Inequity Aversion and Individual Behavior in Public Good Games: An Experimental Investigation

Astrid Dannenberg[a], Thomas Riechmann[b], Bodo Sturm[a], and Carsten Vogt[c]

[a] Centre for European Economic Research (ZEW), Mannheim
[b] Faculty of Economics and Management, University of Magdeburg
[c] Department of Business Administration, Leipzig University of Applied Sciences

Abstract
We present a simple two-step procedure for a within-subject test of the inequity aversion model of Fehr and Schmidt (1999). In the first step, subjects played modified ultimatum and dictator games and were thereupon classified according to their preferences. In the second step, subjects with specific preferences according to the Fehr and Schmidt model were matched into pairs and interacted with each other in a standard public good game and a public good game with punishment possibility. Our results show that the specific composition of pairs significantly influences the subjects' performance in the public good games. We identify the aversion against advantageous inequity and the information about the co-player's type as the main influencing factors for the behavior of subjects.

JEL classification: C91, C92, H41
Keywords: individual preferences, inequity aversion, experimental economics, public goods
Corresponding Author: Astrid Dannenberg, Email: dannenberg@zew.de, Phone: +49 621 1235 332,
Fax: +49 621 1235 226.

1 Introduction

Within experimental economics there are a growing number of stylized facts which contradict the model of rational payoff maximizing actors. People cooperate in social dilemmas such as public good games (Ledyard, 1995), they reject high amounts of money in the ultimatum game (Güth et al., 1982; Camerer, 2003) and last but not least they make positive contributions in the dictator game (Kahneman et al., 1986; Forsythe et al., 1994; Camerer, 2003). The contradiction between the standard economic model of selfish behavior and empirical observations has been a challenge for both theorists and experimentalists. In the last ten years a number of theories that try to close this gap in explanatory power have been developed. Most of these theories are based on the assumption that people have some kind of other-regarding, or social, preferences. These approaches seek to overcome the discrepancies between standard game-theoretical predictions and experimental observations by altering the underlying utility function of subjects, but stick to the assumption that subjects behave rationally.

There are two different concepts of other-regarding preferences for reciprocity-free situations. The first concept goes back to Andreoni's and Miller's (2002) idea of impure altruism and has been re-formulated in a more general framework by Cox et al. (2008), who call the concept 'egocentric other-regarding' preferences. This concept sees the payoff of person A and the payoff of other persons as the direct sources of that person A's (positive) utility. Differences in payoffs between A and others do not matter. The other concept is that of inequity aversion. The models by Bolton and Ockenfels (2000) and Fehr and Schmidt (1999) are prominent examples for this approach. They assume that people are willing to pay money in order to avoid unequal payoff distributions. Consequently, other individuals' payoffs affect a person's utility level, but not necessarily in a positive way. A common property of models with other-regarding preferences is that subjects are assumed to be heterogeneous in their preferences. This implicates that theoretical prognoses about individual behavior may differ between subjects for the same decision problem.

This study aims to investigate the additional explanatory power of models with inequity-averse preferences. Thereby, we focus on the model of inequity aversion by Fehr and Schmidt (1999) in the following F&S. There are two reasons for doing this. First, the F&S model is able to explain an impressive amount of experimental evidence not in line with the standard model of selfish behavior. Second, F&S use a model which from a theoretical point of view is quite parsimonious as only two additional parameters are added to the individual utility function, which is still solely based on monetary payoffs. Moreover, both parameters of the model can be estimated with the help of simple laboratory techniques.

One interesting implication of models with other-regarding preferences such as F&S is that they allow within-subject tests, i.e. controlled experiments with the same subject but different decision problems. Due to the fact that these theories predict – given different preferences – different behavior of subjects, one may test hypotheses at the individual level with the following two-step procedure. In a first step, individual other-regarding preferences are measured by means of appropriately designed games. In a second step, the same subjects interact with each other in a controlled environment under specific rules for which hypotheses regarding the individual behavior have been derived in advance. Under the assumption that preferences are stable at least within a short time period, this approach allows a robust test of such models in the laboratory.[1]

Our study implements such a two-step procedure. In the first step, we measure the individual F&S preferences by means of two simple experiments, a modified ultimatum game (MUG) and a modified dictator game (MDG). In the second step, subjects with specific preferences are matched in pairs and interact with each other in a standard public good game and a public good game with a punishment possibility. We distinguish between three groups of pairs. In particular, we form a group of "fair" subjects where both players of the pair are highly inequity averse, a group of "egoistic" subjects where both players are very little inequity averse, and a "mixed" group where one player is "fair" and the other one is "egoistic". Due to the composition of groups with subjects with specific preferences we are able to derive and test hypotheses according to the F&S model. Furthermore, we control for the information subjects receive about the type of their co-player.

Our approach is based on the method introduced by Blanco et al. (2006) but differs from their approach in some but important aspects. First, in the MUG used by Blanco et al. each subject reacts to a specific proposal of his or her co-player, i.e. there is a distinct element of strategic interaction in this game. In our MUG, however, there is no direct interaction between both subjects and no room for strategic considerations. Second, before playing the public good games subjects were sorted into pairs according to their F&S parameters. Third, we control

---

[1] This approach has already been mentioned by Fehr and Schmidt (1999), p. 847.

for the information subjects receive in the games following the classification. Fourth, we play a repeated public good game while Blanco et al. use the one-shot version of this game. Two recent papers (Gächter and Thöni, 2005; Gunnthorsdottir et al., 2007) study the consequences of group composition of different types of players for the outcome of experimental public good games. Both papers essentially use a player’s former contribution in a public good game to measure her individual 'cooperative disposition'. In a second step, they use this measure of cooperative disposition to form groups composed of different types of players (more or less cooperative ones) and have another (repeated) public good game played in these groups. The common result is the finding that in groups of cooperators the familiar decay of contributions over time is less severe than in other groups. Our method differs from the above. While it is our aim, too, to check whether certain individual behavioral characteristics and a respective composition of groups matters for individual behavior in public good games, our way of classifying individuals is different. In order to classify individuals, we use the well established F&S preferences which we uncover with the help of an ultimatum and a dictator game as mentioned above.

The most important result of our study is that the specific composition of groups significantly influences the subjects' performance in the final period of the public good games: As long as subjects are informed about the type of their co-player, “fair” groups contribute more to the public good than “egoistic” or “mixed” groups. In addition, it turns out that explicit information is a key factor for this difference in behavior: If “fair” subjects are not informed about the fact that their co-player is “fair”, too, they act like “egoistic” subjects. Only the explicit information that they are playing with a “fair” co-player significantly enhances their contributions.

The remainder of the paper is organized as follows. Section 2 sets the stage by describing the F&S model which underlies our experiment and the games we used. Section 3 describes the design of our experiment including treatments and hypotheses. Section 4 presents the experimental results. Section 5 discusses our results and concludes.

2 Theoretical background: The model of Fehr and Schmidt (1999)

2.1 Preferences

According to Fehr and Schmidt (1999) individuals are not exclusively motivated by the absolute payoff they can earn but also by value allocations due to their distributional consequences. Particularly, assuming that individuals suffer from unequal distributions, F&S introduce the following utility function for subject i:

$$U_i\left(\pi_i, \pi_j\right) = \pi_i - \alpha_i \frac{1}{n-1} \sum_{j \neq i}^{n} \max\left\{\pi_j - \pi_i, 0\right\} - \beta_i \frac{1}{n-1} \sum_{j \neq i}^{n} \max\left\{\pi_i - \pi_j, 0\right\} \tag{1}$$

where π_i and π_j denote the absolute payoffs of subjects i and j, respectively, n denotes the total number of players involved in a decision problem, $\alpha_i \geq 0$ measures the impact of disadvantageous inequality on i's utility while $\beta_i \geq 0$ measures the corresponding impact of advantageous inequality. In the two-player case[2] which is particularly relevant for our experimental setting, (1) reduces to

$$U_i\left(\pi_i, \pi_j\right) = \pi_i - \alpha_i \max\left\{\pi_j - \pi_i, 0\right\} - \beta_i \max\left\{\pi_i - \pi_j, 0\right\}. \tag{2}$$

F&S assume $\beta_i < 1$, i.e. players are not willing to "burn“ their money to eliminate advantageous inequality. In addition, they assume that players put a stronger weight on disadvantageous inequality, i.e. $\alpha_i \geq \beta_i$.[3] In our experiment, we will obtain the weights α_i and β_i from the MUG and MDG (see section 3.1).

---

[2] In the following, all conditions are stated for the case of two players. The generalization to the n-player case is straightforward and can be found in Fehr and Schmidt (1999).

[3] This condition is employed by Fehr and Schmidt (1999) in their proof of Proposition 4 (part C, p. 862) in order to facilitate the critical condition for cooperation in a voluntary contribution game (VCG). Proposition 4 states that a fair player chooses to cooperate in a VCG if the following condition is met: $k/(n-1) \leq (a + \beta_j - 1)/(\alpha_j + \beta_j)$. If $\alpha_i \geq \beta_i$ then this is the sole condition that has to be fulfilled. If one abandons $\alpha_i \geq \beta_i$, then a second condition might become binding, namely $k/(n-1) \leq a/2$, where a denotes the marginal per capita return of the public investment.

2.2 Voluntary contribution games

2.2.1 The standard voluntary contribution game

The assumption of inequity aversion has a strong impact on the theoretical predictions of the outcomes in several classes of games. In a public good (PG) game for example, preferences of the F&S type may lead to much higher cooperation rates compared to the predictions derived by standard economic theory. To see this, look at the following voluntary contribution game (VCG). Both players $i = 1,2$ are given some initial endowment y which can be devoted to the production of some public good. Player i's contribution to the public good is denoted by g_i, the production function for the public good is simply given by the sum over all contributions, i.e. $\sum_{i=1}^{2} g_i$. Let us assume that the marginal per capita return of an investment in the public project is given as some constant a with $1/2 < a < 1$. Then the monetary payoff for player i is given by $\pi_i(g_1,...,g_n) = y - g_i + a\sum_{j=1}^{2} g_j$. Thus, for player i it is a dominant strategy to choose $g_i = 0$. Since this holds for both players identically, the unique equilibrium of this game is characterized by contributions $g_j = 0\ \forall j$ and the public good will not be provided at all. However, the provision would be beneficial since the collective marginal return is *2a* which is clearly above the marginal costs of provision. Hence, the social optimum is achieved if both players contribute their entire initial endowment to the public good leading to payoff $\pi_i^{SO} = 2ay$, which is above the payoff players receive in the Nash equilibrium ($\pi_i^{NE} = y$).

F&S show that this result is fundamentally altered if players are endowed with inequality aversion according to (1). They prove the following results:

i) If $a + \beta_i < 1$, then it is a dominant strategy for player i to choose $g_i = 0$.

ii) Let $a + \beta_1 < 1$, but $a + \beta_2 \geq 1$. Although player 2 is relatively strongly averse to inequality, in the unique equilibrium both players choose not to contribute, i.e. $g_i = 0, i = 1,2$.

iii) If, however, $a + \beta_i > 1$ holds for both players, equilibria with positive contributions to the public good exist, i.e. both players choose contribution levels $g_i = g \in [0, y]$.

The intuition behind these results is not too difficult. First, if a player with $a + \beta_i < 1$ invests one monetary unit in the public good his monetary return is a while he gains a maximum non-monetary utility of β_i. Now, if the sum of both returns is less than one it is obviously the best strategy not to invest into the public good, irrespectively of what the other player does. Second, if the other player obeys to $a+\beta_j<1$, player i will not be willing to contribute even if he shows stronger inequality aversion, i.e. for him $a+\beta_i>1$ holds. The reason is that the "fair" player would suffer too much by the disadvantageous inequality caused by the free rider. Only if both players are sufficiently "fair-minded" they are able to sustain a cooperative outcome in the VCG.

2.2.2 The voluntary contribution game with punishment

The idea that punishment of defecting players may increase contribution rates to the public good is straightforward. In a setting with standard preferences, however, punishment is a non credible threat. Imagine a two-stage game: Stage one is the voluntary contribution game as described in the section above. Stage two of the game incorporates the possibility for players to enact some punishment on their opponents. A player can punish his opponent by lowering the opponent's payoff. In order to reduce an opponent's payoff by one unit, the punisher must incur costs of *c*. Since punishment is costly it will not be carried out by rational players interested only in their absolute material payoff on the second stage. Since players anticipate the outcome on the second stage they will not contribute in the first stage of the game.

This outcome is substantially altered if preferences of the F&S type are involved. The existence of at least one individual called a "conditionally cooperative enforcer" may enhance the prospects for cooperation. This individual must show sufficiently strong aversion against advantageous inequality, i.e. its preferences must obey $\beta_i \geq 1-a$. Two possibilities have to be distinguished. i) Only one player obeys $\beta_i \geq 1-a$. If, in addition, this player is also sufficiently averse of disadvantageous inequality, i.e. if $\alpha_i > \frac{c}{1-c}$, then subgame perfect equilibria exist where both players choose positive contribution levels $g_i = g \in [0,...,y]$. The reason is simple: If the other player does not contribute, the enforcer will carry out some punishment $p_{ij} = (g-g_j)/(1-c)$ in the second stage of the game. This threat is credible, because the punishment reduces the disutility the enforcer derives from disadvantageous inequality. In addition, the condition $\beta_i \geq 1-a$ guarantees that the enforcer will prefer to cooperate on the first stage of this game due to his relatively high degree of aversion to advantageous inequality. ii) The same outcome, however, can be achieved, if both players obey $\alpha_i > \frac{c}{1-c}$, irrespectively of their degree of aversion against advantageous inequality. In this case, both players relatively strongly dislike disadvantageous inequality and both players simply police each other.

2.2.3 Introducing uncertainty

The analysis of F&S is based upon the assumption that players know their opponents' type. For this reason, in most of our experimental treatments, we informed the participants previous to the public good games on how their opponent had behaved in the MUG and MDG played before (see section 3.1). Thus, these subjects were principally able to derive the corresponding type of their co-player. In one treatment, however, we did not inform the subjects about their opponent. These subjects were only able to predict their opponent's type with some probability.[4] Allowing for uncertainty, in the two-player case a "fair" player with preference $\beta_j > 1-a$ will choose to contribute to the provision of the public good if the following condition is met:[5]

$$E(k) < \frac{a+\beta_j-1}{\alpha_j+\beta_j} \tag{5}$$

where $E(k)$ denotes the expected value of k, the number of players with $\beta_j \leq 1-a$. Obviously, for $n=2$, k can only take on the values one or zero, i.e. $E(k)$ is the probability a "fair" player attaches to the possibility that his co-player is an "egoistic" type. Note that as a consequence of introducing uncertainty the parameter α_i matters. This is different in the case of perfect knowledge, which can be easily seen by setting $E(k)=0$. Then, (5) reduces to $\beta_j > 1-a$.

---

[4] In the following we assume risk neutral behavior.

[5] This condition is derived from the one for cooperation in the n-player case. F&S show that in this case a player with preferences $\beta_j > 1-a$ only decides to cooperate in the VCG if $\frac{k}{n-1} < \frac{a+\beta_j+1}{\alpha_j+\beta_j}$. Note additionally that, if one skips F&S's assumption $\alpha_i \geq \beta_i$, which might be appropriate for our subject pool (see section 4.1), then the additional condition $E(k) \leq a/2$ must hold. See footnote 2.

However, if $E(k) > 0$ with increasing values of α_j it becomes more difficult to fulfil (5), i.e. to ensure that the condition is still met, players must have higher values of β_j.[6] The intuition behind this is as follows. On the one hand, if a "fair" but uninformed subject contributes to the public good he runs the risk of having an "egoistic" opponent and, therefore, being exploited. On the other hand, if he does not contribute, he runs the risk of having a "fair" opponent and possibly exploiting her. In other words, positive contributions to the public good would increase the risk of disadvantageous inequity and decrease the risk of advantageous inequity. Hence, the subject is only willing to contribute, if his aversion against disadvantageous inequity is sufficiently low and his aversion against advantageous inequity is sufficiently large.

Condition (5) can also be used to illustrate the effect of information. If "fair" players are informed about the behavior of their opponent in the former MUG and MDG this increases their confidence about their opponent's type. Particularly, in this case they should expect their opponent to be a "fair" type with higher probability if he or she behaved accordingly in the former games. Technically, this means because $E(k)$ decreases, condition (5) is more easily met for informed subjects. Hence, we should observe a higher level of cooperation in groups consisting of informed "fair" players than in groups consisting of uniformed "fair" players.

3 Experimental design

3.1 Games

We used four different games (games A, B, C, and D) in our experimental design. Thereby, the purpose of games A and B was to measure each subject's preferences according to the F&S model. After these games, some of the subjects with certain preferences played games C and D in various treatments (see section 3.2). The design of the games is presented in the following.[7]

Game A is designed to measure the subjects' aversion against disadvantageous inequity. The game resembles the responder's basic decision situation in the ultimatum game but abstracts

---

[6] This can be easily seen when (5) is solved for β_j. In this case, condition (5) reads $\beta_j > E(k)\alpha_j / (1 - E(k)) + (1 - a)/(1 - E(k))$.

[7] See Dannenberg et al. (2007) for the instructions that we distributed to our participants.

from strategic interaction, so that we can rule out individual behavior caused by strategic considerations such as intentions or reciprocity.[8] In this game, each subject has to decide in 22 cases (numbered from #1 to #22) in the role of player 1 between two pairs of payoffs (pair I and pair II), each with an amount of money for himself or herself and another subject in the role of player 2. Payoffs (see the left part of Table 1) are chosen in a way that – except for #1 – subjects always have to choose between "pair I", a disadvantageously unequal division of €10.00, and "pair II", an equal distribution with €2.00 for both players. In this game, a purely selfish subject should choose pair I from #1 to #20 and pair II for #21 and #22.[9] A subject strongly disliking disadvantageous inequity, in contrast, would choose pair I in #1 and pair II from #2 to #22. Subjects with other-regarding preferences according to F&S between these two extremes would be expected to switch from choosing pair I to pair II after #2 but prior to #21.

We call individual behavior in game A consistent if (1.) a subject has a unique switching point from pair I to pair II and (2.) the switching point is between #2 and #21. Regarding the first condition, a subject with aversion against disadvantageous inequity consistent with the F&S model who switches for a specific case from pair I to pair II should choose pair II also for all subsequent cases. As the payoffs for player 1 in pair I are arranged in descending order, a switchback to pair I in any of the subsequent decisions is not consistent. This would lead to a lower own payment and to higher disadvantageous inequity than in the case that was rejected before. In relation to the second condition, it is useful to consider the decision cases outside the "consistent range" between #2 and #21. A subject who chooses pair II already in #1 is not regarded as consistent because he or she could attain an equal allocation with higher own payoff by choosing pair I. A subject who chooses pair II in #22 only or never switches to pair II at all has a negative value for the weight of aversion against disadvantageous inequity ($\alpha_i < 0$), i.e. likes disadvantageous inequity and is therefore not consistent with the F&S model. With the subject's switching point we can determine the upper and lower bounds of the individual α_i. We approximate the individual value for α_i by choosing the mean of the corresponding interval (see Table 1).[10]

---

[8] The difference to the payoffs of the original ultimatum game is the fact that the conflict point payoffs (in €) are changed to (2, 2) instead of the original (0, 0). We changed the payoffs in order to avoid zero payoffs and a border solution for the "selfish" switching point from pair I to pair II.

[9] In the following we assume rational behavior of all subjects.

[10] There are two exceptions to this rule. First, we cannot determine an upper bound for α_i of a subject who

Game B – which basically resembles the player's decision problem in the dictator game – is designed to measure the subjects' aversion against advantageous inequity.[11] Again, each subject had to decide in 22 cases (from #1 to #22) between two pairs of payoffs (pair I and pair II) each with an amount of money for himself or herself in the role of player 1 and another subject in the role of player 2 (see the right part of Table 1). Payoffs are chosen in a way that subjects had to choose between "pair I", an extremely unequal but advantageous distribution of €10.00, and "pair II", an equal distribution of different amounts from €0.00 to €21.00. In this game, a purely selfish subject would choose pair I from #1 through #20 and pair II for #22. In the case of #21, this subject would be indifferent between pair I and pair II. A subject strongly disliking advantageous inequity would always choose pair II. Subjects with "fairness preferences" according to F&S would be expected to switch from choosing pair I to pair II after #1 but before #21.

---

switches from pair I to pair II in #2. Therefore, we assign to those subjects the value of the lower bound, $\alpha_i = 2.18$. Second, we assign the value $\alpha_i = 0$ to a subject who switches from pair I to pair II in #21, although the corresponding interval for this case is $-0.08 \le \alpha_i \le 0.04$.

[11] Strictly speaking, game B is equivalent to the dictator game only for decision #11. However, similar to the dictator game, game B creates a trade-off between own monetary payoff which creates advantageous inequity and a lower but equally distributed payoff.

Table 1: Payoffs in game A and game B

		game A					game B				
		pair I		pair II			pair I		pair II		
		payoffs (in €) for player					payoffs (in €) for player				
	#	1	2	1	2	α_i	1	2	1	2	β_i
switching point from pair I to pair II	1	5.00	5.00	2.00	2.00	-	10.00	0.00	0.00	0.00	1.00
	2	4.44	5.56	2.00	2.00	2.18	10.00	0.00	0.50	0.50	0.98
	3	4.42	5.58	2.00	2.00	2.13	10.00	0.00	1.00	1.00	0.93
	4	4.39	5.61	2.00	2.00	2.02	10.00	0.00	1.50	1.50	0.88
	5	4.36	5.64	2.00	2.00	1.90	10.00	0.00	2.00	2.00	0.83
	6	4.32	5.68	2.00	2.00	1.77	10.00	0.00	2.50	2.50	0.78
	7	4.29	5.71	2.00	2.00	1.66	10.00	0.00	3.00	3.00	0.73
	8	4.24	5.76	2.00	2.00	1.54	10.00	0.00	3.50	3.50	0.68
	9	4.19	5.81	2.00	2.00	1.41	10.00	0.00	4.00	4.00	0.63
	10	4.14	5.86	2.00	2.00	1.30	10.00	0.00	4.50	4.50	0.58
	11	4.07	5.93	2.00	2.00	1.18	10.00	0.00	5.00	5.00	0.53
	12	3.92	6.08	2.00	2.00	1.00	10.00	0.00	5.50	5.50	0.48
	13	3.86	6.14	2.00	2.00	0.85	10.00	0.00	6.00	6.00	0.43
	14	3.81	6.19	2.00	2.00	0.79	10.00	0.00	6.50	6.50	0.38
	15	3.68	6.32	2.00	2.00	0.70	10.00	0.00	7.00	7.00	0.33
	16	3.53	6.47	2.00	2.00	0.58	10.00	0.00	7.50	7.50	0.28
	17	3.33	6.67	2.00	2.00	0.46	10.00	0.00	8.00	8.00	0.23
	18	2.85	7.15	2.00	2.00	0.30	10.00	0.00	8.50	8.50	0.18
	19	2.72	7.28	2.00	2.00	0.18	10.00	0.00	9.00	9.00	0.13
	20	2.22	7.78	2.00	2.00	0.10	10.00	0.00	9.50	9.50	0.08
	21	1.43	8.57	2.00	2.00	0.00	10.00	0.00	10.00	10.00	0.03
	22	0.10	9.90	2.00	2.00	-0.14	10.00	0.00	10.50	10.50	0.00

We call individual behavior in game B consistent if (1.) a subject has a unique switching point from pair I to pair II and (2.) this switching point is between #2 and #22., i.e. if the individual weight of aversion against advantageous inequity meets $0 \leq \beta_i < 1$. Relating to the first condition, a subject with aversion against advantageous inequity consistent with the F&S model switching from pair I to pair II at one point should also choose pair II in all cases after the switching point. As the payoffs for player 1 in pair II are arranged in an ascending order, a switchback to pair I in any of the subsequent cases is not consistent. This would lead to the same advantageous inequity than was rejected before but now with higher opportunity costs in terms of equal payoffs for both players. For the second condition, we consider again the decision cases outside the "consistent area" between #2 and #22. A subject choosing pair II already in #1 has $\beta_i \geq 1$, i.e. is willing to "burn" money in order to produce equal payoffs. A subject who does not switch at all displays affection for advantageous inequity. Both behavioral patterns are not consistent with F&S. Similar to game A, we can determine the upper and lower bounds for the individual's β_i by means of a subject's switching point. We approximate the individual value of β_i by choosing the mean of the corresponding interval (see Table 1).[12]

Fehr and Schmidt, 2006 recommend using “strategic games” in order to elicit preference parameters that capture not only traits of inequity aversion but, moreover, strategic considerations like intentions or reciprocity.[13] In a way, this recommendation by Fehr and Schmidt makes their model broader in scope than we regard useful for our aim because they include preferences for equity and strategic considerations. As our main focus is on the effect of inequity aversion, i.e. “pure” preferences net of other behavioral influences, we do not follow this recommendation. Our games are explicitly of non-strategic nature.[14]

Game C is a standard two-player PG game with a voluntary contribution mechanism. Two players get a fixed balance of €3.00 for show-up and are endowed with €10.00 in each round of the game. They decide simultaneously how much (if any) money from the endowment to contribute to a public good. Each monetary unit that the subject keeps for himself or herself raises the individual payoff by exactly that amount. Both subjects receive €0.70 for each €1.00 contributed to the public good, i.e. the marginal per capita return is constant and equal to 0.7. The game was played using a partner design over 10 rounds with the number of rounds as common knowledge. After each round subjects were informed about their own contribution and the contribution of the co-player as well as the payoffs of both players.

Game D consists of two stages. Stage 1 is equivalent to game C, i.e. subjects play the same two-player PG game as described above. Stage 1 in game D is followed by a stage 2. In this stage subjects have the possibility to assign his or her co-player negative points, i.e. a punishment mechanism (Fehr and Gächter, 2000, 2002) is introduced. Each negative point reduces the payoff of the co-player by €1.00. However, the assignment of negative points is costly. Each negative point assigned reduces the punisher’s own payoff by either €0.17 or

---

[12] As before, there are two exceptions to this rule. We assign the value $\beta_i = 0$ to a subject who switches from pair I to pair II in #22, although the corresponding interval in this case is $-0.05 \le \beta_i \le 0$. For subjects switching in #1, $\beta_i = 1$ is chosen.

[13] Following Fehr and Schmidt (2006) we regard a game as “strategic” if each player has an influence on each other player’s material payoff.

[14] Even if we would have followed the recommendation to use strategic games to elicit preferences, this would probably just enhance the effects observed in our experiments.

€0.50 (depending on the treatment). Again, the game was played using a partner design over 10 rounds with the number of rounds as common knowledge.[15]

Subjects were paid separately for games A and B and games C and D. The payments from games A and B were computed as follows: All subjects within a session were randomly matched into pairs of subjects. After this, it was determined (again by chance) whether game A or B would be relevant for the payoffs. After the selection of the relevant game, a random draw selected which number of the payoff list (between #1 and #22) would be relevant. Finally, a random draw decided which person within a pair determined the payoffs, i.e. whose decision as player 1 is to be realized. According to this rule, each of the 22 decision cases in game A and B had the same chance to be relevant for the payment. Subjects were informed about this payoff rule before the start of the experiment and we tried to make sure that everybody clearly understood that there is no sense in doing something like maximizing expected payoffs or efficiency. Moreover, we checked the comprehension of this payoff rule in a detailed quiz before the experiment started. A part of those subjects who had behaved consistently in games A and B (which had been played first) were invited to play games C and D. Subjects playing games A and B were not informed about the fact that further games would follow.[16] The payments from games C and D were determined in a similar way: After both games, a random draw determined which game (C or D) would be relevant. Following this decision, one of the 10 periods was selected randomly and the payments were realized according to the players' decisions in this round. As before, the payoff rule was common knowledge to all participants.

---

[15] Theoretically, due to possible punishment and costs of punishment in game D, it is possible to encounter cases with negative payoffs. We were prepared to handle these cases (payments would have been set to zero). Luckily, though, a case like this did not occur during the experiment.

[16] In order to avoid strategic behavior subjects were not informed about the games C and D in the beginning of the experiment. We believe that this is the only way to test theories which make type-specific predictions. Several authors, see e.g. Ockenfels and Weimann (2002) and Ben-Ner et al. (2004), choose a similar way when they match subjects according to the behavior in previous games without informing subjects in advance about this procedure. Because some "types" of subjects were observed more frequently than others, we had to select appropriate subjects in order to ensure a certain number of observations per treatment. Furthermore, we had to ensure equal distributions of β_i in treatments FAIR and FAIR(ni) (see notes in Table 3). Subjects were selected only by their individual F&S parameters after game B. Those subjects who were not invited to take part in games C and D played a one-shot prisoner's dilemma game. The only purpose of this game was to avoid a zero payment for subjects who didn't play games C and D (there was no show-up fee).

3.2 Subject pool and treatments

We ran 25 sessions with 18 to 20 subjects in each session. All in all, 492 subjects participated in games A and B and 160 of these subjects were invited to play games C and D. Sessions lasted about 90 minutes and the average earning was €5.90 for games A and B and €15.20 for games C and D. The sessions were conducted in November 2006 and May 2007 at the Magdeburg Experimental Laboratory (MaXLab). The experiment was fully computerized and anonymous.[17] In the laboratory, subjects were randomly allocated to separate cabins and had no mutual contact during or after the experiment. The main characteristics of our subject pool are displayed in Table 2. Though the majority of our subjects are students of economics, the fraction of non-economists (41%) is quite high. About two thirds of our subjects had already gained experience in experiments, i.e. had participated in at least one experimental session before.

In order to check the robustness of our design, we implemented two modifications during the experiment. In two of our sessions, we modified the order of play. While in most sessions, subjects first played game A followed by game B, in two sessions (for 40 of 492 subjects), we had subjects playing game B first, then A. Moreover, in seven of our sessions (for 136 of 492 subjects), payoffs in game A were slightly different: The payoff in pair II of game A was €0.00 for both subjects instead of €2.00. Both changes do not lead to significantly different distributions of the values for α_i and β_i at the 5% level (K-S test, two-tailed).[18] Therefore, we felt free to pool the data of all 25 sessions of games A and B conducted.

Table 2: Subject pool – descriptive statistics

	absolute frequency	relative frequency in percent
total	492	100.0
consistent choices	371	75.4
field of study		
economics	291	59.1
non economics	201	40.9
experienced (at least one experiment before)	339	68.9
male subjects	270	54.9
female subjects	222	45.1

---

[17] We used Z-tree for programming. See Fischbacher (2007).

[18] Unless it is explicitly noted, in the following all tests are two-tailed.

In order to analyze the predictive power of the F&S model as well as some comparative static effects we defined several treatments for games C and D, which are described below.[19] Treatments differ with respect to (1.) the preference parameter β_i of the subjects that form a pair in games C and D, (2.) the state of information about the co-player's behavior in games A and B, and (3.) the value of parameter c, which defines the marginal costs of punishment in game D.

Our four treatments in game C differ with respect to the composition of the (two-person) groups playing the game. Details are displayed in Table 3. Thereby, a group of two players (i = 1, 2) makes one statistically independent observation. In treatment EGO, two subjects with $\beta_i < 0.3$ are matched together to form a common group. Treatments MIX, FAIR, and FAIR(ni) are defined correspondingly. In all treatments, except for FAIR(ni), subjects were informed about their co-player's former behavior in games A and B. In FAIR(ni), players did not receive any information about each other's former behavior.

Table 3: Treatments in game C

Treatment	Parameter β_i, i = 1, 2	Information	Observations
EGO	$\beta_i < 0.3$	yes	35
MIX	$\beta_1 < 0.3 \wedge \beta_2 > 0.3$	yes	13
FAIR	$\beta_i > 0.3$	yes	17
FAIR(ni)	$\beta_i > 0.3$	no	15
Σ			80

Notes: (1.) All subjects in FAIR and FAIR(ni) have $\alpha_i = 0$. (2.) No one in our subject pool has $\beta_i = 0.3$. (3.) There are no differences for β_i between FAIR and FAIR(ni) at the 5% level (K-S test for differences in the distributions and MW U test for differences in median values).

All subjects who played game C also completed game D, the PG game with punishment possibility. The different treatments in game D are displayed in Table 4. Within the EGO treatment, we distinguished between high and low costs of punishment. Therefore, pairs of subjects from the EGO treatment in game C were allocated into two separate EGO treatments in game D, namely EGO(h) and EGO(l). As in game C, in most of the treatments, subjects were informed about their co-player's behavior in games A and B, again with the exception of FAIR(h, ni).

---

[19] The definition of treatments is determined by the structure of the distribution of α_i and β_i within our subject pool. Due to the fact that there is virtually no dispersion for α_i (see section 4.1) we had to focus the definition of treatments on β_i.

Table 4: Treatments in game D

Treatment	Parameter β_i, i = 1, 2	Information	Costs of Punishment	Observations
EGO(l)	$\beta_i < 0.3$	yes	low	9
EGO(h)	$\beta_i < 0.3$	yes	high	26
MIX(h)	$\beta_1 < 0.3 \wedge \beta_2 > 0.3$	yes	high	10
FAIR(h)	$\beta_i > 0.3$	yes	high	17
FAIR(h, ni)	$\beta_i > 0.3$	no	high	15
Σ				77

Notes (see also Table 3): (1.) As we have only three independent observations in the MIX(l) treatment it is omitted from hypotheses testing in game D. (2.) None of the subjects with $\beta_i > 0.3$ in MIX(h) fulfills condition (4).

3.3 Hypotheses

Based on the definition of treatments and the methodology developed in section 3, we are able to derive specific hypotheses which follow from the F&S model for our subject pool. Thereby we focus – at least for the non-parametric tests in section 4 – on the last period of both games C and D. In doing so we can exclude any repeated game effects, which may come into play if there is a repeated interaction between only two subjects. In the following we assume that, whenever F&S predict the existence of multiple equilibria, subjects will prefer the Pareto dominant equilibrium, i.e. the one with positive monetary payoff.

As there is no difference in the prognosis regarding the contribution behavior between games C and D we are able to formulate joint hypotheses for both games. Note that there is no treatment in game D where punishment is part of an equilibrium strategy: In MIX(h) none of the subjects with $\beta_i > 0.3$ fulfills condition (4), i.e. there are no subjects who may enforce cooperation. In FAIR(h) cooperation without punishment is an equilibrium because – in the case of mutual cooperation – there are no incentives for subjects with $\beta_i > 0.3$ to punish each other. In this case, punishment would reduce the own payoff and create advantageous inequality.

We can derive the following hypotheses regarding the contributions to the public good in the different treatments for games C and D:

H1: In the EGO and MIX treatments of both games, zero contributions of all subjects should be observed.

H2: In the FAIR treatment with information, positive contributions to the public good should be observed more frequently than in the corresponding EGO and MIX treatments.

H3: In the FAIR treatment with information, positive contributions to the public good should be observed more frequently than in the corresponding FAIR(ni) treatment without information.

These hypotheses regard behavior in the last round of the VCG. Assuming perfect knowledge of the opponent's type the analysis can easily be extended to the whole finitely repeated VCG. Consider, for example, the EGO treatment. By the familiar argument of backward induction, not to contribute is then the perfect equilibrium of the finitely repeated VCG. By the same token, non-cooperation is the sole perfect equilibrium in the MIX treatment and two fair players in the FAIR treatment will always choose to cooperate in all rounds of the VCG.

Note, however, that these results may change if one abandons the assumption of perfect type knowledge. Consider the EGO and MIX treatment. If an egoistic type attaches a possibly small, but positive probability to the possibility that his opponent is a fair type, then it might become rational even for the egoist to cooperate for some period of the game. [20] The reason is simple: It pays to build a reputation as a fair player and to defect for the first time at a rather late stage of the game, because this allows for harvesting the fruits of cooperation until the first time of defection. Hence, if we assume that type knowledge is imperfect we should expect subjects to cooperate to a certain extent also in the EGO and MIX treatments. In the last round, however, we should observe a significant last round effect in the EGO and MIX treatment, because then there is no more rationale for reputation building.

4. Results

The results section consists of two main parts. The first part analyzes the subjects' behavior in games A and B. The second part focuses on games C and D and refers to the treatments and hypotheses described above.

4.1 Behavior in games A and B

As discussed in the previous section, we are able to select subjects with consistent preferences which are in line with the assumptions $\alpha_i \geq 0$ and $0 \leq \beta_i < 1$ in the F&S model. All in all, out of 492 subjects who participated in the experiment 371 (75%) behaved consistently in games

---

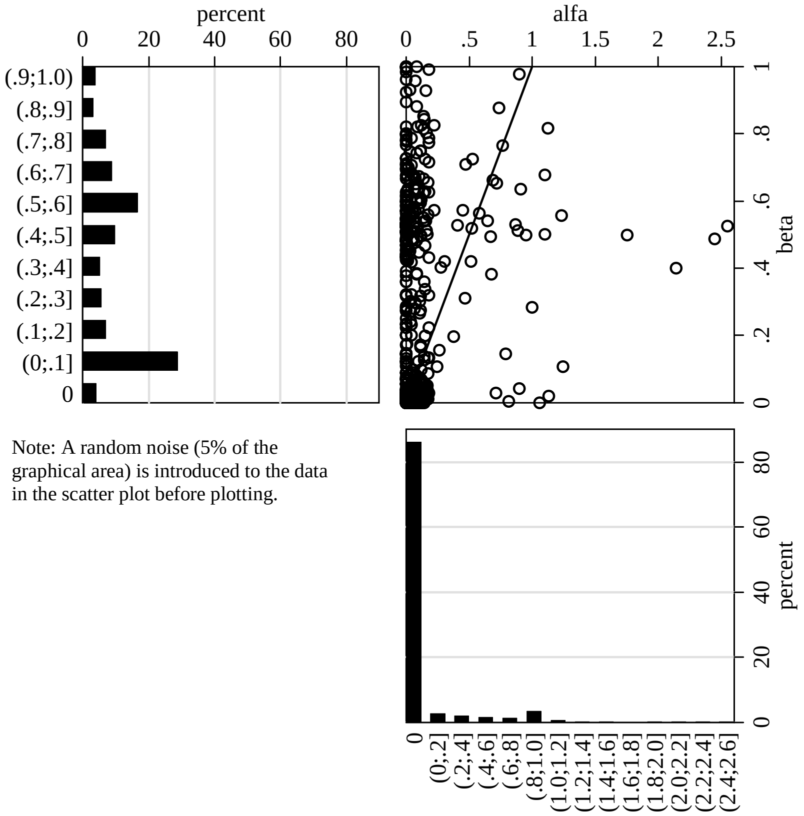
[20] This has been shown in the famous paper by Kreps et al. (1982).

A and B. Thereby, we do not detect any significant correlations between the socio-economic characteristics in Table 2 and the consistency of choices at the 5% level (Spearman's ρ).

Figure 1 presents the distribution of the F&S parameters, α_i and β_i. A brief look at the distribution of the values for α_i, the weight of the aversion against disadvantageous inequity, shows that about 86% of all subjects behave selfishly in game A (lower right in Figure 1). There is a second peak in the range of $0.8 < \alpha_i \leq 1.0$. These are subjects who prefer in #12 of game A the payoff in pair II (€2.00 for both) to the payoff in pair I (€3.92 for themselves and €6.08 for the other subject). The mean value for α_i is 0.102. The median value is 0 indicating that selfish behavior is the dominant pattern in our subject pool.

The distribution of the values for β_i, the weight of the aversion against advantageous inequity, looks quite different (upper left in Figure 1). We observe two peaks in the distribution. First, for $0 \leq \beta_i \leq 0.1$, i.e. for subjects who behave rather selfishly and switch from pair I to pair II in #21 or #22 of game B. Second, there is a peak for the range $0.4 < \beta_i \leq 0.6$, i.e. for subjects who switch from pair I to pair II between #10 and #13 and who prefer a rather equal allocation of the €10 vis-à-vis an advantageous but very unequal allocation. The mean of the β_i values is 0.356, the median is equal to 0.375. The scatter plot in Figure 1 (upper right) shows the joint distribution of the two parameters. We found very few subjects (45 of 371, i.e. 12% of consistent choices) meeting the F&S condition $\alpha_i \geq \beta_i$. The corresponding data points lie below the line in the scatter plot. It is apparent that α_i and β_i are not significantly correlated and a test for correlation confirms this (Spearman's $\rho = 0.345$, p = 0.299).

Figure 1: Distribution of F&S parameters for consistent choices



Note: A random noise (5% of the graphical area) is introduced to the data in the scatter plot before plotting.

Kritikos and Bolle (2001) study games where the decision maker faces choices similar to ultimatum-game second movers, but without any preceding choice by the "proposer", so just as in the MUG to measure α_i here. They find very little behavior corresponding to rejections. Similar results have been found, among others, by Charness and Rabin (2002). Furthermore, Blanco et al. (2006) who use the MUG in a strategic setting observe significantly higher α_i values (see Dannenberg et al. 2007). Therefore, it is not very surprising that the variation in the α_i data in our non-strategic MUG is very low.

Given the quite heterogeneous subject pool in our experiment, we are able to test whether there are any correlations between the socio-economic characteristics in Table 2 and the individual values for α_i and β_i. We observe a small but significant negative correlation between subjects who study Management Science, Economics or a related field (e.g. "Business Information Systems") and the value for β_i (Spearman's $\rho = -0.154$, p = 0.003). This observation is in line with the well-known fact that economics students in general behave more selfishly than other students (e.g. Marwell and Ames, 1981; Frank et al. 1996; Carter and Irons, 1991). However, this is the only significant correlation between α_i or β_i and the socio-economic characteristics at the 5% level.

4.2 Behavior in games C and D

The results section for games C and D consists of five parts. The first and second part analyze the subjects' contributions in the standard PG game (game C) and the PG game with punishment possibility (game D). Here, we are able to test whether the prognoses from the F&S model apply to our subject pool. The third part investigates the effect of information and the fourth part the effect of punishment. Finally, we present Tobit estimates for the contribution to the public good in game C and the punishment behavior in game D.

4.2.1 Hypotheses testing of the F&S model in game C

A total of 160 subjects played the PG game C. The mean contribution per period over all subjects and all periods is €5.90. In order to test our hypotheses derived from the F&S model, we analyze the behavior of subjects in the three treatments with information, namely EGO, MIX, and FAIR (see Table 3). These treatments differ only in their composition of individual types, namely "egoistic", "fair" or both types. The attributes "egoistic" and "fair" arise only from the value of β_i, as in the absence of uncertainty about the co-player's type, the subjects' contributions in the standard PG game solely depend on β_i.[21]

The prognosis of the F&S model is the same for EGO and MIX. Contributions of zero are the only equilibrium. This is the same allocation that standard economic theory predicts for both games. In contrast, according to F&S in the FAIR treatment the "standard" non-contribution equilibrium exists as well as equilibria with positive contributions. Figure 2 shows the mean contributions in game C for the treatments FAIR, EGO, and MIX. The mean contribution per period is €7.10 for FAIR, €6.00 for EGO, and €4.30 for MIX. A Mann-Whitney U test (MW U test) shows that the differences between treatments are not significant at the 5% level except for those between FAIR and MIX ($p = 0.013$). In each period of all treatments the share of subjects contributing a positive amount of money to the public good is significantly

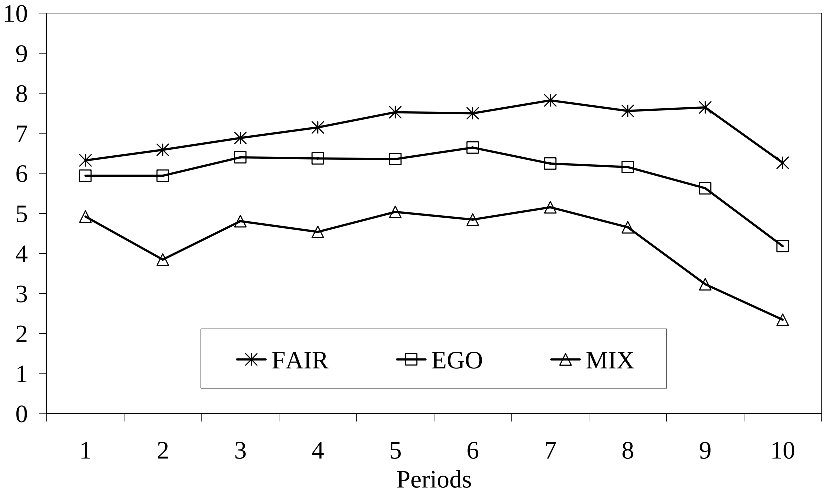
---

[21] If we control for α_i and apply the non-parametric tests to all subjects with $\alpha_i = 0$ (86% of our subjects in games C and D) we find the same results as in the tests including all subjects. Seeing this, α_i seems indeed to have no significant effect on subjects' contribution behavior. We will return to this point later.

greater than zero (Binomial Sign test, one-tailed, p = 0.000). Therefore, we can sum up these observations into the following result.

> *Result 1: The contributions of the subjects in the EGO and the MIX treatment to the public good are significantly higher than the levels predicted by F&S and by the standard model of pure selfishness. The contributions in FAIR are in line with F&S but not with the standard model.*

Figure 2: Contributions in game C



As already explained above (section 3.3), our analysis focuses particularly on the behavior in the final period. The mean contribution in the final period is €6.30 in FAIR, €4.20 in EGO, and €2.40 in MIX. Thereby, the differences in final period contributions are significant between MIX and FAIR (MW U test, p = 0.007) and weakly significant between EGO and FAIR (p = 0.090). There are, in contrast, no significant differences between EGO and MIX. In consideration of our hypothesis *H2*, this is what we expected. According to this, the F&S model seems to have some explanatory power for the behavior in the final period of the standard PG game. In order to check whether this result is robust, we also consider the share of subjects who "defect" in the final period by contributing nothing to the public good. 54% of the subjects in EGO, 58% of the MIX subjects but only 32% of the FAIR subjects defect in the final period. Thus, the shares of defecting subjects are relatively high in EGO and MIX compared to the FAIR treatment. In order to test for differences in shares of defecting subjects between the treatments we apply a χ^2 goodness-of-fit test that defines defection when the mean contribution of both players in a pair is either (i) €0, i.e. both players contribute nothing, or (ii) below €3. According to the first definition we find weakly significant differences between FAIR and EGO ($p = 0.060$). Applying the second definition, we find significant differences between MIX and FAIR ($p = 0.012$) and weakly significant differences between EGO and FAIR ($p = 0.063$). In both cases, there are no significant differences between EGO and MIX.

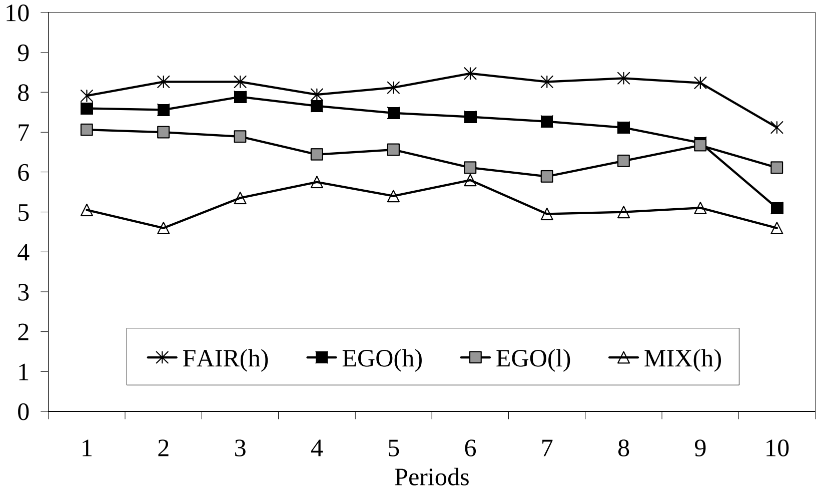
> *Result 2: The F&S model has some explanatory power for the behavior in the final period of the PG game: Subjects in the FAIR treatment contribute more to the public good than subjects in EGO and MIX. Furthermore, the share of subjects who "defect" in the final period is lower in FAIR than in EGO and MIX.*

Figure 2 illustrates the relatively poor performance of the subjects in the MIX treatment. Even though the differences between MIX and EGO are not significant, it is remarkable that the contributions in MIX are always lower. The MIX treatment is the only heterogeneous treatment where two different types of subjects make up a pair, namely a "fair" subject and an "egoistic" subject. In order to find out whether both types choose low contributions or whether one of the types is responsible for this development, we analyze the players' behavior in the first period in more detail. The first period plays an important role because the only information subjects have about their co-player is the behavior in games A and B. The "egoistic" subjects in MIX contribute on average €6.30 in the first period whereas the "fair" subjects contribute only €3.50. This difference is weakly significant (MW U test, $p = 0.057$). After the first period, the contributions quickly converge to a relatively low level so that we do not find any more significant differences between both types. Considering the share of defecting subjects in the first period, 38% of the "fair" subjects chose zero contributions compared to only 8% defecting "egoistic" subjects. The difference between the share of defecting subjects is significant (χ^2 test, $p = 0.024$). We conclude that it is the knowledge that the co-player is an "egoistic" type, which prevents the "fair" subjects from being cooperative in the beginning of the PG game. This suggests that these players believe more in the informative value of the choices in games A and B than they should. Put differently, subjects seem to believe in F&S even though its predictive power is low at this stage of the game.

4.2.2 Hypotheses testing of the F&S model in game D

All of the 160 subjects who played game C also completed game D, the PG game with punishment possibility. The mean contribution per period over all subjects and all periods is €6.90. In order to test our hypotheses for game D we consider the behavior of subjects in the three treatments with information and high costs, namely FAIR(h), EGO(h), and MIX(h) (see Table 4). The F&S prognoses for these treatments are the same as for game C. Zero contributions constitute the only equilibrium for EGO(h) and MIX(h) whereas also equilibria with positive contributions are feasible for FAIR(h). Figure 3 shows the development of the mean per-period contributions over time for all three treatments. The mean contribution over all periods for FAIR(h) is €8.10, for EGO(h) €7.20, and for MIX(h) €5.20. The differences between treatments are not significant at the 5% level except for those between FAIR(h) and MIX(h) (MW U test, p = 0.013). In all periods and in all treatments, the share of cooperating subjects who contribute a positive amount to the public good is significantly greater than zero (Binomial Sign test, one-tailed, p = 0.000), indicating that the contributions of the subjects in EGO(h) and MIX(h) are generally higher than the F&S predictions.

Figure 3: Contributions in game D



Result 3: In the EGO(h) and the MIX(h) treatments, the individual contributions to the public good in the PG game with punishment possibility are significantly higher than the levels predicted by F&S and by the standard model of pure selfishness. Again, contributions in FAIR(h) are in line with F&S but not with the standard model.

Again, we consider the subjects' behavior in the final period in more detail. The mean final period contribution in game D is €7.10 for FAIR(h), €5.10 for EGO(h), and €4.60 for MIX(h). This time, however, the differences between treatments are insignificant at the 5% level (MW U test). Considering the share of defecting subjects, we find that 48% of all subjects in EGO(h) and 45% of the MIX(h) subjects contribute nothing in the final period compared to only 24% of the FAIR(h) subjects. As already observed in game C, pairs consisting of two "fair" players seem to be more likely to cooperate in the final period than "egoistic" or "mixed" pairs. The comparison between treatments via χ^2 test with defection defined as zero contributions of both players exhibits that while the share of defecting subjects does not significantly differ between EGO(h) and MIX(h) it significantly differs between EGO(h) and FAIR(h) ($p = 0.026$) as well as between MIX(h) and FAIR(h) ($p = 0.018$). Applying the defection threshold of €3, we find significantly differences between FAIR(h) and MIX(h) ($p = 0.008$), weakly significant differences between FAIR(h) and EGO(h) ($p = 0.084$) and, again, no significant differences between EGO(h) and MIX(h).

> *Result 4: The F&S model has some explanatory power for the behavior in the final period of the PG game with punishing opportunities, although it is somewhat weaker than in the standard PG game: The share of defecting subjects is significantly lower in FAIR(h) than in the other treatments.*

4.2.3 Effect of information

Except for the treatment FAIR(h, ni), previous to games C and D subjects were informed about how their co-player had behaved in games A and B. As FAIR(h) and FAIR(h, ni) differ only with regard to information and not with regard to the F&S parameters α_i and β_i, the comparison between these two treatments enables us to analyze the isolated effect of information on the contributions to the public good.[22]

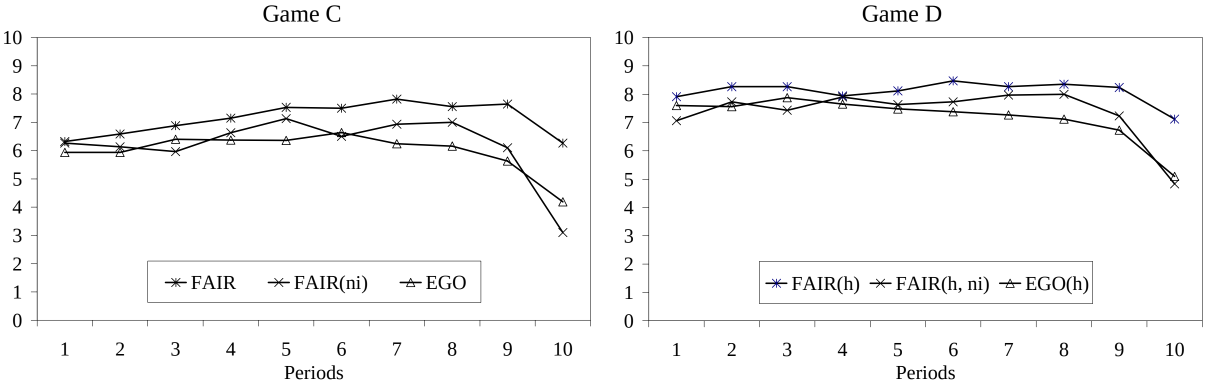
The mean contributions of FAIR(h) and FAIR(h, ni) are shown in Figure 4. For reference, we also insert the contributions of the "egoistic" subjects in EGO and EGO(h). In both games and in all periods the contributions of the uninformed subjects are lower than the contributions of the informed subjects, though the differences between FAIR(h) and FAIR(h, ni) in mean per-period contribution are not significant at the 5% level (MW U test). In the final period of both games, however, the contributions of FAIR(h) and FAIR(h, ni) differ significantly (MW U test, game C: $p = 0.025$, game D: $p = 0.048$). Remarkably, the information about the co-player seems to play an important role mainly in the final period, when strategic incentives to cooperate do not exist any longer. In contrast, we do not find significant differences in final

---

[22] All 64 subjects in FAIR and FAIR(ni) have $\alpha_i = 0$ and $\beta_i > 0.3$. There is no difference in the distributions of β_i between FAIR and FAIR(ni) at the 5% level (K-S test) (see Table 3).

period contributions between FAIR(ni) and EGO as well as between FAIR(h, ni) and EGO(h). It seems that in the final period of both games, the uninformed "fair" subjects behave similar to the "egoistic" subjects rather than to the informed "fair" subjects. Furthermore, the contributions of the informed subjects and uninformed subjects do not converge over time (see Figure 4). Hence, even though the uninformed "fair" subjects have the chance to get to know each other through interaction in the PG games this does not lead to the same level of contributions as observed for the informed subjects. Therefore, we can deduce that the information subjects receive about the type of their co-player before games C and D (based on the decisions in games A and B) is not equivalent to the information subjects are able to gather via direct interaction with their co-player during these PG games.

Figure 4: Contributions – effect of information



Let us now consider the share of defecting subjects in the final periods of both games. In game C, 67% of the FAIR(ni) subjects contribute nothing to the public good compared to only 32% of the FAIR subjects. Thus, the share of defecting subjects in FAIR(ni) is even larger than the share in EGO where 54% defect in the final period. A χ^2 test shows that while the difference in shares of defecting subjects are insignificant between FAIR(ni) and EGO they are significant between FAIR(ni) and FAIR at the 5% level. This result applies to both defection thresholds, zero contributions and contributions below €3 per pair. In the final period of game D, 43% of the FAIR(h, ni) subjects contribute nothing to the public good compared to 24% of subjects in FAIR(h). The difference between informed and uninformed subjects, however, is not significant anymore at the 5% level independently from the underlying defection threshold. To sum up, our observations regarding the behavior of the informed and uninformed subjects generally support hypothesis *H3*, although the evidence for an information effect is weaker in the game with punishment possibilities.

> *Result 5: In the final period of both games, uninformed "fair" subjects contribute significantly less than informed "fair" subjects. Furthermore, in the final period of the standard PG game the uniformed subjects are more likely to defect than the informed subjects.*

4.2.4 Effect of punishment

According to the standard model of purely selfish behavior, the introduction of punishment possibilities does not change the prediction for the contributions. As punishment is costly, a rational individual would not punish and the dominant strategy is still to contribute nothing to the public good. Experimental tests (e.g. Fehr and Gächter, 2000, 2002; Masclet and Villeval, 2008) show, however, that the introduction of punishment usually increases the contributions. This is also what we observe in game D. The mean per-period contribution over all periods is €5.90 in game C and €6.90 in game D.[23] A Wilcoxon Matched-Pairs Signed-Ranks test over all subjects who played C and D confirms that the difference between both games is highly significant (p = 0.000).

The use of two different prices of punishment allows us to analyze the effect of punishment costs on the contributions. Therefore, we compare the mean contributions between the treatments EGO(l) and EGO(h) (see Figure 3). The differences are not significant at the 5% level (MW U test). Thus, the costs of punishment do not seem to have any significant influence on the level of contributions to the public good.[24]

The possibility of punishing the co-player may increase or decrease the payoffs of the subjects. The payoffs will increase if higher contributions outweigh the costs of punishment. They will be lower if the costs of punishment outweigh the increase in contributions. In order to measure the change of payoffs we use a measure of efficiency defined as $Eff = \left(\pi - \pi^{NE}\right)/\left(\pi^{SO} - \pi^{NE}\right)$ with π giving the actual payoff, π^{NE} giving the payoff in the Nash equilibrium, and π^{SO} giving the payoff in the social optimum. In game D, we can furthermore distinguish between gross efficiency and net efficiency. The concept of gross efficiency neglects the fact that individual payoffs may be reduced by punishment received as

---

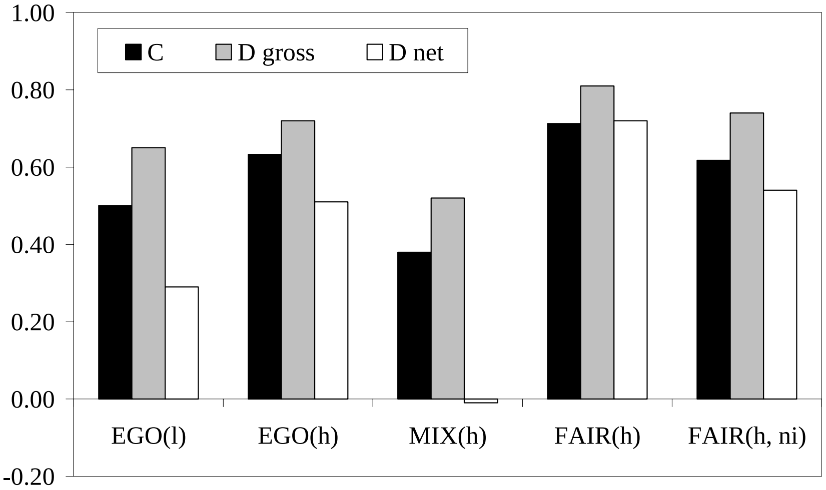
[23] At this point, we refrain from a separate graphical presentation. However, the magnitude of the contribution enhancing effect of punishment is also visible in Figure 5, if one compares the efficiency in game C and the gross efficiency in game D.

[24] See the Tobit estimates in the next section for an analysis of the marginal costs of punishment on the punishment behavior.

well as by costs of punishment imposed on others. Furthermore, gross efficiency in the punishment condition is usually larger than efficiency in the no-punishment condition because punishment tends to increase contributions and, by that, (gross) payoffs. Net efficiency contains the actual payoffs that include both contribution and punishment. Net efficiency in game D can be larger or smaller than efficiency in game C, depending on which of the two effects, change of contributions or punishment, dominates. Note that net efficiency can even be negative. If, for example, two subjects contribute nothing but punish each other at the same time, their payoffs will be lower than the payoff in the Nash equilibrium. Consequently, the net efficiency would be negative in this case.

Figure 5 shows the mean degree of efficiency over all periods in games C and D for all treatments. As expected, gross efficiency in game D always exceeds efficiency in game C. Net efficiency, however, is lower than efficiency in game C for treatments EGO(l), EGO(h), MIX(h), and FAIR(h, ni). This is particularly compelling in MIX(h) where net efficiency is negative. The difference between gains from higher contributions and these gains net of costs induced by punishment is highest in the MIX-groups, i.e. groups consisting of different types of subjects, one “egoistic” and one “fair” type. Homogeneous groups, i.e. groups of only “fair” or “egoistic” types, perform better than mixed groups. One reason for the poor performance of the MIX-groups is the fact that members of these groups tend to react to punishment with retaliation instead of increases in contributions. As a consequence, we often witness strong escalation of punishment over time without any positive effects on contributions.

Figure 5: Efficiency



The comparison between net efficiency of the treatments shows that differences are significant between MIX(h) and EGO(h) (MW U test, p = 0.017) as well as between MIX(h) and FAIR(h) (p = 0.003). Net efficiency between EGO(h) and FAIR(h) does not significantly differ at the 5% level. In contrast to all other treatments, net efficiency of FAIR(h) in the punishment condition approximates efficiency in the no-punishment condition. In this case, punishment does not lead to efficiency losses. Furthermore, the positive effect of information is still at work. The comparison of net efficiency shows that the difference between FAIR(h) and FAIR(h, ni) is significant (MW U test, p = 0.043) whereas it is not significant between EGO(h) and FAIR(h, ni). Thus, the behavior of the uninformed "fair" subjects resembles the behavior of the informed "egoistic" subjects rather than the behavior of the informed "fair" subjects.

> *Result 6: The introduction of an option to punish the co-player significantly increases the mean contribution to the public good. Punishment decreases net efficiency compared to a no-punishment environment for all treatments except FAIR(h). The decrease is particularly distinctive in the heterogeneous MIX(h) treatment. The information effect is still at work.*

4.2.5 Multivariate analysis

In the sections before, we used univariate tests which cannot account for several factors that could affect contributions or punishment behavior. Therefore, in this section we try to investigate the effects of α_i and β_i on the behavior of all subjects in games C and D with a multivariate analysis. In the following we focus on the dependent variables "contribution of subject i in period t" for game C and "punishment points from subject i to j in period t“ for game D. Due to the fact that both dependent variables are censored from below and from above, Table 5 presents Tobit estimates.

In game C, the contribution of the co-player in the previous period has – over all treatments – a strong and significantly positive effect on the contribution of subject i in the current period. Parameter α_i has a negative effect on the contributions of subject i over all treatments which is weakly significant. This effect is quite surprising given the low dispersion for α_i in the data (see Figure 1). Furthermore, β_i has a positive effect on i's contribution, which is significant at the 5% level. Both findings are in line with the F&S model if one assumes that there is, over all treatments, some uncertainty about the co-player’s type (see section 2.2.3). As one could expect from Figure 2, contributions in treatments MIX and FAIR(ni) are significantly below the contributions in the reference treatment (EGO).

> *Result 7: While* α_i *has a weakly significantly negative effect on the contributions to the public good over all treatments, the effect of* β_i *is significantly positive.*

Table 5: Tobit estimates

Independent variables	Dependent variable	
	Game C: contribution of subject *i* in period *t*	Game D: punishment points from subject *i* to *j* in period *t*
Constant	-2.1525 (0.4091)***	-0.4506 (0.1038)***
Contribution of co-player *j* in period *t*		-0.0980 (0.0108)***
Contribution of co-player *j* in period *t*-1	0.9560 (0.0403)***	
Contribution of subject *i* in period *t*		0.0398 (0.0093)***
Punishment points from subject *j* to *i* in period *t-1*		0.0739 (0.0109)***
α_i	-1.0937 (0.6131)*	-0.0611 (0.1512)
β_i	1.7158 (0.7702)**	-0.1352 (0.1995)
Dummy for c = 0.17		0.1782 (0.0857)**
MIX	-0.9537 (0.4114)**	0.2404 (0.0942)**
FAIR	-0.8372 (0.5782)	-0.1589 (0.1450)
FAIR(ni)	-1.5298 (0.5648)***	0.0841 (0.1465)
	N = 1440	N = 1440
	Prob > chi^2 = 0.0000	Prob > chi^2 = 0.0000
	Pseudo R^2 = 0.2168	Pseudo R^2 = 0.1683

Notes: The coefficients are marginal effects on the unconditional expected value of the dependent variable. The marginal effects are obtained running the “dtobit” command in Stata after the corresponding Tobit estimation. Standard errors are in parenthesis. * (**, ***) denotes significance at the 10% (5%, 1%) level. Dummies for periods are included but not indicated. The dependent variable is censored at a lower bound (0) and an upper bound (10 and 12) in game C and D.

Regarding game D, we observe that the contribution of the co-player in the current period has a negative effect on the allocated punishment points by subject i, i.e. higher contributions of j lead to lower punishment from i to j. Remarkably, stronger punishment received from the co-player in the previous period triggers stronger punishment imposed on the co-player, i.e. we observe a kind of escalation or “negative reciprocity” with respect to punishment behavior. Contrary to the estimation for game C, in game D, there are no effects of α_i and β_i on the dependent variable. This result is in line with the F&S model: Given our subject pool and the definition of treatments, there should be no effect of both parameters on punishment behavior (see sections 2.2 and 3.3). Furthermore, lower costs of punishment lead to higher punishment, which is a quite intuitive result. Finally, as one may expect given the results in the previous section (see also Figure 4), subjects punish significantly stronger in MIX than in EGO.

> *Result 8: In the heterogeneous treatment MIX, subjects punish each other significantly stronger than in the homogeneous treatment EGO. Lower costs for punishment lead to higher punishment. There are no significant effects of* α_i *and* β_i *on punishment behavior.*

5 Discussion and conclusion

The paper investigates the effects of heterogeneous other-regarding preferences on individual behavior in social dilemma situations as specified by PG games with and without the opportunity to punish the co-player. The most important result of our study is that the specific composition of pairs significantly influences the subjects' performance in the final period of the PG games: In the standard PG game, “fair” groups contribute more to the public good than “egoistic” or “mixed” pairs. In both PG games (with and without punishing possibility) purely fair pairs are more likely to cooperate in the final period than the other pairs. In addition, it turns out that explicit information is a key factor for the difference in behavior: Uninformed fair subjects in both PG games contribute significantly less than informed fair subjects. In the final period of the standard PG game, the uninformed subjects are more likely to defect than the informed subjects. Thus, as long as fair subjects are not informed on the fact that their co-player is fair, too, they act like egoistic subjects. This indicates that the existence of fair types who are reluctant to free ride at the expense of others is only a necessary but not a sufficient condition for the solution of a social dilemma. Additionally, these fair subjects have to be informed by a credible institution about the fact that subjects in their pair are of the same type. Without this ex ante information fair subjects are not able to solve the dilemma better than egoistic subjects. The absence of any convergence of the behavior of informed and uninformed fair subjects suggests that the information about the co-player’s type cannot be extracted during the PG game.

Our results show that even if we understand the F&S model as purely payoff based, i.e. without any consideration of reciprocity or intentions, at least the weight of aversion against advantageous inequity has some explanatory power for the individual behavior in social dilemma situations. However, the F&S model can explain only a fraction of the individual behavior. More precisely, it cannot explain why also the egoistic and mixed pairs provide clearly positive contributions to the public good. Given our results and in particular the strong effect of information, there are a lot of open questions with respect to within-subject tests of theories for other-regarding preferences.[25] We close this paper mentioning two of these issues which may be the subject of future research.

First, we have learnt that subjects respond very sensitively to uncertainty and information. Uncertainty about the opponent's type induces fair subjects to reduce contributions to the public good right mainly in the last period. In contrast, the knowledge that the opponent is an egoistic type induces fair subjects to contribute low amounts right from the start. This is perfectly in line with F&S but, given the actual behavior of their co-players, it is not the best they could do. Further research might find that subjects have a model similar to that of F&S in their mind even if it does not work that well. People might expect others to behave more consistently than they actually do. Second, we have observed in the repeated PG games that inequity avers partners are better able to maintain cooperation until the end. It remains an open question whether our results would also be valid in a stranger design where subjects meet another co-player in each period. Fehr and Gächter (2000) for example compare the performance of subjects in a PG game (with and without punishment possibility) with partner design and the performance of subjects in a stranger design. Regarding average contribution over all periods, they observe that subjects in the partner treatment contribute significantly more than subjects in the stranger treatment. The partner subjects are also more responsive to the introduction of the punishment opportunity. The dominant behavioral standard changes from full free riding in the no-punishment condition to full cooperation in the punishment condition. The authors suspect that the partner design is more likely to produce a certain behavioral norm because the "common group history provides a better basis for the formation of accurate beliefs about each others' behavior". Relating to our object of investigation it would be very interesting to see whether the informed fair subjects need a "common group history" in order to achieve and, even more importantly, maintain cooperation until the end or whether it is sufficient to know the co-player to be fair, too.

---

[25] For a variety of other points on consistency and stability of other-regarding preferences, see Brosig et al. (2007).

Acknowledgements:

Financial support from the German Science Foundation is gratefully acknowledged. The authors thank Andreas Lange and Dirk Engelmann for helpful comments.

References

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

Ben-Ner A., L. Putterman, F. Kong, and D. Magan (2004) Reciprocity in a Two-Part Dictator Game, Journal of Economic Behavior and Organization, 53, 333-352.

Blanco, M.,D. Engelmann and H.-T. Normann (2006), *A within Subject Analysis of Other-Regarding Preferences*, Royal Holloway College, University of London, working paper October 4, 2006, London.

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

Brosig, J., T. Riechmann and J. Weimann (2007), *Selfish in the End? An Investigation of Consistency and Stability of Individual Behavior*, FEMM Working Paper No. 07005, University of Magdeburg.

Camerer, C. F. (2003), Behavioral Game Theory. Experiments in Strategic Interaction, Princeton, New Jersey.

Carter, J.R. and M.D. Irons (1991), Are Economists Different, and if so, why?, *Journal of Economic Perspectives* 5, 171-177.

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

Cox, J.C., D. Friedman and V. Sadiraj (2008), Revealed Altruism, *Econometrica* 76, 31-69.

Dannenberg, A., T. Riechmann, B. Sturm and C. Vogt (2007), *Inequity Aversion and Individual Behavior in Public Good Games: An Experimental Investigation*, ZEW Discussion Paper 07-034, Mannheim.

Fehr, E. and S. Gächter (2000), Cooperation and Punishment in Public Goods Experiments, *American Economic Review* 90, 980-994.

Fehr, E. and S. Gächter (2002), Altruistic Punishment in Humans, *Nature* 415, 137-140.

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

Fehr, E. and K.M. Schmidt (2006), The Economics of Fairness, Reciprocity and Altruism - Experimental Evidence and New Theories, in: Kolm, S.-C. and J.M. Ythier (Eds.), *Handbook on the Economics of Giving, Reciprocity and Altruism* Vol. 1, Amsterdam, 615-691.

Fischbacher, U. (2007), z-Tree: Zurich Toolbox for Ready-made Economic experiments, *Experimental Economics* 10 (2), 171-178.

Forsythe, R., J.L. Horowitz, N.E. Savin, and M. Sefton (1994), Fairness in Simple Bargaining Experiments, *Games and Economic Behavior* 6, 347–369.

Frank, R.H., T.D. Gilovich and D.T. Regan (1996), Do Economists Make Bad Citizens?, *Journal of Economic Perspectives* 10, 187-192.

Gächter, S. and C. Thöni (2005), Social Learning and Voluntary Cooperation among Like-Minded People, *Journal of the European Economic Association* 3, 303-314.

Güth, W., R. Schmittberger and B. Schwarze (1982), An Experimental Analysis of Ultimatum Bargaining, *Journal of Economic Behavior and Organization* 3, 367-388.

Gunnthorsdottir, A., D. Houser and K. McCabe (2007), Disposition, History and Contributions in Public Goods Experiments, *Journal of Economic Behavior & Organization* 62, 304-315.

Kahneman, D., J.L. Knetsch and R.H. Thaler (1986): Fairness and the Assumptions of Economics, *Journal of Business* 59, 85–300.

Kreps, D., P. Milgrom, J. Roberts, and R. Wilson (1982), Rational Cooperation in the Finitely Repeated Prisoners' Dilemma, *Journal of Economic Theory*, 27, 245-252.

Kritikos, A. and F. Bolle (2001), Distributional Concerns: Equity or Efficiency Oriented?, Economics Letters 73, 333-338.

Ledyard, J.O. (1995): Public Goods. A Survey of Experimental Research, in Kagel, J. H. and A.E. Roth (Eds.), *The Handbook of Experimental Economics*, Princeton, 111-194.

Masclet, D. and M.-C. Villeval (2008): Punishment, Inequality, and Welfare: A Public Good Experiment, *Social Choice and Welfare*, forthcoming.

Marwell, G. and R.E. Ames (1981), Economists Free Ride, Does Anyone Else? Experiments on the Provision of Public Goods IV, *Journal of Public Economics* 15, 295-310.

Ockenfels, A. and J. Weimann (1999), Types and patterns: an experimental East-West-German comparison of cooperation and solidarity, Journal of Public Economics 71, 275–287.