Institute for Empirical Research in Economics
University of Zurich

Working Paper Series
ISSN 1424-0459

Working Paper No. 59

Driving Forces of Informal Sanctions

Armin Falk, Ernst Fehr, Urs Fischbacher

May 2001

Driving Forces of Informal Sanctions

Armin Falk, Ernst Fehr, Urs Fischbacher*

University of Zürich

Instructions included

First Version: September 2000
This Version: May 2001

Abstract: Informal sanctions are a major determinant of a society's social capital because they are key to the enforcement of implicit agreements and social norms. Yet, little is known about the driving forces behind informal sanctions. We systematically examine the determinants of informal sanctions by a large number of experiments. Our experiments allow us to identify the relative importance of three major potential factors: (i) strategic sanctioning for selfish reasons, (ii) non-strategic sanctions driven by spitefulness, and (iii) non-strategic sanctions that are driven by the violation of fairness principles. In addition, the observed sanctioning patterns provide insights into the relevance of different fairness principles.

Our findings show that the violation of fairness principles is the most important driving force of sanctions but, in addition, a non-negligible part of the sanctions is driven by spitefulness. We find surprisingly little evidence for strategic sanctions that are imposed to create future material benefits. While non-strategic sanctions are of major importance in our experiments, strategic sanctions seem to play a negligible role. Within the class of fairness-driven sanctions the motive to harm those who committed unfair actions seems most important.

Keywords: Informal Sanction, Social Norm, Social Capital, Strategic Sanction, Fairness, Reciprocity, Spitefulness.

JEL classification: A13, D63, D23, C92, K42

---

* Institute for Empirical Research in Economics, University of Zurich, Blümlisalpstrasse 10, CH-8006 Zürich, falk@iew.unizh.ch, efehr@iew.unizh.ch, fiba@iew.unizh.ch. Financial support by the Swiss National Science Foundation (Project No. 1214-05100.97), by the MacArthur Foundation (Network on Economic Environments and the Evolution of Individual Preferences and Social Norms), is gratefully acknowledged. This paper is also part of the EU-TMR Research Network ENDEAR (FMRX-CTP98-0238). We also want to thank seminar participants in the Behavioral Economics Seminar at Harvard University, at Pompeu Fabra University in Barcelona and at the MacArthur Preferences Network for helpful comments and discussions.

I. INTRODUCTION

This paper examines the driving forces behind informal sanctions. This topic is important because informal sanctions are key to the enforcement of implicit agreements and social norms. Their importance derives from the fact that the bulk of people's daily interactions is not governed by explicit, enforceable contracts but by informal agreements and social norms. These agreements and norms are major factors in a society's social capital that is crucial for the functioning of democratic institutions as well as for economic and social success (Putnam 1993, Knack and Keefer 1997, Sampson, Raudenbusch and Earls 1997, Bowles and Gintis 2000). How we interact with our colleagues at the workplace, with our friends and neighbors and even with strangers is governed by a set of informally shared rules of appropriate behavior. Examples of such rules abound. For instance, in many countries it is generally viewed as impolite and inappropriate to butt into line in a long queue. Other examples are rate busting in the presence of piece rate incentives, strike breaking during an industrial conflict, shirking in the presence of team incentives and voting behavior. Workers who face the danger of a ratchet effect in a piece rate system frequently develop effort withholding norms (Whyte 1955). Similarly, the peer pressure to join striking colleagues during industrial conflicts is, in general, enormous (Francis 1985) and when workers face team incentives they often tend to establish norms against shirking (Kandel and Lazear 1992, Knez and Simester 1998). Such norms would not be sustained if potential norm violators did not face the disapproval and the contempt, i.e., the informal sanctions, of the people with whom they interact.[1]

There is by now a large body of experimental evidence, which shows the existence of sanctioning behavior. In numerous ultimatum games, e.g., it is observed that responders often sanction proposers who make low offers (e.g., Güth et al. 1982, Roth 1995, Camerer and Thaler 1995). Similarly, Fehr and Gächter (2000) report frequent sanctioning of players in a public good game. In the light of previous evidence, it is not the question anymore *whether* there is informal sanctioning. The problem, which is not yet understood, however,

[1] A drastic example of the sanctioning of norm violations is provided by Frank (1994, p. 31): During the oil crisis in 1979 the Carter administration implemented a system of fuel allocation and price controls that led to long queues at gas stations. Sometimes drivers tried to avoid the costs of waiting by butting into line. As a result of these attempts many motorists were involved in fistfights and shouting matches with one another. One motorist was even shot and killed for butting into line. Another vivid description is given in Francis (1985) who describes the sanctions imposed on strike breakers during the 1984 strike of the British miners: „To isolate those who supported the ‚scab union', cinemas and shops were boycotted, there were expulsions from football teams, bands and choirs and ‚scabs' were compelled to sing on their own in chapel services. ‚Scabs' witnessed their own ‚death' in communities which no longer accepted them". The anti-shirking pressure that evolves under team incentives is described in Rehder (1990).

is *why* people sanction. This question actually consists of two distinct questions. First, what is the evolutionary origin of people's willingness to incur costs to sanction others?[2] Second, what are the proximate reasons, i.e., the motivational forces, that trigger people's willingness to incur costs to sanction others? The present study is devoted to the second question. This does not mean that we regard the first question as unimportant. In fact, we believe that answering the second question is also relevant for answering the first one. Evidence about the proximate forces shaping sanctioning behavior may guide the construction of evolutionary models because these models then have to explain why and how the prevailing proximate forces evolved.

Our investigation is based on a large number of experiments. We conducted several series of experiments with more than 500 subjects who make roughly 4500 sanctioning decisions in 8 different treatment conditions. The contribution of our paper is threefold.

First, we examine the relative importance of three major potential driving forces of informal sanctions – fairness, spite and self-interest in the form of strategic sanctions. Sanctions are *fairness* driven if the violation of fairness principles triggers them. Recently, different theoretical approaches have been developed that provide rigorous, albeit different, definitions of fairness (Rabin 1993, Levine 1998, Dufwenberg and Kirchsteiger 1999, Falk and Fischbacher 1999, Fehr and Schmidt 1999, Bolton and Ockenfels 2000, Charness and Rabin 2000). *Spiteful* sanctions are sanctions that occur because the sanctioning subject values the payoff of the sanctioned subject negatively, irrespective of whether the sanctioned subject behaved fairly or unfairly (see e.g., Mui Vai-Lam 1995, Kirchsteiger 1994). Different from the first two forces, the third force is *strategically* motivated in that people may sanction to induce the future cooperation of others, i.e., they expect a future pecuniary benefit from their sanctions. Strategic sanctions have been modeled in the theory of finitely and infinitely repeated games (e.g., Kreps, Wilson, Milgrom and Roberts 1982, Fudenberg and Maskin 1986). It has been shown that even in random matching games with anonymous interactions the players may achieve cooperation in equilibrium if they use "contagious" sanctions (Kandori 1992, Ellison 1994). Contagious sanctions are also strategic in nature and they suggest that strategic sanctions may be important in a wide variety of circumstances.

[2] For evolutionary explanations of the willingness to impose costly sanctions see, e.g., Bowles and Gintis (1999), Friedman and Singh (2000), Friedman (2001), Gintis (2000), Güth (1995), Henrich and Boyd (2000), Nowak, Page and Sigmund (2000), Sethi and Somananthan (2000a, 2000b).

There has been much research in the past - mainly in the context of bilateral bargaining games - that interprets bargaining behavior in terms of fairness preferences (e.g., Camerer and Thaler 1995). There is, however, little or no research that examines the role of spitefulness and strategic motives. We know, in particular, no study that examines the relative importance of strategic sanctioning behavior.

Second, our examination focuses on the driving forces of informal sanctions in *multilateral* situations of norm enforcement. Almost all past research on sanctioning behavior examines sanctions in two players games. The dominant instrument for the study of sanctions was the ultimatum bargaining game or similar two-person extensive form games.[3] While these games are useful for the examination of bilateral interactions it is problematic to generalize the results that emerged in two-person interactions to the $n \geq 3$-person context. It is far from obvious that behavioral principles that apply in the two-person context will also be present when three or more players interact. We believe that $n \geq 3$-person interactions are particularly important for our study. The reason is that informal sanctions are a key element in the enforcement of social norms and social norms represent an aggregate phenomenon above the dyadic level. Peer pressure, e.g., typically prevails in $n \geq 3$-person groups.

In addition, in $n \geq 3$-person games, that capture the essence of the norm enforcement problem, a completely new set of important questions emerges. For example: Who sanctions whom? Do norm followers sanction the norm violators? Do the norm violators take revenge, i.e., do they also sanction? If the norm violators also sanction, do they preferably sanction the other norm violators or the norm followers? Is the sanctioning of norm violators itself a norm that, if violated, is sanctioned? As we will see, our experiments also provide insights into several of these questions.

Third, by examining the sanctioning patterns in the different experiments we are able to gain important insights into the nature of people's fairness principles. The recently developed theories of fairness and reciprocity provide a rich set of interesting and conflicting predictions about informal sanctions. We also discuss to what extent different notions of fairness are able to capture the major sanctioning regularities in our experiments.

[3] Exemptions are the experiments by Güth and van Damme (1998) and by Kagel and Wolfe (2000). In both experiments there were three players. Yet, the third player was completely passive. He could not make any choice but his payoff was affected by the actions of the other two players. In our 3-person experiments all players have a non-empty strategy set. In particular, all three can sanction the other two players. Charness and Rabin (2000) also examine a few 3-person games but none of these games involves costly sanctions.

The sanctioning pattern in our experiments suggest the following conclusions. *First*, non-strategic forces like fairness and spite are more important than strategic sanctioning forces. In fact, while we were unable to detect a significant impact of strategic forces on sanctioning behavior, non-strategic sanctions are large and significant. *Second*, within the class of non-strategic sanctions fairness driven sanctions are considerably more important than spiteful sanctions. Despite this we found the amount of spiteful sanctions surprisingly large. In one of our main experiments – a three person Prisonner's Dilemma with sanctioning opportunities – 19 percent of the subjects exhibit spiteful sanctions and 38 percent show a sanctioning pattern that is consistent with fairness driven behavior. *Third*, within the class of fairness driven sanctions, the motive to achieve a fair share – as modeled by Bolton and Ockenfels (2000) – cannot explain who is sanctioned. This motive thus seems to be of little importance in our $n \geq 3$-player games. *Fourth*, our results suggest that – within the class of fairness driven sanctions – the retaliation motive, that is the essence of the different reciprocity models (e.g., Rabin 1993, Dufwenberg and Kirchsteiger 1999, Falk and Fischbacher 1999), is more important than the motive to achieve a fair payoff distribution that is the basis of the Fehr-Schmidt (1999) model.

The rest of the paper is organized as follows. Section II presents the whole list of experiments and describes them in the necessary detail. Section III examines the driving forces behind non-strategic sanctions We show to what extent the sanctioning patterns are consistent with fairness and/or spite. In addition we discuss to what extent different notions of fairness are consistent with the sanctioning pattern. In Section IV we study the relative importance of non-strategic and strategic sanctions. Section V summarizes the results and concludes the paper. The appendix provides the instructions for several of our experiments.

II. EXPERIMENTAL DESIGNS

To examine the driving forces behind informal sanctions we conducted three large series of experiments. In all experiments subjects did not know the identities of their interaction partners and all interactions between the subjects were completely anonymous. The reason for maintaining anonymity was that we wanted to study the nature and the relative importance of non-strategic sanctions. In case that subjects know the personal identities of their interaction partners it is, however, not possible to rule out future interactions between them, i.e., sanctioning can always be interpreted as being motivated by future pecuniary

gains. In a first series our research tool was a one-shot 3-person Prisoners' Dilemma (PD) with the possibility to sanction other players. In a second series we used one-shot bilateral bargaining experiments with different treatment conditions as the major research tool. Finally, to study the relative importance of strategic sanctions, we implemented repeated and non-repeated 3-person public good games with sanctioning opportunities. Table 1 lists all the experiments and treatment conditions that were conducted for this paper.

The main reason to choose PDs and public good games with sanctioning opportunities as a research tool was that they capture the problem of norm enforcement in a neat way. Social norms typically arise in situations that are characterized by positive or negative externalities. In Coleman's seminal account of social norms, externalities are even a necessary condition for the existence of social norms because in their absence there is no demand for a social norm (Coleman 1990). When there are externalities, a social norm may act as a substitute for a multilateral agreement that restrains the actions of the involved parties in an efficiency-enhancing way. In the context of our PDs and public good games norm adherence, therefore, means "cooperation" and norm violation means "defection" or "free-riding".

The experimental design of our PD and public good games is inspired by Fehr and Gächter (2000). However, the focus in Fehr and Gächter is very different. They studied the impact of sanctioning opportunities on cooperation behavior and did not examine the sanctioning motives. Their design is, in particular, not suited to analyse the relative importance of strategic and non-strategic sanctions because in all of their treatments the probability of being matched with the same subject in future periods was positive. Strategic sanctioning forces could thus play a role in all of their treatments. Moreover, since they had no pure (i.e., non-repeated) one-shot games, that rule out strategic sanctions with certainty, it is also not possible to examine the different motivational components of non-strategic sanctions in a clean way.

In the following we first describe the PD with sanctions (Treatment 1 and 2 in Table 1). In total 213 subjects participated in this experiment.[4] No subject participated in more than 1 treatment. The experiment was one-shot and each subject was assigned to a group with two other players. The PD was framed in terms of investments into a project. A player

[4] In all experiments described in this paper subjects were undergraduate or graduate students from the University of Zurich or the Federal Institute of Technology in Zurich. All experiments were programmed in z-Tree (Fischbacher 1999) and conducted in the computer laboratory of the Institute for Empirical Research in Economics. The PD experiment lasted roughly 35 minutes and subjects earned CHF 25 (≈ $16) on average.

could either invest nothing (= defection) or he could invest 20 tokens (= cooperation). The experiment had two decision stages. At stage one each of the three players in the group simultaneously decided whether to cooperate or to defect. After this, players moved to the second stage where they made their sanctioning decisions described in more detail below. At the first stage players faced the payoff structure presented in Table 2.

Table 1: List of experiments and treatment conditions

Treatment No.	Description of treatments	No. of subjects	Main research question
1	Prisoner's dilemma with sanctioning possibilities (high-sanction treatment: $f = 2.5$ or 3.33)	120	Who is the sanctioning target? Who sanctions? Fairness versus spite.
2	Prisoner's dilemma with sanctioning possibilities (low-sanction treatment: $f = 1$)	93	Fairness versus Spite. Relevance of different fairness principles.
3	Baseline ultimatum game	96	Serves as a control for treatments 5 and 6.
4	Ultimatum game with an endowment for the responder	90	Does spitefulness trigger rejections?
5	Ultimatum game with constant relative shares	96	Do subjects reject low offers to increase their relative payoff share?
6	Ultimatum game with constant payoff differences	96	Do subjects reject low offers to reduce differences in payoffs?
7	Repeated public good game with a sanctioning opportunity	87	Relative importance of strategic vs. non-strategic sanctions?
8	One-shot public good game with a sanctioning opportunity	87	Serves as a control for treatment 8.

Table 2 makes the PD-structure of the game transparent. Player *i*'s dominant strategy is to defect and he is always better off the more other players cooperate. If all three players in a group defect, each player earns 20 while if all cooperate each earns 36. At the second stage each player could reduce the payoff of each of the other two players by at most 25 tokens. If player *i* wanted to reduce the income of player *j* he had to assign so-called deduction points

to j.[5] If i assigned p_{ij} points to j the first-stage payoff of j was reduced by fp_{ij} ($f \geq 1$) and the first-stage payoff of i was reduced by p_{ij}. Thus, the assignment of deduction points was not only costly for the sanctioned player but also for the sanctioning player. We study two treatments: In the *low-sanction* treatment (Treatment 2) f was set equal to 1. This means that if i deducted 1 point from j this also cost him 1 point. In the *high-sanction* condition (Treatment 1) punishment was more effective. For reasons that will become evident below, the effectiveness of punishments in the high-sanction treatment varied dependent on whether i punished a cooperator or a defector. If i assigned one deduction point to a cooperator, the cooperator's payoff was reduced by 3.33 tokens (f = 3.33), whereas if i assigned a point to a defector, the defector's payoff was only reduced by 2.5 tokens (f = 2.5). Put differently, it was cheaper to achieve a given payoff reduction by punishing the cooperators. In both treatments deducting is costly and a purely selfish player will never sanction at the sanctioning stage, i.e., the sanction must be due to a non-pecuniary motive.

Table 2: Payoffs to player *i*

	Both other players defect	**One of the other two players cooperates**	**Both of the other players cooperate**
Player *i* defects	20	32	44
Player *i* cooperates	12	24	36

At stage two each group member made a *contingent* sanctioning decision before he was informed about the actual first-stage decisions of the other players. That is, each member assigned points to another member for the case that the other member cooperated and for the case that the other defected. This has the advantage that we collect much more information about subjects' sanctioning behavior.[6] The contingent sanctioning decisions were made with the help of four decision screens. On each screen, i had to indicate whether he wanted to assign deduction points to no other player in the group, or to one or to both of the other players and if so, how many points. The four screens represent the four possible choice

[5] It is worthwhile to point out that we carefully avoided value laden terms like "punishment" or "sanction" in the instruction. This is the reason why we used the term "deduction point". In the Appendix we provide an English translation of the instructions for the PD-experiment.

[6] In principle, it is possible that the elicitation of contingent responses induces a different behavior relative to a situation where the subjects face given, known, cooperation decisions. However, Brandts and Charness (1998) and Cason and Mui (1998) report evidence indicating that the elicitation of contingent responses does not induce different behaviors.

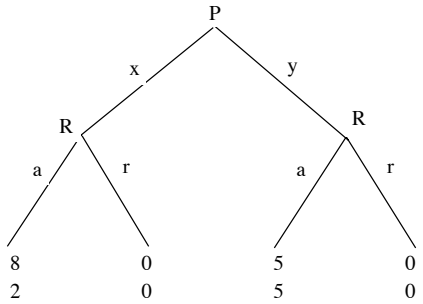
combinations of the other two players: both of the other players defected, both cooperated, the second player cooperated while the third one defected, and the second player defected while the third cooperated. After all 3 players had made their contingent sanctioning decisions they were informed about the behavior of the other players at the first stage and how much they themselves had been sanctioned. The whole procedure of how to make decisions at stage two was carefully explained to the subjects in the instructions. At the end of the instructions we also asked several hypothetical questions to check subjects' comprehension of the procedure. After all subjects had solved all questions successfully the experiment began.

In our second series of experiments (Treatments 3 – 6 in Table 1) we used *four* different variants of the ultimatum game as our research tool. The ultimatum game is a bilateral bargaining game and it is well known that low offers are frequently rejected in this game although the rejection is costly for the individual who rejects (see e.g., Camerer and Thaler 1995 or Roth 1995). The concrete motives driving the rejections in ultimatum bargaining are, however, not yet fully understood: In the standard ultimatum game, e.g., rejections are compatible with the desire to get a fair share, the aim to reduce payoff differences as well as with the desire to harm the proposer. In all bargaining experiments we used a variant of the ultimatum game displayed in Figure 1. In the baseline game (Treatment 3) the proposer could make a 5/5-offer or an 8/2-offer (8 for the proposer and 2 for the responder).[7] Then the responder could accept or reject the offer. In case of acceptance, the offer was implemented and in case of a rejection both players received nil. In all our ultimatum games we implemented the strategy method, i.e., the responder had to indicate an acceptance/rejection decision for both offers before he knew the actual offer.[8]

[7] Each of the ultimatum games in Treatment 3-6 lasted roughly 15 minutes and one token was equal to one CHF. Different subjects participated in the different treatment conditions. The appendix provides the instructions for the responders for one of the ultimatum games. The instructions for all ultimatum game treatments were identical except for the final payoffs in the different treatments.

[8] Note that we applied the strategy method in all our ultimatum games. Therefore, the impact of this method on subjects' behavior, if there is any, is kept constant across conditions. See also footnote 6 for evidence that the elicitation of contingent responses has no behavioral impact relative to the elicitation of responses to given first-mover choices.

Figure 1: Baseline Ultimatum Game (Treatment 3)



Why do many responders reject low offers like the 8/2-offer? Many researchers have argued that responders reject low offers because they resist unfair outcomes (e.g., Bolton and Ockenfels 2000, Fehr and Schmidt 1999). Yet, this raises the question of the determinants of players' fairness perceptions. Is an offer unfair because it gives the responder a low share of the payoff? Or is it unfair because it gives the responder less than the proposer? A further possibility is, however, that responders reject low offers not because they value fairness positively but because they themselves *prefer* unfair outcomes, i.e., outcomes that give them more than the other player.[9]

To examine these questions we varied the above baseline game in two different directions. *First*, in Treatment 4 the payoff of the responder was increased by 10 tokens at *each* final node compared to the baseline game. This is then like an ultimatum game in which the responder has an initial endowment of 10 tokens.[10] Thus, after the proposer chooses the move *x* the responder faces the choice between 8/12 and 0/10 because – relative to the baseline game in Figure 1 – the responder's payoff is now ten units higher at both nodes. Note, that if the responder prefers 0/10 over 8/12 he is willing to give up 2 tokens to increase the payoff difference from 4 to 10 units. The rejections in this game thus provide a measure of the quantitative importance of a preference for inequality or spitefulness. *Second*, we systematically changed the material consequences of a rejection in Treatments 5 and 6. In Treatment 5, e.g., a rejection implied that the proposer and the responder lost 90 percent of the income relative to the proposed amount, i.e., a rejection could not change the relative share. Therefore, fair players with a concern for the relative share will never reject.

[9] Still others have argued that responders do not care for uneven outcomes per se. Instead, they reject because they want to punish unfair intentions (Rabin 1993, Dufwenberg and Kirchsteiger 1999). It is, of course, also possible that both unfair outcomes as well as unfair intentions trigger rejections (Falk and Fischbacher 1999).

[10] We did, however, not frame the responder payoffs in terms of endowments and payoffs that accrue in the game. The proposer's task was to propose one of two feasible income distributions, (8,12) and (5, 15). If the responder accepted, the proposal was implemented. In case of a rejection (0, 10) was implemented.

A similar method was applied in Treatment 6 to examine the quantitative importance of a preference for inequality reduction.

Treatments 1 – 6 implemented pure one-shot games. In these treatments it was made explicit to the subjects that the game is played exactly once. Therefore, the sanctions cannot be based on subjects' attempts to gain a strategic advantage in future interactions or to establish the credibility of a sanctioning threat. To study the quantitative importance of strategic sanctions in repeated games relative to the sanctions that are driven by non-pecuniary motives we conducted treatments 7 and 8. Both treatments are based on a three-person public good game with sanctioning opportunities. In Treatment 7 we implemented stable groups, i. e., the three group members remained together in the same group for 6 periods. In Treatment 8, which also lasted for 6 periods, the group composition changed from period to period such that no subject was matched with any other subject more than once. By comparing the sanctioning behavior across the two treatments we can isolate the relative importance of strategically motivated sanctions. If, e.g., the sanctioning pattern in both treatments is very similar we have an indication that strategically motivated sanctions are not very important in our context.

III. DRIVING FORCES BEHIND NON-STRATEGIC SANCTIONS

To examine the driving forces of non-strategic sanctions we conducted Treatments 1-6. In these treatments players interacted only once so that strategic punishment motives are ruled out. In this section we examine the relative importance of spitefulness and fairness motives for sanctioning behavior. Then we analyze the role of different fairness principles. Since much of this analysis is based on the sanctioning patterns in the 3-person PD it is useful to first describe these patterns in more detail.

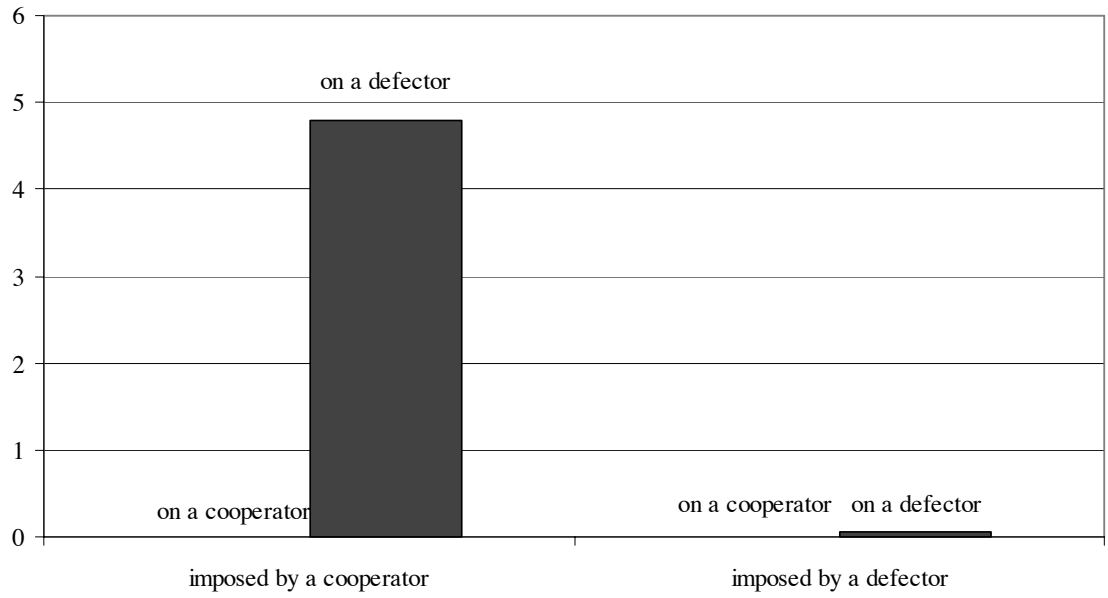
III. 1 BASIC RESULTS IN THE 3-PERSON PD: WHO SANCTIONS WHOM?

In the high-sanction condition (with $f = 2.5$ or 3.33) of our PD 120 subjects participated. 61 percent of them cooperated and the remaining subjects defected. In the low-sanction treatment (with $f = 1$) 93 subjects participated and 51 percent of them cooperated while the others defected. A first indication of the general pattern of sanctions is given in Figures 2a and 2b. Figure 2a depicts the average sanctions in the high-sanction treatment whereas Figure 2b indicates the sanctions in the low sanction treatment. Note that both figures depict the severity of the sanctions imposed on other players from the viewpoint of the subject who

punishes. They do not directly show the average payoff reductions experienced by the sanctioned subjects. Figure 2a shows, e.g., that a cooperator reduces the income of a defector on average by 5.7 tokens. Therefore, if there are two cooperators in a group and one defector, the defector's income is, on average, reduced by 11.4 tokens.

Figure 2a: Average payoff reduction in the high sanction treatment (f = 2.5 or 3.33)

Figure 2b: Average payoff reduction in the low sanction treatment (f = 1)



Figures 2a and 2b exhibit several remarkable features. *First* of all, cooperators tend to punish almost exclusively the defectors. This is also supported by non-parametric Wilcoxon signed rank test. The null hypothesis that cooperators punish defectors and other cooperators with the same probability is clearly rejected ($z = 6.64$, $p < 0.001$).

Second, and this is surprising when viewed through the lens of fairness theories, in the high sanction treatment the defectors do also punish but they tend to punish both defectors and cooperators. Moreover, the strength of the sanctions imposed on other defectors is almost the same as the strength of the sanctions imposed on cooperators, and both cooperators and defectors are equally likely to be punished by the defectors. The null hypothesis that defectors sanction other defectors and cooperators with the same probability cannot be rejected (Wilcoxon signed rank test, $z = 1.34$, $p = 0.179$).

Third, the figures show that by far the strongest sanctions are imposed by the cooperators on the defectors. *Fourth*, in the low-sanction treatment only one source of punishment remains: Cooperators punish the defectors. All other types of punishment are virtually non-existent. Another interesting feature of Figure 2b is that the severity of cooperators' sanctions is almost as high as in Figure 2a. Since in the low-sanction treatment the payoff reduction per assigned deduction point, f, was much lower than in the high-sanction treatment, the punishing cooperators spent much more money on sanctioning in the low-sanction treatment. This indicates that the motives behind the sanctions of the cooperators are very strong.

To assess the relevance of the different motives for the sanctions it is also important to know the fraction of people in the different sanctioning categories. Therefore, in Table 3 we show the fraction of cooperators and defectors who sanction in the two conditions. Since we are interested in the overall quantitative importance of each type of sanction, we have divided the number of people in each sanctioning category by the *total* number of participants in the relevant treatment. In the high-sanction condition, e. g., 42 percent of the 120 participants cooperated and sanctioned the defectors and 16 percent of the 120 subjects defected and sanctioned the defectors. 13 percent of the subjects defected and sanctioned the cooperators. The last row in Table 3 informs us about the percentage of subjects who defected and punished (either a cooperator or a defector) and who cooperated and punished (either a cooperator or a defector). We see that, overall, a surprisingly large fraction of 19 percent defect and punish and 42 percent cooperate and punish in the high-sanction

condition.[11] The remaining 39 percent of the subjects in the high sanction treatment do not punish.

Table 3: Fraction of cooperators and defectors who sanction in the high-sanction and the low-sanction condition

	Sanctioning subject is a			
	defector		cooperator	
Sanctioned subject is a defector	low-sanc. cond.	high-sanc. cond.	low-sanc. cond.	high-sanc. cond.
	1.1 %	16 %	30 %	42 %
Sanctioned subject is a cooperator	low-sanc. cond.	high-sanc. cond.	low-sanc. cond.	high-sanc. cond.
	0.0 %	13 %	0.0 %	4 %
Sanctioned subject is a defector or a cooperator	low-sanc. cond.	high-sanc. cond.	low-sanc. cond.	high-sanc. cond.
	1.1 %	19 %	30 %	42 %

III. 2 FAIRNESS VERSUS SPITE

To what extent are the sanctions that are depicted in Figure 2a and 2b driven by spitefulness and to what extent are they driven by fairness considerations? Spiteful sanctions are sanctions that occur because the sanctioning subject values the payoff of the sanctioned subject negatively, irrespective of whether the sanctioned subject behaved fairly or unfairly, and irrespective of the distribution of pre-sanction payoffs. A particular form of spitefulness prevails if a subject prefers inequality, i.e., if *i*'s utility is increasing in $(\pi_i - \pi_j)$. The sanctions of the cooperators cannot be explained by spitefulness for two reasons. First, spiteful subjects are indifferent with regard to the punishment target. They, therefore, sanction defectors as well as cooperators. Yet, cooperators almost exclusively sanctioned the defectors. Second, spitefulness strongly inhibits cooperation because cooperation means to increase the payoff of the other group members at the expense of one's own material payoff. Thus, cooperators are least likely to be spiteful. The behavior of those subjects who

[11] Note that *all* cooperators who punish other cooperators also punish the defectors. Therefore, the overall percentage of subjects who cooperate and punish (either a cooperator or a defector) is the same as the percentage of subjects who cooperate and punish the defectors. Among the defectors there are some subjects who punish exclusively other defectors or exclusively cooperators. Therefore, the overall percentage of subjects who defect and punish (either a cooperator or a defector) is larger than the percentage of subjects who defect and punish a defector.

cooperate *and* punish the defectors is, therefore, better explained by fairness approaches. This suggests (see Table 3) that in the high-sanction treatment 38 percent of the subjects (45 out of 120) sanctioned in response to the violation of fairness principles while in the low sanction treatment 30 percent of the subjects (28 out of 93) sanctioned for this reason.

However, Table 3 also indicates that 19 percent of the subjects defect *and* punish in the high-sanction treatment. It seems difficult, if not impossible, to reconcile the sanctioning behavior of the defectors with any reasonable notion of fairness. After all, the defectors benefit from the cooperation of the others but refuse to bear "their" part of the cost of the investment project. But not only do these defectors benefit from the project without sharing the burden, they also sanction cooperators and other defectors. Therefore, the sanctions of the defectors point towards motivational forces that have been neglected by most fairness theories.[12] In our view, this type of sanction is better viewed as driven by some form of spitefulness. A first indication for this interpretation is given by the fact that the defectors punish cooperators and other defectors by roughly the same amount and with equal probability (see Figure 2a). Since spitefulness means that the payoff of other subjects is valued negatively irrespective of the behavior of the other subjects one would expect that spiteful subjects do not discriminate between cooperators and defectors when they punish. A second indication for spitefulness emerges from a comparison of Figure 2a with Figure 2b. In fact, Figures 2a and 2b (as well as Table 3) suggest that the sanctions of the defectors are driven by the motive to increase $(\pi_i - \pi_j)$. Recall that in the high-sanction condition a sanctioning subject can increase the payoff difference between himself and the sanctioned subject because $f > 1$. In the low-sanction condition, however, any sanction leaves the payoff difference between the sanctioning and the sanctioned subject unchanged because $f = 1$. In view of this, it is striking that the fraction of subjects who defect *and* punish is 19 percent in the high-sanction condition but virtually zero in the low-sanction condition. Only one subject (= 1.1 percent of 93) defects and punishes in the low-sanction condition. This suggests that the punishment activities of the defectors are driven by the desire to increase payoff differences.

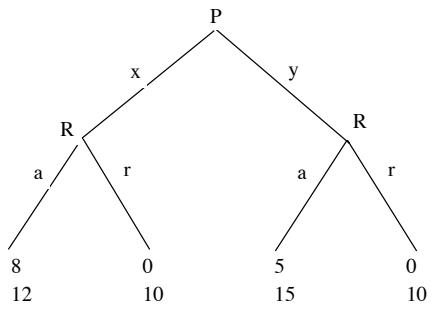
Are fairness driven sanctions or spiteful sanctions more important in our PD? There are two reasons why fairness-driven sanctions are more important. First, the sanctions that the cooperators imposed on the defectors are much stronger than the sanctions that the

[12] An exception is the model by Levine (1998) who explicitly models spiteful types and altruistic types in the same framework. The model of Kirchsteiger (1994) takes into account only spiteful types.

defectors imposed on others. Second, there are much more sanctioning cooperators than sanctioning defectors. A non-parametric Fisher exact test shows that the fraction of cooperators who sanction *only* the defectors (45 out of 120) is significantly higher ($p < 0.001$) than the fraction of defectors who sanction cooperators or defectors (23 out of 120).

In view of the presence of roughly 20 percent of spiteful types in our PD-experiments it is interesting to examine whether similar behaviors show up in other situations. To study the robustness of our finding we conducted the ultimatum game described in Figure 3 (Treatment 4) in which the proposer could make the 8/12- or the 5/15-offer. If the responder rejected an offer the proposer's payoff was always zero while the responder's payoff was reduced by 2 or 5 tokens, respectively. Since both offers give the responder a higher payoff than the proposer they cannot be viewed as unfair from the responder's perspective. Thus, resistance to unfairness cannot explain rejections in this game. However, if responders have spiteful preferences, rejections may occur. We were particularly interested in the rejection rate of the 8/12-offer relative to the 5/15-offer because by a rejection of the 8/12-offer the responder could increase the payoff difference in his favor at a cost of 2 tokens to himself while a rejection of the 5/15-offer leaves the payoff difference intact.

Figure 3: Game tree of the experiment testing for spiteful rejections (Treatment 4)



The results from this game largely confirm what we observed in the PD-experiment. 24 percent of the responders (11 out of 45) rejected 8/12 in favor of 0/10 and only 4 percent (2 out of 45) rejected 5/15 in favor of 0/10. This difference is significant at any conventional significance level (Fisher exact test, $p = 0.006$). Thus, again there is a non-negligible fraction of subjects who are willing to spend money to increase the payoff difference between themselves and the reference agent – which is in this case the proposer. Thus, the presence of this type of spitefulness seems to be a rather robust phenomenon that persists across games.

III. 3 THE RELEVANCE OF DIFFERENT FAIRNESS PRINCIPLES

The results of the previous section suggest that in the class of non-strategic sanctions the fairness motive is more important than spitefulness although spiteful sanctions were surprisingly frequent. It is, therefore, important to know the nature of the fairness principles that drive the sanctions. We turn to this question next.

A. DOES THE CONCERN FOR A FAIR RELATIVE SHARE TRIGGER SANCTIONS?

Fairness concerns are intimately related to the question of how the total surplus is distributed. People very often seem to invoke fairness arguments when they or others do not get what they perceive to be the fair share of the total surplus. The assumption that fair subjects are concerned with getting the fair share of the total surplus is, therefore, a natural starting point.[13] A recent paper by Bolton and Ockenfels (2000, henceforth BO) develops this notion of fairness in a precise way. The theory assumes that fair-minded players care for their relative share σ of the total surplus. If a fair-minded player receives less than the fair relative share σ^*, where σ^* is defined as $\sigma^* = 1/n$, it is assumed that he aims at increasing his payoff share while if the player receives more than σ^* he wants to decrease his payoff share.[14] This theory has been shown to be consistent with a wide variety of puzzling experimental facts like, e.g., fair proposals in bilateral bargaining games and very unfair outcomes in competitive market games.

When evaluating the relevance of the fair share motive according to the specification in BO, it is important to keep in mind that this motive consists of two conceptually distinct ingredients. It means first that, for a fixed n, players evaluate the fairness of an outcome by comparing their payoff with the *average payoff in the group*. Fairness is thus not evaluated by comparing the own payoff with the payoffs of particular other players but by a comparison with the average. This ingredient alone, however, does not yet imply a concern for the relative *share*. It is also possible that fair subjects do make comparisons with the average payoff but are concerned, e.g., with the *difference* between their payoff and the average payoff. Yet, in BO fair players are not concerned with the difference but with the

[13] For instance, if - under team incentives - a hard working employee sanctions a shirking colleague, it seems intuitive that the sanction may be due to the fact that shirking prevents the non-shirker from receiving his fair share.

[14] Formally the utility function of player i in BO is given by $U_i = U_i(\pi_i,\sigma_i)$. The relative share, σ_i, is defined as $\pi_i/(n\pi_a)$ where n is the number of players and π_a denotes the average payoff in the group. If the total surplus $n\pi_a$ is zero, like after a rejection in the Ultimatum game, the theory treats this as if every player received the fair share, i.e., $\sigma_i=1/n$ in this case by assumption. U_i is increasing in σ_i if $\sigma_i<1/n$ and decreasing in σ_i if $\sigma_i>1/n$.

relative *share*.

The assumption that players evaluate the fairness of an outcome in terms of a comparison with the group average has important implications for the sanctioning behavior in our previous experiment. It means, in particular, that a punishing player is *indifferent* between punishing a cooperator and punishing a defector. To see this let π_i be punisher i's payoff before he punishes, let p_{ij} be the cost of punishment for punisher i, $\Sigma\pi_k$ denotes the total group payoff before i's punishment and fp_{ij} is the cost of punishment for the sanctioned individual j where $f \geq 1$. Then i's payoff after his punishment is $\pi_i - p_{ij}$ and the group's average payoff after the punishment is $(1/n)[\Sigma\pi_k - (1+f)p_{ij}]$. It is easy to see that both i's payoff as well as the group's average payoff after punishment is independent of the identity of the punishment target, i. e., whether j is a cooperator or a defector. Therefore, any approach that assumes that the fairness of an individual payoff is evaluated in terms of a comparison with the group average *cannot explain why punishment is predominantly imposed on defectors*. Moreover, a further implication of such an approach is that if it is cheaper to punish the cooperators than the defectors, a punisher will always prefer to punish the cooperators.

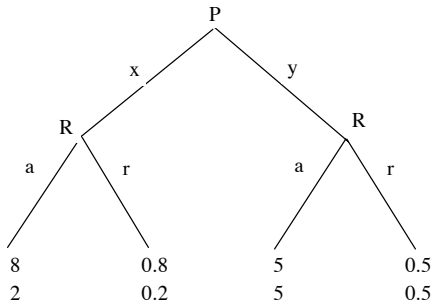
We have tested the latter prediction in the high-sanction condition of the PD-experiment in the following way. If a subject assigned one deduction point to a cooperator, the cooperator's payoff was reduced by 3.33 tokens, whereas if a subject assigned a point to a defector, the defector's payoff was only reduced by 2.5 tokens. Thus, it was cheaper to achieve a given payoff reduction by punishing the cooperators. Therefore, if the fairness evaluation is based on the comparison with the average payoff, the sanctioning player prefers to punish cooperators rather than defectors. This contrasts with the facts, however. Although it was cheaper to reduce the payoff of the cooperators, the sanctions of the cooperators were predominantly imposed on the defectors. 45 of the 73 cooperators punished exclusively the defectors and 5 punished the other cooperators as well as the defectors. Note also that, if the fairness evaluation is based on the comparison with the average payoff, a punishing defector should prefer to sanction the cooperators. In fact, defectors sanction cooperators and other defectors at the same rate (see Figure 2a).

The evaluation of the fairness of an outcome in terms of a comparison with the average group payoff also implies rather implausible sanctioning patterns in the low-sanction game. In this game it was equally costly to punish a defector or a cooperator and,

therefore, this notion of fairness predicts that a punisher is indifferent whether he wants to punish a defector or a cooperator. In fact, we observed that *only* defectors were sanctioned in this condition. If subjects were really indifferent between sanctioning a cooperator or a defector one would have expected that the sanctions were randomly imposed on cooperators and defectors. These facts indicate that any notion of fairness, which relies on the comparison of the own payoff with the average payoff, cannot explain *who is sanctioned.*

Since a concern for the relative share does not lead to satisfactory predictions in our PD, we were also interested in the question to what extent the theory does better in a two person setting. Note that a two person setting gives the theory a much better chance because the comparison with the group average essentially boils down to a comparison with the other player's payoff. We were, therefore, interested whether the pattern of rejection behavior observed in ultimatum games is consistent with this motive. This question is important insofar as the assumption that fair subjects are concerned with their relative share was originally invoked to explain rejection behavior in the ultimatum game. For this purpose we conducted Treatment 3 and Treatment 5. Treatment 3 is the baseline ultimatum game. Treatment 5 is identical to the baseline except that a rejection reduces the proposer's and the responder's payoff by 90 percent and not by 100 percent as in the baseline game (see Figure 4). Therefore, in Treatment 5 the responder's decision does not affect relative shares so that the theory of BO predicts zero rejections of the unfair 8/2-offer in Treatment 5. The rate of rejection of the unfair offer in this game informs us, therefore, about the quantitative importance of those motives for rejections in the UG that are not captured by BO.[15]

Figure 4: Ultimatum Game with Constant Relative Shares (Treatment 5)



[15] The models of Rabin (1993), Levine (1998), Fehr and Schmidt (1999), Dufwenberg and Kirchsteiger (1999), Falk and Fischbacher (1999) and Charness and Rabin (2000) are consistent with rejections in Treatment 5.

The following results were observed in the two treatments. In the baseline game 27 out of 48 subjects (56.3 percent) rejected the unfair offer while in the experiment with constant relative shares 18 out of 48 subjects (37.5 percent) rejected this offer. A Chi-Square test shows that this difference is (weakly) significant at the p = 0.066 level. This suggests that at most one third of all rejections in the baseline game is consistent with the relative share motive while two thirds are not consistent with this motive. Therefore, even in the UG the quantitative importance of this motive seems somewhat limited.

B. INEQUALITY REDUCTION VERSUS RETALIATION

Since fairness concerns are related to distributional questions it seems also quite intuitive that fair subjects are concerned with payoff differences. In principle, payoff differences may play a dual role in a theory of fairness. First, people may want to harm those whose actions created unfair payoff differences. In this case they sanction because they want to retaliate. Retaliatory sanctions are captured by the different reciprocity models (Rabin 1993, Dufwenberg and Kirchsteiger 1999, Falk and Fischbacher 1999, Charness and Rabin 2000). It may be a consequence of retaliatory sanctions that they decrease payoff differences. Second, subjects may *directly* aim at decreasing payoff differences. In the second approach an unfair payoff difference is not only the trigger of sanctioning or rewarding behavior but, in addition, fair subjects also want to establish minimal payoff differences. In the following we start with the second approach. Fehr and Schmidt (1999, henceforth FS) assume that if player *i*'s material payoff differs from those of player *j*, *i* aims at reducing the payoff difference between himself and *j*.[16] In contrast to the BO-model this approach can, in principle, explain why the cooperators predominantly punish the defectors in our PD. The reason is that cooperators have a lower pre-punishment income than the defectors and by punishing the defectors the cooperators can reduce the payoff difference. Thus, the theory is consistent with the cooperators' sanctioning pattern in the high sanction treatment in which the sanctioning subjects could reduce the payoff differences between themselves and the other group members.

The FS-approach can, however, not explain the punishment pattern in the low-sanction condition. Remember that in these experiments it is not possible to change the income differences between the punisher and the punished subject. Therefore, a fair player

[16] Formally, the utility function of player i in Fehr and Schmidt is given by $U_i(\pi_1,..., \pi_N) = = \pi_i - [\alpha_i/(n-1)]max\Sigma_{j\neq i}\{\pi_j-\pi_i,0\} - [\beta_i/(n-1)]max\Sigma_{j\neq i}\{\pi_i-\pi_j,0\}$, with $\beta_i \leq \alpha_i$ and $\beta_i \leq 1$.

in the sense of FS should never punish in this condition. Despite this 30 percent of the subjects, who participated in this experiment, cooperated *and* punished the defectors (see Table 3). In the high-sanction condition 42 percent of the subjects cooperated *and* punished. A non-parametric Fisher exact test indicates that the fraction of sanctioning cooperators is higher in the high-sanction treatment ($p = 0.026$). These facts suggest that at least 30/42, or roughly 71 percent, of the subjects who punish in the high-sanction condition are not motivated by a desire to reduce the payoff differences. Thus, the fraction of punishing cooperators whose punishment choices are driven by a different motive is quantitatively important.

To judge the relevance of this motive it is also important to look at the severity of the sanctions. Table 4 shows the average payoff reduction that a sanctioning subject imposes on the sanctioned subject in the high- and the low-sanction condition.[17] A remarkable feature of Table 4 is that those cooperators who punish in the low-sanction condition impose on average almost the same payoff-reduction (8.0 tokens) as the cooperators who punish in the high-sanction condition (8.3 tokens). Note that this means that a punishing cooperator in the low-sanction condition spends 2.5 times more money on the punishment of the defectors than a punishing cooperator in the high-sanction condition.[18] This indicates that the motive behind the punishment of defectors in the low-sanction condition is rather strong, i.e., the FS-model fails to capture an important motive.

Table 4: Average payoff reduction that a sanctioner imposes on the sanctioned subject in the high-sanction and the low-sanction condition

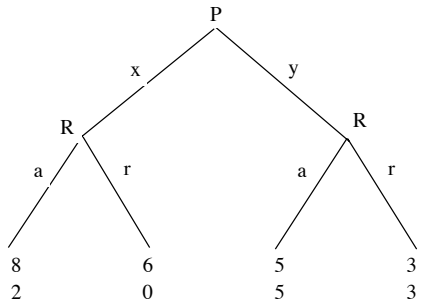
	Sanctioning subject is a			
	defector		cooperator	
Sanctioned subject	low-sanc. cond.	high-sanc. cond.	low-sanc. cond.	high-sanc. cond.
is a defector	2.5	5.0	8.0	8.3
Sanctioned subject	low-sanc. cond.	high-sanc. cond.	low-sanc. cond.	high-sanc. cond.
is a cooperator	no obs.	5.0	no obs.	4.5

[17] Note that the population over which averages are taken is the number of sanctioning subjects in the different sanction categories. In the high-sanction condition, e.g., 50 cooperators punished the defectors and each of them reduced the payoff of a defector, on average, by 8.3 tokens.

[18] Recall that in the low-sanction condition spending one punishment point reduced the payoff of the punished defector by one unit while in the high-sanction condition the reduction is 2.5 units.

As in the case of the relative share motive we were also interested in the question to what extent the motive to reduce payoff differences can explain rejections in the ultimatum game. For this purpose 48 pairs of subjects participated in Treatment 6. This treatment is identical to the baseline except that a rejection does not destroy the whole surplus but decreases each players' payoff by two units (see Figure 5). The FS-model, therefore, predicts that no responder will reject.

Figure 5: Ultimatum Game with Constant Differences (Treatment 6)



It turns out that when rejections cannot affect the payoff differences only 9 of the 48 responders (19 percent) reject the unfair 8/2 offer while in the baseline game 27 responders rejected this offer. The difference is significant at any conventional significance level ($p < 0.0001$, Chi-Square test). Nevertheless the data reveals that one third of the rejections in the baseline game are inconsistent with the FS-model.

Taken together the evidence from the PD and the ultimatum experiments reported above suggests that the motive to reduce payoff differences generates better predictions of the sanctioning behavior compared to the relative share motive. Yet, the data from the low-sanction condition also indicate that the desire to reduce payoff differences is not the major driving force of the cooperators' sanctions. The sanctions in the low-sanction treatment are consistent with the idea that most cooperators sanction the defectors because they view the behavior of the defectors as unfair. This suggests that instead of the motive to reduce unfair payoff differences, the *desire to retaliate* seems to be the driving force of these sanctions. This motive is captured by the different reciprocity approaches (Rabin 1993, Dufwenberg and Kirchsteiger 1999, Falk and Fischbacher 1999, Charness and Rabin 2000). Models that incorporate the desire to retaliate are, therefore, consistent with the punishment pattern in the high- *and* the low sanction condition of our PD as well as with the rejections that occur in Treatment 6.

IV. STRATEGIC VERSUS NON-STRATEGIC SANCTIONS

Until now we have examined those driving forces of informal sanctions that arise in one-shot interactions. Yet, perhaps the forces behind non-strategic sanctions are small and, hence, not very important relative to the additional forces that arise in repeated interactions. To analyze the relative importance of strategic and non-strategic sanctions we conducted 3-person public good games with a sanctioning opportunity. In Treatment 7 stable 3-person groups interacted for six periods. In contrast, Treatment 8 ensures that the probability of being matched with the same subject in future periods is zero. In Treatment 8 subjects also played the public goods game for six periods but the group composition changed from period to period such that no subject ever met any other subject more than once. For convenience, we call in the following Treatment 7 "partner treatment" and Treatment 8 "perfect stranger treatment". The subjects in the partner treatment knew that they were in a stable group for all periods and subjects in the perfect stranger treatment knew that they never meet any group member again. In both treatments subjects knew that the treatment ends after period six.

In total we conducted 4 experimental sessions. In three sessions (S1-S3) we conducted first the perfect stranger treatment.[19] After period six we told the subjects that we will conduct a new experiment. This new experiment was the partner treatment. Subjects in the perfect stranger treatment did not know that there will be a partner treatment afterwards. Subjects in the partner treatment were ensured that after period 6 no other experiment will take place. In one session the order of the treatments was reversed. Again subjects in the first (partner) treatment, did not know that there will be another experiments afterwards, and subjects in the second (perfect stranger) treatment were ensured that after period 6 no other experiment will take place. In the fourth session (S4) 24 subjects participated so that we could form 8 partner groups. Note that in the econometric analysis below we only take the observations from the first treatment in each session because this maximizes the number of independent observations. In total we have then 3 independent perfect stranger-clusters from S1-S3 and 8 independent partner-clusters from S4.[20]

[19] In S1 we had 18, in S2 we had 24, and in S3 we had 21 subjects.

[20] The individual observations within a given perfect stranger treatment may not be independent. Our regression techniques will take this into account because we compute robust standard errors that are based on the assumption that observations across S1-S3 are independent while observations within each of these sessions are not independent. The observations across the 8 partner groups in S4 are clearly independent. Thus, if we pool the data from the perfect stranger treatment in S1-S3 and the partner treatment in S4 we have 11 independent clusters.

IV.1 THE PUBLIC GOOD GAME WITH PUNISHMENT

In both treatments each subject was endowed with 20 tokens in each period and a period consisted of two decision stages. At stage one each group member had to decide simultaneously about the contribution g_i to the project. The contribution had to be between 0 and 20 ($0 \leq g_i \leq 20$). At stage two each subject i was informed about the individual contributions of the other two group members. Then all group members simultaneously made their sanctioning decisions. Subjects were not informed about the other group members' sanctions. Each subject only knew how much he sanctioned the other group members and the *total* amount of punishment that was imposed on him by the two others. The monetary payoff at the end of stage one for each subject i was given by

$$\pi_i = 20 - g_i + 0.5 \sum_{j=1}^{3} g_j .$$

This payoff function implies that full free-riding ($g_i = 0$) is a dominant action in the stage game. This follows from $\partial \pi_i / \partial g_i = -1 + 0.5 < 0$. However, the aggregate payoff $\Sigma \pi_k$ is maximized if each group member fully cooperates ($g_i = 20$) because $\partial \Sigma \pi_k / \partial g_i = -1 + 1.5 > 0$. The punishment technology at stage two was similar to the low-sanction treatment of our PD (with $f = 1$). This means that the cost of punishing another player by one point was one point. Players could assign up to 25 deduction points to each other player.

If the rationality and the selfishness of all subjects is common knowledge we should observe that in the perfect stranger as well as in the partner treatment no subject will ever punish and, hence no subject will contribute. Since punishments are as expensive as in the low-sanction PD treatment, the inequity aversion models of FS and BO also predict no punishments in the two treatments (see our discussion in section III.3). However, since we know already from our previous experiments that there is a substantial fraction of subjects who sanction for non-strategic reasons – and even if they cannot reduce inequity – this prediction is likely to be wrong. The interesting question is, therefore, what treatment differences we may expect when it is known that there are subjects who sanction for non-strategic reasons. In the partner treatment a subject, who sanctions a free-rider, may well have a pecuniary benefit in future periods from this sanction because the free-rider knows *for sure* that in all remaining periods the sanctioning subject will be in his group. By sanctioning in period t the sanctioning subject can thus increase the credibility of the sanctioning threat *for the other members in his own group* in the future periods. This in turn

is likely to increase the cooperation levels of potential free-riders which constitutes a private return for the sanctioning subject. Therefore, in the presence of fair players that can be mimicked, even selfish subjects have an incentive to punish the free-riders in the partner design. Moreover, the fair subjects may have an additional pecuniary reason to punish the free-riders. Only in the final (sixth) period of the partner treatment the pecuniary incentive to sanction vanishes. In the perfect stranger treatment – in contrast – a sanctioning subject does not gain from the future cooperation of a sanctioned subject because the sanctioned subject is never in the same group in the remaining periods. In the partner treatment sanctions can, therefore, be driven by strategic and non-strategic reasons while in the perfect stranger treatment they can only be driven by non-strategic reasons. This means that by comparing the sanctioning patterns in the two treatments we can examine the relative importance of strategic and non-strategic sanctions.

IV.2 RESULTS

It turns out that in both treatments sanctioning is wide-spread.[21] In the partner treatment 30 out of 87 subjects never sanctioned another group member and in the perfect stranger treatment 36 out of 87 subjects never sanctioned. In the partner treatment each subject had 12 opportunities to punish another group member during the six periods, which gave 1044 punishment opportunities for all 87 subjects. In 17.2 percent of these cases the subjects punished. In the perfect stranger treatment the 87 subjects had also 1044 punishment opportunities during the six periods of play and they punished in 20.2 percent of these cases. Thus, the relative frequency of punishment acts is even slightly higher in the perfect stranger treatment. Yet, this fact in itself does not imply that non-strategic sanctions are more important. Perhaps, subjects had more reason to sanction in the perfect stranger treatment because there were more defections.

To take the potential reasons for the sanctions into account we have constructed Figure 6. The figure illustrates the sanctioning pattern for both treatments from the viewpoint of the sanctioning subject. It shows how much the sanctioning subject punishes, on average, another subject in the group *depending on the deviation of the punished subject's contribution from the contribution of the punishing subject*. Notice that in this figure we restrict the analysis to periods 1 to 5 only. The reason is that up to period 5 there

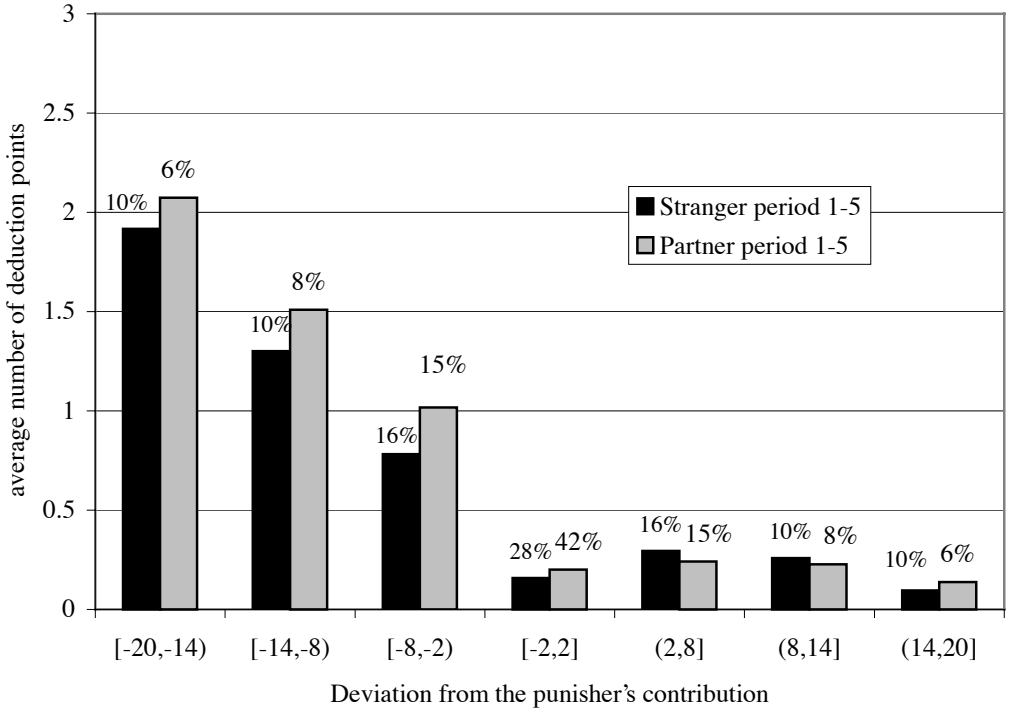
[21] On average a session lasted 1 hours and the earnings per subjects were 20 CHF (US $ 14).

are possible strategic sanctions in the partner treatment while strategic sanctions cannot occur in the final period. The percentage numbers above the bars in Figure 6 show the relative frequency of observations that fall in the respective intervals. For example, in period 1-5 of the perfect stranger treatment 10 percent of the observations exhibit negative deviations between – 20 and –14 from the punisher's contribution level. The numbers above the bars thus indicate the distribution of deviations from the punisher's contribution level in both treatments.

Figure 6 shows that those subjects who deviate negatively from the punisher's contribution level are sanctioned much more heavily than those who contribute more than the sanctioning subject. Nevertheless, as in our PD, we observe again that there are subjects who contribute relatively little (the "defectors") punish those who contribute relatively much (the "cooperators"). This indicates once more the existence of spiteful types. For our purposes, the most interesting fact is, however, that the sanctioning pattern as well as the severity of the sanctions is very similar across treatments. While the punishments are slightly higher in the partner treatment for negative deviations from the punisher's contribution, these differences between the partner and the perfect stranger treatment are rather small. Figure 6 indicates that the strategic sanctioning forces in the partner treatment do neither change the general pattern of punishments nor do they generate considerably more sanctions compared to the non-strategic forces in the perfect stranger treatment.

To examine whether the small differences between the partner and the perfect stranger treatment are significant we ran several regressions with the pooled data of the first five periods of the perfect stranger treatment in S1-S3 and the 8 partner groups in S4 (see Table 5). The first model in Table 5 is our baseline regression where we regress the sanctions subject i imposes on other subjects on the contribution of subject i, g_i, and on the deviation of the sanctioned subject j from the contribution of subject i. Figure 6 indicates that the sanctioning behavior of i is different in the domain of negative deviations (i.e., if $g_j - g_i < 0$) compared to the domain of positive deviations (i.e., if $g_j - g_i > 0$). For this reason our regressions allow for different slopes in the two domains. Moreover, Figure 6 also suggests that there is a discontinuous drop in sanctioning when subject j's contribution is close to the contribution of the sanctioning subject i. We allow for such a discontinuity in our regressions because we use the variable "no deviation" (i.e., if $g_j - g_i = 0$) as a separate regressor.

Figure 6: Average sanctioning decision as a function of the deviation of the punished subject from the contribution of the punisher (periods 1 to 5)



Model 1 shows that a reduction in the contribution level of the sanctioned subject in the domain of negative deviations causes a highly significant increase in the sanctions. For given g_i, a decrease of g_j by 10 units increases the sanction of i by 0.76 on average. In addition, there is a significant drop in the sanction if subject j matches the contribution of subject i. If j increases her contribution in the domain of positive deviations she is also punished significantly less by i, although the decrease in punishment is much smaller than in the domain negative deviations. In Model 1 the sanctioning behavior seems to be independent of the level of i's contribution; only the deviation of j from i's contribution seems to matter.

Model 2 includes, in addition to the previously discussed variables, a dummy P for the partner treatment. This dummy is highly insignificant indicating that we cannot reject the hypothesis of equal sanctions in both treatments. Basically the same message comes out of regression models 3-6. In these regressions we successively interact our deviation variables and the contribution of i with the dummy for the partner treatment. None of these interaction effects turns out to be significant while the baseline effects revealed in Model 1

remain rather robust in models 3-6. In particular, the sign, the size, and the significance level of "negative deviation" is very robust across models.

Taken together Figure 6 and Table 5 suggest that the strategic sanctioning forces are not very strong compared to the non-strategic forces that are already present in the perfect stranger treatment. The lack of significant strategic sanctions is further supported by a closer look at the partner treatment. In Figure 7a we compare the sanctioning pattern in the first five periods of the partner treatment with the pattern in period six. Note that in period six no strategic forces can be operative because subjects knew that period six is the final period. In view of the fact that all punishments in period 6 must be non-strategic in nature, the sanctioning pattern in Figure 7a is quite striking. If anything, subjects impose higher sanctions in the final period than in the previous five periods.[22] The difference between the sanctions in the last period and in the previous five periods is, however, not significant. If we regress the sanctions in the partner treatment on the variables in our baseline regression (i.e., on "negative deviation, no deviation, positive deviation, contribution of the punisher) and include a final period dummy plus interactions of the other variables with the final period dummy, the dummy and the interaction effects are not significant. Nevertheless, the sanctioning pattern in Figure 7a is quite remarkable and suggests – together with the previous evidence – that non-strategic sanctions are of major importance while strategic sanctions do not seem to play a big role in our experiment.

[22] A potential reason for relatively high sanctions in the final period could be that subjects become particularly angry if – after five periods of reasonably high cooperation – a group member defects in the final period. Imagine your response to two different events: (i) a friend betrays you and (ii) a stranger betrays you. We believe that most people would feel more anger if they are betrayed by a friend. While cooperation in a public goods game does not lead to friendship it has been shown to generate social ties (van Dijk and van Winden 1997). Defection in the final period – after successful cooperation in previous periods – may thus cause additional psychic costs for the cooperators in the form of broken social ties.

Table 5: Determinants of sanctioning behavior

Dependent Variable: Sanction of subject *i*.

Independent variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
constant	.476*** (.104)	0.477*** (.106)	.475*** (.111)	.474*** (.112)	.483*** (.126)	.460*** (.127)
negative deviation ($g_j - g_i < 0$)	-.076*** (.017)	-.083*** (.016)	-.074*** (.012)	-.074*** (.013)	-.073*** (.011)	-.072*** (.012)
no deviation ($g_j - g_i = 0$)	-.299*** (.084)	-.291** (.093)	-.207* (.107)	-.203* (.107)	-.226* (.122)	-.232* (.118)
positive deviation ($g_j - g_i > 0$)	-.017*** (.005)	-.017*** (.005)	-.015** (.005)	-.015** (.005)	-.015** (.005)	-.014** (.005)
contribution of punisher (g_i)	-.007 (.007)	-.014** (.005)	-.012* (.006)	-.012* (.006)	-.012* (.005)	-.009 (.006)
dummy for partner: P		.187 (.161)	-.042 (.133)	-.052 (.116)	-.084 (.229)	.114 (.288)
negative deviation x P			-.098 (.056)	-.099 (.057)	-.103 (.066)	-.109 (.066)
no deviation x P					.064 (.249)	.124 (.254)
positive deviation x P				.003 (.017)	.007 (.018)	-.001 (.021)
contribution x P						-.014 (.013)
	N = 870	N = 870	N = 870	N = 870	N = 870	N = 870
	F(4,10) = 12.9***	F(5,10) = 18.8***	F(6,10) = 474.3***	F(6,10) = 52.5***	F(6,10) = 993.0***	F(6,10 = 5.7***
	Adj. R^2 = .11	Adj. R^2 = .11	Adj. R^2 = .12	Adj. R^2 = .12	Adj. R^2 = .12	Adj. R^2 = .12

Notes: Robust standard errors are in parentheses. * denotes significance at the 10-percent level, ** at the 5-percent level, and *** at the 1-percent level. g_j denotes the contribution of the sanctioned subject *j*, g_i is the contribution of the sanctioning subject *i*, and P denotes a dummy for the partner treatment,

To complete the picture we also compare the sanctioning decisions in the final period of the perfect stranger treatment with the sanctioning pattern in the previous five periods of the perfect stranger treatment (see Figure 7b). This is useful to evaluate the argument that subjects in the perfect stranger treatment mistakenly treat the game in the first few periods as a repeated game or that they believe that the sanctioning of defectors in the first few periods will generate indirect pecuniary benefits for them in later periods. Assume, for instance, that subject *i* sanctions *j* in period t, which induces *j* to cooperate in period *t*+1, when *j* is matched with subject *k*. Assume further that the cooperation of *j* in *t*+1 induces subject *k* to cooperate in *t*+2, and that *k* and *i* are matched in *t*+2. If such or similar behavioral chains exist, or are believed to exist, subject *i*'s sanction in the early periods of the perfect stranger treatment may be strategically motivated. If this is the case we should observe less punishment in the final period of the perfect stranger treatment because in the final period sanctions cannot trigger such indirect effects. However, as Figure 7b indicates the general sanctioning pattern as well as the strength of the sanctions in the first five periods is very similar to the sanctioning behavior in the final period.[23] This suggests that subjects sanctions in the early periods of the perfect stranger treatment are not driven by mistaken repeated game beliefs or beliefs in indirect pecuniary benefits.

[23] If we run similar regressions as in Table 5 with the perfect stranger data only, and include final period dummies or interact final period dummies with the other variables, no significant effects emerge. The final period dummy as well as the interaction terms are insignificant.

Figure 7a: Average sanctioning decision in the partner treatment

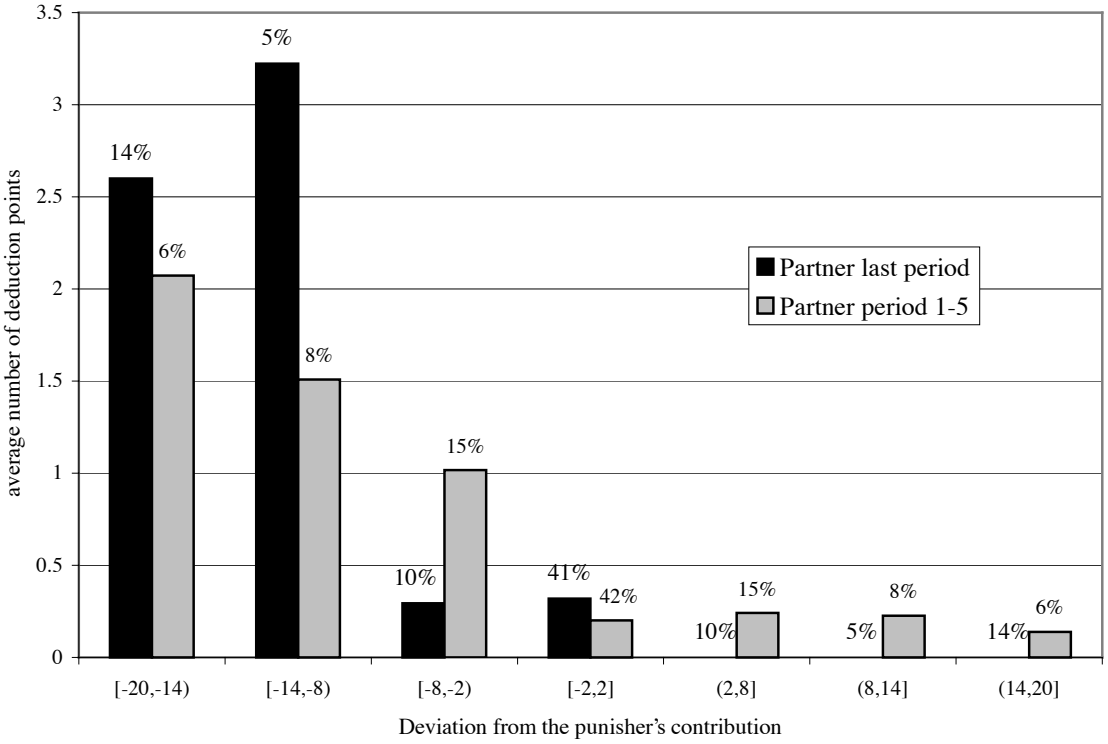
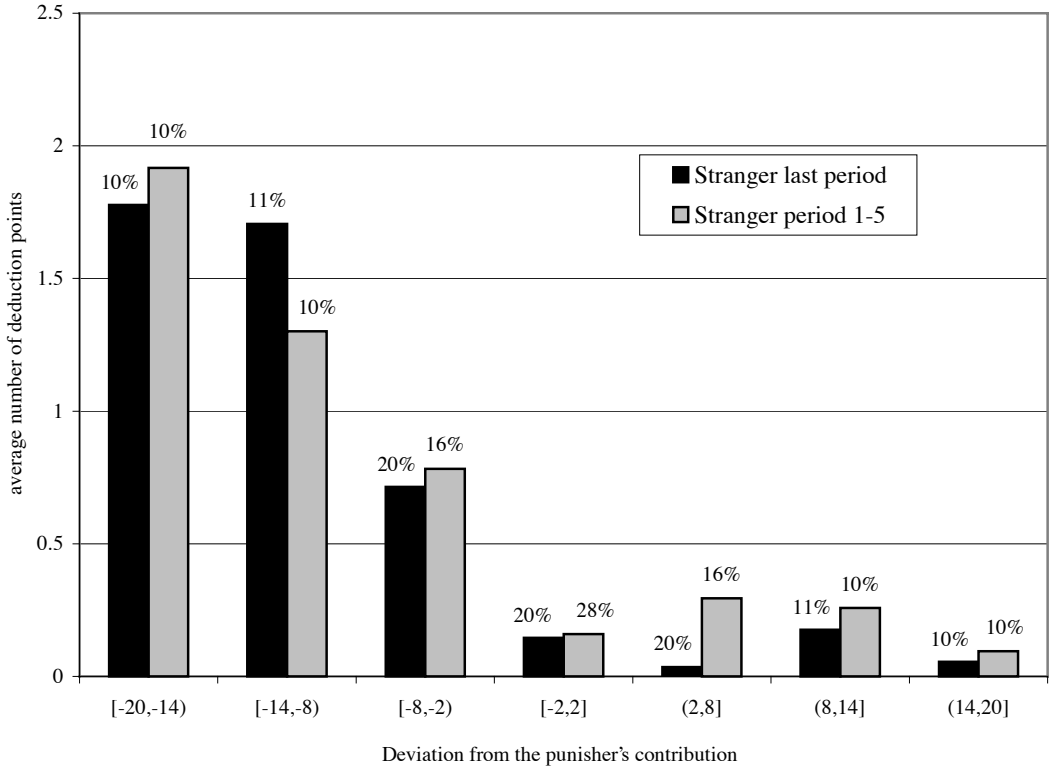


Figure 7b: Average sanctioning decision in the perfect stranger treatment



V. SUMMARY AND CONCLUSIONS

Informal sanctions are the cement of a society's social fabric. They maintain the viability of a myriad of informal agreements that are crucial for social and economic success. They also form the basis of social norms that are essential for the functioning of markets, organizations, the law, families and neighborhoods. The understanding of the driving forces underlying informal sanctions is, therefore, of great importance. We find surprisingly little evidence for strategic sanctions that are imposed to create future material benefits. While non-strategic sanctions are of major importance in our experiments, strategic sanctions seem to play a negligible role. Our results also indicate that – within the class on non-strategic sanctions – the violation of fairness principles is a major driving force of informal sanctions. However, the sanctions of a significant minority of the subjects seem to be driven by a different motive – spitefulness. Our results also suggest that – within the class of fairness-driven sanctions – retaliatory sanctions are most important.

These results have implications for recently developed theories of fairness. Since the violation of fairness principles constitutes the most important source of informal sanctions, descriptively correct fairness theories that capture the major fairness motives are indispensable for the further understanding of informal sanctions and their implications. This supports the relevance of fairness theories. Within the class of fairness theories, those that are based on the notion of retaliation, are most promising and capture the bulk of fairness driven sanctions.

Yet, our results also show that all fairness theories are incomplete to the extent to which they neglect spitefulness. If one does not take into account the presence of spiteful types one is likely to make serious predictive errors. For instance, in the presence of spiteful agents shirking will in general be much more pervasive in principal-agent relations compared to situations in which there are only selfish and fair agents.

We also believe, that the absence of evidence in favor of strategic sanctions, is a challenge for the theory of repeated games because this theory is largely based on the notion that people sanction for strategic reasons. We certainly do not want to claim that the lack of evidence for strategic sanctions in our experiments indicates a general lack of strategic sanctions. However, if the absence of strategic sanctions would turn out to be a robust and replicable fact in future studies, the theory of repeated games would probably need major modifications.

References

Bolton, Gary E., and Axel Ockenfels (2000): "A Theory of Equity, Reciprocity and Competition", *American Economic Review*, XC, 166-193.

Bowles, Samuel and Herbert Gintis (2000): "The Evolution of Strong Reciprocity", Working Paper, Santa Fe Institute.

Bowles, Samuel and Herbert Gintis (2000): "Social Capital and Community Governance", *Economic Journal*, forthcoming.

Brandts, J. and Charness, G. (1998): "Hot versus Cold: Sequential Responses and Preference Stability in Experimental Games", Discussion Paper, Universidad Autonoma de Barcelona.

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

Cason, T. and Mui, V. (1998): "Social Influence in the Sequential Dictator Game", *Journal of Mathematical Psychology*, forthcoming.

Charness, Gary and Matthew Rabin (2000), "Some Simple Tests of Social Preferences and a New Model", Discussion Paper, University of California, Berkeley.

Coleman, James (1990): *Foundations of Social Theory*, The Belknap Press of Harvard University Press.

Dufwenberg, Martin and Georg Kirchsteiger (1999): "A Theory of Sequential Reciprocity", Discussion Paper, CentER, Tilburg University.

Ellison, G. (1994): "Cooperation in the Prisoner's Dilemma with Anonymous Random Matching", *Review of Economic Studies* 61, 567-588.

Falk, A. and Fischbacher, U. (1999): "A Theory of Reciprocity", Working paper No. 6, Institute for Empirical Research in Economics, University of Zurich.

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

Fehr, E. and Schmidt, K. (1999): "A Theory of Fairness, Competition and Cooperation", *Quarterly Journal of Economics*, CXIV, 817-851.

Fischbacher, U. (1999): "Z-tree. Zurich Toolbox for Readymade Economic Experiments – Experimenter's Manual" Working Paper No. 21, Institute for Empirical Research in Economics, University of Zurich.

Frank, Robert (1994): *Microeconomics and Behavior*. New York: W.W. Norton & Company

Francis, Hywel (1985): "The Law, Oral Tradition and the Mining Community", *Journal of Law and Society* 12, 267-271.

Friedman, Daniel and Nirvikar Singh (2000): "On the Viability of Vengeance", Working paper, University of Santa Cruz.

Friedman, Daniel (2001): "Some Issues on Modeling Human Cooperation", Working paper, University of Santa Cruz.

Fudenberg, Drew and Eric Maskin (1986): "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information", *Econometrica* 54, 533-556.

Gintis, Herbert (2000): "Strong Reciprocity and Human Sociality", *Journal of Theoretical Biology* 206, 169-179.

Güth, W., Schmittberger, R. and Schwarze, B. (1982): "An Experimental Analysis of Ultimatum Bargaining", *Journal of Economic Behavior and Organization* 3, 367-88.

Güth, Werner (1995): "An Evolutionary Approach to Explaining Cooperative Behavior by Reciprocal Incentives", *International Journal of Game Theory* 24, 323-344.

Güth, Werner and Eric van Damme (1998): "Information, Strategic Behavior and Fairness in Ultimatum Bargaining: An Experimental Study", *Journal of Mathematical Psychology* 42, 227-247.

Henrich, Joseph and Robert Boyd (2000): "Why People Punish Defectors – Weak Conformist Transmission can stabilize costly Enforcement of Norms in Cooperative Dilemmas", Working Paper, UCLA.

Kagel, John and Katherine Wolfe (2000): Testing between Alternative Models of Fairness: A New Three Person Ultimatum Game, Working Paper, University of Pittsburgh.

Kandel, Eugene and Edward Lazear (1992): "Peer Pressure and Partnerships", *Journal of Political Economy* 100, 801-817.

Kandori, M. (1992): "Social Norms and Community Enforcement", *Review of Economic Studies* 59, 63-80.

Kirchsteiger, Georg (1994): "The Role of Envy in Ultimatum Games", Journal of Economic Behavior and Organization 25, 373-389.

Knez, Marc and Duncan Simester (1998): "Firm-wide Incentives and Mutual Monitoring", Graduate School of Business, University of Chicago.

Knack, Stephen and Peter Keefer (1997): "Does Social Capital Have an Economic Payoff? A Cross-country Investigation", *Quarterly Journal of Economic*s, 112, 1251-1288.

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

Levine, David K. (1998): "Modeling Altruism and Spitefulness in Experiments", *Review of Economic Dynamics* 1, 593-622.

Mui, Vai-Lam (1995): "The Economics of Envy", Journal of Economic Behavior and Organization 26, 311-336.

Nowak, Martin, Karen Page and Karl Sigmund (2000): "Fairness versus Reason in the Ultimatum Game, *Science* 289, 1773-1775.

Putnam, Robert (1993): *Making Democracy Work: Civic Traditions in Modern Italy*, Princeton: Princeton University Press.

Rabin, Matthew (1993): "Incorporating Fairness into Game Theory and Economics", *American Economic Review* 83, 1281-1302.

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

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

Sampson, Robert J., Stephen W. Raudenbusch, and Felton Earls (1997): "Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy", *Science* 277, 918-924.

Sethi, Rajiv and E. Somanathan (2000a): "Preference Evolution and Reciprocity", *Journal of Economic Theory* 97, 273-297.

Sethi, Rajiv and E. Somanathan (2000b): "Understanding Reciprocity", Working paper, Columbia University.

van Dijk, Frans and van Winden, Frans (1997): "Dynamics of social ties and local public good provision", *Journal of Public Economics* 64, 323-341.

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

Appendix 1

This appendix provides an english translation of the instructions for the 3-person Prisonner's Dilemma.

General Instructions for Participants

You are now taking part in an economic experiment, which is financed by various research foundations. Please read the following instructions very carefully. If you have questions, please don't hesitate to ask us.

The instructions are intended solely for your personal information. **During the experiment, communication is not allowed.** Any violation of this rule will lead to the exclusion both from the experiment and from all payments.

During the experiment we talk about points instead of Francs. Your earnings will, therefore, be calculated in points. At the end of the experiment the total amount of points you earn will be converted into Francs at the rate of

1 point = 0.60 Francs.

Upon completion of the experiment, you will be paid out **in cash** the equivalent of the points earned plus 10 Francs for participating in this experiment.

At the beginning of the experiment, all participants are divided into groups of three. Apart from yourself your group will, therefore, consist of two further members. Neither before nor after the experiment you will get to know the identity of the other persons in your group.

The experiment consists of **two stages**. At stage one you decide whether you want to **contribute your points to a project or whether you want to keep your points.** At stage 2 you can decide whether you want to decrease the other group members' income from stage one and, if so, by how much. This is done by allocating **deduction points**. In the following we describe the procedure of the experiment in detail.

Stage one:

At the beginning of stage one each participant receives **20 points**. You will then have to decide if you wish to contribute the 20 points to the **project** or if you wish to keep them. The two other group members have to make the same decision. They also have the choice either to contribute their 20 points or to keep them. Please keep in mind that you have only the two options: either to contribute 20 points to the project or to keep them. As soon as all three group members have made their decision, the incomes earned at stage one are determined. They are calculated as follows:

How is you income calculated at stage one?

The income of each group member is calculated in the same way. The income consists of two parts:

(1) the points kept (**"Income from points kept"**)
(2) the **"Income from the project"**. The income from the project is

0.6 * (total sum of the contributions to the project)

Therefore, your total income is determined according to the following formula:

Your total income =

Income from points kept + Income from the project =

(20 – your contribution to the project) + 0.6*(total sum of contributions to the project).

If you contribute nothing to the project, your "Income from points kept" is 20. If you contribute 20 points to the project, your "Income from points kept" is zero. At the same time the total sum of contributions to the project is increased and so is your "Income from the project".

Specifically, your contribution of 20 points to the project results in the following: You achieve an income from the project of 12 points. At the same time, the other two group members' income also increases by 12 points each, since they receive the same income from the project as you do. Consequently, if you contribute your 20 points to the project, the group's income from the project increases by 36 points. Of course, your income also increases by 12 points if another group member contributes to the project.

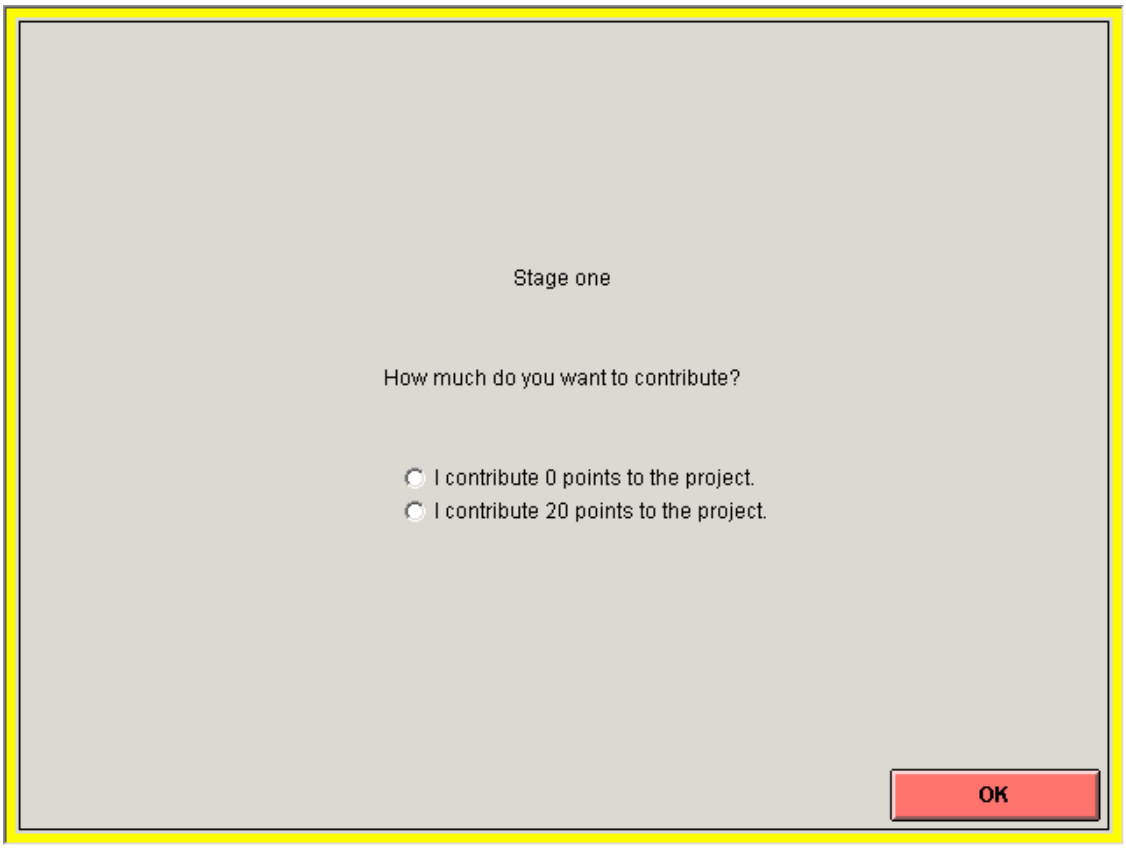
To clarify things we present some examples:

If all three group members contribute 0 points to the project, all three achieve an "Income from points kept" of 20. Since no one contributes, no one gains anything from the project. The total income of each group member is 20 points.
*Calculation: (20 - 0) + 0.6 * (0) = 20*

If all three group members contribute 20 points, a total of 60 points is contributed to the project. Then each member's "Income from points kept" is zero but each member receives an income from the project of 0.6 * 60 = 36 points.
*Calculation: (20 - 20) + 0.6 * (60) = 36*

If you contribute 20 points and a second member also contributes 20 points, but the third member contributes 0 points, the following incomes are realized. Since you and the second member contributed 20 points, both you and the other group member earn 0.6 * 40 = 24 points from the project. The third, non-contributing, member also receives 24 points from the project as well as an additional 20 points from his "Income from points kept", i.e. 20 + 0.6 * (40) = 44 points.
*Calculation for you and the second group member: (20 - 0) + 0.6 * (20 + 20) = 24*
*Calculation for the third group member: (20 - 0) + 0.6 * (20 + 20) = 44*

This is the screen where you make your decision at stage one:



You make your choice on this screen. In the center of the screen you see the following two rows: "I contribute 0 points to the project" and "I contribute 20 points to the project", respectively. Next to both rows is a field where you can click with your mouse. For your decision click on the respective field. When you have made your decision, click on the OK button at the lower right hand side. As long as you did not press this button, you may revise your decision.

As soon as all group members have made their decision, the second stage follows.

Stage two:

At stage two you may **decrease** the income of **every** other group member **by allocating deduction points**. You may also **leave the incomes unchanged**. The other group members can also reduce **your** income. The exact procedure is detailed below. First, we describe the payoff consequences which emerge by allocating deduction points.

How is your income at stage 2 calculated ?

If you allocate deduction points to another group member, this member's income **exactly** decreases by **the number of deduction points you allocated**. Suppose you allocate 5 deduction points to a member. His income will then be decreased by 5 points. If you allocate 9 deduction points, his income will be decreased by 9 points, etc.

If you allocate 0 deduction points to a group member, his income will remain unchanged.

You can at most allocate **25 deduction points to each of the other group members.**

If you allocate deduction points, you have to bear **costs**. For each point you allocate you have to bear costs of one point. Thus if you allocate 11 deduction points, you have costs of 11 points, etc.

If you allocate no deduction points you will, of course, have to bear no costs.

Your total income arising from both stages is as follows:

Your total income at the end of stage 2 =	(Income from stage 1) –
	- (received deduction points) -
	- Cost of deduction points you allocated

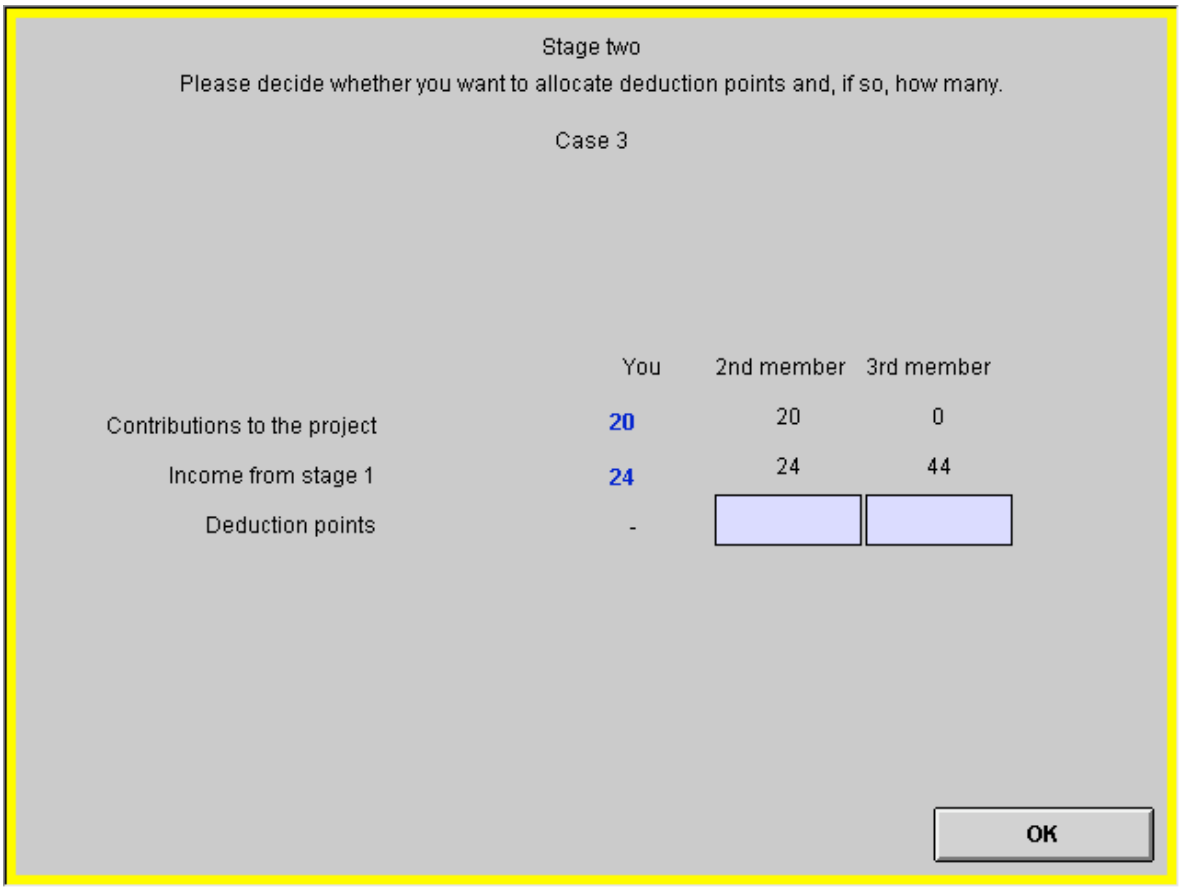
Your total income at the end of stage two consists, therefore, of 3 components: (1) Your income from stage one, (2) the deduction points received from other group members, and (3) the costs of deduction points you allocated.

How do you make your decision at stage two?

At stage one all participants determine their contribution to the project. **Before you get to know the other group members' decisions** (i.e., whether they contributed or not) you must state how many points you wand to deduct for each possible eventuality. There are four possible cases:

Case 1:	the other two group members contributed 0 points.
Case 2:	the second group member contributed 0 points, the third group member contributed 20 points
Case 3:	the second group member contributed 20 points, the third group member contributed 0 points
Case 4:	the other two group members contributed 20 points.

For all four possible cases you have to decide whether you want to allocate deduction points to the other group members, and if so, how many. In order to make your choice you will be given **4 input screens** representing exactly the four cases. As an example take the third screen (case 3):



This screen illustrates case 3 where the second member contributed 20 points and the third member contributed 0 points to the project. As far as your own contribution is concerned, let's assume you contributed 20 points. You now have to decide for this case whether you want to allocate deduction points to the others, and if so, how many. The screen is set up as follows:

The **first row** indicates the "Contributions to the project".

The **second row** ("Income from stage 1") indicates each group member's income which results from the decisions made at stage one. Below the header "You" you see *your* income, and in the second and third columns you see those of the other two group members. Based on the contributions according to case 3, the second member earns an income of 24 and the third member earns an income of 44 points at stage one.

In the **third row** ("Deduction points") you have to make your decision for stage 2: You must decide how many deduction points you want to allocate to **each other** group member. Insert a number between 0 and 25 in each corresponding input field. An input is definitely required. If you don't want to change another group member's income, insert "0". If you want to allocate, say, 11 deduction points, insert "11", etc. **You can allocate a maximum of 25 deduction points to each of the other group members.**

First, you make your decision for case 1. When you have completed your input, the second screen appears (case 2), followed by the third (case 3) and the fourth screen (case 4). Thus, in total you make on four screens **8** decisions regarding the allocation of deduction points.

Of course, **only one combination** of contributions was chosen from the other two group members. This means that only one of the four possible cases actually arises. For your payment it is, therefore, relevant how you decided in this specific case.

Suppose case 1 occurred (both the second and the third group member contributed 0 points): This implies that your decision on the first screen is relevant. This means that the deduction points you chose in case 1 determine your payment and the payments of the other group members. If, on the other hand, case 3 arises, the decisions you made on the third screen are relevant for the payment, etc.

Do you have any questions?

Control-Questionnaire:

You must answer each question. Please write down the whole calculation! If you have questions please ask the experimenters!

1. Each group member has 20 points at his disposal. At stage one, no one (including yourself) contributes to the project. What is

   your income from stage one? .....
   the other members' income from stage one? .....

2. Each group member has 20 points at his disposal. At stage one you contribute 20 points to the project. The other two group members also contribute 20 points to the project at stage one. What is

   your income from stage one? .....
   the other members' income from stage one? .....

3. Each group member has 20 points at his disposal. Both you and the third group member contribute 0 points. The second group member contributes 20 points. What is

   your income from stage one? .....
   the second member's income from stage one? .....
   the third member's income from stage one? .....

4. Each group member has 20 points at his disposal. Both you and the second member contribute 20 points. The third group member contributes 0 points. What is

   your income from stage one? .....
   the second member's income from stage one? .....
   the third member's income from stage one? .....

5. You want to allocate 6 deduction points to a group member who contributed 20 points to the project . What costs do you have to bear?

   costs for allocating the deduction points? .....

6. You want to allocate 10 deduction points to a group member who has contributed nothing to the project . What costs do you have to bear?

   costs for allocating the deduction points? .....

Appendix 2

This appendix provides the translation of the instructions for the responders in our baseline ultimatum game (Treatment 3). The instructions in all other ultimatum games were identical except for the payoff numbers.

Instructions for Participants B

You are now taking part in an economic experiment, which is financed by various research foundations. The instructions are intended solely for your personal information. **During the experiment, communication is not allowed.** If you have questions, please direct them to us. Any violation of this rule will lead to the exclusion both from the experiment and from all payments.

The experiment is very short. You can earn points, which upon completion of the experiment will be converted into Francs at the rate of

1 point = 1 Franc.

At the beginning of the experiment one half of the participants will be "participants A", the other half will be "participants B". **You are a participant B.**

The experiment consists of two stages:

Stage one:
At stage one participant A proposes an income distribution between A and yourself. A has two possibilities:

A can either propose (5:5), i.e., A gets 5 points and you get 5 points, or
A can propose (8:2), i.e., A gets 8 points and you get 2 points.

Stage two:
At stage two you can either **accept** or **reject** A's proposal. If you accept, A's proposed income distribution becomes effective. If you reject, both you and A receive zero points.

There are 4 possible outcomes:

A makes the proposal (5:5), which you accept. Both, you and A get 5 points.
A makes the proposal (5:5), which you reject. Both, you and A get 0 points.
A makes the proposal (8:2), which you accept. You get 2 points and A gets 8 points.
A makes the proposal (8:2), which you reject. Both, you and A get 0 points.

Before you, as a participant B, are informed about A's proposal, you have to indicate **for both possible proposals** whether you want to accept or to reject it. This means that you must decide for both proposal (5:5) and proposal (8:2), whether you want to accept or reject it. Relevant for your payment is only the offer, which A has actually made. **"Your" participant A will get to know your decision only after he has made his proposal.**

At the end of stage two the experiment is over and you get to know your income in points.

When you finished reading these instructions, please raise your hand.