (Berg et al., 1995).

Some authors compare reciprocity and inequality aversion in an ex-ante asymmetric set-up. Studies like Bolton et al. (2000), Abbink et al. (2000) and Falk et al. (2000) present sequential games in which second (and last) movers can evaluate the “kindness” of first movers’ choices by comparing them with the available alternatives, which can be easily ranked in terms of how favorable they are towards opponents. Therefore “reciprocity motivated” second movers should choose to reward (punish) first movers when the latter’s choice was beneficial (damaging) to the former.

In Bolton et al. (2000) we find a strong result in favor of inequality aversion as a stronger motive than reciprocity. They find that when second movers face a given decision node, defined in terms of a set of possible actions and immediate payoff consequences (as their move ends the game), they take approximately the same decision –compatible with a mixture of selfish and inequality averse types in the population– independently of which alternative choice was available for the first movers to the one which led to that node. More specifically, the choice leading to that decision node was the only one first movers could take, in one treatment; in another one it was unambiguously the nicest (from the second mover’s point of view); in a third treatment it was unambiguously the worst. The failure to find statistically significant differences in second movers’ choices across the three treatments indicates a prevalence of distributional preferences over reciprocity.

On the other hand, results in Falk et al. (2000) stress the importance of intentions without, however, neglecting the importance of inequality aversion. In their experiment, one treatment has first movers choosing to transfer or take away payoff units from second movers. The latter can reward or punish first movers by spending some money to increase or decrease their payoffs. The authors find that, unlike standard predictions, second movers are willing to spend in order to punish greed and reward kindness, confirming evidence in Abbink et al. (2000); on the other hand, when the first move is randomly determined, the reactions are mild, although on average inequality reducing second moves are still found to be statistically significant.

The experimental design introduced here shares with the studies just mentioned the focus on second mover behavior, when beneficial, neutral or hurtful choices towards opponents are available, and the “one-shot” nature of the game. The latter allows us to assume away strategic as well as “learning” aspects in the second movers’ choices. Strategic aspects are often ruled out with “stranger” conditions, where players are re-matched at every round making sure the same opponents never meet again. However, the knowledge that the opponent is still going to play

might have an influence in players' mind; moreover, the expectation to repeat the game may provide an incentive to try various strategies in order to "learn" which one best suits the agent's interests.

The ex-ante symmetry, usually present in repeated public good games, is a characteristic of the one-shot game played here. While in the asymmetric sequential games in Abbink et al. (2000) and in Falk et al. (2000) players can tell whether the first move was nicer or not with respect to the alternatives, the definition of *how* kind (or mean) the choice was, in absolute terms, is less clear-cut. Some may find that the taken choice is not only the nicer but also an "obvious" choice that any "decent" individual should take; others may find it exceptionally kind. The subjectivity in judgment can of course be reflected in the variations in the choices different second movers take after the same first move by their respective opponents. However, the experimenter cannot really tell what the evaluation really was just by observing the reaction: some individuals may be strictly revengeful and other may refrain from spending in order to punish even large deviation from "decency".

In contrast to the asymmetric designs mentioned above, our game allows us to observe, on top of the second mover's choice, what his own choice as a first mover was, in the simultaneous public good game played in the first phase of the experiment. As we describe more in detail below, in the first phase all players, matched in couples, play a standard voluntary contribution game and then one for each couple is assigned the role of unique second mover. In other words, no player knows ex ante whether he will be a second mover: after contribution choices are taken *by both*, one opponent is given the option to help or hurt the other. Therefore, players are actually in the same position in the initial moment in which they choose how much to contribute to the public good. As usual, the parameters of the "public good technology" entail that the higher the contribution the more a player is being "nice" towards his opponent. The player who is picked as the second mover has the possibility to take his last decisions according to all the possible opponent's contribution levels (as explained and motivated below, we employed the "strategy method"). A natural way to evaluate opponents' kindness at every possible decision node is implicitly provided here: the second mover can compare his own and his opponent's contribution levels. While in general a possible manner to judge others' actions is to think of what one's own choice would have been in the same circumstances, here this comparison is made explicitly possible: second movers can tell whether in a given node their opponents were more, equally or less "nice" than themselves. The idea of reciprocity assumed here is then –we argue– fairly plausible: you "should" take choices which are beneficial to your opponent (or at least not damaging) if she contributed more than you; vice versa if she contributed less.<sup>2</sup>

As it will be seen, results in the baseline "situation" essentially replicate the ones previously obtained in public good games with punishment and reward, where higher (lower) contributions are rewarded (punished). While this evidence is compatible both with reciprocity and with inequality aversion, other situations in our design are characterized by exogenous assignments by the experimenter, which create decision nodes where these two motives push in opposite directions. More specifically, in some nodes the second mover is poorer despite having contributed less, and in others he is richer despite having contributed more than his opponent. As a consequence, second moves are chosen in nodes where rewarding higher contributors

---

<sup>2</sup> For clarity, we refer to the active player as "he" and to the inactive as "she", from now on.

would increase unfavorable inequality and in others where punishing lower contributors would increase favorable inequality. As anticipated above, evidence in these nodes, as well as in the overall experiments, speak in favor of the robustness of inequality aversion as the main motivation behind “non-selfish” choices.

The way in which second mover’s possibilities are defined also provides insights in relation to the increasingly active debate over the relative importance of equality versus efficiency motives. Experimental results in Charness and Rabin (2002) and Bolle and Kritikos (2001) indicate a general willingness to accept unfavorable inequality if the chosen alternative features high efficiency, defined in terms of total payoffs distributed to the players. In our game, second movers are given the same “technology” in terms of “helping” and “hurting” activities, unlike, for instance, the case in Abbink et al. (2000), Falk et al. (2000) and Sefton et al. (2002) where “sanctioning” devices are more powerful than “rewarding” ones. Specifically, they have to spend  $F$  both if they want to increase and if they want to decrease their opponent’s payoff by  $3F$ . So, a decreasing activity involves a total loss of  $4F$  while an increasing activity provokes a net gain worth  $2F$ . If efficiency motives were strong vis-à-vis equality, they should drive towards more increasing than decreasing activity in general, and in particular towards a tendency to help first movers when reasons to reciprocate (positively or negatively) are absent. According to the definition of reciprocity introduced above, this is assumed to be the case when contribution levels were equal. In order to find further evidence, we also made the same subjects who acted as second movers play another treatment. In that “situation” the first phase is replaced by simple payoff allocations, which can be modified according to the same rules (spend  $F$  to increase or to decrease by  $3F$  your opponent’s payoff). Overall, we find no evidence of efficiency motives; in particular, equal “partial payoffs” are typically left unchanged, both when they come from exogenous allocations and when they are consequence of equal contributions. As argued below, our data stress the importance of inequality aversion especially in relation with other plausible motives such as reciprocity and efficiency.

The next section presents the experimental design. Section 3 presents the different predictions which are derived according to different models of human behavior, starting from “homo economicus” standard assumptions to proceed into the realm of “other-regarding” preferences. Section 4 clarifies the way in which the data are to be analyzed in order to test the relative strength of inequality aversion in the two comparisons we are interested in: versus reciprocity and versus efficiency motives. In section 5 results are presented; in section 6 they are summarized and discussed. Section 7 concludes.

## 2. Experimental Design

The experiment was conducted at the Universitat Autònoma de Barcelona. We ran five sessions which took place in January and February 2003.<sup>3</sup> Overall, 64 students participated; earnings varied from 2 to 15 euro, with an average of 7.6 euros. In every session, students were divided in two groups located in separate rooms. In both rooms instructions (shown in Appendix 1) were read aloud and then a test was administered to ensure proper understanding of the game. During every session of the experiment in both rooms there was a person available to answer questions privately to any players with doubts.

---

<sup>3</sup> Before the sessions which we report here, a pilot involved 14 students and suggested modifications to the instructions and the necessity of a test to ensure the proper comprehension of the rules.

The design makes players interact anonymously in pairs, with each student in a room matched with one in the other, in a one-shot game composed of two phases: a standard public good game in the first one, and the possibility for one player to affect payoff combinations in the second.

More specifically, in the first phase each subject is given an endowment 4 and decides on his contribution  $g$  (an integer between 0 and 4) to a public fund. The returns from the fund are then distributed so that, for player  $i$  with opponent  $k$ , the first phase payoff is:

$$\pi_i = 4 - g_i + 0.75(g_i + g_k)$$

and vice versa for player  $k$ . Notice that this implies that the money in the fund is multiplied by 1.5 before distribution among the two players occurs. The payoff function can also be rewritten as:

$$\pi_i = 4 - 0.25g_i + 0.75g_k$$

(and again vice versa for  $k$ ), which makes it clear that contributions benefit the payoff of your opponent but decrease your own. Subjects were shown in both formulas, as means to understand the determination of payoffs of this phase, also presented in a 5x5 matrix relating own and opponent's choice to the 25 possible payoff distributions. The game just described determines the "first phase results".

In the second phase one player is randomly chosen to be "active", while the other has no more choices to take. Notice that this randomization takes place after the end of the first phase, so that players have equal roles ex ante.

Four possible "situations" compose the second phase: A, B, C, D.

"Partial results" are determined as follows.

In situations A, B, C, they are the sum of the payoffs resulting from the first phase plus the following amounts:

- Situation A: active receives 4, inactive receives 4
- Situation B: active receives 4, inactive receives 0
- Situation C: active receives 4, inactive receives 8

In situation D, instead, "partial results" consist of 10 possible exogenous allocations, independent of first phase.

Every choice by the active player is an expense  $F \in \{0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4\}$  to modify his opponent payoff by  $3F$  or  $-3F$ . For instance, a player who wants to benefit his opponent may choose to spend 1 (or, say, 2.5) to increase her payoff by 3 (7.5); if he wants to hurt her he may spend, say, 0.5 (or, to hurt her more, 2) to decrease her payoff by 1.5 (6); also, he may decide not to spend anything, choosing  $F = 0$  and leaving her unaffected.

The strategy method (Selten, 1967, see also Selten et al., 2003) is used, so that the active player makes choices for every possible situation and, in A, B, C, for every possible first phase



contribution by his opponent; in D, for every possible exogenous allocation.<sup>4</sup> See Appendix 2 for an example of a decision sheet (translated from Spanish) used by an active player –whose first phase contribution was 2– to write his second phase strategy and the possible payoff consequences (the calculation of all possible “final payoffs” was not mandatory, and players typically made some of them to verify their correctness with the experiment assistant).

*After* all contingent decisions by the active player have been collected, a random draw determines the actual situation. If it happens to be A, B, or C, then the actual first phase decision by the inactive player determines which decision by the active player takes effect. For instance, if the opponent of the player whose choices are represented in Appendix 2 has chosen 3, and the draw selects situation B, then the relevant active player’s decision is to spend 1 to increase her payoff by 3; as a consequence, the payoffs distributed are 8.75 for the active and 7.75 for the inactive players. If instead the draw picks situation D, a further random draw determines which one of the ten exogenous allocations is selected. Also in this case the corresponding decision by the active player determines the final outcome.

Some aspects of the setup deserve comments. The choice to make only one player be active in the second phase is made in order to ensure that for every possible payoff combination the subject tells us his choice about how to modify it and, therefore, his favorite payoff combination among the available ones in that particular decision node. If both players were given the possibility to modify the partial payoffs, we could not observe a choice of a final payoff combination, and not even a preference for a specific one: to assess it, we would need to know what a player believes his opponent is choosing. A viable way to obtain information from all players about second phase choices would be to ask for them with the condition that afterwards one player is selected to be active and only his choices are relevant. We preferred not to do so not only to avoid to increase the contingent nature of choices already inherent to the design, but especially to rule out a “coordination” aspect of the game: namely, if both players take increasing choice, they both have a higher expected payoff value. In general, we wanted to exclude any belief-related motivation in second phase decisions, in order to make sure that they really correspond to what players’ preferences dictate in correspondence to every combination of contribution levels and exogenous assignments (in A, B, or C situations) or proposed payoff allocations (situation D).

Situation D was added as a complement to the others in order to observe choices where reciprocity motivations can unambiguously be ruled out, as here first phase contributions do not matter (also recall that opponent’s first phase choices are not disclosed during the second phase). In particular, choices in situation D in case of equal allocations – (10,10) and (8,8) – are compared with choices in situation A when contributions, and therefore “partial payoffs”, are equals; this allows us to determine whether equal contributions actually stimulate no reciprocity or, instead, whether it is the absolute level of opponents’ contributions which prompts positive or negative feelings. Moreover, excluding reciprocity gives the possibility to isolate the interplay of inequality aversion and efficiency, in determining the ways in which players deviate from own-payoff maximizing choices.

---

<sup>4</sup> This method of eliciting strategies provides great advantages in data collection. Whether it affects behavior substantially is an open question. Brandts and Charness (2000) find no effects of this elicitation procedure with respect to the one in which opponents’ choice are disclosed before second movers act.

### 3. Predictions

In this section, we derive predictions for our game, arising from different assumptions on players' preferences. Among them, we consider the standard "own-payoff maximization" but also inequality aversion, reciprocity and efficiency orientation. The fact that such alternative assumptions, which are characterized by different ways of incorporating opponents' payoffs into players' utility functions, could help in predicting and interpreting behavior is not only intuitively plausible but also argued to be empirically supported by substantial experimental evidence of non-selfish behavior. This section, therefore, is focused in particular on pointing out the different consequences each one of these aspects would determine.

#### 3.1. "Homo economicus" Sub Game Perfection

Predictions arising from standard "homo economicus" assumptions and the application of sub game, perfect Nash equilibrium concept are straightforward. In the second phase, the active player picks  $F = 0$ , which implies spending nothing and leaving the opponent's payoff unaltered, for every possible situation and every possible opponent's contribution. Obviously, a positive  $F$  in a possible contingency implies that, in case that contingency is realized, the active player's payoff is lower than in case of choosing  $F = 0$  always. If selfishness is common knowledge, this way of playing in the second phase implies that both players in the first phase (the second phase role yet unknown) contribute zero. In fact, since the public good multiplier (0.75) is higher than  $1/N$  ( $N = 2$  being the number of "group members") and lower than 1, we have the usual "dilemma type" situation which characterizes public good games: social efficiency requires maximal contribution levels to the public fund, while own payoff maximization dictates zero contribution as a dominant strategy so that (0,0) constitutes the unique Nash equilibrium of the game. It is easy to see that for any contribution choice by the opponent, zero contribution is the best reply in terms of maximizing the first phase payoff; in fact the same holds with respect to the final payoff following the assumptions on second phase behavior just described. In particular, a zero contribution by a player, randomly selected to be inactive in the second phase, would not be "punished", nor a high contribution would be "rewarded".

As we focus in particular on active players' behavior in the second phase, we are going to see whether they tend to choose neutral decisions (i.e. zero modification), irrespectively of the decision nodes where they take them. In fact, evidence in situation A is going to reveal whether, as evidence from previous experiments seem to suggest, players are willing to spend in order to help higher and to hurt lower contributors.

#### 3.2. Inequality Aversion

According to the model by Bolton and Ockenfels (BO), for a given level of your payoff, you are happier the closer it is to the payoff average. For two-person games, a possible, simple representation coincides with the one presented in Fehr and Schmidt (FS):

$$u_i(\pi_i, \pi_k) = \pi_i - \alpha_i \max\{\pi_k - \pi_i, 0\} - \beta_i \max\{\pi_i - \pi_k, 0\}$$

Notice that  $\alpha_i$  measures how averse player  $i$  is to be "behind" his opponent  $k$ ,  $\beta_i$  to be "ahead".

If inequality aversion models are good predictors for this game, we should observe frequent non-zero choices of F when partial payoffs are not equal; active players would multiply by 3 when the opponent is behind, by -3 when he or she is ahead, independently of why this occurs, be it for contribution differentials or situation effects. More specifically, increasing the inactive's payoff when lower decreases inequality, as a transfer occurs from the "richer" to the "poorer" player, at least as long as the sign of inequality is not reversed (it is in fact possible in some cases that for some particularly high F a larger, opposite signed inequality level is reached). Decreasing the inactive's payoff when higher than the active's decreases inequality since the loss for the inactive is 3 times the loss for the active. Again, inequality is reduced for sure as long as the inequality signed is not reversed, while if reversal happens inequality could end up larger.<sup>5</sup> As inequality adverse players should tend to help the poorer and hurt the richer, this would imply a substantial shift in choices across situations A, B and C.

### 3.3. Reciprocity

Players who are reciprocity motivated tend to reward (punish) high (low) contributions by their opponents. This is compatible with the logic of "intention oriented" models such as Rabin (1993) and Dufwenberg and Kirchsteiger (2004). While these models are rather complex, the standard proposed here as a definition of "high" or "low" contribution level is the comparison with one's own contribution. As argued in the introduction, while in ex ante asymmetric settings second movers might speculate over how they would have acted in the first mover's shoes, here they can immediately compare their own contribution levels with the different hypothetical choices by their opponents. It seems therefore natural that a reciprocity oriented player should consider "nicer" than himself an opponent who has contributed more, and "meaner" an opponent who has contributed less, and therefore spend something to increase her payoff in the former case, and to decrease it in the latter.

While this informal definition of reciprocity does not allow more precision in deriving quantitative predictions, we can conclude that a "categorical imperative" in Kantian sense that a reciprocity-minded agent should fulfill, can be expressed as follows: "never hurt anybody who contributed more than yourself, never help anybody who contributed less than yourself". That is, decreasing a higher contributor's payoff and increasing a lower contributor's constitute unambiguous violations of reciprocity imperatives.

Summing up, reciprocity oriented players tend to favor higher and hurt lower contributors, in all situations A, B and C.

### 3.4. Social Welfare and "Efficiency": Anormal and "Reciprocity Augmented"

The formulation in Charness and Rabin (2002) is arguably the clearest example of "efficiency-minded" preferences.

The idea is simple: besides your own payoff, you also care about the following social welfare function:

$$W_i(\pi_1, \dots, \pi_i, \dots, \pi_N) = \delta \min\{\pi_1, \dots, \pi_i, \dots, \pi_N\} + (1 - \delta) \sum_j \pi_j$$

---

<sup>5</sup> The inequality sign was rarely reverted. Only 2 players (out of 32) took choices (twice and once) which reverted and increased inequality in some decision nodes.

The social welfare function for two players gets reduced to the weighted average of the sum of the two results (the *efficiency* component) and the lowest one of the two (the *Rawlsian* component).

A player putting weight  $\lambda_i$  on this social component maximizes:

$$u_i(\pi_1, \dots, \pi_i, \dots, \pi_N) = (1 - \lambda_i)\pi_i + \lambda_i W_i(\pi_1, \dots, \pi_i, \dots, \pi_N)$$

A prediction arising from this model is the absence of any payoff reduction activity: if you spend to reduce your opponent's payoff, both your payoff and the social welfare part get lower. What we could observe, instead, are actions of giving, since efficiency increases if you spend F to give 3F. In particular, giving to the opponent when she is behind in the partial results should be frequent. Of course, to derive clear-cut predictions over the chosen quantity one needs to know the exact value of the parameters involved. In the simplest version of efficiency-oriented preferences  $\delta = 1$ , i.e. the *Rawlsian* or *maximin* part would be removed. In fact, this is the idea referred to, for instance, in Bolle and Kritikos (2001).

In the “reciprocity augmented” version, Charness and Rabin's model is integrated with an evaluation of how social minded the opponent is, i.e. how big is her own  $\lambda_i$ , according to her behavior. In our design, this would push choices in the direction specified in the previous subsection, so that “punishment” towards lower contributors could occur and the “reward” component should augment increasing activities towards higher contributors. If positive and negative reciprocity are assumed to be similar in intensity, then the tendency to increase non active's payoffs should still be stronger than its opposite.

To sum up, social welfare models predict no decreasing activity in any of the four situations. When reciprocity enters into play, hurtful choices could arise, but never towards higher contributors, nor in any nodes in situation D.

## 4. Comparisons

We are going to observe whether non-zero choices are substantially present in our data. If that is the case, we get a confirmation that selfish preferences do not lead to good predictions of actual behavior. As pointed out above, it suffices to see that in situation A contribution levels matter in driving second phase decisions.

Once we agree on the relevance of “other regarding” preferences, we are interested in exploring in our data which motivational aspect is more robust. Our design is suitable for two pair-wise comparisons, where the previously mentioned predictions are not compatible.

### 4.1. Inequality Aversion Versus Reciprocity

In previous public good games where a second phase was added, it usually happens, that reciprocity and inequality aversion push behavior towards the same direction. Namely, as the lowest contributors are the richest, a decision to “punish” actually reduces inequality, as long as the negative effect on the payoff of the “punished” is larger than the expense incurred by the “punisher”; rewards favoring the highest contributors also reduce inequality as these players are poorer than their opponents.

In our design, there are in fact decision nodes in which this compatibility occurs, in particular that is always the case in “situation A”, where the exogenous assignments following the first phase are equal (4 to each player) and therefore if a player is richer (poorer) than his opponent, that is because his contribution was lower (higher) than hers. In these circumstances it is not possible to determine which tendency is stronger, between inequality aversion and reciprocity, in case behavior deviates (as typically does) from standard “selfishness” assumptions. Situations B and C were therefore devised in order to “impose” the sign of inequality: in situation B, the active player’s payoff is always at least as high as the non active’s, vice versa in situation C.

A simple way to analyze whether inequality aversion matters is to observe whether behavior changes across the three situations in front of the same choice by the opponent. For instance, if the choice by both players is the same, does it stimulate the same reaction in situations A, B and C? If so, we could conclude that payoff differentials exogenously determined do not matter, and the response to a given choice is purely “reciprocal”.

Also, using reciprocity in the way we defined it, in which the standard of comparison used by active players to evaluate opponents’ contribution is precisely their own contribution, we are going to focus in particular on the following decision nodes, where unlike in the previous examples in the literature the directions in which reciprocity and inequality aversion recommend to modify opponents’ payoffs are opposite.

- Inactive player’s payoff is lower despite the fact that her contribution was lower. This can happen only in situation B. In particular, decreasing her payoff is incompatible with inequality aversion but compatible with reciprocity; vice versa, increasing her payoff is compatible with inequality aversion but incompatible with reciprocity.
- Inactive player’s payoff is higher despite the fact that her contribution was higher. This can happen only in situation C. In particular, decreasing her payoff is compatible with inequality aversion but incompatible with reciprocity; vice versa, increasing her payoff is incompatible with inequality aversion but compatible with reciprocity.

## 4.2. Inequality Aversion Versus Efficiency

As already argued, if a “drive to efficiency” motive exists, it should foster helping behavior in the second phase of the game, which increases the total size of the “pie” as the benefit for the inactive player is three times as large as the sacrifice by the active, and prevent hurting behavior, which obviously reduces the total “pie” as both players’ payoff is lowered. Of course in our design there are circumstances where efficiency interacts with both, other tendencies we are considering, inequality aversion and reciprocity. With the goal in mind to compare the relative strength of equality and efficiency as attractive concepts in players’ mind, we want to focus on decision nodes where reciprocity motives are not prompted.

We are going to observe behavior in situation A, with particular focus on choices towards equal contributors, i.e. agents who put the same contribution level in the first phase. If efficiency motives are at work, prevalence of increasing activity should emerge, while equality motives reinforce 0 as the best choice (which also coincides with the “selfish” choice). The rationale for this observation is straightforward, based on the assumption that equal contribution should not stimulate nice nor bad feelings. In situation C, in case of equal contributions the inactive player is 4 payoff units richer. In this case, a drive towards equality calls for decreasing her payoff, which is not compatible with neither with efficiency orientation, nor with “reciprocity augmented” social welfare models.

A natural objection to this approach could be that high contributions could prompt rewarding behavior also by players who had taken the same choice, and the opposite for low contributions. This possibility is actually controlled for in situation D, where first phase behavior does not enter into the determination of “partial payoffs” and therefore reciprocity motives are absent by definition (recall that the opponent’s choice is not disclosed).

In situation D, we are going to observe whether behavior when exogenous allocations are equal (both 10 and both 8) is significantly payoff-increasing; the alternative hypothesis, zero choices, is compatible with inequality aversion, although also with selfishness, as pointed out above. An obvious way to sort out the two possible reasons to choose zero is observing what happens in the other choices in situation D, where exogenous allocations are unequal. There, decreasing when behind and increasing when ahead is of course incompatible with pure selfishness, while it is consistent with inequality aversion. Therefore, this kind of behavior would preclude a “pure selfishness” explanation as justification for zero choices in equal allocations in situation D, and after equal contributions in situation A. Moreover, those nodes in situation D where the inactive player is richer provide evidence where inequality aversion drives towards hurtful behavior which, as already underlined, is incompatible with social welfare or efficiency maximization.

## 5. Results

We present results in A, B and C situations, first. We especially focus on the relationship between contribution differentials and behavior in each one of them. Then we further organize data collected in A, B and C into “cross-situational” results, to see which differences emerge, among them. Afterwards we show results from situation D. Notice that we talk about positive and negative decisions, to refer to the ones which would increase and decrease inactive players’ payoffs if implemented. Also note that we simply refer to affecting higher, lower or equal contributors when talking about choices, although it should be kept in mind that only one for each active player is actually implemented, after the random draw takes place and the real inactive player’s contribution is verified.

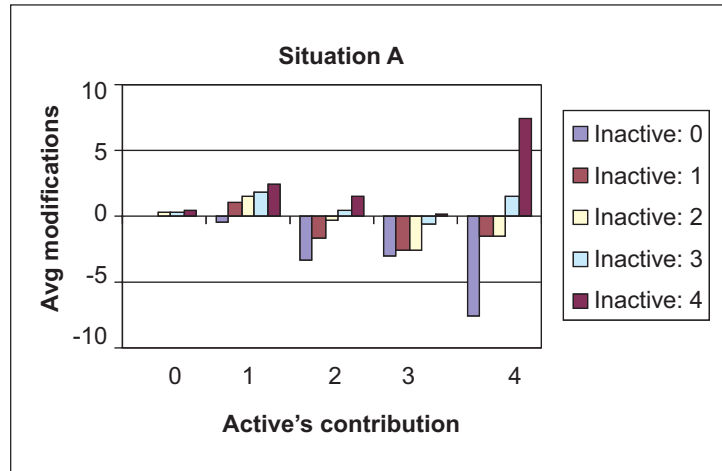
### 5.1. Situation A

Result A1: overall average modification decision is slightly negative (-0.3375) this mean is not significantly different than 0, according to the  $t$  test (95% confidence interval); 15 players’ average decision is negative, 10 players’ positive. This difference in frequency is not statistically significant at any conventional level.

Result A2: average choices towards lower contributors are always strictly negative (i.e. decreasing), with the only exception of 4-contributors (only two participants) choice towards 3-contributors. On the other hand, averages towards higher contributors are always strictly positive, with the exception of 0-contributors towards 1-contributors (all choices were 0) (Figure 1).

**Figure 1**

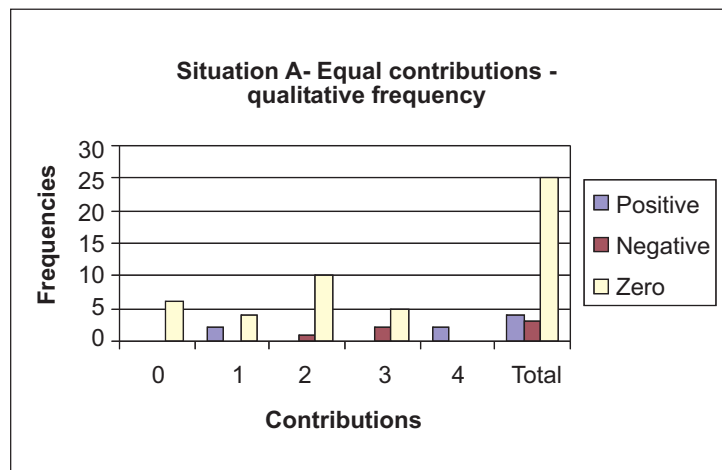
Average modifications according to contribution combinations in situation A



Result A3: equal contributions prompt zero choices, in 25 out of 32 cases. Zero choices are, therefore, significantly more frequent than non-zero choices (binomial test,  $p < 0.01$ ). Moreover, the frequency of positive and negative choices, 4 and 3, is not significantly different (Figure 2). Not surprisingly, a  $t$  test (confidence interval 95%) does not reject the hypothesis that the average choice is zero.

**Figure 2**

Modifications following equal contributions in situation A: qualitative frequencies



Result A4: lower contributions prompt decreasing choices. Out of the 26 players who contributed more than zero in the first phase, 18 take, on average, negative choices and only 1 positive choices. The frequency of negative choices is significantly higher (one-sided binomial test,  $p < 0.001$ ).<sup>6</sup>

Result A5: higher contributions prompt significantly more increasing (12) than decreasing (1) average choices (one-sided binomial test,  $p < 0.005$ ). Zero averages (17), however, are higher here.

Result A6: 23 players out of 32 feature positive averages towards higher contributors and/or negative average towards lower contributors. Of these 23, only 1 took one positive choice towards a lower contributor, none a negative choice towards a higher contributor. Even excluding this case, the frequency of players hurting lower and/or helping higher contributors is significantly higher than the frequency of players choosing an all-zero schedule (7) (two-sided binomial test,  $p < 0.01$ ). The frequencies of players choosing average negative choices both towards higher and towards lower contributors (1) and players taking positive average choices in both cases (1) are extremely low.

Result A7: towards 2 contributors, modifications by higher contributors were less favorable than the ones by lower contributors (one-sided Mann-Whitney test,  $p < 0.01$ ).

Result A8: own contribution is inversely correlated with average decision (-0.48 Spearman correlation rank, significant at 1% level).

Overall, results point out the tendency to take positive and negative decisions towards higher and lower contributors, respectively. Moreover, results A7 and A8 show that such decisions are indeed based on contribution differentials, rather than on absolute levels. In fact, the lower one's own contribution, the nicer his average behavior towards inactive players.

## 5.2. Situation B

Recall that in this situation the active and the inactive players receive 4 and 0 payoff units, respectively, from the experimenter, on top of their first phase results. Therefore, the active player is never poorer.

Result B1: 17 players took, on average, positive choices, 3 players negative ones. The difference in these frequencies is significant (one-sided binomial test,  $p = 0.001$ ). The overall average (0.86) is also found to be significantly positive by a  $t$  test (95% confidence interval).

Result B2: all averages are positive except for choices by 4-contributors towards 1-contributors (zero) and 0-contributors (negative) and by 3 and 1-contributors towards 0-contributors (negative) (Figure 3).

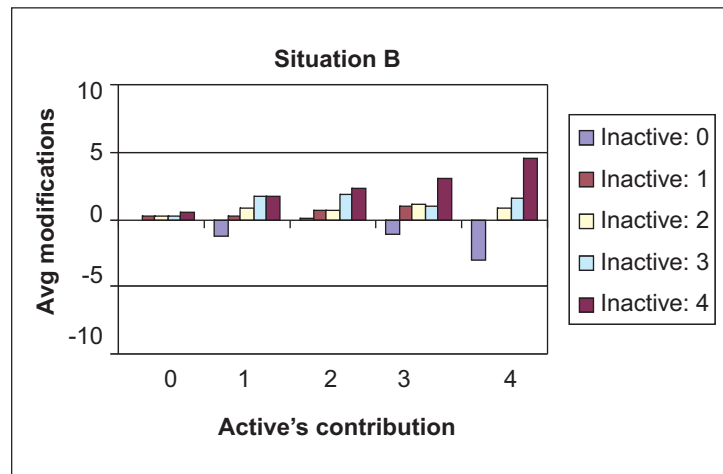
---

<sup>6</sup> Of the 18 averages, all but one are composed by non-positive choices. All the 7 zero averages are composed by zero choices only.



**Figure 3**

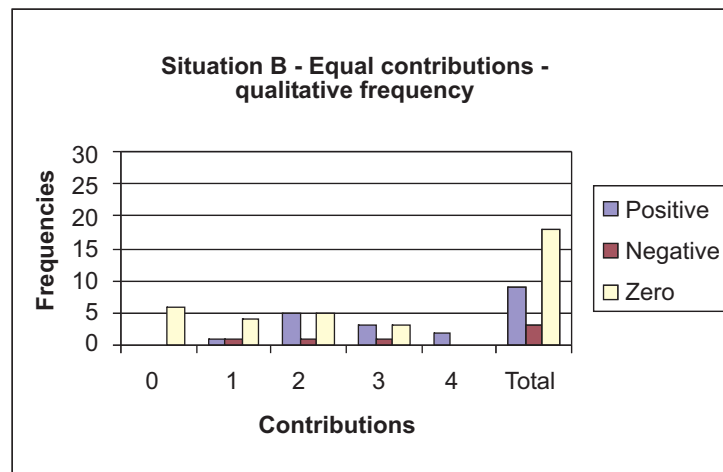
Average modifications according to contribution combinations in situation B



Result B3: In case of equal choice, positive choices (11) are significantly more frequent than negative (3) (one-sided binomial test,  $p < 0.05$ ); 18 players chose zero (Figure 4).

**Figure 4**

Modifications following equal contributions in situation B: qualitative frequencies

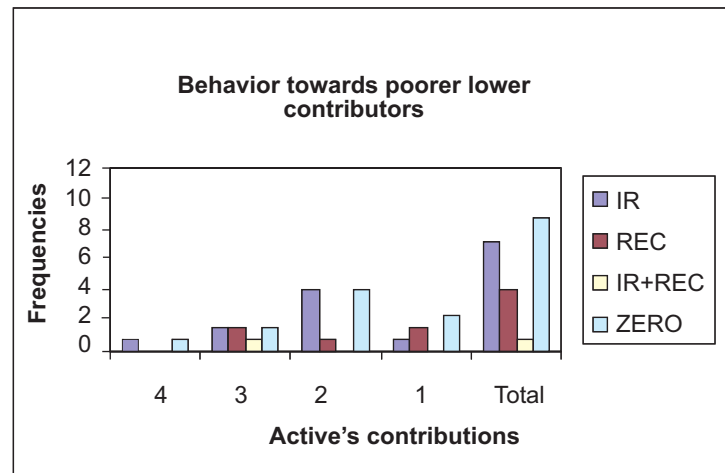


Result B4: towards lower contributors, 9 players take, on average, positive choices and 6 negative choices (these frequency differences are not statistically significant); neither averages include choices of opposite signs. The remaining 11 players took all zero choices, except for one who took positive and negative choices, resulting in zero average.

See Figure 5 where, like in others which follow, we refer to IR and REC as “inequality reducing” and reciprocal behavior, respectively. In this case, 9 players are of “IR” type.

## Figure 5

Situation B: frequencies of inequality reducing and reciprocal behaving players, according to behavior towards lower contributors with lower partial payoffs



Result B5: towards higher contributors, 17 players out of 30 take on average positive choices, and 1 negative choices. The difference in frequency is significant (one-sided binomial test,  $p=0.001$ ).

These results indicate a general tendency to increase inactive players' payoffs. Notably, the players who hurt lower contributors are fewer than the ones who help them.

### 5.3. Situation C

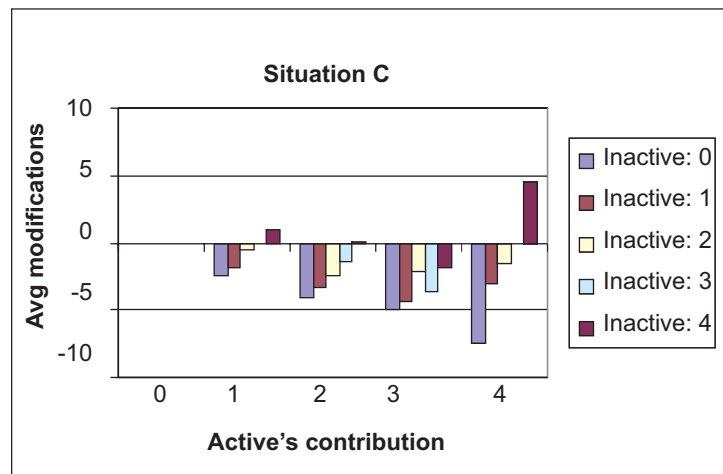
Recall that in this situation the active and the inactive players receive 4 and 8 payoff units, respectively, from the experimenter, on top of their first phase results. Therefore, the active player is never richer.

Result C1: 16 players took, on average, negative choices, 4 players positive ones. The difference in these frequencies is significant (one-sided binomial test,  $p<0.01$ ). The overall average (-1.82) is also found to be significantly negative by a  $t$  test (95% confidence interval).

Result C2: all averages are negative except for positive values towards 4-contributors by 1 and 4-contributors (Figure 6) and zero values always issued by 0 contributors towards all choices.

**Figure 6**

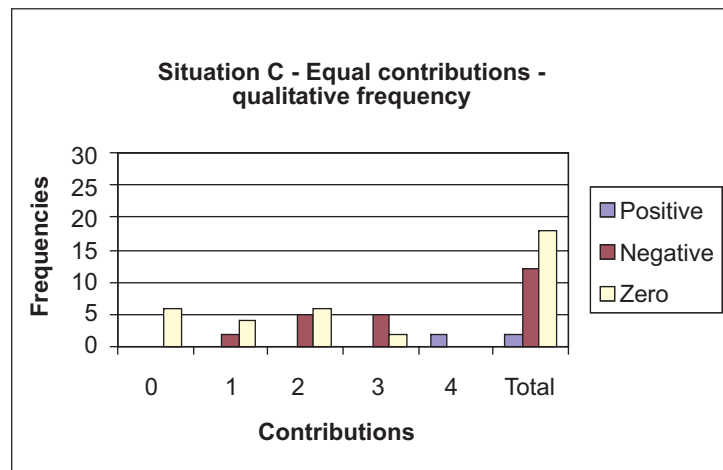
Average modifications according to contribution combinations in situation C



Result C3: towards equal contributors, 2 chose to help, 12 to hurt, 18 zero (Figure 7). Positive choices are significantly less frequent than negative ones (one-sided binomial test,  $p < 0.01$ ).

**Figure 7**

Modifications following equal contributions in situation C: qualitative frequencies

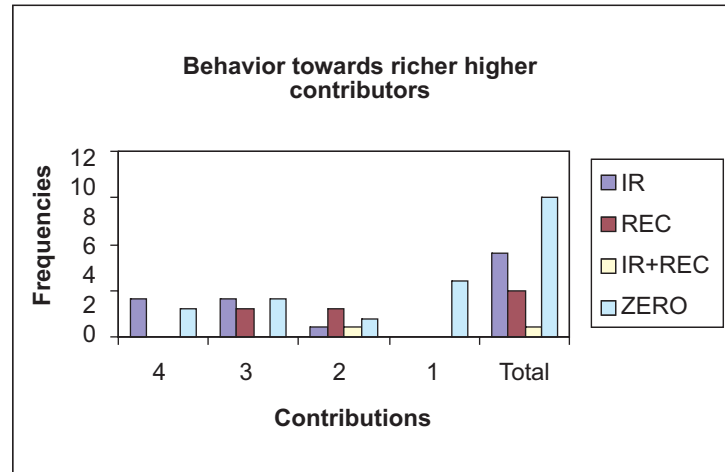


Result C4: towards lower contributors, no single positive choice was taken. 19 players (out of 30) took, on average, negative choices.

Result C5: towards higher contributors, 5 took, on average, positive choices, 9 negative ones (Figure 8). This difference in frequencies is not statistically significant.

## Figure 8

Situation C: frequencies of inequality reducing and reciprocal behaving players, according to behavior towards higher contributors with higher partial payoffs



Here the general tendency is to reduce inactive players' payoff. As we have just seen, players who hurt higher contributors are more than the ones who help them.

### 5.4. Cross-situational Results

Result ABC1: players' average choices are decreasing in the exogenous assignment towards inactive players, i.e. they decrease if we order them from B to A to C, according to the Page test for ordered alternatives ( $p < 0.01$ ).

Result AB1: 20 players decrease their average choices in B, with respect to A; 5 increase them. The first case is significantly more frequent (one-sided sign test,  $p < 0.005$ ). Considering also the magnitude of such changes, we find that average choices in A are lower than average choices in B, according to the Wilcoxon rank-order test for pair-wise replicates (one-sided  $p < 0.01$ ) and the  $t$  test (95% confidence interval).

Result AC1: 21 players increase their average choices in B, with respect to A; 4 increase them. The first case is significantly more frequent (one-sided sign test,  $p < 0.001$ ). Considering also the magnitude of such changes, we find that average choices in A are higher than average choices in C, according to the Wilcoxon rank-order test (one-sided  $p < 0.01$ ) and the  $t$  test (95% confidence interval).

Result AB2: choices towards equal contributors are significantly more helpful in situation B than in A. In particular, the frequency of helpful choices is higher in B at 5% level of significance.

Result AC2: choices towards equal contributors are significantly more hurtful in situation C than in A. In particular, the frequency of hurtful choices is higher in C at 5% level of significance.

Result AB3: towards lower contributors, frequency of negative average choices are higher in A (one-sided binomial test,  $p < 0.05$ ) than in B; frequency of positive average choices are higher in

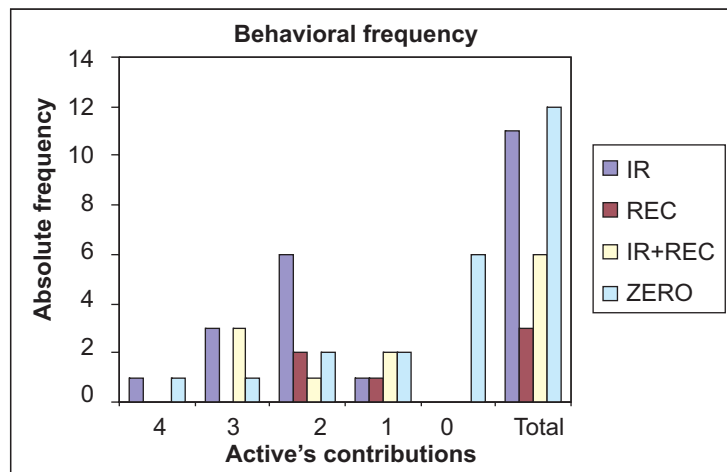
B (one-sided binomial test,  $p < 0.05$ ). 16 out of 26 players changed the sign of their averages in the two treatments: in 14 cases, the sign is more favorable to inactive players in B (more frequent than the opposite, one-sided binomial test,  $p < 0.005$ ).

Result AC3: towards higher contributors, the frequency of positive average choices is higher in A than in C, but only at 10% level of significance (one-sided binomial test,  $p = 0.72$ ); the frequency of negative average choices is higher in C (one-sided binomial test,  $p < 0.05$ ). 16 out of 30 players changed the sign of their averages in the two treatments: in 15 cases, the sign is less favorable to inactive players in C (more frequent than the opposite, one-sided binomial test,  $p < 0.001$ ).

Result BC1: 18 out of 32 players chose either to help lower contributors in B or to hurt higher contributors in C or both; 10 chose either to hurt lower contributors in B or to help higher contributors in C or both; 11 displayed only the former kind of choice, 3 only the latter (Figure 9). The probability of the former “player type” is found to be higher than the probability of the latter at 5% level of significance (one-sided binomial test).

### Figure 9

Situations B&C: inequality reducing and reciprocity oriented players, according to choices towards lower contributors with lower partial payoffs and choices towards higher contributors with higher partial payoffs



These results indicate substantial differences in behavior across the situations A, B and C, based on the tendency to be nice towards inactive players when the exogenous assignments make them relatively poorer, and to hurt them when those assignments make them richer. The last result highlights a significantly higher willingness to help lower contributors or to hurt higher contributors than to help richer or to hurt poorer players.

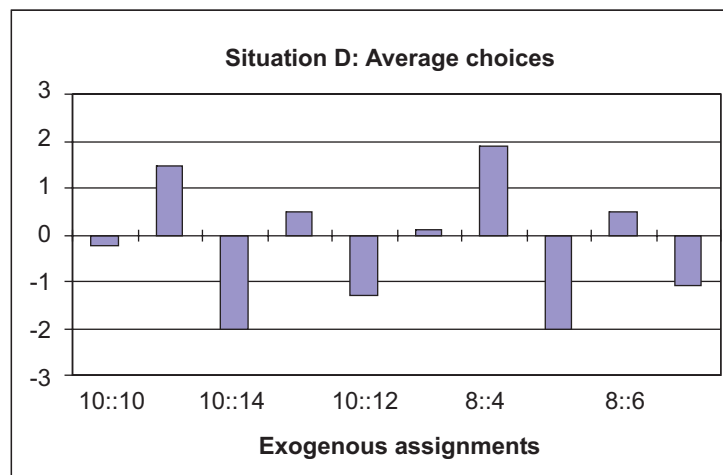
### 5.5. Situation D

Result D1: equality in partial payoffs prompts zero modifications; 28 out of 32 players chose zero both in (10:10) and in (8:8) exogenous assignments. Zero choices are therefore significantly more frequent than non-zero choices (binomial test,  $p < 0.01$ ). Moreover,  $t$  test using 95% confidence interval does not reject the zero mean hypothesis.

Result D2: inequality prompts choices which typically reduce it. Frequency of positive choices is always found to be higher at 1% level of significance, using the binomial test, than negative choices, in all the decision nodes where the inactive player is poorer in the initial distribution; conversely, negative choices are more frequent, again at 1% level of significance, whenever the inactive player is initially richer. In particular, averages are positive when the opponent's partial payoff is lower, and negative when it is higher, using a *t* test with 95% confidence interval (Figure 10).

**Figure 10**

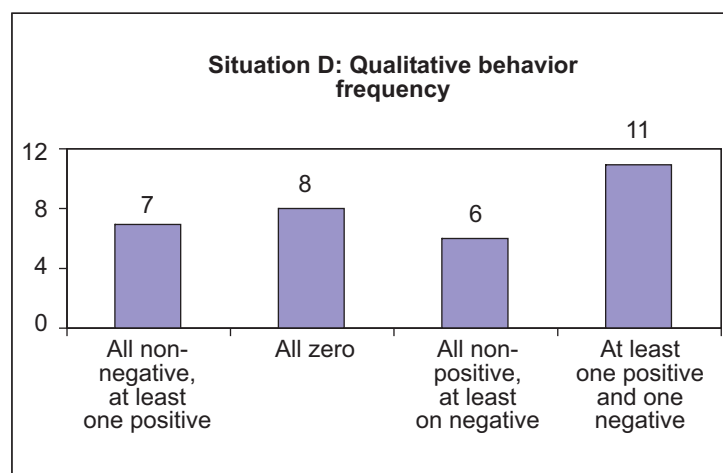
Situation D: average choices for every assignment



Result D3: one quarter (8/32) of players choose an all-zero schedule; among the others, 7 players never hurt and 6 players never help; 11 do both things at *least* once for each (Figure 11).

**Figure 11**

Situation D: qualitative behavior frequencies, according to each player's set of choices



Result D4: 19 players took at least one inequality reducing, and no inequality increasing decisions. Defining these players as inequality reducing types, their frequency is higher than the selfish (all zero) type (binomial test,  $p < 0.05$ ).

## 6. Social Preferences and Our Evidence: Discussion

### 6.1. Non-selfish Preferences

The overall picture from situation A constitutes a replication of the tendency found in previous papers where a reward/punishment phase is added to a standard public good game: there is a substantial demand for modification of opponents' payoffs when contribution levels –and therefore “partial results”– are different. In particular result A3 indicate that only 7 out of 32 players choose the “all zero” schedule which maximizes their expected payoff, and among the others the vast majority (25) takes at least one decision in the direction which is consistent both with inequality aversion and with reciprocity.

Evidence from situation D shows that even after removing the effects of the first phase and, especially, the possibility to condition decisions on opponent's contributions, substantial “inequality reducing” activity is still displayed by active players. These results reinforce the observation that most players behave in a way which is not compatible with predictions arising from standard assumptions.

### 6.2. Inequality Aversion Versus Efficiency Orientation

In order to observe whether efficiency matters in players' behavior, in particular when put in contrast with equality, we need to see whether or not increasing activity is substantially more frequent than decreasing. In situation A, where payoff differentials before the second phase only depend on contributions, we find no evidence of a tendency to increase inactive players' payoffs; indeed, result A1 rather indicates a slight tendency in the opposite direction. Notice that this occurs despite the fact that average contribution choice by active players (1.81) is below 2, the median value among possible contributions. On the other hand, results A4 to A8 show that players do correct inequality. This behavior could actually be related to reciprocity orientation; on the other hand, all the results in situation C clarifies, with choices taken with respect to the same contribution levels, that a substantial fraction of players behave in a way which is not compatible with social welfare, or efficiency based, models.

The evidence of no role of efficiency is strengthened by the data from those circumstances where reciprocity motives are assumed to be absent. That happens, according to the simple reciprocity concept we are using here, when the two contribution levels are the same. In that case, result A3 tells us that players typically choose not to modify opponents' payoffs. That indicates that efficiency does not drive behavior and players would only spend when inequality aversion and/or reciprocity motives are stimulated.

However, this reciprocity concept may be debatable; in particular, high contributions could stimulate nice feelings per se, and low contribution the desire to hurt even by equally low contributors, although results A7 and A8 seems to be consistent with our intuition. Anecdotic evidence by the two active 4-contributors suggests such a thing; however, that is not the case for 3-contributors where an average choice by seven players was negative. In order to have more conclusive data on what happens in the –less debatable– absence of reciprocity motives, we refer to results from situation D. There we find that equal assignments stimulate no expense (result D1); moreover, average reductions of opponents' payoffs are, in absolute value, bigger than increases in the correspondent payoff differences. A tendency towards non-neutral choices can be found, instead, when assignments are unequal (D2 and D4). In particular, 17 players take

hurtful choices when the inactive player is richer, a behavior which is incompatible with any definition of efficiency or social welfare maximization in a one-shot interaction set-up.

Finally, if we take as observations single players, we find (D3) that just a minority, take only efficiency oriented decisions, i.e. present at least one and no negative choices, while most players do display a tendency to “correct” inequality.

Overall, we find that efficiency considerations did not play any role in players’ behavior, which makes it easier to focus on the next comparison.

### 6.3. Inequality Aversion Versus Reciprocity

Overall results in situation A, B and C, as well as the related cross situational results, show that behavior differ substantially across the three treatments, in each one of which choices are taken with respect to all the five possible contribution levels. Results B1 and C1 show an unambiguous tendency to hurt the richer and to help the poorer; consistently, moreover equal contributors are treated differently in the three situations. These results together with, the already mentioned evidence from situation D, indicates that inequality aversion does matter.

Perhaps more interesting is what we find when this motive is put in contrast with reciprocity. Result B4 tells us that there are more players who behave towards poorer lower contributors in the way recommended by inequality aversion (i.e. helping them) than in the way prompted by negative reciprocity (i.e. hurting them). Result C5 tells us that that inequality aversion also drives more players than positive reciprocity, when opponents are richer higher contributors. When we take the overall B&C schedule, we find (result BC1) that 11 players are inequality averse and 3 reciprocity oriented<sup>7</sup> when we observe the overall circumstances described (behavior towards poorer lower contributors and towards richer higher contributors), while 6 took at least one choice in both directions.

Our data indicate that, when in front of a trade-off, there are significantly more players who follow inequality aversion than players who follow reciprocity concerns. We should point out that –unlike in the case for efficiency– this comparison does not indicate the absence of reciprocity motives. 10 out of 32 players (see the previous footnote), in fact, took at least one decision that can only be explained by reciprocity motives. The case for inequality aversion is stronger, as a slight majority, (18 out of 32) players took at least one choice that can only be explained by inequality aversion.

## 7. Conclusions

Our experimental design provides a test of the robustness of different aspects of “other regarding” preferences as driving forces behind non-selfish human behavior. In fact, the design is made to potentially discover, not only egalitarian tendencies, but also the drive for

---

<sup>7</sup> They become 4 if we also take into account a hurtful choice by a 4-contributor towards a 0-contributor in situation B. However, in that circumstance the 0-contributor was not poorer, but had an equal partial payoff. This equality was broken by the hurtful choice. Increasing to 4 the “reciprocity oriented” brings up the p value in the one-sided binomial comparison to 0.059.



“efficiency”, in terms of total payoff or “social surplus” arising from possible choices, and reciprocity, called upon –we argue– in a natural way, based on comparing contribution levels.

The potential “social surplus”, which efficiency oriented agents would pursue, can be easily perceived by observing that for every payoff unit spent to “help” your opponent, she receives three times as much. The proportion by which you can “hurt” your opponent is the same, but in such a case both players lose, albeit the opponent’s loss is three times larger. It is therefore evident the “wasteful” nature of hurting decisions, and the “social payoff augmenting” nature of helping choices.

Our results clearly speak against any effects of efficiency considerations, in decisions where these present a trade-off with respect to equality. In this respect they confirm evidence found, among the others, in Güth et al. (2003), but contradict results found in Charness and Rabin (2001) and Bolle and Kritikos (2001). A main feature which distinguishes their experimental designs and the one presented here is the binary choice the former present to subjects. For instance, most subjects in Charness and Rabin’s article prefer (400,700) (own payoff first) to (400,400). Here subjects are presented with a number of possible payoff combinations and each one of them have various possibilities. A possible effect of this difference could be a status quo bias, in the sense that one possibility is presented as the temporary one that a subject could modify. However, the main differences in results, in terms of the direction of the change, cannot be explained in terms of status quo bias, which would simply make agents less inclined to affect opponents’ payoffs.

The other main result highlights further the robustness of inequality aversion as a social phenomenon. We find that a substantial number of subjects are willing to take inequality reducing decisions even if these move against what reciprocity would suggest, once we take, as a standard of evaluating the kindness of one’s peer, the comparison between her behavior and one’s own in terms of contribution levels. However, our results also indicate that reciprocity, both negative and positive, does play a role in driving average population behavior. As a matter of fact we do find instances, although less frequently, of choices which increase inequality in order to reward (punish) nicer (meaner) agents. Our results in terms of comparative effects of inequality aversion and reciprocity are more in line with respect to what is found in Bolton et al. (2000) than in Falk et al. (2000). In common with the latter we find that both aspects matter, although here we find a clear dominance of equality principles; therefore our results can be seen as a less extreme evidence, with respect to the one in Bolton et al. (2000) where reciprocity aspects were found to be inexistent.

Summing up, results in a game where subjects play in ex ante symmetric roles highlight inequality aversion is a robust phenomenon which persists when put in contrast to reciprocity principles and does not appear to be softened by efficiency considerations. When seen in the broad perspective of ongoing studies on these issues, these results do provide new clarifications but at the same time call for further investigation. In our view, future research should shed more light on which reference group is taken by individuals sharing particular behavioral motives, inequality aversion in particular, and then on how these influence outcomes in economically and socially relevant situations. For instance, studies incorporating social preferences may help clarifying how incentive and control schemes affect the evolution of cooperation in groups, or how cartels are maintained among limited number of firms. Clearly there is great abundance of issues which makes this line of research stimulating and experimental testing an increasingly important tool.

## References

- Abbink, K., Irlenbusch, B. and Renner, E. (2000), "The Moonlighting Game - An Experimental Study on Reciprocity and Retribution", *Journal of Economic Behavior and Organization*, 42, pp. 265-277.
- Andreoni, J., Harbaugh, W., and Vesterlund, L. (2003), "The Carrot and the Stick: Rewards, Punishments and Cooperation", *American Economic Review*, 93 (3), pp. 893-902.
- Berg, J., Dickhaut, J., and McCabe, K. (1995), "Trust, Reciprocity and Social History", *Games and Economic Behavior*, 10, pp. 122-142.
- Bolle, F., and Kritikos, A. (2001), "Distributional Concerns: Equity or Efficiency Oriented?", *Economics Letters*, 73, pp. 333-338.
- Bolton, G.E., Brandts, J., and Ockenfels, A. (2000), "Measuring Motivations for the Reciprocal Responses Observed in a Simple Dilemma Game", *Experimental Economics*, 3, pp. 207-221.
- Bolton, G.E., and Ockenfels, A. (2000), "A Theory of Equity, Reciprocity and Competition", *American Economic Review*, 100, pp. 166-193.
- Brandts, J. and Charness, G. (2000), "Hot and Cold Decisions and Reciprocity in Experiments with Sequential Games", *Experimental Economics*, 2, 3, pp. 227-238.
- Charness, G. and Rabin, M. (2002): "Understanding Social Preferences with Simple Tests", *Quarterly Journal of Economics*, 117, pp. 817-869.
- Dufwenberg, M. and Kirchsteiger, G. (2004), "A Theory of Sequential Reciprocity", *Games and Economic Behavior*, pp. 268-298.
- Falk, A., Fehr, E., and Fischbacher, U. (2000), "Testing Theories of Fairness - Intentions Matter", Working Paper No. 63, University of Zurich.
- Fehr, E. and Gaechter, S. (2000), "Cooperation and Punishment in Public Goods Experiments", *American Economic Review*, 90, 4, pp. 980-994.
- Fehr, E., Kirchsteiger, G., and Riedl, A. (1993), "Does Fairness Prevent Market Clearing? An Experimental Investigation", *Quarterly Journal of Economics*, 108, pp. 437-460.
- Fehr, E. and Schmidt, K.M. (1999), "A Theory of Fairness, Competition and Cooperation", *Quarterly Journal of Economics*, 114, pp. 817-868.
- Güth, W., Kliemt, H., and Ockenfels, A. (2003), "Fairness versus Efficiency: An Experimental Study of (Mutual) Gift Giving", *Journal of Economic Behavior and Organization*.
- Ostrom, E., Walker, J., and Gardner, R. (1992), "Covenants With and Without a Sword: Self-Governance is Possible", *American Political Science Review*, Vol. 86, No. 2, pp. 404-417.
- Offerman, T., (2002), "Hurting Hurts More Than Helping Helps", *European Economic Review*, 46, pp. 1423-1437.
- Rabin, M. (1993), "Incorporating Fairness into Game Theory and Economics", *American Economic Review*, 83 (5), pp. 1281-1302.

- Sefton, M., Shupp, R., and Walker, J. (2002), "The Effect of Rewards and Sanctions in Provision of Public Goods", CEDEX Research Paper, University of Nottingham.
- Selten, R. (1967), "Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopol-experiments", in Sauer mann. H. (ed), "Beiträge zur experimentellen Wirtschaftsforschung", J.C.B. Mohr (Paul Siebeck), Tübingen, 136-168.
- Selten, R., Abbink, K., Buchta, J., and Sadrieh, A. (2003), "How to play 3X3-games: a strategy method experiment", *Games and Economic Behavior*, 45, 1, pp.19-37.

## Appendix 1

Instructions (translated from Spanish)

Thank you very much for participating in this experiment on economic decisions. During the session you will be randomly matched with another participant, whose identity will not be revealed at any time. According to the following rules, your decisions and your partner's decisions will determine the amount of money you are going to receive after the experiment.

### 1. First Phase

In the first phase of the experiment, you and your partner are endowed with 4 “experimental units” each. Both of you decide independently how many units to allocate in a “fund” (the quantity must be an integer between 0 and 4). At the end of the first phase, each player gets the units he/she did not allocate in the fund plus  $\frac{3}{4}$  (three fourth) of the sum of the units both players allocate in the fund. Denote  $x$  your contribution in the fund and  $z$  your partner's contribution, your first phase result will be:

$$4-x+\frac{3}{4}(x+z)=4-\frac{1}{4}x+\frac{3}{4}z$$

and your partner's first phase result will be:

$$4-z+\frac{3}{4}(x+z)=4-\frac{1}{4}z+\frac{3}{4}x$$

The attached Table 1 indicates the results corresponding to all possible combinations of the two contributions, yours and your partner's. These results constitute the basis for the determination of partial results, as explained in the 2.1 paragraph of these instructions.

### 2. Second Phase

A random draw determines which participant between you and your partner is “active” and which one is “inactive”.

The “inactive” player has not any more decisions to be taken, and at the end of the experiment will be informed of his or her partner's relevant decisions in the first and in the second phase, the final results and the corresponding payment.

The “active” player takes his or her decisions, as explained in the 2.2 paragraph, according to which the PARTIAL RESULTS will be converted into the FINAL RESULTS.

#### 2.1. Determination of the Partial Results

In the second phase there are four possible “situations” - A, B, C, D - which affect the PARTIAL RESULTS. After the “active” player takes his or her decisions for each possible situation, a random draw determines which one takes place.

In situations A, B and C the “active” receives 4 additional units, which are added to the first phase results to determine his/her “PARTIAL RESULT”; to the “inactive” player 4 units in

situation A, 0 in situation B, 8 in situation C are added to his/her first phase result to determine his/her "PARTIAL RESULT". That is to say:

SITUATION A: "active" receives 4; "inactive" receives 4

SITUATION B: "active" receives 4; "inactive" receives 0

SITUATION C: "active" receives 4; "inactive" receives 8

The PARTIAL RESULTS in situations A, B and C, therefore, equal the first phase results plus these assignments.

In situation D, instead, the possible PARTIAL RESULTS are predetermined, and do not depend on first phase decisions.

## 2.2. "Active" Player's Decisions

Each decision by the active player consists in increasing, decreasing or leaving unchanged his partner's result.

According to the following rules, the "active" player takes, for situation A, a decision for every one of the five (0,1,2,3,4) first phase contribution leaves chosen by his/her partner (he/she is not yet informed about the actual one) with the corresponding PARTIAL RESULT (the value will be shown in your decision sheet in place of the \* in your DECISION SHEET - EXAMPLE on your desk). If afterwards the random draw picks situation A, the "active" player's decision affecting the final results will be the one in correspondence with the actual first phase decision taken by the active player. The same holds for situations B and C: so the "active" player takes five decisions for each one of the three situations A, B and C.

In the sheet "SITUATION D", instead, of which you have an example on your desk, the "active" player takes a decision for each one of the possible predetermined assignments. If the random draw picks situation D, a further random draw determines which assignment determines, together with the corresponding "active" player's decision, the final results.

As mentioned above, each decision by the "active" player consists in increasing, decreasing or leaving unchanged his/her partner's result. To vary it, he/she spends a quantity which we denote F. The value of F is to be chosen among the following ones:

0; 0.5; 1; 1.5; 2; 2.5; 3; 3.5; 4

This quantity is multiplied by 3, if the "active" player can increase his partner's result, and by "-3" if he/she wants to decrease it. Notice that the "active" player's expense is the same in both cases. So, the "active" player spends 0 if he/she does not modify his/her partner result; 0.5 if he/she increases it by 1.5; 0.5 if he decreases it by 1.5; 1 if he/she increases it by 3; 1 if he/she decreases it by 3; and so forth until the maximum expense, 4, to increase or decrease his/her partner's result by 12. The expense F makes the "active" player's final result be "PARTIAL RESULT - F" and the "inactive" player's result be "PARTIAL RESULT + 3F" or "PARTIAL RESULT - 3F" depending on the "active" player's choice.

### 2.3. Randon Draw of the Situation and Determination of Final Results

After all decisions have been taken by the “active” player for every possible situation and every possible first phase decision by his/her partner or, in situation C, for every possible predetermined assignment, a random draw determines the actual situation. If the situation is A, B or C, the actual first phase decision by the “inactive” player determines which one of the “active” player’s decision is combined with the corresponding PARTIAL RESULT to determine the FINAL RESULT. If the situation is D, a further draw determines the PARTIAL RESULT, and the corresponding decision by the “active” player determines the FINAL RESULT.

#### Conversion

After the experiment, you receive individually 3 euros for your participation and the result you obtained will be converted according to following rate:

$$1 \text{ “experimental unit”} = 0.5 \text{ euros}$$

No communication is allowed among participants. If you have questions now or at any moment during the experiment, please raise your hand and will be attended personally.

Good Luck!

#### Table 1

Your Choice: X

Your Partner’s Choice: Z

First Phase Results

Your Points: Your Partner’s Points

		<b>Z</b>				
		<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
<b>X</b>	<b>0</b>	4::4	4.75::3.75	5.5::3.5	6.25::3.25	7::3
	<b>1</b>	3.75::4.75	4.5::4.5	5.25::4.25	6::4	6.75::3.75
	<b>2</b>	3.5::5.5	4.25::5.25	5::5	5.75::4.75	6.5::4.5
	<b>3</b>	3.25::6.25	4::6	4.75::5.75	5.5::5.5	6.25::5.25
	<b>4</b>	3::7	3.75::6.75	4.5::6.5	5.25::6.25	6::6

## Appendix 2

The parts filled up by the players are in bold

DECISION SHEET		ACTIVE PLAYER
<b>YOU HAVE CHOSEN 2</b>		
<b>SITUATION A:</b>	<b>YOU RECEIVE 4</b>	<b>YOUR PARTNER RECEIVES 4</b>
<b>IF YOUR PARTNER HAS CHOSEN...</b>	<b>PARTIAL RESULT YOU::YOUR PARTNER</b>	<b>YOU SPEND F (INTEGER OR ".5",MAX4)</b>
0	7.5::9.5	0.5
1	8.25::9.25	0
2	9::9	0
3	9.75::8.75	0
4	10.5::8.5	0.5
Your final result equals your partial result MINUS F. Your partner's final result equals his/her partial result PLUS 3F or MINUS 3F, according to your decision in the previous column.		
	<b>YOUR PARTNER GETS (3F OR -3F)</b>	<b>FINAL RESULT YOU::YOUR PARTNER</b>
	-1.5	7::8
	0	8.25::9.25
	0	9::9
	0	9.75::8.75
	1.5	10::10

SITUATION B:		YOU RECEIVE 4	YOUR PARTNER RECEIVES 4
<b>IF YOUR PARTNER HAS CHOSEN...</b>	<b>PARTIAL RESULT YOU::YOUR PARTNER</b>	<b>YOU SPEND F (INTEGER OR ".5",MAX4)</b>	<b>YOUR PARTNER GETS (3F OR -3F)</b>
0	7.5::5.5	0.5	1.5
1	8.25::5.25	0.5	1.5
2	9::5	0.5	1.5
3	9.75::4.75	1	3
4	10.5::4.5	1	3
Your final result equals your partial result MINUS F. Your partner's final result equals his/her partial result PLUS 3F or MINUS 3F, according to your decision in the previous column.			
	<b>YOUR PARTNER GETS (3F OR -3F)</b>	<b>FINAL RESULT YOU::YOUR PARTNER</b>	
	1.5	7::7	
	1.5	7.75::6.75	
	1.5	8.5::6.5	
	3	8.75::7.75	
	3	9.5::7.5	

SITUATION C:		YOU RECEIVE 4		YOUR PARTNER RECEIVES 4	
IF YOUR PARTNER HAS CHOSEN...	PARTIAL RESULT YOU::YOUR PARTNER	YOU SPEND F (INTEGER OR ".5", MAX4)	YOUR PARTNER GETS (3F OR -3F)	FINAL RESULT YOU::YOUR PARTNER	FINAL RESULT YOU::YOUR PARTNER
0	7.5::13.5	2	-6	5.5::7.5	5.5::7.5
1	8.25::13.25	2	-6	6.25::7.25	6.25::7.25
2	9::13	2	-6	7::7	7::7
3	9.75::12.75	2	-6	7.75::6.75	7.75::6.75
4	10.5::12.5	0.5	-1.5	10::11	10::11

Your final result equals your partial result MINUS F. Your partner's final result equals his/her partial result PLUS 3F or MINUS 3F, according to your decision in the previous column.

SITUATION D: RANDOM ASSIGNMENT			
PARTIAL RESULT YOU::YOUR PARTNER	YOU SPEND F (INTEGER OR ".5", MAX4)	YOUR PARTNER GETS (3F OR -3F)	FINAL RESULT YOU::YOUR PARTNER
10::10	0	0	10::10
10::6	1	3	9::9
10::14	-1.5	-4.5	8.5::9.5
10::8	0.5	1.5	9.5::9.5
10::12	-0.5	-1.5	9.5::10.5
8::8	0	0	8::8
8::4	1	3	7::7
8::12	-1.5	-4.5	6.5::7.5
8::6	0.5	1.5	7.5::7.5
8::10	-0.5	-1.5	7.5::8.5