No. 1812

A THEORY OF FAIRNESS, COMPETITION AND COOPERATION

Ernst Fehr and Klaus M Schmidt

INDUSTRIAL ORGANIZATION

Centre for Economic Policy Research

ISSN 0265-8003

A THEORY OF FAIRNESS, COMPETITION AND COOPERATION

Ernst Fehr and Klaus M Schmidt

Discussion Paper No. 1812
February 1998

Centre for Economic Policy Research
90–98 Goswell Rd
London EC1V 7DB
Tel: (44 171) 878 2900
Fax: (44 171) 878 2999
Email: cepr@cepr.org

CEPR Discussion Paper No. 1812

February 1998

ABSTRACT

A Theory of Fairness, Competition and Cooperation*

There is strong evidence that people exploit their bargaining power in competitive markets but not in bilateral bargaining situations. There is also strong evidence that people exploit free-riding opportunities in voluntary cooperation games. Yet, when they are given the opportunity to punish free riders, stable cooperation is maintained although punishment is costly for those who punish. This paper asks whether there is a simple common principle that can explain this puzzling evidence. We show that if a fraction of the people exhibits inequality aversion the puzzles can be resolved.

JEL Classification: C78, C90, D43, H41
Keywords: experimental economics, utility theory, bargaining, public goods, reciprocity

Ernst Fehr
Institute for Empirical Research in Economics
Universität Zürich
Blümlisalpstrasse 10
CH–8006 Zürich
SWITZERLAND
Tel: (41 1) 634 3709
Fax: (41 1) 634 4907
Email: efehr@iew.unizh.ch

Klaus M Schmidt
Department of Economics
Universität München
Ludwigstrasse 28 (Rgb)
D–80539 München
GERMANY
Tel: (49 89) 2180 2250
Fax: (49 89) 2180 3510
Email: klaus.schmidt@lrz.uni-muenchen.de

*This paper is produced as part of a CEPR research programme on *Market Structure and Competition Policy*, supported by a grant from the Commission of the European Communities under its Human Capital and Mobility Programme (no. ERBCHRXCT940653). The authors would like to thank seminar participants at Bonn, Harvard and Princeton and at the CEPR/Studienzentrum Gerzensee European Summer Symposium in Economic Theory, 28 June–11 July 1997, Gerzensee, for helpful comments and suggestions. They are particularly grateful to Drew Fudenberg who

pointed out an error in an earlier version of this paper. The first author also gratefully acknowledges support from the Swiss National Science Foundation (project number 12–43590.95) and the Network on the Evolution of Preferences and Social Norms of the MacArthur Foundation.

Submitted 22 December 1997

NON-TECHNICAL SUMMARY

Almost all economic models assume that *all* people are *exclusively* pursuing their material self-interest and do not care about 'social' goals *per se*. This may be true for some people, but it is certainly not true for everybody. By now we have substantial evidence that fairness motives affect the behaviour of many people. For example, there are many well-controlled bilateral bargaining experiments which indicate that a non-negligible fraction of the subjects do not *only* care about material pay-offs. These experiments are designed such that, according to the self-interest model, one player has all the bargaining power so that the other player is predicted to receive nothing. Contrary to this prediction the vast majority of 'powerful' players are willing to offer the 'weak' player between 40% and 50% of the bargaining surplus.

Although the facts from bargaining experiments suggest that fairness motives matter, it would be wrong to assume that *all* people are motivated by fairness considerations. Most of us know at least some people whom we consider as rather self-interested. In addition, experimental evidence indicates that there are not only those conditions in which most subjects propose a relatively egalitarian distribution of the bargaining pie. When there is competition among people as, for example, in some market games, one gets the impression that fairness motives do not play much of a role. The most rigorous evidence in this regard comes from competitive experimental markets in which a well-defined, homogeneous good is traded. Numerous experiments show that these markets converge to the 'standard' competitive equilibrium under a wide variety of circumstances. By 'standard' we mean the equilibrium that is derived from the premise that *all* subjects care *only* for their material self-interest. Even if this equilibrium is very unfair by almost any conceivable definition of fairness, i.e. if one side of the market reaps the *whole* gains from trade, convergence to the standard competitive equilibrium occurs.

There is similarly conflicting evidence with regard to cooperation. Social reality provides many examples indicating that humans are more cooperative than is assumed in the standard self-interest model. Well known examples are that many people vote, pay their taxes honestly, participate in unions and protest movements, or work hard in teams even when the pecuniary incentives go in the opposite direction. Although, there are also those conditions in which a vast majority of subjects completely defect as predicted by the self-interest model.

There is thus a bewildering variety of evidence, some of which suggests that many people are driven by fairness considerations; some that virtually all people behave as if completely selfish and still other types of evidence suggest that cooperation motives are crucial. In this paper we ask whether this conflicting evidence can be explained by a *single simple* model. Put differently: is it possible to explain the conflicting evidence on the basis of a common underlying principle? Our answer to this question is affirmative if one is willing to assume that, in addition to purely self-interested people, there is a fraction of people who are also motivated by *self-centred inequality aversion*. No other deviations from the standard economic approach are necessary to account for the evidence. In particular, we do not relax the rationality assumption.

According to our definition a person exhibits inequality aversion if they dislike, to some extent, being better-off than relevant others and/or if they dislike, to some extent, being worse-off than relevant others. Inequality aversion is thus based on a social comparison process in which a person's *relative* material pay-off affects their behaviour. It is self-centred for two reasons: (i) inequality *between other people* is assumed to be motivationally *un*important. It is only the inequality between the person and each of the other people that matters motivationally. (ii) Based on evidence from psychological studies we assume that people dislike inequality more if it is to their disadvantage than if it is to their advantage. Thus we suppose a kind of self-serving bias in inequality aversion which can be interpreted as loss aversion with regard to social comparisons.

Since there are both situations in which the vast majority of the people behave as if they are solely motivated by material self-interest and situations in which the vast majority of people behave fairly or cooperatively, we face the challenge of explaining under what circumstances the aggregate outcome is shaped by the purely self-interested types and when it is determined by those who also are inequality averse. A major result of our paper is that the economic and institutional environment determines which preference type is decisive for the prevailing behaviour in equilibrium. In particular, we show that 'fair' outcomes prevail in simple bargaining games if there is at least a small fraction of subjects who are sufficiently motivated by inequality aversion. In contrast, in monopolistic market games the outcome is largely independent of the distribution of types and there is a unique equilibrium where the monopolist gets all of the available surplus. In public good games our theory predicts that a rather small fraction of selfish players renders cooperation impossible. On the other hand, if the players have an opportunity to punish each other at some cost after the contributions to the public good have been made, a radically different equilibrium outcome occurs. Now full cooperation can be

enforced if there is a relatively small fraction of players that care sufficiently about inequality.

The above results suggest that the interaction between the economic environment and a population with a fraction of people who exhibit self-centred inequality aversion goes a long way to explain 'fair' and 'cooperative' as well as 'competitive' and 'non-cooperative' behavioural patterns in a coherent framework. Of course, we do not believe that other types of motivations, such as altruism or reciprocity, are unimportant and should be disregarded. We compare our approach to these alternative concepts in Section VIII. The discussion in this section also shows that the frequently raised criticism 'that one can explain everything by choosing the appropriate utility function' need not be appropriate. Of course, if we postulated *different* utility functions for each of the situations we consider, the criticism would apply. But, we keep the utility function, and in particular the distribution of utility functions in the population, constant across situations. Moreover, as the discussion in Section VIII shows, it is possible to design experiments that can discriminate between explanations based on different motivational assumptions. Therefore, in view of the explanatory power of self-centered inequality aversion across many different situations, we believe that it should be considered as a serious motive that deserves further analysis.

I. Introduction

Almost all economic models assume that *all* people are *exclusively* pursuing their material self-interest and do not care about "social" goals per se. This may be true for some (may be many) people, but it is certainly not true for everybody. By now we have substantial evidence that fairness motives affect the behavior of many people (Kahneman, Knetsch and Thaler 1986). For example, there are many well-controlled bilateral bargaining experiments which indicate that a nonnegligible fraction of the subjects does not *only* care about material payoffs (Güth and Tietz 1990, Roth 1995, Camerer and Thaler 1995). These experiments are designed such that, according to the self-interest model, one player has all the bargaining power so that the other player is predicted to receive nothing. Contrary to this prediction the vast majority of "powerful" players is willing to offer the "weak" player between 40 and 50% of the bargaining surplus.

Although the facts from bargaining experiments suggest that fairness motives matter it would clearly be equally wrong to assume that *all* people are motivated by fairness considerations. Most of us know at least some people whom we consider as rather self-interested. In addition, experimental evidence indicates that there are not only those conditions in which most subjects propose a relatively egalitarian distribution of the bargaining pie. When there is competition among people, as, e.g., in some market games, one gets the impression that fairness motives do not play much of a role. The most rigorous evidence in this regard comes from competitive experimental markets in which a well-defined, homogeneous good is traded. Numerous experiments show that these markets converge to the "standard" competitive equilibrium under a wide variety of circumstances (Smith 1982, Plott 1989, Davis and Holt 1993). By "standard" we mean the equilibrium that is derived from the premise that *all* subjects care *only* for their material self-interest. Even if this equilibrium is very unfair by almost any conceivable definition of fairness, i.e., if one side of the market reaps the *whole* gains from trade, convergence to the standard competitive equilibrium occurs. (Smith and Williams 1990, Kachelmeier and Shehata 1992). Similar results can be observed in experiments in which a monopolistic seller faces many competing buyers or a monopsonistic buyer faces many competing sellers. In these situations the "powerful" subjects show few scruples to extract all of the rents from the "weak" subjects. Likewise, the subjects who are in competition with each other exhibit little cooperation to achieve positive shares in the gains from trade [Roth, Prasnikar, Okuno-Fujiwara and Zamir (1991), henceforth RPOZ; Güth, Marchand and Rulliere (1997), henceforth GMR].

There is similarly conflicting evidence with regard to cooperation. Social reality provides many examples indicating that humans are more cooperative than is assumed in the standard self-interest model. Well known examples are that many people vote, pay their taxes honestly, participate in unions and protest movements, or work hard in teams even when the pecuniary

incentives go in the opposite direction.[1] That people seem to be more cooperative than predicted by the standard self-interest model is also supported by laboratory experiments (Dawes and Thaler 1988, Ledyard 1995). Under some conditions it has even been shown that subjects achieve nearly complete cooperation although the self-interest model predicts complete defection (Isaac and Walker 1988 and 1991, Ostrom and Walker 1991, Fehr and Gächter 1996).[2] However, as we will see in more detail in Section IV there are also those conditions in which a vast majority of subjects completely defects as predicted by the self-interest model.

There is thus a bewildering variety of evidence. Some pieces of evidence suggest that many people are driven by fairness considerations, other pieces indicate that virtually all people behave as if completely selfish and still other types of evidence suggest that cooperation motives are crucial. In this paper we ask whether this conflicting evidence can be explained by a *single simple* model. Put differently: Is it possible to explain the conflicting evidence on the basis of a common underlying principle? Our answer to this question is affirmative if one is willing to assume that, in addition to purely self interested people, there is a fraction of people who are also motivated by *self-centered inequality aversion.* No other deviations from the standard economic approach are necessary to account for the evidence. In particular, we do not relax the rationality assumption[3].

According to our definition a person exhibits inequality aversion if it dislikes, to some extent, being better off than relevant others and/or if it dislikes, to some extent, being worse off than relevant others. Inequality aversion is thus based on a social comparison process in which a person's *relative* material payoff affects its behavior. It is self-centered for two reasons: *(i)* Inequality *between other people* is assumed to be motivationally *un*important. It is only the inequality between the person and each of the other people that matters motivationally. *(ii)* Based on evidence from psychological studies we assume that people dislike inequality more if it is to their disadvantage than if it is to their advantage. Thus we suppose a kind of self-serving bias in inequality aversion which can be interpreted as loss aversion with regard to social comparisons.[4]

People who exhibit inequality aversion are willing to give up some material payoff to achieve less inequality between them and relevant others. In Section II we argue that many forms

---

[1] On voting see Mueller (1989). Skinner and Slemroad (1985) show that the standard self interest model substantially underpredicts the number of honest taxpayers. Successful team production in, e.g., Japanese-managed auto factories in North America is described in Rehder (1990). Whyte (1955) discusses how workers establish "production norms" under piece rate systems.

[2] Isaac and Walker and Ostrom and Walker allow for cheap talk while in Fehr and Gächter subjects could punish each other at some cost.

[3] The issue of rationality is further discussed in Section VIII below.

[4] For the relevance of loss aversion in other contexts see Tversky and Kahneman (1992).

of actual behavior, like e.g. charitable giving, voluntary labor, bargaining breakdowns and, more generally, costly retaliation can easily be interpreted in terms of self-centered inequality aversion.

Since there are *(i)* situations in which the vast majority of the people behave as if they are solely motivated by material self-interest *and (ii)* situations in which the vast majority behaves fairly or cooperatively we face the challenge to explain under what circumstances the aggregate outcome is shaped by the purely self interested types and when it is determined by those who also are inequality averse. A major result of our paper is that the economic and institutional environment determines which preference type is decisive for the prevailing behavior in equilibrium. In particular, we show the following results:

(i) In simple bargaining games the prevalence of "fair" outcomes in equilibrium depends on the full distribution of types. We show, however, that a relatively small fraction of subjects who are sufficiently motivated by inequality aversion is sufficient to induce even very selfish players to make "fair" bargaining offers.

(ii) In a monopolistic market in which one seller proposes a share of the surplus to many competing buyers ("responder competition") a *single* selfish buyer is sufficient to trigger the extreme outcome where the seller reaps the whole gains from trade. Even if *all other* buyers are extremely inequality averse their inequality aversion has no impact on the final outcome. In equilibrium every buyer behaves as if *only* motivated by material self-interest.

(iii) In case of a monopsonistic buyer facing many potential sellers who compete in prices ("proposer competition") an even stronger result obtains. Irrespective of the degree of inequality aversion of *all* players, the most inegalitarian result, where the buyer reaps the whole gains from trade, is the unique equilibrium outcome. No matter how inequality averse, every player behaves as if only interested in his own material payoff. Note that the distribution of types is completely irrelevant here. Results *(ii)* and *(iii)* explain why under competitive conditions selfish behavior prevails.

(iv) In a public goods game in which players have no opportunity to punish each other and in which no external agency enforces contributions a relatively *small fraction* of egoistic players is sufficient to guarantee the existence of a *unique* equilibrium in which *everybody fully* defects. For example, in a typical experiment with four players a *single* egoistic player is sufficient to trigger this result. Even if all other players are very inequality averse they are not capable of sustaining cooperation. This follows from the fact that inequality averse people are "conditional cooperators". They are willing to forgo the material gains from free-riding *provided* that sufficiently many of the other players cooperate as well. In the presence of too many selfish players this condition is, however, not met so that inequality averse players prefer to free-ride, too.

(v) A radically different equilibrium outcome can be sustained in a public good game in which participants have the opportunity to punish each other at some cost after everybody has observed the vector of all contributions. Note that a selfish player will never punish since punishing is costly. Therefore, the punishment opportunity is behaviorally irrelevant if *all* players are selfish. However, a single player who cares sufficiently about equality can sustain an equilibrium in which *everybody* cooperates fully. Even if *all other* players are completely selfish a single "enforcer" can ensure that nobody defects in equilibrium.

(vi) In a simultaneous move prisoner's dilemma *both* players have to pass a threshold value of inequality aversion to render mutual cooperation an equilibrium. In contrast, in a sequential prisoner's dilemma cooperation can be an equilibrium even if one of the players *does not* pass this threshold. Therefore, the prospects for cooperation are much better if the players move sequentially. This result provides a potential rationale for the relatively high degree of cooperation that has been observed in gift exchange markets.[5]

The above results suggest that the interaction between the economic environment and a population with a fraction of people who exhibit self-centered inequality aversion goes a long way to explain "fair" and "cooperative" as well as "competitive" and "non-cooperative" behavioral patterns in a coherent framework. Of course, we do not believe that other types of motivations, such as altruism or reciprocity, are unimportant and should be disregarded. We will compare our approach to these alternative concepts in Section VIII below. The discussion in this section also shows that the frequently raised criticism "that one can explain everything by choosing the appropriate utility function" need not be appropriate. Of course, if we would postulate *different* utility functions for each of the situations we consider, the criticism would apply. But, we keep the utility function, and in particular the distribution of utility functions in the population, constant across situations. Moreover, as the discussion in Section VIII shows, it is possible to design experiments that can discriminate between explanations based on different motivational assumptions. Therefore, in view of the explanatory power of self-centered inequality aversion across many different situations, we believe that it should be considered as a serious motive that is worth being analyzed further.

The rest of the paper is organized as follows. In Section II we present our model of inequality aversion. Section III applies this model to bilateral bargaining and market games. In Section IV cooperation games with and without punishments are considered. In Section V we show that, on the basis of plausible assumptions about preference parameters, the majority of individual choices in the ultimatum *and* market *and* cooperation games considered in the previous sections can be explained. Section VI deals with the dictator game and with various other games,

[5] In gift exchange markets firms typically pay supracompetitive wages and workers respond with cooperative effort choices despite pecuniary incentives to the contrary (Fehr, Kirchsteiger and Riedl 1993; Fehr, Gächter and Kirchsteiger 1997). The results from experimental gift exchange markets are in sharp contrast to the competitive outcomes that arise in experimental markets in which a well defined, homogeneous, good is traded.

including the one-shot and the finitely repeated prisoner's dilemma as well as trust- and gift exchange games. In Section VII we compare our model to alternative approaches and conclude.

II. A Simple Model of Self-centered Inequality Aversion

By inequality aversion we mean that in a given situation a person has a positive willingness to pay for a reduction in inequality between himself and some relevant others who are affected by his actions. In general a reduction in inequality may take two forms. First, a person may voluntarily give away material resources in order to *increase* the well-being of other people. This kind of transfer is observed in many natural environments as well as in laboratory experiments. For example, according to Kelly (1997) there are 115000 organisations engaged in charitable fundraising in the USA. Weisbrod (1988) estimates that the total value of voluntary labor in the U.S. is $ 74 billion annually. In the laboratory the so-called dictator game is the prototypical example indicating that a nonnegligible fraction of the better off subjects has a willingness to redistribute money to the worse off. In this game an individual (the dictator) is endowed with a certain amount of money and has to decide what fraction of the money to transfer to another, unknown, stranger who has no choice to make. The dictator knows only that the receiver has not been given any money by the experimenter. Typically there are between 20 and 40% of dictators who keep the entire amount but also between 20 and 40% of subjects who share the money equally. The rest of the subjects transfer more than zero but less than 50%. (Forsythe, Horowitz, Savin and Sefton 1994, henceforth FHSS; Andreoni and Miller 1995).

A second form of inequality reduction can be achieved if a person gives up resources to *reduce* the monetary payoff of others by more than the reduction of his own monetary payoff. Many forms of costly retaliation take this form. For example, in bilateral bargaining experiments people often destroy the whole bargaining surplus by rejecting *positive*, yet uneven, offers. Thus, they are loosing money by the rejection but the bargaining opponent looses even more. Likewise in public good experiments with costly punishment opportunities some people choose to punish and to reduce the payoff of other players even though this is costly to themselves. We will discuss these experiments in more detail in Sections III and IV below.

In addition to the above mentioned *behavioral* evidence in favor of inequality aversion there also exists questionnaire evidence from psychological studies. In an interesting paper by Loewenstein, Thompson and Bazerman (1989) the authors ask subjects to ordinally rank outcomes that differ in the distribution of payoffs between the subject and a comparison person. On the basis of these ordinal rankings the authors estimate how *relative* material payoffs enter the person's utility function. The results show that subjects exhibit a strong and robust aversion against disadvantageous inequality: For a given own income x_i subjects rank outcomes in which a comparison person earns more than x_i substantially lower than an outcome with equal material

payoffs. Many subjects also exhibit an aversion against advantageous inequality although this effect seems to be significantly weaker than the aversion against disadvantageous inequality. The results of Loewenstein, Thomson and Bazerman, thus, suggest that there is a nonnegligible fraction of people who are willing to reduce inequality relative to comparison persons. Yet, the willingness to pay for a reduction in disadvantageous inequality is, in general, substantially higher than the willingness to pay for a reduction in advantageous inequality.[6]

Formally, we define self-centered inequality aversion as follows. Consider a set of n players indexed by $i \in \{1,...,n\}$ and let $x=(x_1,...,x_n)$ denote the vector of monetary payoffs. The utility function of player $i \in \{1,...,n\}$ is given by

$$(1) \qquad U_i(x) = x_i - \alpha_i \frac{1}{n-1} \sum_{j \neq i} \max \{x_j - x_i, 0\} - \beta_i \frac{1}{n-1} \sum_{j \neq i} \max \{x_i - x_j, 0\} ,$$

where we assume $\beta_i \leq \alpha_i$ and $0 \leq \beta_i < 1$. In the two-player case (1) simplifies to

$$(2) \qquad U_i(x) = x_i - \alpha_i \max \{x_j - x_i, 0\} - \beta_i \max \{x_i - x_j, 0\}, \quad i \neq j.$$

The second term in (1) or (2) measures the utility loss from disadvantageous inequality, while the third term measures the loss from advantageous inequality. Figure 1 illustrates the utility of player i as a function of x_j for a given income x_i. Given his own monetary payoff x_i, player i's utility function obtains a maximum at $x_j = x_i$. The utility loss from disadvantageous inequality ($x_j > x_i$) is larger than the utility loss if player i is better off than player j ($x_j < x_i$).

To evaluate the implications of this utility function let us start with the two-player case. For simplicity we assume that the utility function is *linear* in inequality aversion as well as in x_i. This implies that the marginal rate of substitution between monetary income and inequality is constant. This may not be completely realistic, but we will show that surprisingly many experimental observations that seem to contradict each other can be explained on the basis of this very simple utility function already. However, we will also see that some observations in dictator experiments suggest that there is a nonnegligible fraction of people who exhibit nonlinear inequality aversion in the domain of advantageous inequality (see Section VI below).

---

[6] In addition to the study of Loewenstein, Thompson and Bazerman there is a literature in psychology that indicates that human subjects are prone to self-serving biases (Hastorf and Cantril 1954, Sanitiosa, Kunda and Fong 1990). Of particular interest for our purposes is the evidence in favor of self-serving biases in fairness judgements (Messick and Sentis 1979, Babcock, Loewenstein, Issacharoff and Camerer 1995, Roth and Murningham 1982). There is also a large literature on relative deprivation in sociology (e.g. Runciman 1966) which is based on the idea that being *relatively* worse off in material terms is particularly harmful for the people.

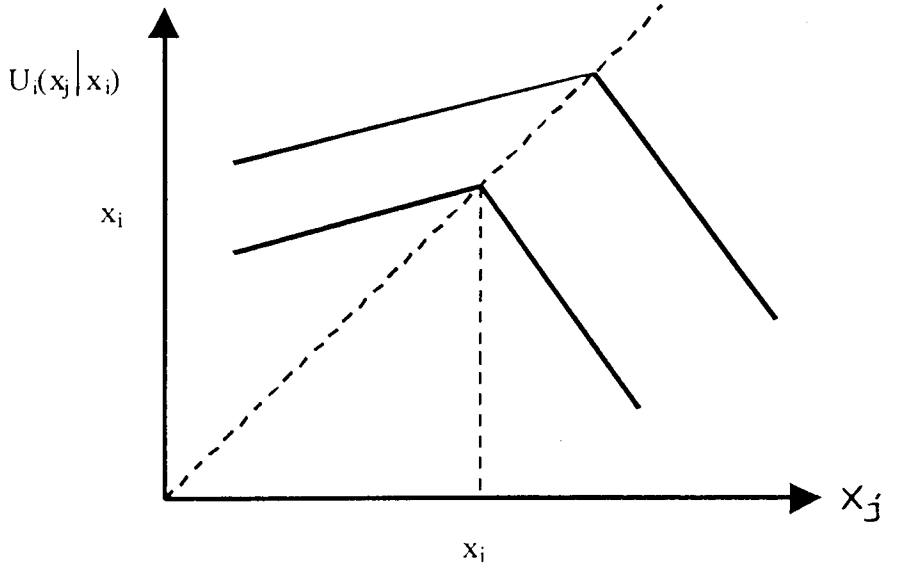


Figure 1: Preferences with Inequality Aversion

Furthermore, on the basis of the evidence in favor of self-serving biases in fairness judgements and on the basis of the empirical results of Loewenstein, Thompson and Bazerman (1989) we assume $\alpha_i \geq \beta_i$, i.e., a player suffers more from inequality that is to his disadvantage. We also assume that $\beta_i < 1$. To interpret this restriction suppose that player i has a higher monetary payoff than player j. In this case $\beta_i = 0.5$ implies that player i is just indifferent between keeping 1 Dollar to himself and giving this Dollar to player j. If $\beta_i = 1$, then player i is prepared to throw away 1 Dollar in order to reduce his advantage relative to player j which seems very implausible. This is why we do not consider the case $\beta_i \geq 1$. On the other hand, there is no justification to put an upper bound on α_i. To see this suppose that player i has a lower monetary payoff than player j. In this case player i is prepared to give up one Dollar of his own monetary payoff if this reduces the payoff of his opponent by $(1+\alpha_i)/\alpha_i$ Dollars. For example, if $\alpha_i = 4$, then player i is willing to give up one Dollar if this reduces the payoff of his opponent by 1.25 Dollars. We will see that observable behavior in bargaining and public good games suggests, that there are at least some individuals with such high α's.

If there are $n > 2$ players, player i compares his income to *all other* $n - 1$ players.[7] In this

[7] Since we restrict attention to the analysis of experimental games with a well defined set of players who interact in the laboratory, it seems natural to assume that each player compares himself to each of his opponents. In naturally

case the disutility from inequality has been normalized by dividing the second and third term by $n-1$. This normalization is necessary to make sure that the relative impact of inequality aversion on player i's total payoff is independent of the number of players. Furthermore, we assume for simplicity that the disutility from inequality is self-centered in the sense that player i compares himself to each of the other players, but he does not care *per se* about inequalities within the group of his opponents.

III. Fairness, Retaliation and Competition - Ultimatum and Market Games

In this section we apply our model to a well known simple bargaining game - the ultimatum game - and to simple market games in which one side of the market competes for an indivisible good. As we will see below a considerable body of experimental evidence indicates that in the ultimatum game the gains form trade are shared relatively equally while in market games very unequal distributions are frequently observed. Hence, any alternative to the standard self-interest model faces the challenge to explain both "fair" outcomes in the ultimatum game and "competitive" and rather "unfair" outcomes in market games.

III.1. The Ultimatum game

Consider a single buyer and a single seller who bargain for the price of an indivisible good. Without loss of generality we normalize the gains from trade to one. The buyer's share is denoted by s and the seller's share by $1-s$. The bargaining rules stipulate that the seller proposes a price that gives the buyer a share $s \in [0,1]$ of the surplus. The buyer can accept or reject s. In case of a rejection no trade takes place while in case of acceptance the good is traded at the proposed share. This yields a (normalized) monetary payoff $x_1 = 1-s$ for the seller and $x_2 = s$ for the buyer. In case of a rejection both players receive a monetary return of zero. The self-interest model predicts that the buyer accepts any $s \in (0,1]$ and is indifferent between accepting and rejecting $s = 0$. Therefore, there is a unique subgame perfect equilibrium in which the seller proposes $s = 0$, which is accepted by the buyer.[8]

---

occurring environments it may be more difficult to define the group of relevant comparison persons. This focus does not imply, however, that we consider our model inapplicable to real world issues. It merely acknowledges the fact that well-controlled laboratory experiments provide particularly clean data to judge the adequacy of a theory. Experimental games are particularly useful when assumptions about motives are critical.

[8] Given that the seller can choose s continuously, any offer $s>0$ cannot be an equilibrium offer since there always exists an s' with $0 < s' < s$ which is also accepted by the buyer and yields a strictly higher payoff to the seller. Furthermore, it cannot be an equilibrium that the seller offers $s=0$ which is rejected by the buyer with positive probability. In this case the seller would do better by slightly raising his price in which case the buyer would accept with probability 1. Hence, the only subgame perfect equilibrium is that the seller offers $s=0$ which is accepted by the buyer. If there is a smallest money unit ε, then there exists a second subgame perfect equilibrium in which the

By now there are numerous experimental studies from different countries, with different stake sizes and different experimental procedures, that clearly refute this prediction (for overviews see Güth and Tietz 1990, Camerer and Thaler 1995, Roth 1995). The following regularities can be considered as robust facts: (i) There are virtually no offers above 0.5. (ii) The vast majority of offers in almost any study is in the interval [0.4, 0.5] (see Table 1).

Table 1

Percentage of offers below 0.2 and between 0.4 and 0.5 in the Ultimatum Game

Study (Payment method)	Number of observations	Stake Size (Country)	Percentage of offers with $s < 0.2$	Percentage of offers with $0.4 \leq s \leq 0.5$
Cameron (1995) (all Ss paid)	35	Rp 40.000 (Indonesia)	0	66
Cameron (1995) (all Ss paid)	37	Rp 200.000 (Indonesia)	5	57
FHSS (1994) (all Ss paid)	67	$5 and $10 (USA)	0	82
Güth et al. (1982) (all Ss paid)	79	DM 4 - 10 (Germany)	8	61
Hoffman, McCabe and Smith (1996) (all Ss paid)	24	$10 (USA)	0	83
Hoffman, McCabe and Smith (1996) (all Ss paid)	27	$ 100 (USA)	4	74
Kahneman, Knetsch and Thaler (1986) (20 % of Ss paid)	115	$10 (USA)	?	75[a]
RPOZ (1991) (random payment method)	116[b]	approx. $10 (USA, Slovenia, Israel, Japan)	3	70
Slonim & Roth (1997) (random payment method)	240[c]	SK 60 (Slovakia)	0.4[d]	75
Slonim & Roth (1997) (random payment method)	250[c]	SK 1500 (Slovakia)	8[d]	69
Aggregate result of all studies[e]	875		3.8	71

a: percentage of equal splits, b: only observations of the final period, c: observations of all 10 periods, d: percentage of offers below 0.25, e: without Kahneman, Knetsch and Thaler (1986).

(iii) There are almost no offers below 0.2 (see Table 1). (iv) Low offers are frequently rejected and the probability of rejection decreases, in general, with s. (v) A further interesting fact

---

buyer accepts any $s \in [\varepsilon, 1]$ and rejects $s = 0$ while the seller proposes ε.

concerns the behavior of sellers if the power of the buyers is varied. FHSS (1994) show that if the buyer's option to reject is removed, so that the buyer *must* accept any offer, the sellers become significantly less "generous". Zamir and Winter (1996) conducted experiments in which sellers faced computerized buyers although they were led to believe that they face human opponents. They show that the frequency of high offers increases if computers play "tough" and decreases if computers accept relatively low offers. This suggests that the buyers' rejection behavior imposes a binding constraint for at least some sellers.

Regularity (i) to (iv) also hold true for rather high stake sizes, as indicated by the results of Cameron (1995), Hoffman, McCabe and Smith (1996) and Slonim and Roth (1997). The 200.000 Rupia in the second experiment of Cameron (see Table 1) are, e. g., equivalent to three months' income for the Indonesian subjects. Overall, roughly 60 - 80% of the offers in Table 1 fall in the interval [0.4, 0.5] while only 3% are below a share of 0.2.

To what extent is our model capable of accounting for the stylized facts of the ultimatum game? To answer this question let the seller's preferences be represented by (α_1, β_1) and assume that the seller does not exactly know the preference parameter (α_2, β_2) of a randomly matched buyer but only the distribution of buyers' preferences. In fact, as we will see, the seller needs to know only the distribution $F(\alpha_2)$. We assume that α_2 has support $[\underline{\alpha}, \overline{\alpha}]$ with $0 \leq \underline{\alpha} < \overline{\alpha} < \infty$.

Proposition 1: It is a dominant strategy for every buyer to accept any offer $s \geq 0.5$, to reject s if

$$s < s'(\alpha_2) \equiv \alpha_2/(1+2\alpha_2) < 0.5,$$

and to accept $s > s'(\alpha_2)$. From the perspective of the seller the probability of acceptance of an offer $s < 0.5$ by a randomly selected buyer is given by

$$(3) \qquad p = \begin{cases} 1 & \text{if } s \geq s'(\overline{\alpha}) \\ F(s/(1-2s)) \in (0,1) & \text{if } s'(\underline{\alpha}) < s < s'(\overline{\alpha})) \\ 0 & \text{if } s \leq s'(\underline{\alpha}) \end{cases}.$$

In a perfect Bayesian equilibrium the seller offers

$$(4) \qquad s^* \begin{cases} = 0.5 & \text{if } \beta_1 > 0.5 \\ \in [s'(\overline{\alpha}), 0.5] & \text{if } \beta_1 = 0.5 \\ \in (s'(\underline{\alpha}), s'(\overline{\alpha})] & \text{if } \beta_1 < 0.5 \end{cases}$$

Proof: If $s \geq 0.5$ the utility of a buyer from accepting s is $U_2(s) = s - \beta_2(2s-1)$, which is always positive for $\beta_2 < 1$ and thus better than a rejection which yields a payoff of 0. The point is that the buyer can achieve equality only by destroying the entire surplus which is very costly to him

if $s \geq 0.5$, i.e., if the inequality is to his advantage. For $s < 0.5$ a buyer accepts the offer only if the utility from acceptance, $U_2(s) = s - \alpha_2(1-2s)$, is nonnegative which is the case only if s exceeds the acceptance threshold

$$s'(\alpha_2) \equiv \alpha_2/(1+2\alpha_2) < 0.5$$

Therefore, from the perspective of the seller an offer s to a randomly selected buyer will be accepted with probability $F(s/(1-2s))$ which is equal to 1 if $s \geq \overline{\alpha}/(1+2\overline{\alpha})$ and equal to 0 if $s \leq \underline{\alpha}/(1+2\underline{\alpha})$.

At stage 1 a seller never offers $s > 0.5$. This would reduce his monetary payoff as compared to an offer of $s=0.5$, which would also be accepted with certainty and which would yield perfect equality. Therefore, a seller's utility can be written as

(5) $$U_1(s) = [(1-s) - \beta_1(1-2s)]\, F(s/(1-2s)).$$

If $\beta_1 > 0.5$ his utility is strictly increasing in s for all $s \leq 0.5$. This is the case where the seller prefers to share his resources rather than to maximize his own monetary payoff. If $\beta_1 = 0.5$ he is just indifferent between giving one Dollar to the buyer and keeping it to himself, i.e., he is indifferent between all offers $s \in [s'(\overline{\alpha}), 0.5]$, but he would not offer $s < s'(\overline{\alpha})$ which would be rejected with positive probability. If $\beta_1 < 0.5$, the seller strictly prefers to keep the money to himself. In this case he may offer $s < s'(\overline{\alpha})$ even if this is rejected with positive probability, but he will never offer $s \leq s'(\underline{\alpha})$. Hence, in this case there exists an optimal offer $s \in (s'(\underline{\alpha}), s'(\overline{\alpha})]$. *Q.E.D.*

Proposition 1 accounts for many of the above mentioned facts. It shows that there are no offers above 0.5, that offers of 0.5 are always accepted, and that very low offers, i.e., those below the lowest acceptance threshold $s'(\underline{\alpha})$, are rejected with probability one. Furthermore, the probability of acceptance, $F(s/(1-2s))$, is increasing in s for $s < s'(\overline{\alpha}) < 0.5$. Note also that the acceptance threshold $s'(\alpha_2) = \alpha_2/(1+2\alpha_2)$ is nonlinear and has some intuitively appealing properties. It is increasing and strictly concave in α_2 and it converges to 0.5 if $\alpha_2 \to \infty$. Furthermore, relatively small values of α_2 already yield relatively large thresholds. For example, $\alpha_2 = 1/3$ implies $s'(\alpha_2) = 0.2$ and $\alpha_2 = 0.75$ implies $s'(\alpha_2) = 0.3$.

Consider now different beliefs of the seller about the "weakness" of the population of buyers. Let $w \in [\underline{w}, \overline{w}]$ be a parameter that characterizes $F(\alpha_2, w)$ in the sense that a higher value of w represents a "weaker" distribution of types. More precisely, $w_1 > w_2$ implies that $F(\alpha_2, w_2)$ dominates $F(\alpha_2, w_1)$ in the sense of first order stochastic dominance. We assume that $F(\alpha_2, w)$ is twice continuously differentiable in both arguments.

Corollary 1: Suppose that $\beta_I < 0.5$ and that for any $w \in [\underline{w}, \overline{w}]$ there exists a unique optimal offer $s^*(w)$ that maximizes the expected payoff of the seller. If $s^*(w) < s'(\overline{\alpha})$ then an increase in w strictly reduces $s^*(w)$, i.e., the seller will make a lower offer to a weaker population.

Proof: Since we assumed that $s^*(w)$ is unique, any $s^*(w) < s'(\overline{\alpha})$ is uniquely characterized by

(6) $$dU_I/ds \equiv U_s = (-1 + 2\beta_I)F(.) + [(1 - s) - \beta_I (1-2s)]F_\alpha(.) = 0$$

Implicitly differentiating (6) yields

(7) $$ds^*/dw = - (1/U_{ss})(-1 + 2\beta_I)F_w < 0.$$

Q.E.D.

Corollary 1 accounts for the fact that weakening the power of buyers, whatever its cause may be, induces some sellers with $\beta_I < 0.5$ to make less generous offers.

In Section V we go beyond the predictions implied by Proposition 1. There we ask whether there is a distribution of preferences that can explain not just the major facts of the ultimatum game but also the facts in market and cooperation games which will be discussed in the next sections.

III.2. Market Game with Proposer Competition

It is a well established experimental fact that a broad class of market games has the empirical property to quickly converge to the competitive equilibrium. (Smith 1982, Davis and Holt 1993). By the competitive equilibrium we mean a price at which the supply and demand curve, that have been derived from *purely* selfish preferences, intersect. For our purposes, the interesting fact is that convergence to the competitive equilibrium can be observed even if that equilibrium is very "unfair" by almost any conceivable definition of fairness, i.e., if all of the gains from trade are reaped by one side of the market. This empirical feature of competition can be demonstrated in a simple market game in which many sellers want to sell one unit of a good to a single buyer who demands only one unit of the good.[9]

Such a game has been implemented in four different countries by RPOZ (1991): Suppose that there are $n - 1$ sellers who simultaneously propose a share $s_i \in [0,1]$, $i \in \{1, ..., n-1\}$, to the buyer. The buyer has the opportunity to accept or reject the *highest* offer $\overline{s} = max_i \{s_i\}$. If there

[9]We deliberately restrict our attention to simple market games for two reasons: (i) The potential impact of inequality aversion can be seen most clearly in such simple games. (ii) They allow for an explicit game-theoretic analysis. In particular, it is easy to establish the identity between the competitive equilibrium and the subgame perfect equilibrium outcome in these games. Notice that some experimental market games, like e.g. the continuous double auction as developed by Smith (1962), have such complicated strategy spaces that no complete game-theoretic analysis is yet available. For attempts in this direction see Friedman and Rust (1993).

are several sellers who offered $\bar{s}$ one of these sellers is randomly selected with equal probability. If the buyer rejects $\bar{s}$ no trade takes place and all players receive a monetary payoff of zero. If the buyer accepts $\bar{s}$ her monetary payoff is $\bar{s}$ and the successful seller earns $1 - \bar{s}$ while unsuccessful sellers earn zero.[10] If players are only concerned about their monetary payoffs this market game has a straightforward solution: The buyer accepts any $\bar{s} > 0$. Hence, for any $s_i \leq \bar{s} < 1$ there exists an $\varepsilon > 0$ such that seller i can strictly increase his monetary payoff by offering $\bar{s} + \varepsilon < 1$. Therefore, any equilibrium candidate must have $\bar{s} = 1$. Furthermore, in equilibrium a seller i who proposed $s_i = 1$ must not have an incentive to lower his offer. Thus, there must be at least one other player j who proposes $s_j = 1$, too. Hence, there is a unique subgame perfect equilibrium outcome in which at least two sellers make an offer of one, and the buyer reaps all gains from trade.[11]

RPOZ have implemented a market game in which nine players simultaneously proposed s_i while one player accepted or rejected $\bar{s}$. Experimental sessions in four different countries have been conducted. The empirical results provide ample evidence in favor of the above prediction. After a few periods the subgame perfect equilibrium outcome was reached in each experiment in each of the four countries. To what extent can our model explain this observation?

Proposition 2: Suppose that the utility functions of the players are given by (1). For any parameters (α_i, β_i), $i \in \{1, \ldots, n\}$, there is a *unique* subgame perfect equilibrium outcome in which at least two sellers offer $s = 1$ which is accepted by the buyer.

The formal proof of the proposition is relegated to the Appendix, but the intuition is quite straightforward. Note first that, for similar reasons as in the ultimatum game, the buyer must accept any $\bar{s} \geq 0.5$. Suppose that he rejects a "low" offer $\bar{s} < 0.5$. This cannot happen on the equilibrium path either since in this case seller i can improve his payoff by offering $s_i = 0.5$ which is accepted with probability 1 and gives him a strictly higher payoff. Hence, on the equilibrium path $\bar{s}$ must be accepted. Consider now any equilibrium candidate with $\bar{s} < 1$. If there is one player i offering $s_i < \bar{s}$, then this player should have offered slightly more than $\bar{s}$. There will be inequality anyway, but by winning the competition player i can increase his own monetary payoff and he can turn the inequality to his advantage. A similar argument applies if all players offer $s_i = \bar{s} < 1$. By slightly increasing his offer player i can increase the probability of winning the competition from $1/(n-1)$ to 1. Again, this increases his expected monetary payoff and it turns the inequality towards the other sellers to his advantage. Therefore, $\bar{s} < 1$ cannot be part of a subgame perfect equilibrium. Hence, the only equilibrium candidate is that at least two

[10] In the RPOZ-experiments those who could propose a division of the gains from trade were called buyers. However, from a game theoretic point of view this is not essential. The crucial feature is that the "proposers" are competing against each other.

[11] Note that there are many subgame perfect equilibria in this game. As long as two sellers propose $s = 1$ any offer distribution of the remaining sellers is compatible with equilibrium.

sellers offer $\bar{s} = 1$. This is a subgame perfect equilibrium since all sellers receive a payoff of 0 and no player can change this outcome by changing his action.

Proposition 2 shows that inequality aversion does not affect the subgame perfect equilibrium outcome relative to the prediction of the standard model that assumes purely selfish preferences. It demonstrates the power of competition. Even if all players strongly dislike inequality, competition makes it impossible for any of them to enforce an egalitarian outcome. The proposition also shows that competition between proposers renders the distribution of preferences completely irrelevant. It does not matter for the outcome whether there are many or only a few subjects who exhibit strong inequality aversion. By the same token it also does not matter whether subjects know or do not know the prevailing amount of inequality aversion. The outcome of the game is in a sense "preference-free", i.e., the institution of proposer competition alone shapes the equilibrium outcome. This also explains why markets in *all four* countries quickly converged to the competitive outcome even though the results of the ultimatum games, that have also been conducted in these countries, are consistent with the view that the distribution of preferences differs across countries. [12]

III. 3. Market Game with Responder Competition

In this section we apply our model of inequality aversion to a market game for which it is probably too early to speak of well established stylized facts since only one study with a relatively small number of independent observations (Güth, Marchand, Rulliere 1997) has been conducted so far. The game concerns a situation in which many buyers compete for a single indivisible good that can be delivered by only one seller. Since this situation can be considered as the prototype of a supply monopoly it is interesting to see how inequality aversion affects the behavior of the monopolist and the competing buyers. The rules of the game are as follows. The seller, which we denote as player 1, is the price maker, i.e. he proposes a share $s \in [0,1]$ to the buyers. There are $2,...,n$ buyers who observe s and decide simultaneously whether they accept or reject s. Then a random draw selects with equal probability one of the accepting buyers. In case that all buyers reject s all players receive a monetary payoff of zero. In case of acceptance the seller receives $1-s$ and the randomly selected buyer gets paid s. All other buyers receive zero. Note that in this game there is competition in the second stage of the game ("responder" competition) while in Section III.2 we had competing players in the first stage ("proposer" competition).

The prediction of the standard model with purely selfish preferences for this game is again straightforward. Buyers accept any positive s and are indifferent between accepting and rejecting $s=0$. Therefore, there is a unique subgame perfect equilibrium outcome in which the

[12] Rejection rates in Slovenia and the USA were significantly higher than rejection rates in Japan and Israel.

seller offers $s=0$ which is accepted by at least one buyer.[13] The results of GMR (1997) show that the standard model captures the regularities of this game rather well. The acceptance thresholds of buyers quickly converged to very low levels. Although the game was repeated only five times, in the final period the *average* acceptance threshold is well below 5% of the available surplus with 71% of the buyers stipulating a threshold of exactly zero and 9% a threshold of $s' = 0.02$.[14] Likewise, in period five the average offer declined to 15% of the available gains from trade. In view of the fact that sellers had not been informed about buyers' previous acceptance thresholds such low offers are remarkable.[15] In the final period *all* offers were below 25% while in the ultimatum game such low offers are very rare.

To what extent is this apparent willingness to make and to accept extremely low offers compatible with inequality aversion? As the following propositions show inequality aversion can account for the above regularities.

Proposition 3: There exists an equilibrium in which all buyers accept any $s \geq 0$ and the seller offers $s = 0$ if and only if $\beta_i < (n-1)/n$.

The full proof of this proposition is given in the appendix. To see the intuition for it note first that all buyers must accept any offer $s \geq 0.5$ and that the seller will never make such an offer for similar reasons as in the ultimatum game. Next consider any offer s with $0 \leq s < 0.5$. Suppose that there is at least one buyer i who accepts this offer. In this case there will be inequality anyway, so buyer $j \neq i$ should accept s as well. This increases his expected monetary payoff and it may turn some of the disadvantageous inequality vis à vis buyer i into inequality to his advantage. Moreover, for $s > 0$ acceptance may also reduce the amount of disadvantageous inequality vis à vis the seller. Thus, for any $s \geq 0$ there is a continuation equilibrium where s is accepted by all buyers. But then it is easy to see that there is also an equilibrium where the seller offers $s = 0$ which is accepted by everybody.

However, the seller refrains from making a very low offer if he cares very much about inequality to his advantage. If there are n players altogether, than giving away one Dollar to one of the buyers reduces inequality by $1 + [1/(n-1)] = n/(n-1)$ Dollars. Thus, if the nonpecuniary gain from this reduction in inequality, $\beta_i[n/(n-1)]$, exceeds the cost of 1, player 1 prefers to give

---

[13] In the presence of a smallest money unit there exists an additional, slightly different equilibrium outcome: The seller proposes $s = \varepsilon$ which is accepted by all the buyers. To support this equilibrium all buyers have to reject $s = 0$. We assume, however, that there is no smallest money unit.

[14] The gains from trade were 50 French Francs. Before observing the offer s each buyer stated an acceptance threshold. If s was above the threshold the buyer accepted the offer, if it was below she rejected s. The advantage of eliciting acceptance thresholds is that the experimenter gains more information about buyers' behavior. The potential disadvantage is that asking for acceptance thresholds may inflate rejection rates because buyers are more willing to psychologically commit themselves to reject low offers. This renders the low acceptance thresholds even more remarkable.

[15] Due to the gap between acceptance thresholds and offers we conjecture that the game had not yet reached a stable outcome after five periods. The strong and steady downwards trend in all previous periods also indicates that a steady state had not yet been reached. Recall that the market game of RPOZ (1991) was played for ten periods.

away money to one of the buyers. Recall that in the bilateral ultimatum game the seller proposed an equal split if $\beta_I > 0.5$. An interesting aspect of our model is that an increase in the number of buyers renders $s = 0.5$ less likely because it increases the threshold β_I has to pass.

Proposition 3 shows that competition among buyers ensures the existence of an equilibrium in which all the gains from trade are reaped by the seller irrespective of the prevailing amount of inequality aversion. However, as we will see next, there may also exist other equilibria in which $s > 0$ is offered.

Proposition 4: Suppose $\beta_I < (n-1)/n$. Then the highest offer s that can be sustained in a subgame perfect equilibrium is given by

$$\bar{s} = min_{i\in\{2,\ldots,n\}}\left\{\frac{\alpha_i}{(1-\beta_i)(n-1)+2\alpha_i+\beta_i}\right\} < \frac{1}{2}. \tag{8}$$

Proof: See Appendix.

Clearly, a positive s can be sustained in a subgame perfect equilibrium only if all buyers threaten to reject any $s' < s$. Why is this threat credible? Suppose that $s < 0.5$ has been offered and that this offer is being rejected by all other responders $j \neq i$. In this case player i can enforce an egalitarian outcome by rejecting the offer as well. Rejecting reduces not only the inequality towards the other buyers but also the disadvantageous inequality towards the seller. Therefore, buyer i is willing to reject this offer if nobody else accepts it and if the offer is sufficiently small, i.e., if the disadvantageous inequality towards the seller is sufficiently large. More formally, given that all other buyers reject, buyer i prefers to reject as well if and only if the utility of acceptance obeys

$$s - \frac{\alpha_i}{n-1}(1-2s) - \frac{n-2}{n-1}\beta_i s \leq 0. \tag{9}$$

This is equivalent to

$$s \leq s_i \equiv \frac{\alpha_i}{(1-\beta_i)(n-1)+2\alpha_i+\beta_i}. \tag{10}$$

Thus, an offer $s > 0$ can be sustained if and only if (8) holds for *all* buyers.

It is interesting to note that the highest sustainable offer does not depend on the full distribution of the parameters α_i and β_i but only on the inequality aversion of the buyer with the lowest acceptance threshold s_i'. In particular, if there is only one buyer with $\alpha_i = 0$, Proposition 4 implies that there is a unique equilibrium outcome with $s=0$. Furthermore, the acceptance

threshold is decreasing with n. Thus, the model makes the intuitively appealing prediction that for $n \to \infty$ the highest sustainable equilibrium offer converges to zero whatever the prevailing amount of inequality aversion.

IV. Cooperation and Retaliation - Cooperation Games

In the previous section we have shown that inequality aversion can account for the relatively "fair" outcomes in the bilateral ultimatum game as well as for the rather "unfair" or "competitive" outcomes in games with seller or buyer competition. In this section we investigate the conditions under which cooperation can flourish in the presence of inequality aversion. We show that inequality aversion improves the prospects for voluntary cooperation relative to the predictions of the standard model. In particular, we show that there is an interesting class of conditions under which the selfish model predicts complete defection while in our model there exist equilibria in which everybody cooperates fully. Under some conditions, however, the predictions of our model coincide with the predictions of the standard model.

Consider the following public good game. There are $n \geq 2$ players who decide simultaneously on their contribution levels $g_i \in [0,y]$, $i = \{1,...,n\}$, to the public good. Each player has an endowment of y. The monetary payoff of player i is given by

$$x_i(g_1,...,g_n) = y - g_i + a \sum_{j=1}^{n} g_j, \quad 1/n < a < 1, \tag{11}$$

where a denotes the constant marginal return to the public good $G \equiv \sum_{j=1}^{n} g_j$. Since $a < 1$, a marginal investment into G causes a monetary loss of $(1-a)$, i.e., the dominant strategy of a completely selfish player is to choose $g_i = 0$. Thus, the standard model predicts $g_i = 0$ for all $i = \{1,...,n\}$. However, since $a > 1/n$, the aggregate monetary payoff is maximized if each player chooses $g_i = y$.

Consider now a slightly different public good game that consists of two stages. At stage 1 the game is identical to the previous game. At stage 2 each player i is informed about the contribution vector $(g_1,...,g_n)$ and can simultaneously impose a punishment on the other players, i.e., player i chooses a punishment vector $p_i = (p_{i1},...,p_{in})$ where $p_{ij} \geq 0$ denotes the punishment player i imposes on player j. The cost of this punishment to player i is given by $c \sum_{j=1}^{n} p_{ij}$, $0<c<1$. Player i may, however, also be punished by the other players which generates an income loss to i of $\sum_{j=1}^{n} p_{ji}$. Thus, the monetary payoff of player i is given by

$$x_i(g_1,...g_n, p_1,...,p_n) = y - g_i + a \sum_{j=1}^{n} g_j - \sum_{j=1}^{n} p_{ji} - c \sum_{j=1}^{n} p_{ij}, \tag{12}$$

What does the standard model predict for the two-stage game? Since punishments are

costly a selfish player will never punish. Therefore, if selfishness and rationality are common knowledge each player knows that the second stage is completely irrelevant. As a consequence, players have exactly the same incentives at stage 1 as they have in the one-stage game without punishments, i.e., each player's best choice is given by $g_i = 0$.

To what extent are these predictions of the standard model consistent with the data from public good experiments. For the one-stage-game there are, fortunately, a large number of experimental studies. They investigate the contribution behavior of subjects under a wide variety of conditions. Since it is unreasonable to expect that subjects jump to an equilibrium in their first interaction we concentrate in the following on the *final* period behavior of subjects who had the opportunity to learn in iterated trials (see Table 2).[16]

The striking fact revealed by Table 2 is that in the final period the vast majority of subjects plays the equilibrium strategy of complete free-riding. If we average over all studies 73 percent of all subjects choose $g_i = 0$ in the final period. It is also worth mentioning that in addition to those subjects who play *exactly* the equilibrium strategy there is very often a nonnegligible fraction of subjects who play "close" to the equilibrium. In view of the facts presented in Table 2 it seems fair to say that the standard model "approximates" the choices of a big majority of subjects very well.

However, if we turn to the public good game with punishment there emerges a radically different picture although the standard model predicts the same outcome as in the one-stage game. Figure 2 shows the distribution of contributions in the final period of the two-stage game conducted by Fehr and Gächter (1996). Note that the *same subjects* generated the distribution in the game without and in the game with punishment.[17] Whereas in the game without punishment most subjects play close to complete defection a strikingly large fraction of roughly 75 % cooperates *fully* in the game with punishment. Fehr and Gächter report that lower contribution levels are associated with higher received punishments. Thus, defectors do not gain from free-riding because they are being punished.

[16] In some of these studies the group composition was the same for all T periods (partner condition). In others the group composition randomly changed from period to period (stranger condition). However, in the last period subjects in the partner condition play also a true one-shot public goods game. Therefore, Table 2 presents the behavior from stranger as well as from partner experiments.

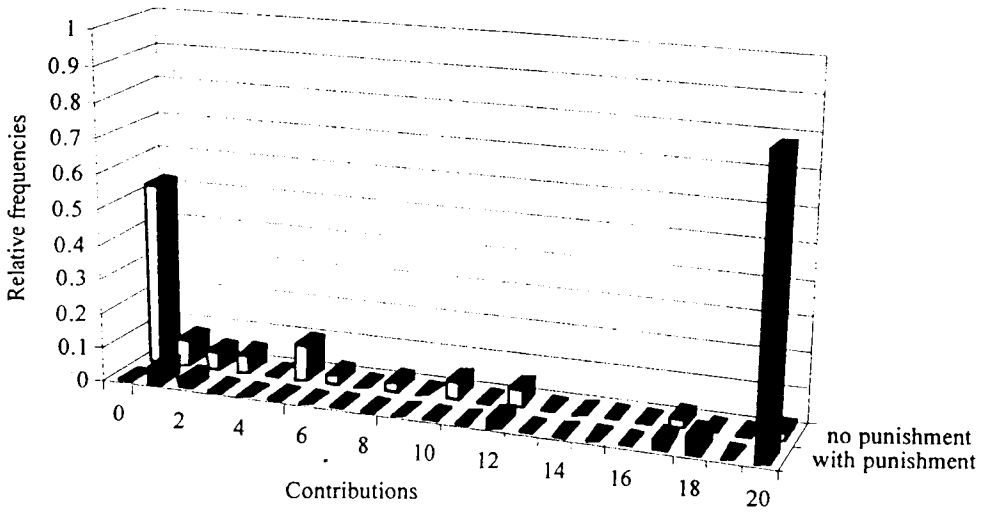
[17] Subjects in the Fehr and Gächter study participated in both conditions, i.e. in the game with punishment *and* in the game without punishment.

Table 2

Percentage of subjects who *free-ride completely* in the *final period* of a repeated public good game

Study	Country	Group Size (n)	Marginal pecuniary return (a)	Total number of subjects	Percentage of free-riders ($g_i = 0$)
Isaac and Walker 1988	USA	4 and 10	0.3	42	83
Isaac and Walker 1988	USA	4 and 10	0.75	42	57
Andreoni 1988	USA	5	0.5	70	54
Andreoni 1995a	USA	5	0.5	80	55
Andreoni 1995b	USA	5	0.5	80	66
Croson 1995	USA	4	0.5	48	71
Croson 1996	USA	4	0.5	96	65
Keser and van Winden 1996	Holland	4	0.5	160	84
Ockenfels and Weimann 1996	Germany	5	0.33	200	89
Burlando and Hey 1997	UK, Italy	6	0.33	120	66
Falkinger, Fehr, Gächter & Winter-Ebmer 1997	Switzerland	8	0.2	72	75
Falkinger, Fehr, Gächter & Winter-Ebmer 1997	Switzerland	16	0.1	32	84
Total number of subjects in all experiments and percentage of complete free-riding				1042	73

Figure 2: Distribution of contributions in the final period of public good games with and without punishment (Endowment y = 20). Source: Fehr and Gächter 1996.



The behavior in the game with punishment represents an unambiguous rejection of the standard model. This raises the question whether our model is capable of explaining both the evidence of the one-stage public good game and of the public good game with punishment. Consider the one-stage public good game first. The prediction of our model is summarized in the following proposition:

Proposition 5:

(a) If $1\text{-}a > \beta_i$ for player i, then it is a dominant strategy for that player to choose $g_i = 0$.

(b) Let k denote the number of players with $1\text{-}a > \beta_i$, $0 \leq k \leq n$. If $k/(n\text{-}1) > a/2$, then there is a unique equilibrium with $g_i = 0$ for all $i \in \{1, ..., n\}$.

(c) If $k/(n\text{-}1) < (a+\beta_j\text{-}1)/(\alpha_j+\beta_j)$ for all players $j \in \{1, ..., n\}$ with $1\text{-}a < \beta_j$, then it is an equilibrium if all k players with $1\text{-}a > \beta_j$ choose $g_i = 0$ while all other players contribute $g_j = g \in [0,y]$. Furthermore, $(a+\beta_j\text{-}1)/(\alpha_j+\beta_j) < a/2$.

The formal proof of Proposition 5 is relegated to the Appendix. To see the basic intuition for the above results consider a symmetric contribution vector where *everybody* contributes $g_i = g \in [0,y]$. Note that according to part (c) this is an equilibrium if *all* players obey the condition $1\text{-}a < \beta_i$, i. e., if k = 0. Clearly, no player wants to contribute more than g because this reduces his monetary payoff by $1\text{-}a$ and increases the payoff of the other players, i.e., it causes

disadvantageous inequality. If $1-a<\beta_i$, then player i will not reduce his contribution, either, because the monetary gain of $1-a$ from the reduction of his contribution by one Dollar is more than offset by the disutility of β_i from the (advantageous) inequality towards the other n-1 players. Hence, if this inequality holds for all players, this is indeed a Nash equilibrium. This shows that players with $1-a<\beta_i$ are „conditionally cooperative". They are willing to contribute if others contribute, too.

By a similar argument one can show that if $1-a > \beta_i$ holds for player i, then it is a dominant strategy for i to contribute nothing. Despite the existence of some players who are not going to contribute, it may be an equilibrium that all the other players contribute to the public good. In the Appendix we show how such an equilibrium can be constructed if k, the number of players with $1-a > \beta_i$, is sufficiently small. Yet, according to part (c) the condition for the preference parameters of the conditional cooperators becomes now more demanding because $k/(n-1) < (a+\beta_j-1)/(\alpha_j+\beta_j)$ has to be met. Since, in the presence of free-riders, inequality averse players suffer from inequality to their disadvantage their α_j becomes important. The higher α_j the less likely they are willing to cooperate. Or put differently: The greater the aversion against being the sucker the more difficult it is to sustain cooperation.

According to part (c) the inequality $(a+\beta_j-1)/(\alpha_j+\beta_j) < a/2$ is always met. By part (b), if $k/(n-1) > a/2$, there is a unique equilibrium in which no player contributes to the public good. Note that in almost all experiments $a \leq 1/2$ (see Table 2). Thus, if the fraction of players with $1-a>\beta_i$ is larger than $1/4$, then there is no equilibrium with positive contribution levels. This is consistent with the very low contribution levels that have been observed in these experiments. Finally, it is worthwhile to mention that the prospects for cooperation are weakly increasing with the marginal return a.

Consider now the public good game with punishment. To what extent is our model capable of accounting for the very high cooperation in the public good game with punishment. In the context of our model the crucial point is that free-riding generates a distributional advantage relative to those who cooperate. Since $c < 1$, cooperators can reduce this distributional disadvantage by punishing the free-riders. Therefore, if those who cooperate are sufficiently upset by the inequality to their disadvantage, i.e., if they have sufficiently high α's , then they are willing to punish the defectors even though this is costly to themselves. Thus, the threat to punish free-riders may be credible, which may induce potential defectors to contribute at the first stage of the game. This is made precise in the following proposition.

Proposition 6: Suppose that there is a group of n' "conditionally cooperative enforcers", $1 \leq n' \leq n$, with preferences that obey $\beta_i \geq 1 - a$ and

$$(13) \qquad c < \frac{\alpha_i}{(n-1)(1+\alpha_i) - (n'-1)(\alpha_i+\beta_i)} \quad \text{for all } i \in \{1,...,n'\},$$

whereas all other players do not care about inequality, i.e. $\alpha_i = \beta_i = 0$ for $i \in \{n'+1,...,n\}$. Then the following strategies form a subgame perfect equilibrium:

- In the first stage each player contributes $g_i = g \in [0,y]$
- If each player does so there are no punishments in the second stage. If one of the players $i \in \{n'+1,...,n\}$ deviates and chooses $g_i < g$, then each enforcer $j \in \{1, ..., n'\}$ chooses $p_{ji} = (g - g_i)/(n' - c)$ while all other players do not punish. If one of the "conditionally cooperative enforcers" chooses $g_i < g$, or if any player chooses $g_i > g$, or if more than one player deviated from g, then one Nash-equilibrium of the punishment game is being played.

Proof: See Appendix.

Proposition 6 shows that full cooperation, as observed in the experiments by Fehr and Gächter (1996), can be sustained as an equilibrium. To see how the equilibrium works consider the "conditionally cooperative enforcers". For them $\beta_i > 1 - a$, so they are happy to cooperate if *all others cooperate as well* (this is why they are called "conditionally cooperative"). In addition, condition (13) makes sure that they care sufficiently about inequality to their disadvantage. Thus they can credibly threaten to punish a defector (this is why they are called "enforcers"). The punishment is constructed such that the defector gets the same monetary payoff as the enforcers. Since this must be less than what he would have received if he had chosen $g_i = g$, a deviation is not profitable.

V. Predictions across Games

In this section we examine whether the distribution of parameters that is consistent with experimental observations in one game also is consistent with the experimental evidence from other games. We would like to stress that it is not our aim here to explain 100% of the individual choices. Rather, our objective is to offer a relatively crude test in order to see, whether there is a chance that our theory is consistent with the *quantitative* evidence from different games. We are content if our theory is consistent with the bulk of individual choices.

It is well known in experimental economics that in interactive situations one cannot expect the subjects to play an equilibrium in the first period already. Yet, if subjects have opportunities to repeat their choices and if they receive appropriate information feedback, then very often rather stable behavioral patterns, that may differ substantially from first-period-play,

emerge. Therefore, whenever available, we take the data of the final period as the facts to be explained. The questions is, whether the bulk of individual choices in the final period can be interpreted as equilibrium choices in a game in which a fraction of players suffer from inequality aversion. We start with the ultimatum game in order to roughly pin down the distribution of preferences. Then we ask whether this distribution of preferences is consistent with the observations made in the other experiments we discussed so far.

There is a large body of experimental evidence on the ultimatum game (see Table 1 above, Roth 1995, and the references given there). We first search for a distribution of α that is consistent with the behavior of the buyers in the ultimatum game. We know that for any given α_i, there exists an acceptance threshold $s'(\alpha_i) = \alpha_i/(1+2\alpha_i)$ such that player i accepts s if and only if $s \geq s'(\alpha_i)$. In all experiments there is a fraction of people who reject offers even if they are very close to an equal split. Let us (conservatively) assume that 10% of the subjects have α=4 which implies an acceptance threshold of s'=4/9=0.444. Another, typically much larger fraction of people insists on getting at least one third of the surplus, which implies a value of α which is equal to 1. These are at least 30% of the population. Note that they are prepared to give up one Dollar if this reduces the payoff of their opponent by two Dollars. Another, say, 30% of the subjects insists on getting at least one quarter, which implies $\alpha=0.5$. Finally, the remaining 30% of the subjects do not care very much about inequality and are happy to accept any positive offer ($\alpha=0$). Given this distribution of α, it is straightforward to compute the optimal offer of a seller with a given inequality parameter β. The optimal offer is given by

$$s^*(\beta) = \begin{cases} 0.5 & \text{if} \quad \beta_i > 0.5 \\ 0.\overline{4} & \text{if} \quad 0.235 < \beta_i < 0.5 \\ 0.\overline{3} & \text{if} \quad \beta_i < 0.235 \end{cases} \tag{14}$$

Note that it is never optimal to offer less than one third of the surplus, even if the seller is completely selfish. If we look at the actual offers made in the experiments, the evidence is compatible with the following: There are roughly 40% of the subjects who suggest an equal split. Another 30% offer s$\in$[0.4,0.5), while 30% offer less than 0.4. There are hardly any offers below 0.25. This gives us some information about the distribution of β in the population. The following Table 3 summarizes a distribution of α and β that is consistent with the experimental evidence from the ultimatum game.

Table 3:

Assumptions on the distribution of inequality aversion

DISTRIBUTION OF α'S AND ASSOCIATED ACCEPTANCE THRESHOLDS OF BUYERS			DISTRIBUTION OF β'S AND ASSOCIATED OPTIMAL OFFERS OF SELLERS (BASED ON (14))		
α = 0	30%	s'(0) = 0	β = 0	30%	s* = 1/3
α = 0.5	30%	s'(0.5) =1/4	β = 0.25	30%	s* = 4/9
α = 1	30%	s'(1) = 1/3	β = 0.6	40%	s* = 1/2
α = 4	10%	s'(4)=4/9			

Let us now take this distribution of preferences in order to see whether it is consistent with the observed behavior in other games. Clearly, we have no problem to explain the evidence on market games with proposer competition. Any distribution of α and β yields the competitive outcome which is observed by RPOZ in all their experiments. Similarly, in the market game with responder competition, we know from Proposition 4 that if there is at least one responder who does not care about disadvantageous inequality (i.e., $\alpha_i=0$), then there is a unique equilibrium outcome with $\bar{s}=0$. With five responders in the experiments by GMR (1997) and with the distribution of types from Table 3, the probability that there is at least one such player in each group is given by $1-0.7^5 = 83\%$. This is roughly consistent with the fact that 71% of the players accepted an offer of 0, and 9% had an acceptance threshold of $s' = 0.02$ in the final period.

Consider now the public good game. We know by Proposition 5 that cooperation can be sustained as an equilibrium outcome only if the number k of players with $a+\beta_i < 1$ obeys $k/(n-1) < a/2$. Thus, our theory predicts that there is less cooperation the smaller a which is consistent with the empirical evidence of Isaac and Walker (1988) presented in Table 2.[18] In a typical treatment $a=0.5$ and $n=4$. Therefore, if there is at least one player with $a+\beta_i < 1$, then there is a unique equilibrium with $g_i=0$ for all players. Given the distribution of preferences of Table 3, the probability that there are four people with $\beta > 0.5$ is equal to $0.4^4 = 2.56\%$. Hence, we should observe that, on average, almost all individuals fully defect. A similar result holds for most other experiments in Table 2. Except for the Isaac and Walker experiments with n = 10 a

[18] For a = 0.3 the rate of defection is substantially larger than for a = 0.75. The Issac and Walker experiments were explicitely designed to test for the effects of variations in a.

single player with $a+\beta_i < 1$ is sufficient for the violation of the necessary condition for cooperation, $k/(n-1) < a/2$. Thus in all these experiments our theory predicts that randomly chosen groups are almost never capable of sustaining cooperation. Table 2 indicates that this is not quite the case although 73% of individuals choose indeed $g_i = 0$. Thus, it seems fair to say that our model is consistent with the bulk of individual choices in this game.[19]

Finally, the most interesting experiment from the perspective of our theory is the public good game with punishment. While in the game without punishment most subjects play close to complete defection a strikingly large fraction of roughly 75% cooperates *fully* in the game with punishment. Moreover, this high cooperation rate is sustained by the actual punishments that contributing players imposed on defectors. This indicates that players with a high β-value have also a high α-value. To what extent can our model explain this phenomenon?

We know from Proposition 6 that cooperation can be sustained if there is a group of n' "conditionally cooperative enforcers" with preferences that satisfy (13) and $\beta_i \geq 1-a$. If we want to compute the probability that these conditions are met simultaneously, we have to make an assumption about the correlation between α_i and β_i. We mentioned already, that the empirical evidence suggests that these parameters are positively correlated. For concreteness we assume that the correlation is perfect. Thus, in terms of Table 3, all players with $\alpha=1$ or $\alpha=4$ are assumed to have $\beta \geq 0.6$. This is clearly not fully realistic, but it simplifies the analysis dramatically.

In the Fehr-Gächter experiment the relevant parameters are a=0.4, n=4, and (roughly[20]) c=0.2. The following table summarizes the conditions on α_i and β_i implied by Proposition 6 for a group of $n' \in \{1, ..., 4\}$ conditionally cooperative enforcers:

[19] When judging the accuracy of the model one should also take into account that there is in general a significant fraction of the people who plays close to complete free-riding in the final round. A combination of our model with the view that human choice is characterised by a fundamental randomness (McKelvey and Palfrey 1995) may explain much of the remaining 25% of individual choices. This task is, however, left for future research.

[20] The cost function in Fehr-Gächter is actually convex, so that we have to slightly simplify their model. Yet, the vast majority of actual punishments occurred where c = 0.2.

Table 4:

Conditions under which full cooperation can be sustained in the public good game with punishment

n'=1	$\alpha_i \geq 1.5$	$\beta_i \geq 0.6$
n'=2	$\alpha_i \geq 1 - 0.\bar{3}\beta_i$	$\beta_i \geq 0.6$
n'=3	$\alpha_i \geq 0.75 - 0.5\beta_i$	$\beta_i \geq 0.6$
n'=4	$\alpha_i \geq 0.6 - 0.6\beta_i$	$\beta_i \geq 0.6$

The conditions on α_i and β_i have to hold simultaneously which can only be the case if

- there is at least one player with $\alpha_i = 4$ and $\beta_i = 0.6$, or
- there are at least two players with $\alpha_i = 1$ and $\beta_i = 0.6$, or
- both.

Given the numbers of Table 3, it is easy to show that the probability that one of these cases applies is equal to 61.12%. Thus, our model is roughly consistent with the experimental evidence of Fehr and Gächter.[21]

VI. Dictator Games, Prisoner's Dilemma, Trust- and Gift Exchange Games

The preceding sections have shown that our very simple model of linear inequality aversion can explain the most important facts in ultimatum, market and cooperation games. One problem of our approach is, however, that it yields too extreme predictions in some other games, such as the "dictator game". The dictator game is a two person game in which only player 1, the "dictator", has to make a decision. Player 1 has to decide what share $s \in [0,1]$ of a given amount of money to pass on to player 2. For a given s monetary payoffs are given by $x_1 = 1 - s$ and $x_2 = s$,

[21] In this context one has to take into account that the total number of available individual observations in the game with punishment is much smaller than for the game without punishment or for the ultimatum game. Future experiments will have to show whether the Fehr-Gächter results are the rule in the punishment game or whether they exhibit unusually high (low) cooperation rates.

respectively. Obviously, the standard model predicts $s = 0$. In contrast, in the experimental study of FHSS (1994) only about 20% of subjects chose $s = 0$, 60% chose $0 < s < 0.5$, and again roughly 20% chose $s = 0.5$. In the study of Andreoni and Miller (1995) the distribution of shares is again bimodal but puts more weight at the "extremes": There approximately 40% of the subjects gave $s = 0$, 20% gave $0 < s < 0.5$, and roughly 40% gave $s = 0.5$. Shares above $s = 0.5$ were practically never observed.

Our model predicts that player 1 offers $s = 0.5$ if $\beta_1 > 0.5$ and $s = 0$ if $\beta_1 < 0.5$. Thus we should observe *only* very "fair" or very "unfair" outcomes, a prediction that is clearly refuted by the data. However, there is a straightforward solution to this problem. We assumed that the inequality aversion is piecewise *linear*. The linearity assumption was imposed in order to keep our model as simple as possible. If we allow for a utility function that is concave in the amount of advantageous inequality, there is no problem to generate optimal offers that are in the interior of $[0,0.5]$.

It is important to note that non-linear inequality aversion does not affect the qualitative results in the other games we considered. This is straightforward in market games with proposer or responder competition. Recall that in the context of proposer competition there exists a unique equilibrium outcome in which the buyer receives the whole gains from trade *irrespective of the prevailing amount of inequality aversion.* Thus, it also does not matter whether linear or nonlinear inequality aversion prevails. Likewise, under responder competition there is a unique equilibrium outcome in which the seller receives the whole surplus if there is at least one buyer who does not care about disadvantageous inequality. Obviously, this proposition holds irrespective of whether the inequality aversion is linear or not. Similar arguments hold for public good games with and without punishment. Concerning the public good game with punishment, for example, the existence of *nonlinear* inequality aversion obviously does not invalidate the existence of an equilibrium with full cooperation. It only renders the condition for the existence of such an equilibrium, i.e., condition (13), slightly more complicated.

What are the implications of our model for the one-shot prisoner's dilemma (PD)? The PD is just a special case of the public good game analysed in Section IV for $n = 2$ and $g_i \in \{0,y\}$, $i = 1,2$. Therefore, Proposition 5 applies, i.e., cooperation is an equilibrium if *both* players meet the condition $\beta_i + a > 1$. Yet, if only one player meets this condition defection of both players is the unique equilibrium. If the prisoner's dilemma game is finitely repeated and if preferences are common knowledge, a similar result obtains. Cooperation in every period is an equilibrium if $\beta_i + a > 1$ holds for both players, otherwise defection in each period is a unique subgame perfect equilibrium. This need not be the case, however, if each player does not know the inequality aversion of his opponent. Since both players gain if cooperation can be sustained for some time, a player with $\beta_i + a < 1$ may have an incentive to mimic the behavior of a player with $\beta_i + a > 1$ in the beginning of the game in order to induce his opponent to cooperate as well. This is the well

known "reputation" story of Kreps et.al. (1982).[22] From this perspective it is interesting to note that in most public good experiments in which the same subjects play against each other for a finite number of times there is a substantial amount of cooperation in the first periods, even though most subjects defect in the final period.

The prospects to sustain cooperation as an equilibrium also are much better if the game has a sequential structure. Suppose that player 1 has to contribute first. Player 2 observes this contribution and decides thereafter on his own contribution level. Suppose that preferences are common knowledge and that player 2 satisfies $\beta_2 + a > 1$. In this case player 1 knows that his contribution will be matched by player 2. Therefore, player 1 has an incentive to induce reciprocal cooperation by contributing himself even if he does not care at all about equality.

Reciprocal cooperation has also been observed in a large class of games called trust- or gift exchange games (see e. g., Fehr, Kirchsteiger and Riedl 1993; Berg, Dickhaut and McCabe 1995; Fehr, Gächter and Kirchsteiger 1997). Although these games differ in the details from our sequential public good game the basic structure is very similar: Two players have to move sequentially. If both players cooperate, they can achieve a Pareto improvement in material payoffs. However, if each player is only concerned about his own monetary payoff, cooperation cannot be an equilibrium. The empirical evidence reported in the above cited contributions clearly refutes the selfish prediction: Many first movers choose to cooperate and many second movers respond to cooperation by cooperating as well. In the context of a labor market such reciprocal cooperation leads to systematic deviations from the competitive equilibrium: Wage setting firms pay noncompetitive efficiency wages to induce workers to choose high effort levels. Our model of inequality aversion also offers a potential explanation for these observations.

VII. Summary and Discussion

The data pattern in ultimatum games, public good games with punishment as well as in dictator games, trust and gift exchange games constitute clear evidence against the standard self-interest model while the data from market games and public good games without punishment provide support for this model. We have shown that this conflicting evidence can be reconciled if a fraction of the population exhibits self-centered inequality aversion. The standard self-interest model rests on two crucial assumptions: *(i)* Subjects are fully rational. *(ii)* Subjects are only interested in their material payoffs. In principle one can give up either of these assumptions in an attempt to explain the evidence. We have chosen to relax Assumption (ii) because we believe that there is nothing complicated in, e.g., the dictator game or the subgame of the ultimatum game

[22] Note that Kreps et al. (1982) simply assume that there is some positive probability for each of the players to be "irrational" and to cooperate in each period. In contrast, our model of inequality aversion offers a more natural explanation of why some types may be willing not to defect.

where the responder has to decide upon acceptance or rejection. Likewise, it is straightforward to understand that punishing in the public good game with punishment reduces the own monetary payoff. There also is ample evidence that subjects in the first stage of the ultimatum game or public good game with punishment take into account the expected actions at the second stage.[23] Hence, we are reluctant to interpret sharing in the dictator game, rejecting positive offers in the ultimatum game or engaging in costly punishment as "irrational" or "boundedly rational" actions. Yet, if bounded rationality is ruled out we have to consider more complicated motivational structures and look for utility functions that are not only defined on a person's material well-being. We want to emphasize, however, that maintaining the rationality assumption for the explanation of relatively simple games does not imply that this is also a useful strategy for the explanation of more complicated games. For example, it is well possible that most subjects are unable to go through a long backward induction argument in a complicated multi-stage game. With regard to the games considered in this paper we doubt this, however.

To our knowledge there are three other approaches which try to account for persistent deviations from the prediction of the self-interest model in terms of different motives. One is based on the assumption of altruistic motives (Andreoni and Miller 1995), the other on the assumption that people are motivated by reciprocity (Rabin 1993, Levine 1996) while the third one is also based on a kind of inequality aversion (Bolton and Ockenfels 1997). There are many different notions of altruism, but all of them involve that the material well-being of player i has a *strictly positive* impact on the utility of player j, so j may be prepared to give up resources in order to help i. Positive reciprocity is defined as the desire to be kind to those who signal kindness through their actions or who are expected to be kind in the future. Negative reciprocity is the desire to hurt those who signal hostility through their actions or who are expected to be hostile. Rabin (1993) and Levine (1996) incorporate reciprocity into the rational actor framework although their models differ in many details.

Altruism can explain voluntary giving in dictator and public good games. But it is inconsistent with the fact that low offers are frequently rejected in the ultimatum game. It also seems difficult to reconcile the difference between the very unfair outcomes in some market and the rather fair outcomes in ultimatum games with pure altruism. Nor can it explain the behavioral differences between public good games with and without punishment. Our model allows for

---

[23]For example, the results of FHSS (1994) and Zamir and Winter (ZW 1996), which we discussed in Section III.1, show that proposers quickly adapt their offers when the responders' degree of "toughness" varies. Remember that ZW show that the frequency of high offers increases if computers play "tough" and decreases if computers accept relatively low offers. Interestingly, they also show that responders do *not* change their acceptance behavior if the degree of "toughness" of computerized proposers changes. The first result of ZW suggests, that proposers rationally take into account responders likely reaction. The stability of responder behavior in view of changing degrees of proposers' "toughness" suggests that a more deeply rooted motive drives responders' behavior. With regard to the public good game with punishment, Fehr and Gächter (1996) provide evidence indicating that subjects were well aware that defections triggered punishments so that it was in their self-interest to cooperate.

altruism but it limits altruism to the domain of advantageous inequality. Thus we rule out that the "poor" are altruistic towards to the "rich". Instead, we assume that in the domain of disadvantageous inequality the material payoff of player *i* has a *strictly negative* impact on player *j*'s utility. As we have seen in the previous section, if altruism is restricted to the domain of advantageous inequality and if the aversion against disadvantageous inequality is added the above phenomena can be explained.

Inequality aversion can also explain the phenomenon of "conditional cooperation", i.e., that people are often willing to be cooperative or to make a transfer to another person if this person is willing to do so as well. For example, Proposition 5 has shown that positive contributions can be sustained as an equilibrium outcome if sufficiently many other players are going to contribute. Of course, conditional cooperation can also be explained by reciprocity. However, models of reciprocity have difficulties to explain voluntary transfers when there is no chance for the responder to reciprocate, as, e.g., in the dictator game. Since the second player in the dictator game cannot respond at all there is no reason to give for somebody who is motivated solely by reciprocity.

The approach by Bolton and Ockenfels (BO, 1997) is similar to our model although there are some important differences in the details: (i) In BO people compare their material payoff to the material *average* payoff of the group. (ii) BO do not assume that inequality aversion is asymmetric. We believe, however, that the data are incompatible with symmetric inequality aversion. For example, it seems impossible to explain the high fraction of offers between 0.4 and 0.5 in the ultimatum game without invoking the assumption that for some people $\alpha_2 > 1$ (i. e., that $s'(\alpha_2) > 1/3$).[24] (iii) In BO the marginal disutility of small deviations from equality is zero. Therefore, if subjects are non-satiated in their own material payoff they will never *propose* an equal split in the dictator game. Likewise, they will - in case of nonsatiation in material payoffs - never propose an equal split in the ultimatum game unless $\alpha_2 = \infty$ for sufficiently many responders. Typically the modal offer in most ultimatum game experiments is, however, the equal split. In addition, assumption (iii) implies that complete free-riding is the unique equilibrium in the public good game without punishment for *all* $a < 1$ and *all* $n \geq 2$. The BO approach thus rules out the possibility that only a fraction of the people cooperates.

To what extent is there evidence that allows to discriminate between the reciprocity approach and our model? We know of three experimental studies that speak to this question: Blount (1995) conducted ultimatum games, Charness (1996) implemented gift exchange games and Bolton, Brandts and Katok (1996) conducted sequential public good games. Note that all these studies implemented a sequential move structure so that the second mover could respond to the actions taken by the first mover. The crucial idea in these studies is that in one condition

[24] Alternatively, one could assume that roughly 70 percent of the proposers have $\beta_i > 0.5$ which is, however, obviously not true. With this assumption one should also observe 70 percent equal splits in the dictator game.

(the „usual" condition) the first mover could signal an intention by taking an action that is more or less fair. In a second condition the experimenter removes the first movers option to signal an intention by forcing the first mover to take a particular action. Whenever the first mover cannot signal an intention there is no basis for rewarding or punishing intentions and, as a consequence, the second mover should behave completely selfish according to the reciprocity approach. In contrast, a second mover who is motivated by inequality aversion is still ready to reject low offers in the ultimatum game and to behave reciprocally cooperative in the gift exchange or sequential public good game. This follows from the fact that for inequality averse players intentions are behaviorally irrelevant - only final payoff consequences count.

In all three studies rejections or reciprocal cooperation occur also when first movers *cannot* signal an intention. This is consistent with our approach and cannot be explained by pure reciprocity. A comparison of the behavior across conditions shows that in one of the no-intention-conditions rejection rates are substantially lower in the ultimatum game (Blount 1995). Yet, in Charness (1996) reciprocal cooperation is only slightly lower in the no-intention-condition while in Bolton, Brandts and Katok (1996) it is even higher albeit not significantly higher. Thus, all of the above mentioned studies are consistent with the view that inequality aversion shapes bargaining, gift-giving and contribution behavior. Some of them also assert a role for reciprocity, while the last study denies such a role. These studies also show that it is possible to deal with preference questions in a disciplined scientific manner. They invalidate the frequently expressed view that changing the utility function necessarily is an arbitrary and "ad hoc" enterprise that has no scientific merit.

In our view the study of Bolton, Brandts and Katok is probably not the last word in the debate on reciprocity versus inequality aversion. The idea that an action is not only judged in terms of its consequences but also in terms of its intentions, which lies at the heart of the reciprocity approach, certainly has some appeal. It is, for example, impossible to understand many decisions of courts without accepting the notion that intentions matter. It matters whether somebody is forced, by the threat of loosing his life, to kill an innocent person or if somebody kills the innocent voluntarily. However, having said this, the final pecuniary consequences and, in particular, the distribution of material payoffs, also seem to be crucial to understand actual human behavior. We hope that our contribution induces new experiments that shed additional light on these issues.

APPENDIX

Proof of Proposition 2: Consider an offer $\bar{s} \geq 0.5$. The buyer (player n) accepts this offer if and only if

(A1) $$\bar{s} - \frac{1}{n-1}\beta_n(\bar{s}-1+\bar{s}) - \frac{n-2}{n-1}\beta_i(\bar{s}-0) \geq 0$$

which is equivalent to

(A2) $$(n-1)\bar{s} \geq \beta_i(n\bar{s}-1)$$

Since $\beta_i \leq 1$, this inequality clearly holds if

(A3) $$(n-1)\bar{s} \geq n\bar{s} - 1$$

which must be the case since $s \leq 1$. Hence, the buyer prefers to accept $\bar{s} \geq 0.5$.

The buyer must also accept any $\bar{s} < 0.5$ in equilibrium. Suppose not. Then seller *i* could deviate and offer $s_i = 0.5$ which will be accepted with probability one and which yields

(A4) $$U_i(s=0.5) = \frac{1}{2} - \frac{n-2}{n-1}\beta_i\frac{1}{2}$$

$$= \frac{1}{4(n-1)}\left[(2-\beta_i)(n-1)+\beta_i\right] \geq 0$$

to player *i*. Hence, such a deviation would be profitable.

It cannot be an equilibrium that $\bar{s} < 1$. Suppose not. Two cases have be distinguished. Suppose first that there is at least one player j who is supposed to offer $s_j < \bar{s}$. This player could deviate and offer $\tilde{s}_j = \bar{s} + \varepsilon < 1$. Such a deviation is profitable if and only if

(A5) $$1-\bar{s}-\varepsilon - \frac{\alpha_j}{n-1}(2\bar{s}+2\varepsilon-1) - \frac{n-2}{n-1}\beta_j(1-\bar{s}-\varepsilon) > 0 - \frac{\alpha_j}{n-1}\bar{s} - \frac{\alpha_j}{n-1}(1-\bar{s})$$

(A6) $$\Leftrightarrow (1-\bar{s}-\varepsilon)(n-1) - \alpha_j(2\bar{s}+2\varepsilon-1) - (n-2)\beta_j(1-\bar{s}-\varepsilon) > -\alpha_j$$

(A7) $$\Leftrightarrow (1-\bar{s}-\varepsilon)\left[(n-1)(1-\beta_j)+\beta_j\right] > -2\alpha_j(1-\bar{s}-\varepsilon).$$

Since the left hand side is strictly positive while the right hand side is strictly negative, this deviation is profitable. Suppose now that all sellers are supposed to offer $s = \bar{s}$. If nature selects player i, his payoff is

(A8) $$1-\bar{s} - \frac{\alpha_i}{n-1}(2\bar{s}-1) - \frac{n-2}{n-1}\beta_i(1-\bar{s}) = (1-\bar{s})\left[1 + \frac{2\alpha_i}{n-1} - \frac{n-2}{n-1}\beta_i\right] - \frac{\alpha_i\bar{s}}{n-1}$$

If nature does not select player i, his payoff is

(A9) $$-\frac{\alpha_i}{n-1}\bar{s}-\frac{\alpha_i}{n-1}(1-\bar{s})=-\frac{\alpha_i}{n-1}$$

Since $\beta_i \leq 1$, (A8) is strictly bigger than (A9) for all $\bar{s}<1$. But then there exists an $\varepsilon>0$ sufficiently small, such that player i could offer $\bar{s}+\varepsilon$, receive $1-\bar{s}-\varepsilon$ with certainty, such that he is strictly better off. Thus, this cannot be an equilibrium either.

Hence, the unique equilibrium candidate is a situation in which $\bar{s}=1$. It cannot be an equilibrium that only one player offers $s_i = 1$. In this case he could do strictly better by slightly lowering his offer. Therefore, it must be the case that at least two players offer $s_i = s_j = 1$. This is indeed an equilibrium since all sellers receive a monetary payoff of 0 and no player can change this outcome by changing his action. *Q.E.D.*

Proof of Proposition 3: Note first that any offer $s \geq 0.5$ will be accepted by all buyers. The argument is exactly the same as the one in the beginning of the proof of Proposition 1. The following lemma will be useful:

Lemma 1 *For any $s < 0.5$ there exists a continuation equilibrium in which everybody accepts s.*

Given that all other players accept s player i prefers to accept as well if and only if

(A10) $$s-\frac{1}{n-1}\alpha_i(1-s-s)-\frac{n-2}{n-1}\beta_i(s-0)\geq 0-\frac{1}{n-1}\alpha_i(1-s)-\frac{1}{n-1}\alpha_i s$$

which is equivalent to

(A11) $$(1-\beta_i)(n-1)+2\alpha_i+\beta_i \geq 0.$$

Since we assume $\beta_i \leq 1$, this inequality must hold.□

Consider now the seller. Clearly, it is never optimal to offer s > 0.5. Such an offer is always dominated by s = 0.5 which yields a higher monetary payoff to player 1 and less inequality. On the other hand, we know by Lemma 1 that for any $s \leq 0.5$ there exists a continuation equilibrium in which this offer is accepted by everybody. Thus, we only have to look for the optimal s from the point of view of the seller, given that s will be accepted. His payoff function is

(A12) $$U_1(s)=1-s-\frac{1}{n-1}\beta_1(1-s-s)-\frac{n-2}{n-1}\beta_1(1-s)$$

Differentiating with respect to s yields

(A13) $$\frac{dU_1}{ds}=-1+\frac{2}{n-1}\beta_1+\frac{n-2}{n-1}\beta_1$$

which is independent of s and is smaller than 0 if and only if

(A14) $$\beta_1 \leq \frac{n-1}{n}$$

Hence, if this condition holds, it is an equilibrium that the seller offers $s = 0$ which is accepted by all buyers. *Q.E.D.*

Proof of Proposition 4: The following lemma will be useful.

Lemma 2 *Suppose that $s < 0.5$ has been offered. There exists a continuation equilibrium in which this offer is rejected by all buyers if and only if*

(A15) $$s < \frac{\alpha_i}{(1-\beta_i)(n-1)+2\alpha_i+\beta_i} \forall i \in \{2,\ldots,n\}.$$

Given that all other buyers reject s, buyer i will reject s as well if and only if

(A16) $$0 \geq s - \frac{\alpha_i}{n-1}(1-2s) - \frac{n-2}{n-1}\beta_i s$$

which is equivalent to (A15). Thus, a necessary condition for a continuation equilibrium in which s is rejected by everybody is that (A15) holds for all $i \in \{2,\ldots,n\}$.

Suppose now that (A15) is violated for at least one $i \in \{2,\ldots,n\}$. Then buyer i prefers to accept s if all other buyers reject it. Suppose now that at least one other buyer accepts s. In this case buyer i prefers to accept s as well if and only if

(A17) $$s - \frac{\alpha_i}{n-1}(1-2s) - \frac{n-2}{n-1}\beta_i s \geq 0 - \frac{\alpha_i}{n-1}(1-s) - \frac{\alpha_i}{n-1}s$$

The right hand side of this inequality is smaller than 0. We know already that the left hand side is greater than 0 since (A15) is violated. Therefore, buyer i prefers to accept s as well. We conclude that if (A15) does not hold for at least one i, then at least one player will accept s.□

If $\beta_1 < \frac{n-1}{n}$ an equilibrium offer must be sustained by the threat that any smaller offer $\tilde{s}$ will be rejected by everybody. But we know from Lemma 2 that an offer $\tilde{s}$ may be rejected only if (A15) holds for all i. Thus, the highest offer s that can be sustained in equilibrium is given by (8).

Q.E.D.

Proof of Proposition 5:

(a) Suppose that $1-a > \beta_i$ for player i. Consider an arbitrary contribution vector $(g_1,\ldots,g_{i-1}, g_{i+1},\ldots,g_n)$ of the other players. Without loss of generality we relabel the players such that $i=1$ and $0 \le g_2 \le g_3 \le \ldots \le g_n$. If player 1 chooses $g_1=0$, his payoff is given by

$$U_1(g_1=0) = y + a\sum_{j=2}^{n} g_j - \frac{\beta}{n-1}\sum_{j=2}^{n} g_j \tag{A18}$$

Note first, that if all other players choose $g_j=0$, too, then $g_1=0$ is clearly optimal. Furthermore, player 1 will never choose $g_1 > max\{g_j\}$. Suppose that there is at least one player who chooses $g_j>0$. If player 1 chooses $g_1>0$, $g_1 \in [g_k, g_{k+1}]$, $k \in \{2, \ldots, n\}$, then his payoff is given by

$$\begin{aligned}
U_1(g_1>0) &= y - g_1 + ag_1 + a\sum_{j=2}^{n} g_j - \frac{\beta_1}{n-1}\sum_{j=k+1}^{n}(g_j - g_1) - \frac{\alpha_1}{n-1}\sum_{j=2}^{k}(g_1 - g_j) \\
&< y - g_1 + ag_1 + a\sum_{j=2}^{n} g_j - \frac{\beta_1}{n-1}\sum_{j=k+1}^{n}(g_j - g_1) + \frac{\beta_1}{n-1}\sum_{j=2}^{k}(g_1 - g_j) \\
&= y - g_1 + ag_1 + a\sum_{j=2}^{n} g_j - \frac{\beta_1}{n-1}\sum_{j=2}^{n} g_j + \frac{\beta_1}{n-1}(n-1)g_1 \\
&= y - (1-a-\beta_1)g_1 + a\sum_{j=2}^{n} g_j - \frac{\beta_1}{n-1}\sum_{j=2}^{n} g_j \\
&< y + a\sum_{j=2}^{n} g_j - \frac{\beta_1}{n-1}\sum_{j=2}^{n} g_j \\
&= U_1(g_1=0)
\end{aligned}$$

Hence, $g_i=0$, is indeed a dominant strategy for player i.

(b) It clearly is an equilibrium if all players contribute nothing because to unilaterally contribute more than zero reduces the monetary payoff and causes disadvantageous inequality. Suppose that there exists another equilibrium with positive contribution levels. Relabel players such that $0 \le g_1 \le g_2 \le \ldots \le g_n$. By part (a) we know that all k players with $1-a>\beta_i$ must choose $g_i=0$. Therefore $0 = g_1 = \ldots = g_k$. Consider player $l > k$ who has the smallest positive contribution level, i.e., $0 = g_{l-1} < g_l \le g_{l+1} \le \ldots \le g_n$. Player l's utility is given by

$$\begin{aligned}
U_l(g_l) &= y - g_l + ag_l + a\sum_{j=l+1}^{n} g_j - \frac{\beta_l}{n-1}\sum_{j=l+1}^{n}(g_j - g_l) - \frac{\alpha_l}{n-1}\sum_{j=1}^{l-1} g_l \\
&= y + a\sum_{j=l+1}^{n} g_j - \frac{\beta_l}{n-1}\sum_{j=l+1}^{n} g_j - (1-a)g_l + \beta_l\frac{n-l}{n-1}g_l - \alpha_l\frac{l-1}{n-1}g_l \\
&= U_l(0) - (1-a)g_l + \beta_l\frac{n-l}{n-1}g_l - \alpha_l\frac{l-1}{n-1}g_l,
\end{aligned} \tag{A19}$$

where $U_l(0)$ is the utility player l gets if he deviates and chooses $g_l=0$. Since $\alpha_l \geq \beta_l$, $l \geq k+1$, and $\beta_l < 1$, we have

(A20)
$$\begin{aligned} U_l(g_l) &\leq U_l(0) - (1-a)g_l + \beta_l \frac{n-l}{n-1} g_l - \beta_l \frac{l-1}{n-1} g_l \\ &\leq U_l(0) - (1-a)g_l + \beta_l \frac{n-2(k+1)+1}{n-1} g_l \\ &< U_l(0) - (1-a)g_l + \frac{n-2k-1}{n-1} g_l \\ &= U_l(0) - \frac{(1-a)(n-1)-(n-2k-1)}{n-1} g_l . \end{aligned}$$

Thus, if

(A21)
$$\frac{(1-a)(n-1)-(n-2k-1)}{n-1} \geq 0$$

player l prefers to deviate from the equilibrium candidate and to choose $g_l=0$. But this inequality is equivalent to

(A22)
$$\begin{aligned} &(1-a)(n-1) \geq n-2k-1 \\ \Leftrightarrow\ & a \leq 1 - \frac{n-2k-1}{n-1} \\ \Leftrightarrow\ & a \leq \frac{n-1-n+2k+1}{n-1} = \frac{2k}{n-1} \\ \Leftrightarrow\ & \frac{k}{n-1} \geq \frac{a}{2} \end{aligned}$$

which is the condition given in the proposition.

(c) Suppose that the conditions of the Proposition are satisfied. We want to construct an equilibrium in which all k players with $1-a > \beta_i$ contribute nothing, while all other $n-k$ players contribute $g \in [0,y]$. We only have to check that contributing g is indeed optimal for the contributing players. Consider some player j with $1-a < \beta_j$. If he contributes g his payoff is given by:

(A23)
$$U_j(g) = y - g + (n-k)ag - \frac{\alpha_j}{n-1} kg$$

It clearly does not pay to contribute more than g. So suppose player j reduces his contribution level by $\Delta > 0$. Then his payoff is

$$U_j(g-\Delta) = y - g + \Delta + (n-k)ag - \Delta a - \frac{\alpha_j}{n-1}k(g-\Delta) - \frac{\beta_j}{n-1}(n-k-1)\Delta$$
$$= y - g - (n-k)ag - \frac{\alpha_j}{n-1}kg + \Delta\left(1 - a + \frac{\alpha_j}{n-1}k - \frac{\beta_j}{n-1}(n-k-1)\right)$$
$$= U_j(g) + \Delta\left(1 - a + \frac{\alpha_{ji}}{n-1}k - \frac{\beta_j}{n-1}(n-k-1)\right)$$

Thus, a deviation does not pay if and only if

$$1 - a + \frac{\alpha_j}{n-1}k - \frac{\beta_j}{n-1}(n-k-1) \le 0$$

which is equivalent to

(A24)
$$\frac{k}{n-1} \le \frac{a + \beta_j - 1}{\alpha_j + \beta_j}$$

Thus, if this condition holds for *all (n-k)* players j with $1\text{-}a < \beta_j$, then this is indeed an equilibrium. It remains to be shown that $(a+\beta_j-1)/(\alpha_j+\beta_j) \le a/2$. Note that $\alpha_j \ge \beta_j$ implies $(a+\beta_j-1)/(\alpha_j+\beta_j) \le (a+\beta_j-1)/(2\beta_j)$. Furthermore

$$\frac{a + \beta_j - 1}{2\beta_j} \le \frac{a}{2} \Leftrightarrow a + \beta_j - 1 \le \beta_j a \Leftrightarrow a(1-\beta_j) \le 1 - \beta_j \Leftrightarrow a \le 1$$

which proves our claim.

Q.E.D.

Proof of Proposition 6: Suppose that one of the players $i \in \{n'+1,...,n\}$ chooses $g_i < g$. If all players stick to the punishment strategies in stage 2, then deviator i gets the same monetary payoff as each enforcer $j \in \{1,...,n'\}$. In this case monetary payoffs of i and j are given by

$$x_i = y - g_i + a[(n-1)g + g_i] - n'\frac{(g - g_i)}{(n'-c)} \tag{A25}$$

$$x_j = y - g + a[(n-1)g + g_i] - c\frac{(g - g_i)}{(n'-c)} - \frac{n'-c}{n'-c}(g_i - g_i) \tag{A26}$$

$$= y - g_j + a[(n-1)g+g_i] - (n' - c + c)\frac{(g - g_i)}{(n'-c)} = x_i$$

Thus, given the punishment strategy of the enforcers, deviators cannot get a higher payoff than the enforcers. However, they get a strictly lower payoff than the non-enforcers who did not deviate. We now have to check that the punishment strategies are credible, i.e. that an enforcer cannot gain from reducing his p_{ij}. If an enforcer reduces p_{ij} by ε he saves $c\varepsilon$ and experiences less disadvantageous inequality relative to those $(n - n' - 1)$ players who chose g but do not punish. This creates a non-pecuniary utility gain of $[\alpha_i(n - n' - 1)c\varepsilon]/(n - 1)$. On the other hand, the enforcer also has non-pecuniary costs because he experiences now disadvantageous inequality relative to the defector and a distributional advantage relative to the other $(n' - 1)$ enforcers who punish fully. The latter generates a utility loss of $\beta_i(n' - 1)c\varepsilon/(n - 1)$ whereas the former reduces utility by $\alpha_i(1 - c)\varepsilon/(n - 1)$. Thus the loss from a reduction in p_{ij} is greater than the gain if

$$\frac{1}{(n-1)}[\alpha_i(1 - c)\varepsilon + \beta_i(n' - 1)c\varepsilon] > c\varepsilon + \alpha_i(n - n' - 1)\frac{c\varepsilon}{(n-1)} \tag{A27}$$

holds. Some simple algebraic manipulations show that condition (A27) is equivalent to condition (13). Hence, the punishment is credible.

Consider now the incentives of one of the enforcers to deviate in the first stage. Suppose he reduces his contribution by $\varepsilon > 0$. Ignoring possible punishments in the second stage for a moment, player i gains $(1-a)\varepsilon$ in monetary terms but incurs a non-pecuniary loss of $\beta_i\varepsilon$ by creating inequality to all other players. Since $1-a < \beta_i$ by assumption this deviation does not play. If his defection triggers punishments in the second stage, then this reduces his monetary payoff which cannot make him better off than he would have been if he had chosen $g_i = g$. Hence, the enforcers are not going to deviate at stage 1 either. It is easy to see that choosing $g_i > g$ can not be profitable for any player either, since it reduces the monetary payoff and increases inequality.

Q.E.D.

References

Andreoni James (1988); „Why Free Ride? - Strategies and Learning in Public Goods Experiments", Journal of Public Economics, Vol. 37, 291-304.

Andreoni James (1995a); „Cooperation in Public-Goods Experiments: Kindness Or Confusion"; American Economic Review, Vol. 85, 891-904.

Andreoni, James (1995b); „Warm Glow versus Cold Prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments", Quarterly Journal of Economics, Vol. 110, 1-21.

Andreoni, James and Miller, John H. (1996); „Giving according to GARP: An Experimental Study of Rationality and Altruism", SSRI Working Paper, Univ. of Wisconsin, Madison.

Babcock, Linda, George Loewenstein, Samuel Issacharoff, and Colin Camerer (1995); „Biased Judgements of Fairness in Bargaining", American Economic Review, Vol. 85,1337-1343.

Berg,Joyce, John Dickhaut and Kevin McCabe (1995); "Trust Reciprocity and Social History", Games and Economic Behavior, Vol. 10, 122-142.

Blount, Sally (1995); „When Social Outcomes aren't Fair: The Effect of Causal Attributions on Preferences", Organizational Behavior and Human Decision Processes, Vol. 63, No. 2, 131-144.

Bolton, Gary E., Jordi Brandts, and Elena Katok (1997); „A Simple Test of Explanations for Contributions in Dilemma Games", Discussion Paper, Penn State University.

Bolton, Gary E. and Axel Ockenfels (1997); „A Theory of Equity, Reciprocity and Competition", Discussion Paper, Penn State University.

Burlando, Roberto and John D. Hey (1997); „Do Anglo-Saxons free-ride more?", Journal of Public Economics, Vol. 64, 41-60.

Camerer, Colin and Richard Thaler (1995); "Ultimatums, Dictators, and Manners", Journal of Economic Perspectives, vol. 9, 209-219.

Cameron, Lisa (1995); "Raising the Stakes in the Ultimatum Game: Experimental Evidence from Indonesia", Discussion Paper, Princeton University.

Charness, Gary (1996); "Attribution and Reciprocity in a Labor Market: An Experimental Investigation", Mimeo, University of California at Berkeley.

Croson, Rachel T. A. (1995); „Expectations in Voluntary Contributions Mechanisms", Discussion Paper, Wharton School, Univ. of Pennsylvania.

Croson, Rachel T. A. (1996); „Partners and Strangers Revisited", Economic Letters, Vol 53, 25-32.

Davis, Douglas and Charles Holt (1993); Experimental Economics, Princeton University Press, Princeton, New Jersey.

Dawes, Robyn M. and Richard Thaler (1988); "Cooperation", Journal of Economic Perspectives, Vol.2, No. 3, 187-197.

Falkinger Josef, Ernst Fehr, Simon Gächter and Rudolf Winter-Ebmer (1995); "A Simple Mechanism for the Efficient Private Provision of Public Goods - Experimental Evidence" Discussion Paper 9517, Dept. of Economics, Univ. of Linz.

Fehr, Ernst, Georg Kirchsteiger and Arno Riedl (1993); "Does Fairness prevent Market Clearing? An Experimental Investigation", Quarterly Journal of Economics, Vol. 108, 437-460

Fehr, Ernst and Simon Gächter (1996); "Cooperation and Punishment - An Experimental Analysis of Norm Formation and Norm Enforcement", Discussion Paper, Institute for Empirical Research in Economics, University of Zürich

Fehr, Ernst, Simon Gächter and Georg Kirchsteiger (1996); "Reciprocity as a Contract Enforcement Device", Econometrica, Vol. 65, No. 4, 833-860.

Forsythe, Robert, Hoel L. Horowitz, N. E. Savin, and Martin Sefton (1988); " Fairness in Simple Bargaining Games", Games and Economic Behavior, Vol. 6, 347-369.

Friedman, Daniel and John Rust (1993); The Double Auction Market - Institutions, Theories and Evidence, Addison Wesley Publishing Company, Readings MA.

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze (1982); "An Experimental Analyses of Ultimatium Bargaining", Journal of Economic Behavior and Organization, vol. 3, no. 3, 367-88.

Güth, Werner and Reinhard Tietz (1990); "Ultimatum Bargaining Behavior - A Survey and Comparison of Experimental Results", Journal of Economic Psychology, vol. 11, 417-449.

Güth, Werner, Nadège Marchand, and Jean-Louis Rulliere (1997); „On the Reliability of Reciprocal Fairness - An Experimental Study", Discussion Paper, Humbold University Berlin.

Hastorf, Albert and Hadley Cantril (1954); „They Saw a Game: A Case Study", Journal of Abnormal and Social Psychology, Vol. 159, 129-134.

Hoffman, Elisabeth, Kevin McCabe, and Vernon Smith (1996);."On Expectations and Monetary Stakes in Ultimatum Games", International Journal of Game Theory, Vol. 25, 289-301.

Isaac, Mark R. and James M. Walker (1988); „Group Size Effects in Public Goods Provision: The Voluntary Contribution Mechanism", Quarterly Journal of Economics, Vol. 103, 179-199.

Isaac, Mark R. and James M. Walker (1991); „Costly Communication: An Experiment in a Nested Public Goods Problem", in Thomas R. Palfrey, ed., Laboratory Research in Political Economy, Ann Arbor: University of Michigan Press.

Kachelmeier, Steven J. and Mohamed Shehata (1992); "Culture and Competition: A Laboratory market Comparison between China and the West", Journal of Economic Organization and Behavior, vol. 19, 145-168.

Kahneman, Daniel, Jack L. Knetsch, and Richard Thaler (1986); "Fairness as a Constraint on Profit Seeking: Entitlements in the Market", American Economic Review, vol 76, no. 4, 728-41.

Kelly, Kathleen (1997); „From Motivation to Mutual Understanding: Shifting the Domain of Donor Research", in Dwight Burlingame, ed., *Critical Issues in Fundraising*, NY: John Wiley,

Keser, Claudia and Frans van Winden (1996); „Partners Contribute More to Public Goods than Strangers: Conditional Cooperation", Discussion Paper, Univ. of Karlsruhe.

Kreps, David, Robert Wilson, Paul Milgrom, and John Robert (1982); "Rational Cooperation in the Finitely-Repeated Prisoner's Dilemma", Journal of Economic Theory, 27, 245-252.

Ledyard, John (1995); "Public Goods: A Survey of Experimental Research", in: J. Kagel and A. Roth (eds.): Handbook of Experimental Economics, Princeton University Press

Levine, David K. (1996); „Modeling Altruism and Spitefulness in Experiments", Discussion Paper, Dept. of Economics, UCLA.

Loewenstein, George F. and Leigh Thompson and Max H. Bazerman (1989); "Social Utility and Decison Making in Interpersonal Contexts", Journal of Personality and Social Psychology, Vol. 57, No. 3, 426-441.

McKelvey, Richard D. and Thomas R. Palfrey (1995); „Quantal Response Equilibria for Normal Form Games", Games and Economic Behavior, Vol. 10, 6-38.

Messick, David and Keith Sentis (1979); „Fairness and Preference", Journal of Experimental and Social Psychology, Vol. 15, 418-435.

Mueller, Denis (1989); *Public Choice II*, Cambridge: Cambridge University Press.

Ockenfels, Axel and Joachim Weimann (1996); „Types and Patterns - An Experimental East-West Comparison of Cooperation and Solidarity", Discussion Paper, Dept. of Economics, Univ. of Magdeburg.

Ostrom, Elinor and James M. Walker (1991); „Cooperation without External Enforcement", in Thomas R. Palfrey, ed., Laboratory Research in Political Economy, Ann Arbor: University of Michigan Press.

Plott, Charles R. (1989); "An Updated Review of Industrial Organisation: Applications of Experimental Methods", in: Schmalensee, R. and Willig, R.: Handbook of Industrial Organisation, Vol. 2, North-Holland, 1109-1176.

Rabin, Matthew (1993); "Incorporating Fairness into Game Theory and Economics", American Economic Review, Vol. 83, No. 5, 1281-1302.

Rehder, Robert (1990); „Japanese Transplants: After the Honeymoon", Business Horizons, 87-98.

Roth, Alvin E., Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir (1991); "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study", American Economic Review, Vol. 81, 1068-95.

Roth, Alvin E. (1995); "Bargaining Experiments", in: J. Kagel and A. Roth (eds.): Handbook of Experimental Economics, Princeton, Princeton University Press.

Roth, Alvin E. and Keith Murningham (1982); „The Role of Information in Bargaining: An Experimental Study", Econometrica, Vol. 50, 1123-1142.

Runciman, Walter G. (1966); *Relative Deprivation and Social Justice*, NY: Penguin.

Sanitiosa, Rasyid, Ziva Kunda and Jeffrey T. Fong (1990); „Motivated Recruitment of Autobiographical Memories", Journal of Personality and Social Psychology, Vol. 59, 229-241.

Skinner, Jonathan and Joel Slemroad (1985); „An economic Perspective on Tax Evaluation", National Tax Journal, Vol. 38, 345-353.

Slonim, Robert and Alvin E. Roth (1997); "Financial Incentives and Learning in Ultimatum and Market Games: An Experiment in the Slovak Republic", forthcoming in: Econometrica.

Smith, Vernon L. (1962); „An Experimental Study of Competitive Market Behavior", Journal of Political Economy, Vol. 70, 111-137.

Smith, Vernon L. (1982); "Microeconomic Systems as an Experimental Science", American Economic Review, Vol. 72, No. 5, 923-955.

Smith, Vernon L. and Arlington W. Williams (1990); "The Boundaries of Competitive Price Theory: Convergence Expectations and Transaction Costs", in: L. Green and J. H. Kagel (eds.), Advances in Behavioral Economics, Vol. 2, Ablex Publishing Corporation, Norwood, New Jersey.

Tversky, A. and Kahneman, D. (1992); "Loss Aversion in Riskless Choice: A Reference Dependent Model", Quarterly Journal of Economics, vol. 106, 1039-1062.

Weisbrod, Burton A. (1988); The Nonprofit Economy, Cambridge, MA: Harvard University Press.

Whyte, William F. (1955); Money and Motivation, New York: Harper and Brothers.

Zamir, Shmuel and Eyal Winter (1996); „Ultimatum Bargaining in a Changing Environment", Discussion Paper, Hebrew University, Jerusalem.