The economic implications of reciprocity in teams and markets

Elke Renner

2005

DISSERTATION

zur Erlangung eines Grads einer Doktorin der Wirtschaftswissenschaft (Dr. rer. pol.) der

Universität Erfurt
Staatswissenschaftliche Fakultät

Gutachter: Prof. Dr. Bettina Rockenbach
Prof. Dr. Manfred Königstein

Datum der Disputation: 19.12.2005

urn:nbn:de:gbv:547-200600171
[http://nbn-resolving.de/urn/resolver.pl?urn=nbn%3Ade%3Agbv%3A547-200600171]

Acknowledgements

Many thanks to Klaus Abbink, Christoph Albrecht, Andrea Bäcker, Simon Gächter, Mario Gruppe, Özgür Gürerk, Sebastian Händschke, Bernd Irlenbusch, Manfred Königstein, Jürgen Kumbartzki, Thomas Lauer, Rosemarie and Peter Meyer, Mark Peacock, Christiane Pilz, Bettina Rockenbach, Karim Sadrieh, Paul and Dorothea Schindegger, Martin Sefton, Martina Stamm, Jean-Robert Tyran, Anne Ulmke, Katrin Uschmann, Gerhard Wegner, Tim Wenniges and Egon Zimpel.

Contents

1	Introduction and overview	6
1.1	References	8
2	Leading by example in the presence of free rider incentive	9
2.1	Introduction	9
2.2	Related literature	10
2.3	Design and procedures	10
2.3.1	Design	10
2.3.2	Procedures	12
2.4	Leading by example in a one-shot experiment – a benchmark	13
2.4.1	Results	13
2.5	The dynamics of leadership and voluntary cooperation	15
2.5.1	Does the leader matter?	15
2.5.2	Are there good and bad leaders?	17
2.5.3	The impact of increased gains from cooperation	19
2.5.4	Leading by example?	22
2.5.5	Shall I be a bold or timid leader?	23
2.6	Concluding remarks	24
2.7	References	25
3	When in doubt... – Cooperation and punishment under incomplete information	27
3.1	Introduction	27
3.2	Experimental design and procedures	29
3.3	Results	30
3.3.1	Contribution behaviour	30
3.3.2	Punishment behaviour	33
3.4	Welfare implications	37
3.5	Conclusions	38
3.6	References	40
4	Group size and social ties in microfinance institutions	41
4.1	Introduction	41
4.2	Related experimental studies	43
4.3	Model and experimental design	44
4.3.1	The model	45
4.3.2	Treatments	46
4.3.3	Experimental procedures	47
4.4	Results	48

4.4.1	The effect of the group size	50
4.4.2	Dynamics of play	50
4.4.3	The effect of social ties	51
4.4.4	Gender effects	52
4.5	Summary and discussion	52
4.6	References	54
5	Price rigidity in customer markets	56
5.1	Introduction	56
5.2	An experimental approach to price rigidity in customer markets	57
5.3	Experimental design	58
5.3.1	General description of the experimental design	58
5.3.2	Procedures, parameters and information conditions	59
5.3.3	Predictions	60
5.4	Results	62
5.5	Concluding remarks	68
5.6	References	69
Appendix 1.A: Instructions		71
Appendix 1.B: Tables 1.a and 1.b		77
Appendix 2.A: Instructions		79
Appendix 2.B: Tables 2.a – 2.d		85
Appendix 3.A: Instructions		89
Appendix 4.A: Instructions		91
Appendix 4.B: Screens announcing the cost shock in period 3		97
Appendix 4.C: Figure 4.a		98

1 Introduction and overview

The question of how to design efficient economic organizations is a key issue in economic analysis. Traditionally economists have built this analysis on theoretical models assuming fully rational and purely selfish agents. This assumption has a clear advantage in its simplicity which allows building tractable mathematical models. However, it has often been criticised as unrealistic. In the last few decades a growing number of researchers in economics have investigated actual human behaviour in controlled laboratory experiments. These experiments explore the boundaries of human rationality in search for alternative behavioural assumptions that can feed back into the development of new models with more explanatory power.

One central finding in this field is the prominent impact of reciprocity. Many experiments have demonstrated that people are not unboundedly selfish but care about others' well being. They reward kind acts (positive reciprocity) and they punish unkind acts (negative reciprocity). The existence of positive reciprocity has been demonstrated in gift exchange and trust games (e.g. Berg, Dickhaut and McCabe, 1995 and Fehr, Kirchsteiger and Riedl, 1993). Negative reciprocity can be observed in rejection behaviour in ultimatum games (Güth, Schmittberger and Schwarze, 1982) and in punishment behaviour in public good games (Fehr and Gächter, 2000). In the moonlighting game (Abbink, Irlenbusch and Renner, 2000) positive and negative reciprocity are observed simultaneously. From these and many other experiments reciprocal behaviour can be considered as a stylized fact (Camerer 2003).

This thesis investigates the economic implications of reciprocity in teams and markets. The environments studied in the next four chapters share several common features. First, in all of them, traditional theory suggests opportunistic behaviour by individuals will undermine efficiency. In the team and group settings investigated efficiency requires individuals to cooperate, but they have private incentives to free ride; in the market setting we investigate efficiency requires consumers to trust sellers not to supply sub-standard products, but moral hazard problems give consumers good reason to distrust sellers. Second, in the four environments we study, the inefficiency problem may be mitigated through reciprocity. These chapters together show that whether in fact reciprocity considerations can overcome team inefficiency and market failure depends on several factors. First, and most obviously, individuals must be sufficiently reciprocal in nature. Second, they must have the information that allows them to act on their reciprocal motives in a way that improves group outcomes. Third, they must have instruments at their disposal to effectively monitor and discipline individuals who act opportunistically.

A more detailed outline of the specific research questions is as follows: First, does reciprocity give rise to leading by example in teams and do leaders make a difference? Second, can peer punishment enhance cooperation in teams when it is difficult to identify free riders? Third, can reciprocity sustain group lending schemes in microfinance institutions? Fourth, can reciprocal loyalty help overcome market failures in markets with incomplete information about quality? Each of these four questions will be addressed in separate chapters. The next section gives a brief overview.

The second chapter examines the impact of positive reciprocity in the form of "leading by example" in the context of a simple team cooperation problem with free rider incentives. In a public goods game one player was assigned to the role of the leader and had to choose how much to contribute to a linear public good. His contribution was announced to three followers who then decided simultaneously on their contribution decisions. If followers are sufficiently strongly motivated by reciprocity, their contributions will increase with the leader's contribution. By allowing the leader to set an example, we provide the leader with an instrument to harness these motivations to benefit the whole group. We find that leader and follower contributions are positively correlated even in a one-shot game, testifying to the relevance of reciprocity. Despite this, we find that the presence of a leader in a repeated leadership game did not enhance the overall level of cooperation relative to a situation with an absent leader. Even though on average it pays to be a bold leader because sufficiently many people follow the leader's example, the leaders who do so earn less than followers and leaders are not willing to tolerate this: they resist being taken for a sucker and the only way they can resist is to reduce their contributions.

The third chapter examines negative reciprocity in the form of peer punishment in a public goods game. Previous research has shown that reciprocally motivated players are willing to punish free riders, and so the opportunity to punish provides an instrument for disciplining free riders and improving group outcomes. Here we study the effectiveness of this instrument when it is difficult to identify free riders. The experimental setup discriminates between the ability to observe and the ability to evaluate others' contributions to a public good. In both treatments contributions are observable but only in one treatment are the individual endowments publicly known before punishment decisions are taken. When endowments are publicly known punishment exerts a strong disciplining effect and cooperation rates are high. In contrast, when endowments are private information, cooperation rates are substantially lower and punishment has virtually no effect. Surprisingly this is not due to a reluctance to punish; rather those who are punished do not change their behaviour in response to punishment. We conclude that transparency about others' endowments is crucial for successful cooperation.

The fourth chapter demonstrates that reciprocity can sustain group lending schemes in microfinance programmes. These programmes provide poor people with small loans given to jointly liable self-selected groups. In a typical microfinance scheme borrowers with individual risky projects form groups which apply for loans together. The whole group is liable if any group members default. Thus, joint liability provides insurance against individual risks but relies on reciprocal solidarity alone. We show that reciprocity based group lending outperforms individual lending. This result is robust with respect to group size and social ties.

The last chapter analyses reciprocity in customer markets. Customer markets are characterized by long-term relationships between buyers and sellers that evolve if buyers trust sellers to provide high quality and if sellers are trustworthy. However, changes in the terms of this implicit contract may antagonize customers and disrupt the relationship. We experimentally show that mutually beneficial long-term relationships frequently prevail in markets for experience goods. However, we also show that price rigidity after a temporary cost shock is much more pronounced if price increases cannot be justified by cost increases. Hence, long-term relationships in customer markets mitigate market failure of the "lemons" type, but are prone to price stickiness.

Taken together the four chapters of this thesis provide evidence for the power and limitations of reciprocity and illustrate the significant effect alterations in institutional rules can exert on outcomes. Positive reciprocity is strong enough to sustain microfinance institutions and helps to mitigate market failure in customer markets. Substantial positive reciprocity is also observed in leadership games. However, the effect is not strong enough to attain efficient team outcomes. Substantial negative reciprocity helps to sustain cooperation in peer punishment games. Yet, if there is a lack of transparency about others' capabilities to cooperate punishment is detrimental.

1.1 References

Abbink, Klaus, Bernd Irlenbusch, and Elke Renner (2000): The Moonlighting Game – An Experimental Study on Reciprocity and Retribution. Journal of Economic Behavior and Organization, 42, 265-277.

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

Camerer, Colin (2003): Behavioral Game Theory: Experiments in Strategic Interaction, Princeton University Press

Fehr, Ernst and Simon Gächter (2000): Cooperation and Punishment in Public Goods Experiments, American Economic Review, 90 (4), 980-94.

Fehr, Ernst, Georg Kirchsteiger, and Arno Riedl (1993): Does Fairness Prevent Market Clearing? An Experimental Investigation. Quarterly Journal of Economics, 108, 437-459.

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze (1982): An Experimental Analysis Ultimatium Bargaining. Journal of Economic Behavior and Organization, 3 , 367-88.

2 Leading by example in the presence of free rider incentive[1]

„Once you as a CEO go over the line, then people think it‘s okay to go over the line themselves.“
Lawrence Weinstein, Head of Unisys
(quoted after The Economist, July 27, 2002, p.58)

2.1 Introduction

A central question in leadership research concerns how a leader can effectively influence his followers. Of the many types of such influence one of the most common and elementary is leading by example. Public figures often demonstrate exemplary behaviour in order to motivate people to carry out unpleasant but socially desired activities: acts of blood donation and giving to charity are often publicized in order to motivate other people to follow suit. Role modelling is a widespread phenomenon frequently used in order to enhance followers' motivation, especially if the desired behaviour cannot be enforced by other means like rewards or penalties. Setting a bad example, on the other hand, is held to detract from the willingness of others to cooperate: a boss cannot leave the office before the end of the working day while demanding that his staff works overtime; neither can a manager ask employees to defer expected pay increases while awarding him a rise.

In this paper we approach the behavioural impact of "leading by example" with a series of experimental studies. Our basic design involves a simple team cooperation dilemma (or linear public goods game) in which complete free riding is a dominant strategy but the total surplus would be maximized if all players contribute their whole endowment. We implement leading by example as a sequential decision process: In treatments with a leader, one member of a group is selected to be the person who decides first on his contribution to the team project. The decision of the leader is conveyed to the followers, who then decide privately on their own contribution. To separate the effect of leadership experiments without a leader, where all members of a group decide independently serve as control treatments. It is straightforward that the mere existence of a leader does not affect incentive structures. Thus, assuming selfish, money maximizing individuals, the game theoretic prediction of non-cooperation is the same for both settings and leadership should be irrelevant.

We believe that our simple leadership game allows us to isolate one important, and hitherto rather neglected aspect of leadership, namely the role of example setting in the presence of free rider incentives. Our simple game isolates exactly this aspect without confounding it with other features of leadership that surely are important in reality, like managerial skills, charisma, coordination, signalling etc.

We ran two series of experiments, a strictly one-shot leader game and a finitely repeated leader game. The former gives us a benchmark on the importance of a non-strategic following of the leader's example. The second experiments will allow us to assess the dynamics of leadership.

The main objective of our experiments is to examine i) whether this leader-effect can really be observed ii) whether it positively affects the behaviour of the leader and subjects in the leader role raise their level of cooperation and iii) whether the presence of a leader enhances efficiency and groups with leader achieve higher cooperation rates.

Our main results are as follows. We find that many followers indeed follow the leader's example and contribute the more to the team project the more the leader puts on the line. This happens already in a strictly one-shot experiment where strategic incentives for following the example are absent. This observation testifies the potential relevance of leading by example. We also find that this result carries over to the repeated leader game, which we set up to study the dynamics of leadership. Given the effectiveness of leading by example, we expected to find an improved efficiency in the repeated

---

[1]This chapter is based on the working paper "Leading by example in the presence of free rider incentives", joint work with Simon Gächter.

leadership game. Yet, it turned out that leaders were not able to improve the efficiency of cooperation relative to a situation where leaders were absent. The reason is that being a leader means being a sucker, since followers follow their leader's example only half-heartedly, i.e. they contribute systematically less than the leader. As a consequence leaders reduce their cooperation over time, with the consequence that followers reduce as well. The question arises whether it pays to be a bold leader. Empirically, it turns out, that the correlation of leader and follower behaviour is strong enough to make it worthwhile on average to be bold and contribute high amounts. But a large variance remains.

We proceed as follows. In the next section we put our work into perspective of the previous literature. Section 2.3 describes the basic decisions situations and the experimental procedures. Section 2.4 provides the one-shot leadership experiment that will provide the benchmark for the relevance of leading by example. Section 2.5 will investigate the dynamics and difficulties of leading in the presence of free rider incentives. Section 2.6 concludes.

2.2 Related literature

Although there is a huge literature on various aspects of leadership, there is to our knowledge only one economic approach to the phenomenon of leading by example. Benjamin Hermalin (1998) provides a signalling model in which a leader can induce following by moving first. If the leader is privately informed on the productivity parameter of his team's project, he can increase welfare through expanding his effort and thus, leads by example. Meidinger, Sinclair-Desgagné and Villeval (2002) provide experimental evidence on Hermalin's model and find that leadership works more through reciprocity than through signalling. Weber, Camerer, Rottenstreich and Knez (2001) study leadership in a coordination game. Although they primarily focus on the attribution of leadership quality and whether subjects exhibit the fundamental attribution error, they also find that a leaders' speech only slightly increases efficiency in large groups, whereas small groups coordinate well, independent of the presence of a leader.

Since we focus on a type of leadership that induces following not by solving coordination games or incentive problems assuming perfectly rational agents, we mainly consider the experimental literature on related issues. The studies most akin to our set-up are the experiments by Moxnes and van der Heijden (2000) and van der Heijden and Moxnes (2000). They introduce leadership in a public bad context using - similar to our study - a sequential decision protocol. They find that followers invest less in a public bad when a leader is introduced but this effect is attenuated if the leader faces no costs for leadership.

There is also a literature that investigates the importance of leadership giving charitable fund raising (Andreoni 2001; Potters, Sefton and Vesterlund 2001), in politics (Arce M. 2001; Rhodes and Wilson 1999) and strategic management (Miller 1990; Foss 2001).

2.3 Design and procedures

2.3.1 Design

Our basic design involves a simple team cooperation dilemma (or linear public goods) game. Each of the four team members has to decide on how many out of 20 tokens to keep and how many tokens to contribute to a team project. For simplicity, the size of the team project is just the sum of all contributions to it. The payoff for each subject is the following:

$$\pi_i = 20 - g_i + \alpha \sum_{j=1}^{4} g_j \qquad 0 < \alpha < 1 \tag{1}$$

In our experiments, the marginal per capita return of a contribution to the team project, α, was either $\alpha = 0.4$ or $\alpha = 0.6$ (for details see below). Comparing the marginal payoffs it is obvious that a rational individual has an incentive to free ride on the other group member's contributions. Regardless of what the other group members contribute, every individual is better off by keeping all tokens for him- or herself, since the marginal payoff of a token contributed to the team project is only $\alpha < 1$ compared to a marginal payoff of 1 from keeping the token. Thus, assuming that subjects maximize their own monetary income, the dominant strategy in this game is to contribute nothing. A social dilemma arises, since it is in the joint interest of the group to contribute the whole amount to the team project but each team member has an incentive to free ride.

The leader game is a very simple variation of this standard public goods game. One subject that we henceforth will call the "leader" decides first in his or her team. The other three group members (called the "followers") learn about the leader's contribution to the team project and then decide simultaneously about their contributions to the team project.[2] The payoff functions of all team members, including the leader, are identical and equal to (1). It is straightforward that the equilibrium strategy of full free riding is independent of the decision of others. The only Nash equilibrium is full free riding by all subjects.

All our experiments were conducted with two basic treatment variations: Experiments without a leader provide the baseline to compare the effect on cooperation in the presence of a leader. Table 2.1 gives an overview of all treatments that we employed in this study.

Tab. 2.1: Treatment overview.

	Benchmark experiment: *One-shot* public goods game with or without leader α = 0.4			
Treatment	Criterion for leader selection	With leader	No leader	No. of subjects (groups)
0	Randomly	20 groups	8 groups	112 (28)
	Dynamic leadership experiments: *Repeated* public goods game with and without a leader α = 0.4			
Treatment	Criterion for leader selection	Period 1-10	Period 11-20	No. of subjects (groups)
1	Randomly	With leader		48 (12)
2a	Highest contributor in the group in period 1-10 ('pro-social')	No leader (Control for treatment 1)	With leader	36 (9)
2b	Lowest contributor in the group in period 1-10 ('free rider')	No leader (Control for treatment 1)	With leader	36 (9)
	Dynamic leadership experiments: *Repeated* public goods game with and without a leader α = 0.6 – The impact of increased gains from cooperation on leaders and followers			
Treatment	Criterion for leader selection	Period 1-10	Period 11-20	No. of subjects (groups)
3	Randomly	With leader		36 (9)
4a	Highest contributor in the group in period 1-10 ('pro-social')	No leader (Control for treatment 3)	With leader	20 (5)
4b	Lowest contributor in the group in period 1-10 ('free rider')	No leader (Control for treatment 3)	With leader	20 (5)

[2] In the experiment we never talk of "leaders" and "followers", but instead of the subject that decides first in his or her group. See the instructions in appendix B for further details.

Treatment 0 is a benchmark study. Here we investigate the leader game in a purely one-shot situation and compare it to a simultaneous one-shot public goods game without a leader. Since this experiment is just played once, there is no strategic reason whatsoever for the followers not to free ride, if they are interested in maximizing their payoffs. A leader who anticipates this, has as well no reason to contribute to the team project. Thus, our one-shot game measures the extent and effectiveness of leading by example in a very simple team leader game that requires voluntary cooperation to yield efficiency.

Treatments 1 – 4 take the one-shot leader game into a strategic context of a repeated team game. The same group of team members plays the same team game for ten times. This study allows us to assess the dynamics of leadership in our context. We can investigate whether leaders are able to lead their groups toward efficiency over time. To check for the robustness of our findings, we also vary the gains from cooperation by changing α: In treatments 1 and 2, $\alpha = 0.4$; in treatments 3 and 4, $\alpha = 0.6$.

The main design of our experiments in treatments 1 – 4 is again very simple. Our subjects played a sequence of ten team games with a leader and a sequence of ten team games without a leader. During the whole experiment groups remained unchanged. The subject who had to decide first in a group (i.e., the “leader”) was the same person during all leader periods.

Treatment 1 is an experiment with a leader.[3] The subjects who had to decide first in their group were chosen randomly before the first period started.

Our treatment 2 will change the perspective of ‘equally qualified leaders’ and investigate whether some leaders are better than others. To that end, subjects first played a no-leader and then a leader sequence. This has the added advantage that the groups without leaders from the first sequence can serve as control groups for the groups with leaders in treatment 1. To address the impact of different types of leaders on the leadership problem, the leaders in the second treatment were chosen according to their contribution in the first sequence. Before period 11, subjects with either the highest or the lowest average contribution in their group in the previous periods were assigned to the leader role. We will discuss the exact rationale for these design features below.

The main purpose of treatments 3 and 4 is to check for the impact of increased gains from cooperation on leaders’ and followers’ contribution behaviour. The experiments were conducted in exactly the same way as in treatments 1 and 2. The only difference was that $\alpha = 0.6$, which increased the social gains from cooperation ($4\times\alpha$) from 1.6 to 2.4.

2.3.2 Procedures

All experiments were conducted in the computerized experimental laboratory at the University of Erfurt. In total 308 students from various disciplines participated in our experiments. In total we had 77 independent groups of four subjects. The last column of table 2.1 contains information how many participants (and groups of four subjects) took part in the various treatments.

The subjects were encouraged to participate through posters on the campus. We used the experimental software z-tree developed by Urs Fischbacher (1999). At the beginning of each session the participants were randomly assigned to one of the booths in the laboratory. The booths visually separated the participants and ensured that every individual made his or her decision anonymously and independently. The written instructions were handed out in which the above social dilemma situation and the experimental procedures were explained in detail. The groups were randomly assigned, such that subjects did not know which of the other participants were in their group. To make sure that every participant understood the decision problem, they had to answer a set of control questions. We did not start the experiment before all subjects had correctly answered all control questions.

In all treatments 1 – 4 subjects received the instruction only for the condition in the first ten periods

[3] We also played ten periods with no leader after the leader experiments were finished. We do not report these results here, since we do not need the observations from these experiments for our analysis.

(with or without a leader respectively). Before the second sequence began, they were informed that the experiment will continue for another ten periods and the corresponding instructions were handed out. In our benchmark experiment (treatment 0) subjects received only one set of instructions, since they made only one decision in either the game with or without a leader.

Immediately after each session subjects were paid anonymously in cash. On average, our subjects earned € 10.3 in the one-shot experiments and € 9.5 in treatments 1 – 4. The one-shot experiments lasted approximately 40 minutes and the repeated game experiments took 75 minutes.[4]

2.4 Leading by example in a one-shot experiment – a benchmark

The set of experiments that we describe in this section will provide the benchmark for our analysis. In these experiments the leader game is played exactly once. One of the four team members ("the leader") is randomly selected to make the first contribution decision. The rationale for the random leader selection is straightforward. In our experiment, all team members have equal incentives (see (1)) and no further qualifications to be a leader are necessary. Therefore, ex ante, each team member is equally qualified to be the leader. By the one-shot design we rule out any reputation considerations, which might be decisive in repeated interactions. Thus we have a very strong test for a leadership effect since followers can free ride on the leader's contribution without fearing any influence on future contribution behaviour of either the leader or any other member in their group.

2.4.1 Results

Figure 1a shows the average contribution of followers as a function of their leader's contribution (i.e., each dot represents a group). The more the leaders in this one-shot experiment contribute, the higher are the followers' contributions. An interesting observation is that the followers, in some groups, contribute on average more than the leader (i.e., their mean contribution lies above the diagonal). Yet, this holds only for contributions ≤ 10. For contributions above ten, followers in all groups contribute less than the leader. On average, followers contribute more for leader contributions above ten than below ten. Overall, the correlation of leaders' and follower's contributions is positive. A simple linear regression of mean follower contribution on the respective leader contribution yields a highly significantly positive coefficient (t = 3.76). The more conservative Spearman rank correlation coefficient also yields ρ = 0.56, which is significantly positive at p = 0.01.

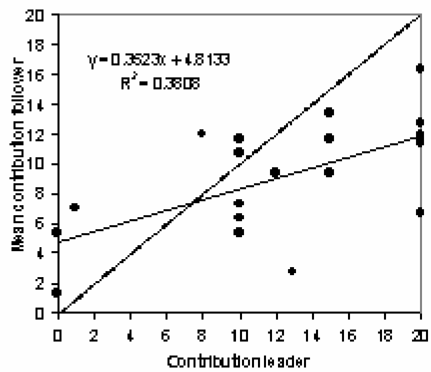


Figure 1a. Mean contributions of followers after the observation of their leader's contribution. Trend line of linear regression.

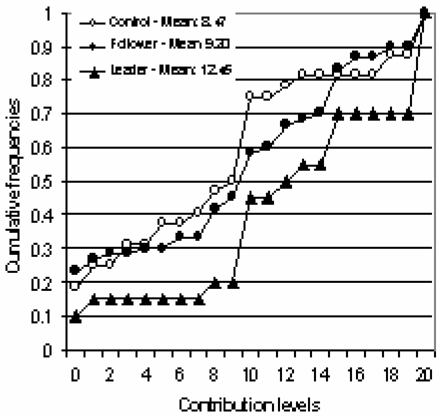


Figure 1b. Cumulative contribution functions of contributions in the control-treatment without a leader and of followers and leaders in experiments with a leader

[4] We deliberately had relatively high stakes in the one-shot game to ensure that our subjects take their one-shot decision very seriously.

The observation from Figure 1a is consistent with recent findings in public goods experiments that many subjects are willing to cooperate if others cooperate as well (see, e.g., Croson 1998; Keser and van Winden 2000; Clark and Sefton 2001; Fischbacher, Gächter, and Fehr 2001; Fischbacher and Gächter 2003).

Does this leader effect induce leaders to raise their level of cooperation and thus, can leading by example lead to higher overall cooperation rates? Figure 1b gives the answer by showing the cumulative contribution functions of subjects in the control treatment without a leader and of followers and leaders in the teams with a leader. On average, subjects in the teams without a leader contributed 8.5 tokens; the followers in the teams with a leader invested 9.2 tokens and the leaders put 12.5 tokens into the project. The cumulative distribution function of leaders' contributions lies below the cumulative distributions of both the control experiment and the follower contributions, which indicates that leaders contribute more than followers and control subjects. A nonparametric Wilcoxon rank-sum test shows that the contributions of subjects in the leader role are significantly higher compared to the contributions of subjects in the control groups ($p = 0.028$, two-sided). As can be seen from Figure 2, the modal choice of the leaders was to contribute the maximum possible amount of 20 tokens, which was chosen by 30 percent of the leader subjects, compared to only 12.5 percent of full contribution choices in the control treatment. Only 10 percent of the leaders contributed zero to the team project; whereas the followers and the control subjects did so in 19 and 23 percent of the cases, respectively. On average, followers contributed only slightly more to the team project than the benchmark subjects. Moreover, the followers' average contribution fell short of the leaders' contribution by an average of 24 percent. This difference is significant (Wilcoxon matched-pairs test, $p = 0.01$, two-sided). The average contribution only slightly but insignificantly ($p = 0.401$, Wilcoxon rank-sum test, two sided) increased from 8.5 tokens in the control groups to 10.0 tokens in the teams with a leader. Hence we find a tendency to higher contributions but no significant effect of leadership on the overall cooperation rates. We summarize these results from our benchmark experiment as

Observation 1: Followers follow their leader's example in the one-shot team game: Contributions of followers and leaders are significantly positively correlated. Leaders contribute more than subjects in control groups and also more than their followers. Followers contribute insignificantly more than control subjects. Overall cooperation rates in one-shot interactions are slightly but insignificantly higher in the presence than in the absence of a leader.

The fact that leaders contributed more than control subjects suggests that leaders even in the one-shot experiment correctly anticipated that their contribution will influence their followers' contribution. How profitable was it for leaders to choose their respective contribution? Did it pay to be a bold leader who contributes the whole amount or would timidness and low contributions have paid off? As Figure 1 shows there is indeed a positive correlation between the leaders' contribution and the average follower contribution. Yet, is the observed relation strong enough to secure profits above 20 (which is the minimum payoff from full free riding)? From the payoff function we can calculate the necessary strength of followers' contribution behaviour such that a leader's mean payoff increases in his contribution to the team project. Assume that the estimated linear relation between a leader's and his or her average follower's contribution is $\bar{g}_f = c + \beta g_l$. Plugging this relation into the payoff function (1) gives us the following relation between the leader's payoff, π_l, and his or her own contribution g_l:

$$\pi_l = 20 + 3\alpha c + (3\alpha\beta + \alpha - 1)g_l \qquad (2)$$

From (2) we can easily calculate the critical c^* and β^*, such that (i) the leader's mean payoff is at least 20 (which he would get from full free riding) and that (ii) the leader's payoff increases in g_l. Setting $d\pi_l/dg_l = 0$ and solving for β gives

$$\beta^* = \frac{1-\alpha}{3\alpha} \qquad (3)$$

With $\alpha = 0.4$ is straightforward to see that $c^* = 0$ and $\beta^* = 0.5$; i.e., if $\beta > \beta^*$ it pays for the leader on average to contribute his or her endowment; if $\beta < \beta^*$ he or she earns more if by contributing nothing.[5]

[5] Of course, if $\alpha < 0$, the critical β^* has to increase to secure a mean payoff > 20 for all leader

As Figure 1 shows, c = 4.81 and β = 0.35. From this it follows that the average leader would have fared better, in terms of his material payoff, if he or she had contributed zero tokens to the team project.

From the viewpoint of a follower, there are two sources of "successful" leading by example. First, he directly benefits from the leader's contribution g_l. Secondly, if another follower follows the leader's example and increases his contribution (on average according to $g_f = c + \beta g_l$ with $\beta > 0$), then all followers benefit from an increased contribution of other followers. Both arguments are straightforward from the follower's payoff function (adapted from (1) and setting $\alpha = 0.4$):

$$\pi_f = (20 - 0.6g_f) + 0.4g_l + 2\times 0.4(c + \beta g_l). \qquad (4)$$

The follower's mean payoff net of the leader's contribution (which is $\pi_f - 0.4g_l$) is simply increasing in β if $\beta > 0$; the mean marginal follower payoff in an increase in g_l. is 0.8β (i.e., in our case (see Figure 1) this is 0.8×0.35 = 0.28). Of course, (4) says just that the mean follower payoff will be increasing in g_l, and nothing about significance. The Spearman rank correlation between the leader's contribution and the followers' average payoff net of the leader's contribution is 0.56, which is highly significant (p = 0.0111).

Observation 2: Timid leaders earned more than bold leaders. Follower's payoff net of the leader contribution increased significantly in the leader's contribution.

In summary, leading by example works even in a one-shot context. Yet, timid and greedy leaders fare better than bold and generous leaders in terms of the payoff they can secure for themselves.

How does leading by example work in case of repeated interactions, which characterize many leadership problems in reality? What is the dynamics of leadership in the presence of free rider incentives? We will investigate these issues in the following section. Given that (i) there is a positive correlation between the contributions of leaders and followers even in a one-shot game, and (ii) leaders contributed more than control subjects, which suggests that they were aware of their impact on followers' contribution behaviour, there should be the possibility that the leaders, by setting the right example, can lead their groups toward more efficient contributions than in simultaneous choice setups. In the repeated games this should really be feasible, since we have seen that already in the one-shot game contributions are (albeit insignificantly) higher in the presence than in the absence of leaders. Yet, we have also seen that this will be no easy task, since in the one-shot experiment timid leaders, who did not set a good example and contributed little or nothing, fared better than generous leaders. Even if the leaders are willing to set a bold example, the benchmark experiment has shown that they are in jeopardy of being "exploited" by their followers, if the followers only half-heartedly cooperate. Thus, the real dynamics of leadership in the presence of free rider incentives and conditional cooperators is an empirical question, which we try to enlighten in the remainder of this paper.

2.5 The dynamics of leadership and voluntary cooperation

In the discussion of our results from the experiments in treatments 1 – 4, we will proceed as follows. We will first look at how cooperation evolved over time in the presence and absence of a leader in treatments 1 and 2. Second, we will investigate whether there are different types of leaders and whether the type of leader makes a difference. Third, we analyze the impact of increased gains from trade by looking at treatments 3 and 4. Fourth, we will check whether we find evidence for leading by example and how the results from the one-shot experiment carry over to the repeated games. As a final step we will take up again the question whether it pays more to be a bold or a timid leader.

2.5.1 Does the leader matter?

We start with the evolution of cooperation in the presence and absence of a leader in treatment 1, where the leader is randomly selected. We expect that contributions in the presence of a leader are higher than in their absence. Figure 2 shows the average evolution of contributions in groups where in the leader was randomly determined. For the sake of comparison, we also include the cooperation levels of the control groups (from treatments 2a and 2b) where there was not yet a leader in the first

---

contributions $g_l > 0$.

sequence (see table 2.1).

Leaders contributed more than subjects in the control groups and more than their followers. On average a leader contributed 10.8 tokens. Control subjects contributed only 6.68 and followers 6.54 tokens. A Wilcoxon rank-sum test and a Wilcoxon matched-pairs test rejects the null hypothesis of equal contribution of leaders compared to subjects in the control groups and to followers respectively (significant at $p < 0.01$, one-sided).

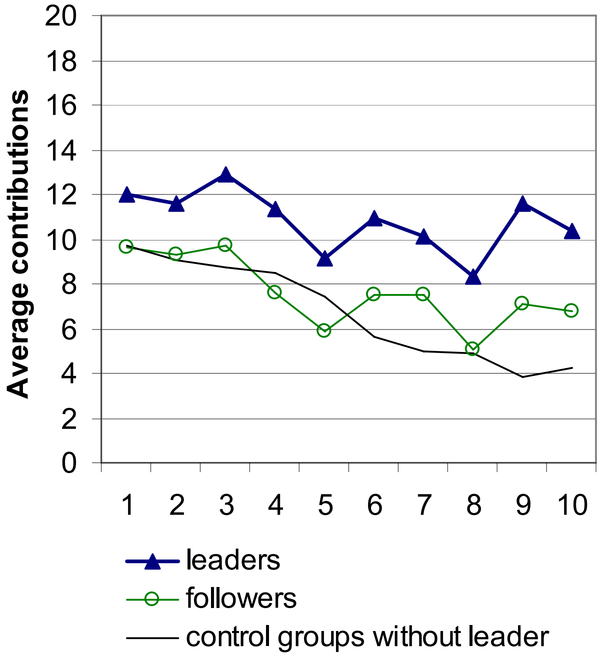


Figure 2. Average contribution over time of leaders, followers, and control groups without a leader

Overall, and contrary to our expectations, the presence of a leader seems not to have mattered much. This is very remarkable, given the correlation of leader and follower contribution that we have observed in the one-shot experiment. The average cooperation rate in the experiments with a leader was not significantly higher than in the control experiment (Mann-Whitney test, two-sided, $p = 0.80$). In the control experiments, subjects contributed 6.67 tokens to the team project; in the experiments with a leader contributions amounted to 7.61 tokens. We summarize our results as

Observation 3: Randomly determined leaders contribute more than followers and also more than the subjects from the control teams without a leader. Yet, the overall cooperation rate is not significantly higher in the presence than in the absence of a leader.

Observation 3 seems to have a sober implication: leaders do not matter in the sense that, in their presence, efficiency of team production is not increased relative to a team situation where they are absent. However, one might object that in reality leaders are usually not randomly selected, but pre-selected according to some pre-defined criteria. Yet, in contrast to many real life situations, leaders in our setup need no special managerial "leadership skills". In our experiment, every subject seems to be equally skilled ex ante. This justifies our random leader selection. Nevertheless, the equal skill argument might not be true because some leaders may in fact even in this simple situation be better suited than others to be the leader. The reason for this conjecture is that evidence suggests that people differ in their cooperative inclinations. Some people are free riders and others are conditional cooperators. For instance, Fischbacher et al. (2001) and Fischbacher and Gächter (2003) find that a quarter to a third of the subjects can be classified as free riders, and roughly 50 percent are conditional cooperators.[6] These types of subjects may differ in their 'leadership qualities'.

[6] See Ockenfels (1999); Gächter, Herrmann and Thöni (2003); Falk and Fischbacher (2002), and

Here is why. As the above results show, randomly determined leaders contributed more than followers. The follower's contribution in all periods fell short of the leader's contribution. In other words, the leaders got "suckered" in each period. Different types of players may be differently suited to stand this situation. Intuitively, one might assume that those subjects who are willing to cooperate in the simultaneous team games (the 'pro-socials'), might be the better leaders, because they are maybe more willing than others to sacrifice payoffs for the sake of the group, or because they feel less suckered than others. Another reason could be that they are conditional cooperators, who cooperate when others cooperate and who are therefore willing to set the example of being a cooperator. Apparently, setting an example works on average, as our results so far have shown. By contrast, free rider types might not be willing to set an example and to contribute to the team project. Remember that more than 20 percent of our followers and control subjects in the one-shot experiments contributed exactly zero to the team project. Maybe, if they were the leaders, low contributions would emerge. One may argue that making a free-rider type the leader of the group is like "setting the fox to keep the geese". When the leader is randomly selected, it is inevitable that we sometimes "select a fox", and sometimes a cooperative leader. Our result about ineffective leaders might then be due to an aggregation effect that hides the true differences in relevant leadership skills. Our following experiments test these arguments.

2.5.2 Are there good and bad leaders?

To test the argument about 'leader quality', we look at the experiments of treatments 2a and 2b in which, in the first ten periods, subjects played a simultaneous team game with no leader (see table 2.1). We then ranked the subjects in each group according to their average contribution to the team project of their group over all ten rounds. We call a subject with the highest contribution over all ten periods in his or her group the 'pro-social' type. Subjects with the lowest contribution are called 'free rider' types.[7] In treatment 2a we made the pro-social the leader in the second sequence; in treatment 2b the "free rider" became the leader. Subjects were not aware of this ranking and the determination of the leader. All subjects were only informed that they would learn whether they are the leader or not. This avoids any influence of the leader determination rule on the followers' behaviour, and allows us to determine the mere leader effect that is not confounded with the information about the leader's (in-)famous past. For a similar reason, subjects did not know about the leader experiments in the second sequence, which avoids confounding different contribution types and their anticipation of possible consequences of the leader or follower role.

Figure 3 depicts the evolution of cooperation in the first and second sequence. In the first sequence we show the average contribution paths of our two different future leader types. We also show the average contribution path of our future followers. The differences in average contribution levels between pro-socials and free riders are remarkable. In the initial period, average contributions differ by almost ten points. Pro-socials always contribute more than free riders, and even strongly increase their contributions in the final period. The difference in contribution patterns suggests that our classification of types is sensible.

Kurzban and Houser (2001) for related results.

[7] This procedure implies that the classification of types is relative to the average contribution of a particular group. This is justified because group composition stays the same over all 20 periods, i.e., with and without a leader.

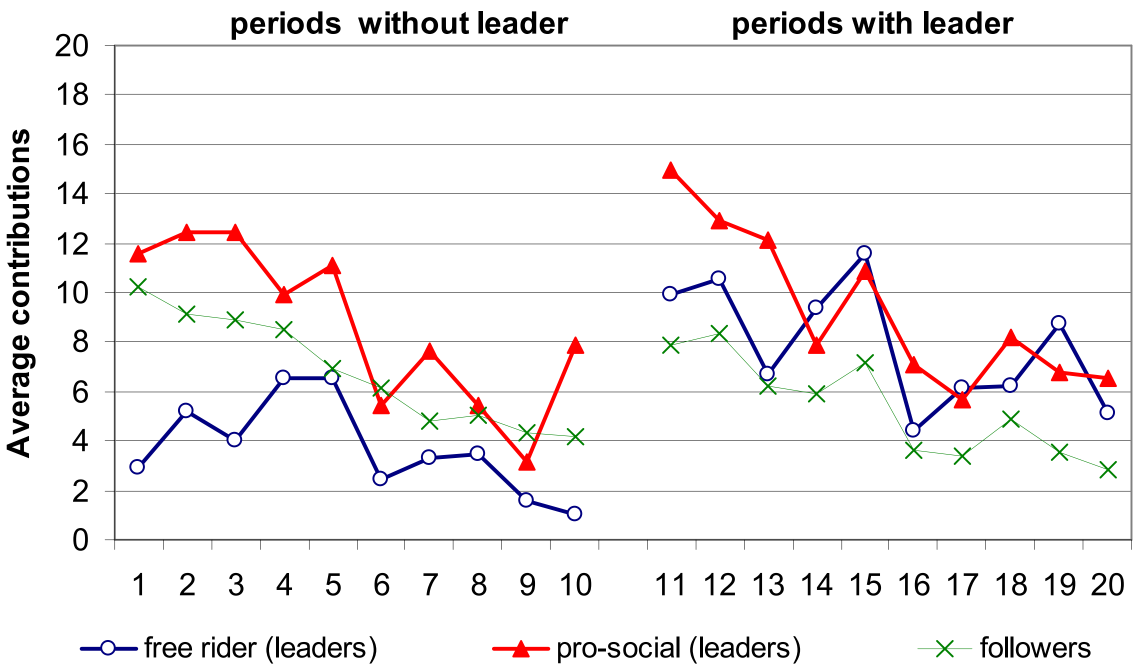


Figure 3. Average contributions in treatments 2a and 2b of different leader types (free riders and pro-socials) and followers in periods 1-10 (before they are leaders and followers, resp.) and period 11-20 (when they are leaders and followers, resp.).

The second sequence is the one we are mainly interested in here because there is now a team leader in periods 11-20. We show the average contribution of the pro-social and free rider leader, respectively, and the average follower contribution.[8] A first main observation is that both types of leaders contribute more than followers, which is consistent with our previous findings.

But who is the better leader? Is it true that making a former free rider the leader is like 'setting the fox to keep the geese'? Much to our surprise, free riders behaved more like 'wolves in sheep's clothing'. Once they became leaders, they immediately and significantly changed their contribution behaviour: on average they contributed 3.7 tokens in the first sequence and jumped upward to 9.9 tokens in period 11. When we compare this to the first period, then the immediate increase in contributions is even more remarkable. The change is highly significant (p = 0.0079, two-sided Wilcoxon signed ranks test). Thus, this change is more than just a restart effect. On average, free rider leaders contributed 7.87 tokens in the second sequence, which is an increase of almost 114 percent compared to the first sequence! The null hypothesis that the average contribution of the free-rider leaders is the same in the first and the second sequence can be rejected at p = 0.01 (Wilcoxon matched-pairs test).

Are pro-socials better leaders? Figure 3 suggests not. When pro-socials become leaders, they do not change their behaviour substantially. They only slightly but insignificantly increase their contributions from an average of 8.7 in the first to 9.3 tokens in the second sequence (Wilcoxon signed ranks test, p = 0.441). For pro-socials we cannot exclude that their increase in contributions is just a restart effect. When we compare period 1 and period 11 contributions, the change is positive, but insignificant (Wilcoxon signed rank test, two-sided, p = 0.127). We keep these findings as

Observation 4: Free riders strongly increase their contribution rate when they become leaders, whereas the pro-socials only slightly increase their contributions, once they are leaders.

A comparison of pro-social leaders and free-rider leaders shows that the former contribute slightly more (on average 1.43 tokens), but this difference is not significant (Mann-Whitney test, two-sided, p = 0.453). In other words, once a leader, the '(in-)famous past' does not matter for the leader's behaviour. The average contribution behaviour of pro-social and free rider leaders is also not significantly

[8] We pool the follower contributions, because they are basically identical.

different from the contribution of randomly selected leaders (Kruskal-Wallis test, $p = 0.236$). Thus, we have no evidence for the existence of 'good' and 'bad' leaders, when we define 'good' and 'bad' by the past contribution behaviour.

As with randomly determined leaders, cooperation rates are not significantly higher than in the control experiments without a leader. This holds for both leader types.[9] All types of leaders are apparently rather ineffective in stimulating higher levels of voluntary cooperation than is achieved in the absence of a leader. We summarize our results in the following

Observation 5: There are no evident differences between leader types: former free riders show similar contribution rates as former pro-socials, or randomly selected leaders. Neither type of leader is able to induce significantly higher cooperation levels than in simultaneous team games.

2.5.3 The impact of increased gains from cooperation

How robust are the results of Observations 3 – 5 with respect to gains from cooperation? To investigate this question we look at the results from treatments 3 and 4, which were exactly identical to treatments 1 and 2 except that the marginal per capita return α is 0.6 instead of 0.4 in the former treatments. The social gains from cooperation are therefore increased from 1.6 to 2.4. Yet, to free ride is still a dominant strategy, which implies that there should not be any difference in cooperation behaviour between treatments 1 and 2 and 3 and 4, respectively. However, it is known from previous experiments that the higher α is the higher are contributions to the public good (see, e.g., Ledyard 1995). Thus the cooperation problem is eased when the gains from cooperation increase. The question therefore is how increased gains from cooperation change the results from Observations 3 – 5.

Three arguments suggest that there will be a significant leader effect on cooperation rates. First, if we can replicate the positive influence of α on contributions, it should be easier for followers to contribute since the 'loss' they incur from contributing, $1 - \alpha$, is lower if $\alpha = 0.6$ than if $\alpha = 0.4$. Second, the same argument holds of course also for the leader who, in addition, should feel less suckered if followers have a higher willingness to contribute. Third, if $\alpha = 0.6$ the critical β (see (3)) is $\beta^* = 0.22$ compared to $\beta^* = 0.5$ if $\alpha = 0.4$. Put differently, since β measures the followers' marginal increase to a contribution by the leader, already a rather small β suffices to make it worthwhile on average for the leader to contribute his or her whole endowment. Since we have observed that in the one-shot benchmark experiments $\beta = 0.35$ (see Figure 1), the threshold of $\beta^* = 0.22$ should be easily met.

We report the results of the experiments of treatments 3 and 4 in Figures 4 and 5, respectively. Several interesting observations can be made from Figures 4 and 5. Qualitatively, the main findings from the $\alpha = 0.4$ experiments carry over to the $\alpha = 0.6$ experiments. As in treatment 1 the randomly determined leaders of treatment 3 (Figure 4) contribute more on average than followers and subjects in the control experiments.

[99] We look at the group averages in sequence 1 and compare the group averages of groups with a pro-social leader and groups with a free-rider leader, respectively. The Wilcoxon-matched pairs test returns $p = 0.549$ for groups with a pro-social leader and $p = 0.515$ for groups with a free-rider leader.

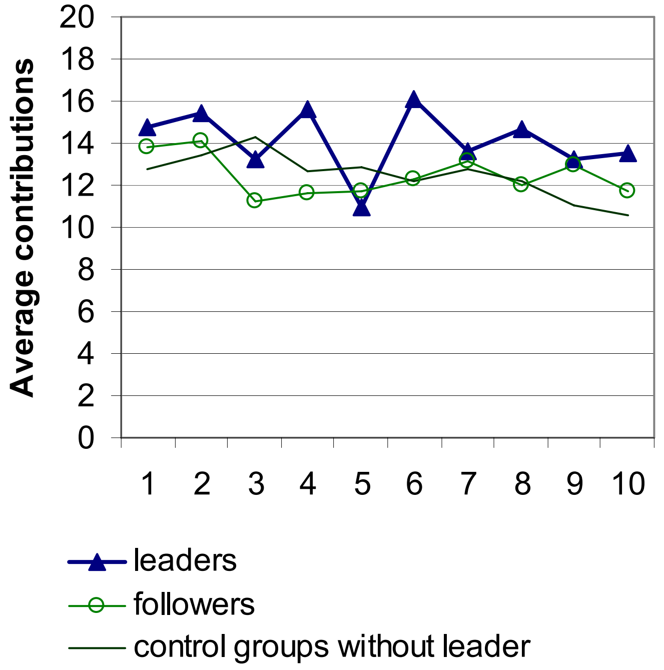


Figure 4. Average contribution over time of leaders, followers, and control groups without a leader

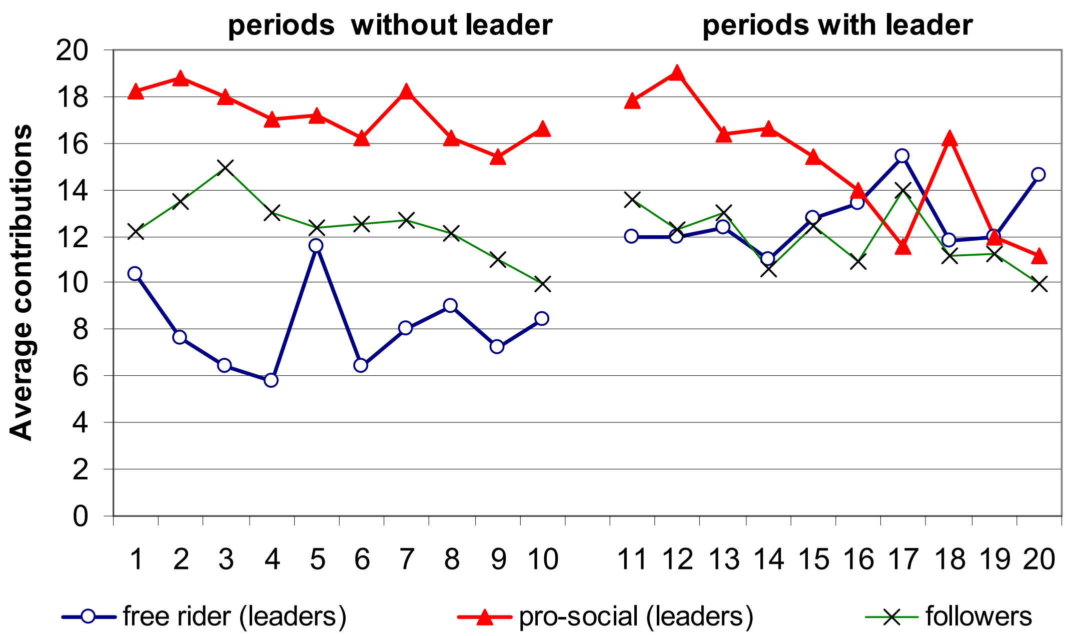


Figure 5. Average contributions in treatments 4a and 4b of different leader types (free riders and pro-socials) and followers in periods 1-10 (before they are leaders and followers, resp.) and period 11-20 (when they are leaders and followers, resp.) when $\alpha = 0.6$.

However, the difference in contributions of leaders and followers is clearly diminished in the experiments of treatment 3. Statistical tests reveal that leaders in treatment 3 contribute significantly more than followers ($p = 0.008$) but only insignificantly more than control subjects ($p = 0.400$). This stands in contrast to the results from treatment 1 where leaders contribute significantly more than followers and control subjects.

In treatments 4a and 4b (Figure 5), we find that, as in treatments 2a and 2b, pro-socials and free riders show strongly different cooperation patterns in sequence 1. As in treatments 2a and 2b we find that, once they are leaders, the pro-socials do not change their contributions significantly ($p = 0.225$), whereas the free-riders significantly increase their contributions ($p = 0.042$). Leaders contribute insignificantly more than followers, but significantly more than control groups.

A comparison of Figures 4 and 5 with Figures 3 and 4 suggests that there is a strong level effect in contributions. The increased gains from cooperation if $\alpha = 0.6$ exert a strong influence on the overall level of cooperation. Mann-Whitney U-tests that compare cooperation levels find that in groups with and without a leader, average cooperation is significantly higher in $\alpha = 0.6$ experiments than in $\alpha = 0.4$ experiments (all p-values < 0.05).

In summary, consistent with our conjectures, we find that the leadership problem is reduced if the gains from cooperation increase. First, as we have just seen, the overall cooperation rate is significantly higher when $\alpha = 0.6$ than when $\alpha = 0.4$, irrespective of whether there is a leader or not. Second, the leader is indeed less of a sucker under high gains from cooperation. In the $\alpha = 0.4$ experiment the followers' contribution falls short of the leaders' contribution by 3.74 and 3.21 tokens, respectively. By contrast, in the $\alpha = 0.6$ experiments the difference between the leaders' and the followers' average contribution is reduced to 1.67 in sequence 1 and 1.97 in sequence 2.

Figure 6 looks at the contribution of followers relative to the leaders as a function of the leader's contribution. In this figure we distinguish between the $\alpha = 0.4$ and the $\alpha = 0.6$ experiments. Figure 6 reveals an interesting difference between treatments. When the gains from cooperation are low ($\alpha = 0.4$) followers contribute about 30 percent less than leaders. This relation holds for all contribution levels of leaders, except if the leader contributes only 1 token. If gains from cooperation are high ($\alpha = 0.6$), followers often contribute more than leaders, if the leader makes a low (i.e., smaller than 11) contribution to the team project. For leader contributions above 11, followers contribute less than leaders, but still more than in the $\alpha = 0.4$ experiments. Thus, with low gains from cooperation leaders are the sucker for all contribution levels, whereas with high gains from cooperation they are only suckered if they make a high contribution.

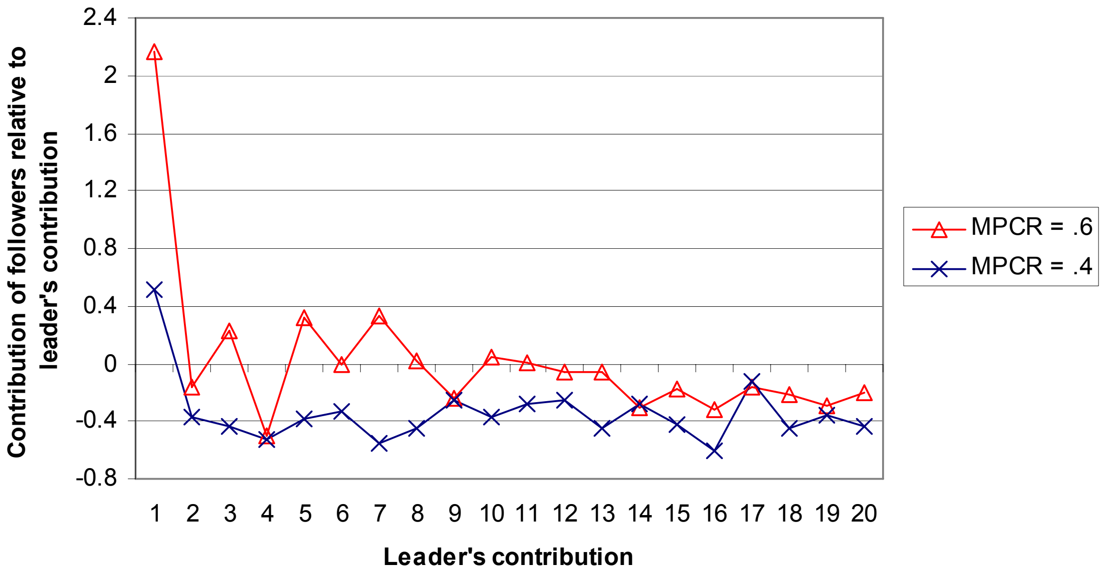


Figure 6. Followers' relative contribution as a function of leaders' contribution (for leader contributions > 0 and $\alpha = 0.4$ and $\alpha = 0.6$, respectively).

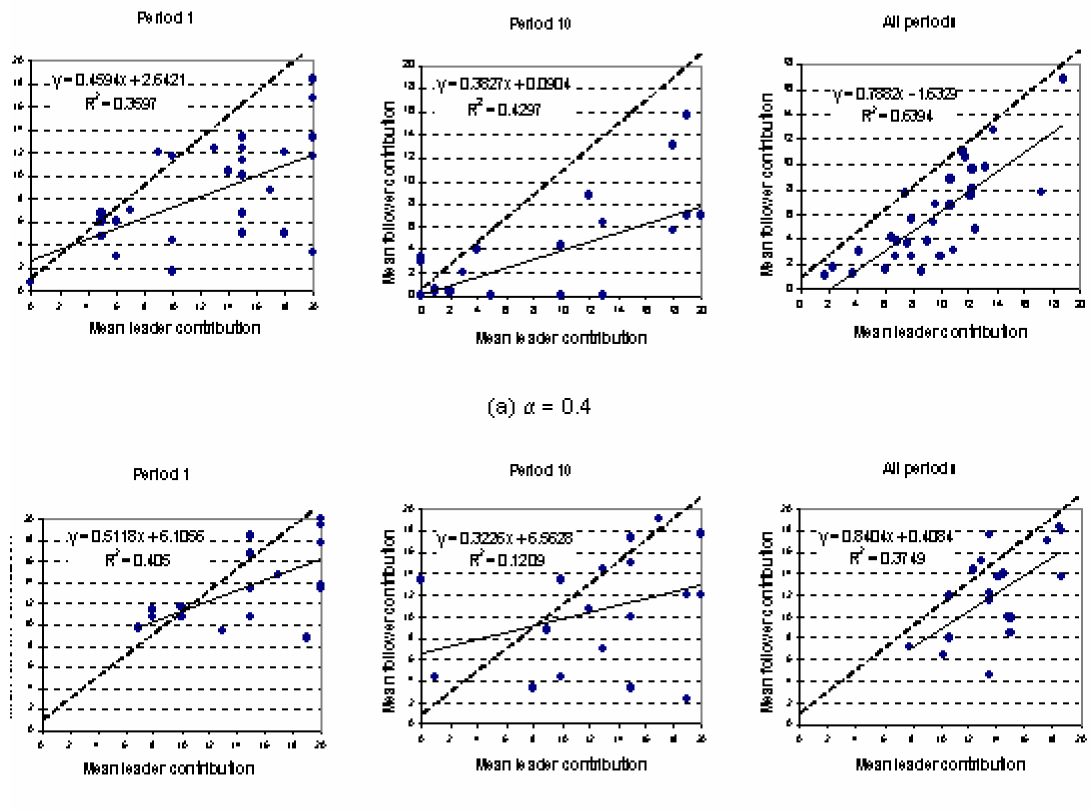
However, despite the increase in the cooperation levels in the experiments with higher gains from cooperation and despite the reduced sucker problem, we still get no effect of leaders on contribution levels. This holds for both

Observation 6: Qualitatively, all results from Observations 3 to 5 carry over to team games with high gains from cooperation. Quantitatively, the overall cooperation level is strongly increased and the leader is less systematically suckered than in team games with low gains from cooperation.

Maybe the reason for Observation 6 is that the positive correlation of leader and follower contributions observed in the one-shot games did not 'survive' the repeated game, which is why we do not observe an impact of the presence of a leader on overall cooperation levels? We investigate this question in the following subsection.

2.5.4 Leading by example?

Figure 7 shows that leading by example – in the sense of positively correlated leader and follower contributions – is well and alive in all repeated team games. The figure plots the mean follower contribution as a function of the mean leader contribution. Each data point represents one group and we do not distinguish between different leader types (i.e., we pool the leader sequences from treatment 1 and 2). We show the relation between average leader and follower contributions for period 1, period 10, and pooled over all ten periods, and separately for $\alpha = 0.4$ (Figure 7a) and $\alpha = 0.6$ (Figure 7b).



(a) $\alpha = 0.4$

(b) $\alpha = 0.6$

Figure 7. Mean follower contribution as a function of mean leader contribution: (a) $\alpha = 0.4$ and (b) $\alpha = 0.6$.

The positive correlation between the leaders' and the followers' contribution holds already in the first period. The Spearman rank correlation coefficient between the leaders' contributions and the average followers' contribution is $\rho = 0.54$, which is significant (p=0.0021). Likewise, in period 10, we get $\rho = 0.59$ (p = 0.0007). Correlations are positive and significant at least at p = 0.002 in all other periods as well. Pooled over all ten periods we get ρ = 0.79, which is significant at any level.[10]

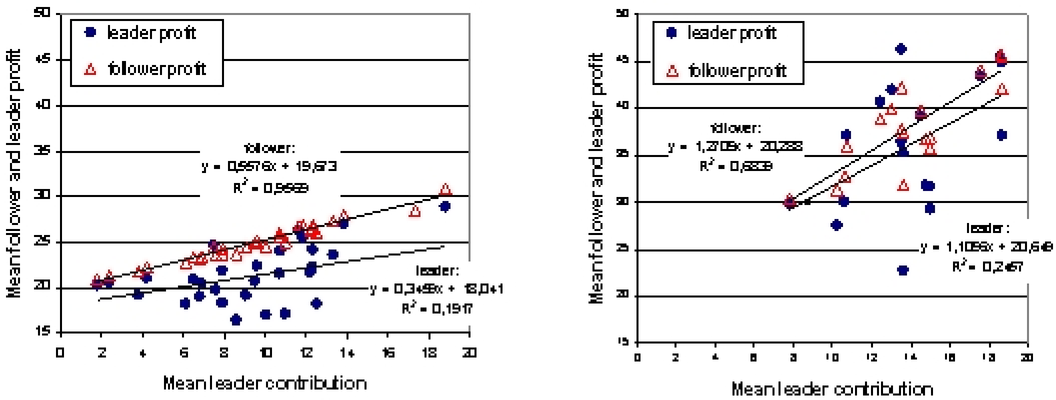
Moreover, we find positive correlation coefficients not only in the aggregate data but also for the single independent groups. Tables 1.a and 1.b in Appendix 2.B report these coefficients. Since there is no negative coefficient, the binominal test highly significantly rejects the null hypothesis that positive and negative correlation coefficients are equally likely. Thus, we conclude:

Observation 7: The ineffectiveness of leaders to affect overall contribution levels is clearly not due to a general lack of following the leader's example, in the sense of having significantly positively correlated contributions. Observation 1, made in the one-shot experiments, carries over to the repeated interactions.

How can we square Observation 6 and Observation 7? The follower's behaviour might give some hints. As Figures 1 to 7 show from different angles, it is a tough life being a leader. Although, on average, the followers will follow the leader's example and increase their contribution, they do so only half-heartedly, in particular in team games with low gains from cooperation. Given this situation, is it better to be bold or timid?

2.5.5 Shall I be a bold or timid leader?

In Figure 8 we look at this issue. In Figure 8a we show the leader and the follower payoffs as a function of the leader's mean contribution over all periods to the team project in team games with α = 0.4. Figure 8b does the same for games with α = 0.6. The black circles indicate the payoffs of each individual leader, and the triangles represent the mean payoff of a group of three followers. By construction, the triangle that represents a particular group lies either exactly above or below the circle that represents the leader of this particular group.



(a) α = 0.4 (b) α = 0.6

Figure 8. Mean follower and mean leader payoff as a function of mean leader contribution in games with low and high gains from cooperation.

Several interesting observations can be made from Figure 8. We start by discussing Figure 8a. First, the follower's mean payoff strongly increases in the leader's contribution. This is inevitable, for two reasons, as we have shown in Section III (see equation (4)). By virtue of a public good, a follower's payoff directly increases in the leader's contribution to the team project. Moreover, as long as the sensitivity of followers' reaction to an increase in the leader's contribution is non-negative, the followers will also indirectly benefit from a triggered increase in other follower's contribution to the

[10] For these calculations we have pooled the data from the two sequences. When we do the calculations separately for the two sequences, we get ρ = 0.79 if the leader sequence is played first and ρ = 0.72 if the leader sequence comes second. Both are highly significant ($p<0.0001$).

team project.[11] Second, the follower's payoff is in all 30 groups, except one, strictly higher than the leader's payoff. This just reflects again that the leader is the sucker in this game. Third, and most interesting for our main question at hand, on average it pays to be bold and contribute high amounts. The more a leader contributes on average the higher is his average payoff. Remember that this can only happen if the followers sufficiently strongly follow the leader's example, which in our context means that $\beta > 0.5$. Over all periods, the actual $\beta = 0.7882$ (see Figure 7), i.e., on average this suggests that it pays to be a bold leader. The Spearman rank order correlation between the leader's average contribution and his realized mean payoff is indeed significantly positive ($\rho = 0.368$, $p = 0.0452$, two-sided). A linear regression (see Figure 7) confirms this result. Fourth, the variance in payoffs is very large for leaders and very small for followers. For a third of the leaders this resulted in average payoffs that were lower than 20, which is the secure payoff from full free riding. Look at the unlucky guy in the lower right part of Figure 8a. He contributed on average 17.3 tokens to the team project, but still ended up with an average payoff of 18.3, whereas his followers earned a mean payoff of 28.9. Good and bad luck can be very close. The top earning leader in this experiment contributed slightly more than his unlucky colleague (18.8 on average) and took home a mean payoff of almost 29.

Figure 8b depicts the results from the experiments with high gains from cooperation. We conclude this section by recording

Observation 8: On average, it pays to be a high contributing leader in this experiment. The leader's payoff increases in his contribution because the followers increase their contributions sufficiently strongly to make leading by example worthwhile on average. Yet, the variance in payoffs is very large. Leaders therefore face a trade off between the risk and return of leading by example.

Can we draw some advice for leaders from Observation 8? If there is one, then, as so often in economics, it relates to one's risk and return preferences. Given that you will be suckered anyway, make up your mind how bold you really want to be.

2.6 Concluding remarks

It is a tough life being a leader in the presence of free rider incentives. Ironically, the reason, as we have shown, is not in fact that everyone is a free rider. That would make the problem easy. The difficulty arises exactly because, on average, people are following the leader's example.

Interesting questions for further research emerge for which our simple experiment can serve as a template. First, what is the role of beliefs about other follower's contribution as a function of the leader's contribution? We know from previous experiments that there is a positive correlation between a contributor's belief about the average contribution of others and his or her own contribution to the public good. An interesting issue therefore is, how a leader shapes these beliefs about other follower's contribution behaviour. Second, an apparent problem is that the leader gets suckered and most leaders cannot stand this disadvantage. Therefore, what would the optimal compensation for the leader be to solve the problem, if it can be solved at all? Paying the leader better than the followers may not solve the problem because it may give the followers a further excuse to rip off the leader. Third, would the problem be solved if the leader would have a sanctioning device at his disposal? Fourth, how does the problem change with team size? And so on. All these problems are interesting because many people don't free ride, even if they easily could.

[11] If we just look at the indirect effect, i.e., if we subtract the direct payoff from the leader's contribution, we find that the indirect benefit from an increased contribution of the other followers, as a function of the leader's contribution, is significantly positive ($\rho = 0.792$, $p < 0.0001$).

2.7 References

Abbink, Klaus, Bernd Irlenbusch and Elke Renner (2000): The moonlighting game – An experimental study on reciprocity and retribution, Journal of Economic Behavior & Organization, Vol. 42, 265–277.

Andreoni, James (1988): Why free ride? Strategies and learning in public goods experiments, Journal of Public Economics, 37, 291–304.

Andreoni, James (2001): Leadership Giving in Charitable Fund-Raising, Mimeo, University of Wisconsin-Madison.

Andreoni, James (2003): Public Goods Experiments Without Confidentiality: A Glimpse into Fund-Raising, Mimeo, University of Wisconsin-Madison.

Arce, M. Daniel, G. (2001): Leadership and the aggregation of international collective action, Oxford Economic Papers, Vol. 53, 114-137.

Berg, Joyce, John Dickhaut and Kevin McCabe (1995): Trust, Reciprocity and Social History,. Games and Economic Behavior, 10, 122-42.

Clark, Ken, and Martin Sefton (2001): The Sequential Prisoner's Dilemma: Evidence on Reciprocation,. Economic Journal Vol. 111, 51-68.

Cookson, R. (2000): Framing Effects in Public Goods Experiments. Experimental Economics, Vol. 3(1), 55-79.

Croson, Rachel T.A. (1996): Partners and strangers revisited." Economics Letters 53, 25-32.

Croson, Rachel T.A. (1998): Theories of altruism an reciprocity: evidence from linear public goods games, Working Paper 98-11-04, The Wharton School of the University of Pennsylvania.

Falk, Armin and Urs Fischbacher (2002): 'Crime' in the lab – Detecting social interaction, European Economic Review, 46, 859-869.

Fischbacher, Urs (1999): ztree. Zürich Toolbox for Readymade Economic Experiments, University of Zürich, Institute for Empirical Research in Economics, Working Paper No. 21.

Fischbacher, Urs, Simon Gächter and Ernst Fehr (2001): Are people conditionally cooperative? Evidence from a public goods experiment, Economics Letters, Vol. 71, 397-404

Fischbacher, Urs, and Simon Gächter (2003): Heterogeneous motivations in voluntary contributions to public goods, Mimeo.

Gächter, Simon, Benedikt Herrmann, and Christian Thöni (2003): Understanding Determinants of Social Capital: Trust, Cooperation, and Informal Sanctions in a Cross-Societal Perspective, Mimeo.

Foss, Nicholas (2001): Leadership, Beliefs and Coordination: An Explorative Discussion, Industrial and Corporate Change, Vol. 10(2), 357-388.

Heijden, Eline van der and Erling Moxnes (2000): Leadership and following in a public in a public bad experiment, SNF Report 13/00, Bergen, Norway: Foundation for Research in Economics and Business Administration.

Hemphill, J.K., and Coons, A.E. (1957): Development of the leader behavior description questionnaire, in: R.M. Stogdill and A.E. Coons (Eds.):Leader behavior: It's description and measurement, pp.6-38, Columbus: Bureau of Business Research, Ohio State University.

Hermalin, Benjamin E. (1998): Toward an Economic Theory of Leadership: Leading by Example, American Economic Review, Vol. 88 (5), 1188-1206.

Keser, Claudia, and Frans van Winden (2000): Conditional cooperation and voluntary contributions to public goods, Scandinavian Journal of Economics Vol. 102 (1), 23-39.

Kurzban, Robert, Houser, Daniel (2001): Individual Differences in Cooperation in a Circular Public Goods Game, European Journal of Personality, Vol. 15, 37-52.

Ledyard, John (1995): Public Goods: A Survey of Experimental Research. In: Kagel, John & Roth, Alvin E. (Eds.), Handbook of Experimental Economics. Princeton University Press, Princeton.

Miller, Gary J. (1990): Managerial Dilemmas: Political Leadership in Hierarchies, in: Karen S. Cook and Margret Levi (Eds): The Limits of Rationality, Chicago: The University of Chicago Press.

Moxnes, Erling and Eline van der Heijden (2000): The effect of leadership in a public bad experiment, SNF Report 15/00, Bergen, Norway: Foundation for Research in Economics and Business Administration.

Ockenfels, Axel (1999): Fairness, Reziprozität und Eigennutz – ökonomische Theorie und experimentelle Evidenz. Die Einheit der Gesellschaftswissenschaften, Bd. 108, Mohr Siebeck, Tübingen.

Potters, Jan, Martin Sefton, and Lise Vesterlund (2001): Why Announce Leadership Contributions? An Experimental Study of the Signaling and Reciprocity Hypotheses,. CentER DP 2001-100, University of Tilburg.

Rhodes, Carl M. and Rick K. Wilson (1999): Leaders, Followers, and the Institutional Problem of Trust, Mimeo.

Weber, Roberto, Colin Camerer, Yuval Rottenstreich and Marc Knez (2001): The Illusion of Leadership: Misattribution of Cause in Coordination Games, Organization Science Vol. 12, No. 5, 582.598.

Yukl, Gary, and W. Nemeroff (1997): Identification and measurement of specific categories of leadership behavior: A progress report, in: J.G. Hunt and L.L. Larson (Eds.): Crosscurrents in leadership, Carbondale: Southern Illinois University Press, 164-200.

3 When in doubt... – Cooperation and punishment under incomplete information

I say that individuals should be held accountable for their degrees of effort but not their levels of effort. The consequence of an action for a person should not depend on his circumstances, but it should depend on how hard he tries.
(John E. Roemer, 1998, p.18)

3.1 Introduction

Recent experimental studies have shown that peer monitoring is a particularly effective means of attenuating incentive problems and alleviating free riding behaviour (e.g., Ostrom, Gardner and Walker, 1992; Fehr and Gächter, 2000 and 2002; Sefton, Shupp and Walker, 2002; Masclet, Noussair, Villeval and Tucker 2003; Page and Putterman, 2000; Carpenter 2003). While in standard experimental social dilemma games cooperation is substantial in early rounds but significantly erodes over time, in games with additional punishment opportunities cooperation increases and is sustained at remarkably high levels. In these studies members of peer groups can identify free riders and their willingness to punish them successfully disciplines those who deviate from a group's cooperative norm. Peer punishment thus seems to be a major key to persistent cooperation.

Almost all previous studies investigate voluntary cooperation and punishment in environments in which group members are homogeneous both with respect to their endowments and their preferences. This symmetry is a natural way to begin studying cooperation and punishment behaviour. Yet, homogeneity is not a feature of many everyday social settings: members of work teams or communities most likely differ with respect to wealth, ability, experience, strength, health, or a host of other resource endowments that limit the extent to which they are capable of contributing to their group.

In this paper we are interested in how cooperation and punishment are affected if team members differ with respect to their "capabilities" (wealth, abilities, etc). We model this situation as a public goods problem in which group members have different endowments. Specifically, the basic decision problem is as follows (for details see the next section): in each of the ten periods each subject receives an endowment from a uniformly distributed set of potential endowments. Subjects simultaneously decide how much of this endowment they to contribute to the public good.

If group members are heterogeneous with respect to their endowments information may become crucial. For instance, it is often easy to observe what others' are doing, but not what they could be doing. It is, for example, obvious that a colleague is absent from work but this might be due to a sickness or malingering. People who live on social welfare might do so because they need it or because they want to exploit welfare. Low contributions to a work team's project can be undeserved or due to low effort. In such situations it difficult to judge others' behaviour: the mere observation of others' levels of cooperation allows no inference of their cooperativeness.

To capture these effects, we vary the information condition in two experimental treatments. In one treatment, which we call PUBLIC, every group member is informed about the endowment of all other group members. In PRIVATE, subjects only know their own endowment in a particular group. This setup allows us to answer our first research question: How does the information about the endowments influence cooperation rates, i.e., what is the difference between PUBLIC and PRIVATE?

Our second research interest concerns the power of punishment in this environment. Recall that previous experiments have shown that punishment in homogeneous groups can strongly mitigate the free rider problem inherent in the public goods game. Therefore, we are interested in (i) how punishment is used when groups are heterogeneous with respect to their endowments and (ii) how is it affected by incomplete information?

Heterogeneous endowments make the punishment decision more complex: First, in PUBLIC absolute or relative contributions (i.e., absolute contributions as a proportion of the endowment) might trigger punishment. Second, in PRIVATE a potential punisher lacks the information about the actual endowment of other group members; he or she sees only the absolute contributions of the other group members. If he or she decides not to punish a certain ('low') contribution he or she might commit a "type I error": the punisher might believe that the subject has a low endowment but actually has a high endowment and nevertheless decided to contribute little – a behaviour that under PUBLIC would have led to punishment. Of course, a "type II error" is also possible: if a subject decides to punish a certain contribution he or she might actually hurt someone who contributed a large fraction of his or her endowment.

Our main results are the following: First, as soon as the endowment is not publicly known, subjects take advantage of this situation and contribute considerably lower proportions of their endowments. Second, the presence of punishment opportunities promotes cooperation. Yet, in PRIVATE this effect is quite small. Surprisingly, low cooperation rates are not due to a reluctance to punish. Punishment expenses are not significantly different in PRIVATE and PUBLIC. Third, punishment has negative welfare implications in both treatments: even with a strong increase in cooperation rates in PUBLIC teams are better off if they cannot punish each other.

We believe that our results are of practical and theoretical relevance. In reality groups certainly differ with respect to their capabilities to cooperate. It is therefore important to understand how heterogeneity and the information about it affect cooperation. Our results show that in PUBLIC people largely base their punishment of other group members by "how hard they try". Thus, in PUBLIC people's behaviour is consistent with John Roemer's normative request in the opening quote. By contrast, in PRIVATE punishment can do considerable harm because punishment may hurt people who actually contributed their whole endowment (which may happen to be small). Therefore, information about others' endowments is crucial. This finding is also of theoretical relevance because it provides a robustness check of the power of punishment to mitigate the free rider problem with respect to heterogeneous endowments and incomplete information.

The remainder of this chapter is organized as follows: in the next section we describe the design and the experimental procedures. Section 3.3 presents the results. Section 3.4 discusses welfare implications. Section 3.5 concludes.

3.2 Experimental design and procedures

Our experimental set up is closely related to the design of Fehr and Gächter (2002). The basic decision situation involves a simple linear public goods game in the first stage and a punishment decision in the second stage. In the first stage each of four team members receives an endowment of tokens and can allocate these either to a team project or to oneself. The first- stage earnings of group member i are

$$\pi_i = e_i - g_i + \alpha \sum_{j=1}^{4} g_j \qquad (1)$$

where e_i denotes individual i's endowment and g_i is her contribution to the team project. The marginal per capita return of a contribution to the team project is set at $\alpha = 0.4$, which implies that it is socially optimal if all group members contribute their entire endowment to the team project: each token invested yields a net return of 4*0.4 − 1 = 0.6 to the group. In contrast to this socially optimal solution however, it is individually not rational to contribute since every token contributed yields a private net return of 0.4 − 1 = − 0.6. Therefore the unique equilibrium consists of all group members free-riding in the first stage if subjects are rational and only care about their own monetary payoff. When punishment opportunities are available, the first stage is followed by a punishment stage in which each subject can assign up to ten punishment points to each of the other group members. Each point spent on punishment reduces the punishing member's payoff by one point and the punished member's payoff by three points. Since punishment is costly it is not a credible threat and the subgame perfect equilibrium predicts no punishment in the second stage and zero contributions in the first stage.

As in Fehr and Gächter (2002), some of our subjects first played a sequence of ten public good games with punishment and then another sequence of ten public good games without punishment opportunities. For other subjects we reversed the order of the sequences to obtain control groups without punishment in the first sequence.

In the present study we implement uncertainty about the capability of individual group members to contribute by randomly determining endowments. At the beginning of stage one each individual's endowment was independently drawn from a uniform distribution on {0, 1, …, 40}. Thus the expected endowment is 20 tokens for each individual. By this feature we model group members with the same expected capabilities; however the actual capability of a given subject can vary across games.[12]

To investigate the impact of information about endowments on contribution and punishment behaviour we conducted two treatments with different information conditions: the public information treatment (PUBLIC) and the private information treatment (PRIVATE). In the PUBLIC condition subjects learn their individual endowment before stage one. After they have made their contribution decisions and before they have to decide whether and how many punishment points they wish to assign to each of the other team members they also learn the endowments of every single group member and how many tokens out of this endowment a person contributed to the team project. An overview of our design is presented in table 3.1.

---

[12] We drew two sets of 40 numbers (set A and set B) which determined the endowments for each of the four group members for ten periods. In 9 out of 18 groups in the first and 10 out of 19 of the groups in the second treatment we used the random set A in the first sequence and B in the second sequence and reversed this order in the remaining groups. By this procedure we control for random differences in the endowment structure between the two treatments, which enables us to directly compare the results.

.

Tabelle 3.1:. The treatments of the experiments

Information on individuals endowment	Sequence I (10 periods)	Sequence II (10 periods)	No. of independent groups
PRIVATE	with punishment without punishment	without punishment with punishment	9 10
PUBLIC	with punishment without punishment	without punishment with punishment	10 8

The PRIVATE treatment differs only with respect to the information available in the punishment stage. At the beginning of the punishment stage each group member is informed of the other subjects' contributions, but is not informed of their endowments. Neither the random determination of the endowments, nor the information on the endowments, affects the equilibrium solution described above: if it is common knowledge that subjects maximise own earnings then subgame perfection requires that in every period no subject punish another, and that all subjects free ride.

The experiments were conducted in the experimental laboratory at the University of Erfurt using the experimental software z-tree (Urs Fischbacher, 1999). A total of 148 students from various disciplines participated, thus we collected a total 37 independent observations. The subjects were encouraged to participate through posters on the campus and we made sure that each subject participated only once. At the beginning of each session the participants were randomly assigned to one of the booths in the laboratory. The booths visually separate the participants and ensure that every individual makes his decision anonymously and independently. The groups were randomly assigned, so that subjects did not know which of the other participants were in their group. Written instructions were handed out in which the decision situation of the first sequence of ten periods and the experimental procedures were explained in detail (a translation of the original instructions is provided in Appendix 2.A). The instructions were written in a neutral language. The order in which the other group members' decisions (and endowments in the case of PUBLIC) were displayed after stage one was randomly shuffled in each period, such that it was not possible to attribute the current decision of a person to his or her behaviour in previous periods. We used this procedure to exclude reputation effects.

After period ten subjects were informed that another sequence of 10 periods will follow. Subjects then received the corresponding instructions in which the modified decision situation was explained. Immediately after each session subjects were paid privately in cash. Average earnings were € 13.20 (about US $13 at that time). An experimental session lasted approximately 75 minutes.

3.3 Results

We start with an overview of contribution behaviour in our two information conditions before we scrutinize punishment behaviour and the reactions to punishment.

3.3.1 Contribution behaviour

Figures 1a and 1b plot the evolution of contribution rates over time.[13] Figure 1a shows the results from sessions with punishment opportunities in the first sequence; figure 1b shows the results from sessions with punishment opportunities in the second sequence. Figures 1a and 1b clearly indicate that the average contribution rates are substantially lower when the individual endowments are not displayed to the other group members. This difference is significant regardless of whether punishment opportunities are available or not (significant at $p < 0.01$ for sequences without punishment and $p = 0.000$ for sequences with punishment, using two-sided Mann-Whitney U-tests applied to group

[13] Most of our data analysis is based on contribution rates. For each subject and period we compute the contribution rate as g_i/e_i. Figure 1 displays the average of these across all subjects within the relevant period and treatment.

averages).

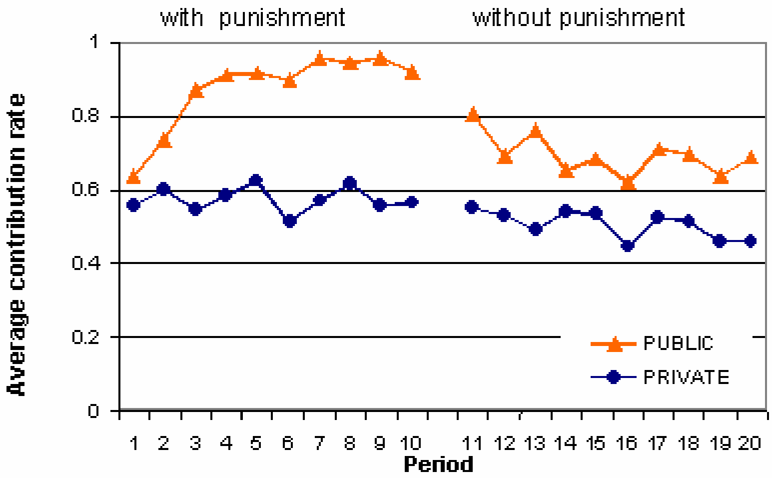


Figure 1a. Evolution of average contribution rates

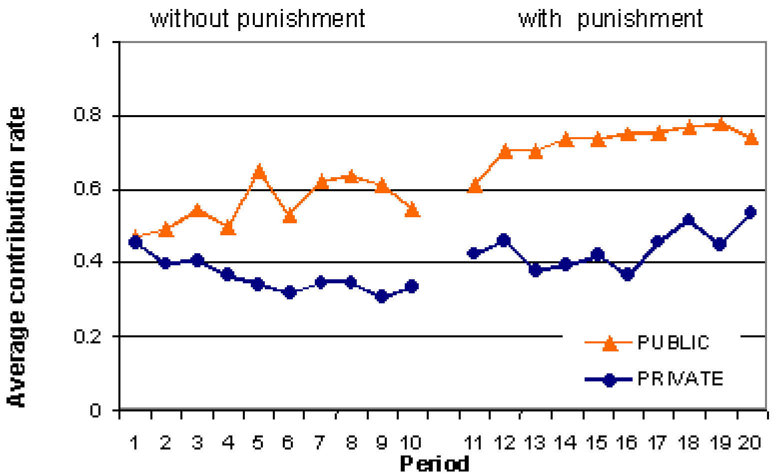


Figure 1b. Evolution of average contribution rates

Figure 2, which depicts the average contribution rate for different endowment intervals, shows that this difference is observed across the entire range of endowments. For all endowment intervals the contribution rates in PUBLIC are well above those in PRIVATE.

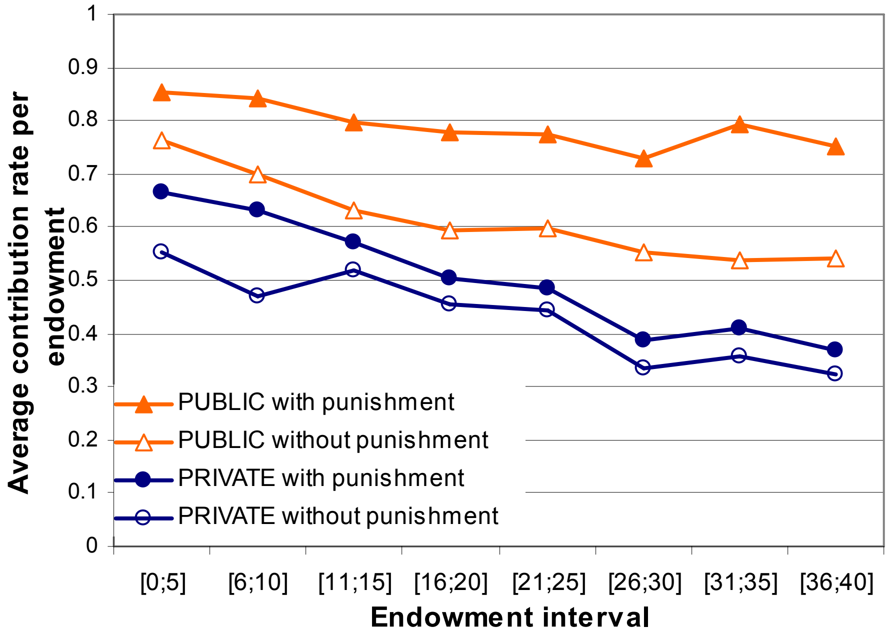


Figure 2. Average contribution rates subject to the endowments

We summarize these observations in our first result:

Result 1: Contribution rates are lower when endowments are private information.

Figure 2 also suggests that in both information conditions subjects with higher endowments contribute a lower proportion of their endowment. Spearman rank correlation coefficients between the endowment interval and the average contribution rates in the group provide statistical evidence for this decreasing trend at the independent group level (the results are reported in detail in table 2.a in Appendix 2.B). In PUBLIC we obtain 17 negative and 3 positive coefficients and 16 negative and 4 positive for the sequences without and with punishment respectively. In PRIVATE we obtain 18 negative and 1 positive coefficients for both sequences. Using these coefficients as a summary statistic two-sided binomial tests reject the null hypotheses that positive and negative coefficients are equally likely for all sequences ($p < 0.03$). We record these findings as our second result:

Result 2: Subjects with higher endowments contribute at a lower rate.

We next turn to the impact of punishment opportunities on cooperation behaviour. It is immediately evident from inspection of figures 1a and 1b that the availability of sanctions has a strong positive impact on contribution rates in the PUBLIC condition: the availability of punishment opportunities increase the average contribution rates by 16.66 (when punishment is available in the first sequence) and by 17.95 percentage points (when punishment is available in the second sequence). A two-sided Wilcoxon matched-pairs test, applied to independent group averages, is significant at $p = 0.036$ if the punishment sequence is played first and at $p = 0.028$ if the punishment sequence is played second. In PRIVATE however, the impact of punishment opportunities is considerably weaker. The average contribution rates increase by only 7.84 percentage points if the punishment sequence is played first and by only 6.93 percentage points if the punishment sequence is played second. The associated two-sided Wilcoxon matched-pairs tests, again applied to independent group averages, are insignificant at $p = 0.093$ and $p = 0.169$ respectively.

Figures 1a and 1b also show that contribution rates in the PUBLIC treatment tend to increase over time when punishment is possible. In PRIVATE however, there is no apparent time trend. To gain a statistical representation for the evolution of contributions for the independent groups we calculate Spearman rank correlation coefficients between average contribution rates and periods (the results are reported in detail in table 2.b, Appendix 2.B). In PUBLIC 14 groups yield positive and 4 groups negative coefficients in the punishment sequence and we can reject the null hypothesis that positive and negative coefficients are equally likely (binomial test, significant at $p = 0.031$). In PRIVATE the coefficient is positive in 11 groups and negative in 8 group and we cannot reject the null hypothesis that positive and negative coefficients are equally likely (binomial test, insignificant at p = 0.648).

3.3.2 Punishment behaviour

Looking again at figure 2, we see that the effect of punishment opportunities on cooperation rates is stronger in PUBLIC at all endowment intervals. In PUBLIC it is further remarkable that the presence of punishment opportunities exerts a particularly strong disciplining effect on subjects with high endowments: average contribution rates increase by more than 20.9 percentage points in the endowment interval (31;35) and by 25.4 percentage points in the (36;40) interval. In PRIVATE however, for all endowments above 10 tokens, the possibility of being punished increases cooperation rates by only about 5 percentage points. In PRIVATE only subjects with endowments below 10 tokens exhibit a slightly higher increase. Our third result summarizes these findings:

Result 3: In PUBLIC the presence of punishment opportunities significantly increases cooperation rates and high cooperation rates are sustained. In PRIVATE the presence of punishment opportunities leads to a small, but insignificant, increase in contribution rates.

Why is punishment less effective when endowments are private information? A first conjecture is that subjects in PRIVATE simply refrain from punishment. However, there is little evidence in the data to support this conjecture. Punishment is slightly less frequent in PRIVATE than in PUBLIC: in PRIVATE 36.3% and in PUBLIC 39.7% of all decisions are punished. Altogether, subjects spend slightly more on punishment in PUBLIC (1.14 tokens per period) than in PRIVATE (0.86 tokens per period), although the difference is not significant (two-sided Mann-Whitney U-test, p = 0.126). This is our fourth result:

Result 4: Subjects spend slightly less on punishment in the PRIVATE treatment compared to the PUBLIC treatment, but the difference is not significant.

If readiness to punish is similar across the two treatments, then what else might account for the moderate impact of punishment on cooperation rates in PRIVATE? The reason is that subjects who are punished do not change their behaviour in response to the sanctions. To measure the reaction of a person we calculate the relative change in the contribution rates from period *t* to *t+1* for each subject.[14] In PUBLIC we find clear statistical support for a disciplining effect of punishment. Spearman rank correlation coefficients between the sum of punishment points spent by a group in period t and the average relative change in a group's contribution rate in period *t+1* yield positive coefficients in 15 groups and negative coefficients in only 2 groups (in one group we did not calculate the coefficient since the observed punishment levels were less than 3; see table 2.c, Appendix 2.B for details). Thus, in the vast majority of the groups, more points spent on punishment lead to higher increases in contribution rates in the next period. In PRIVATE however, the opposite is the case for 10 groups, where higher sanctions leads to decreasing contribution rates. Only in 5 of the PRIVATE groups is the coefficient positive. (We could not calculate coefficients in 4 groups in which the observed punishment levels were less than 3). Hence, the null hypothesis that positive and negative correlation coefficients are equally likely can be rejected for PUBLIC (two-sided binomial test, p = 0.002) but not for PRIVATE (p = 0.302). This is our fifth result:

Result 5: In PUBLIC punishment has a disciplining effect: subjects respond to higher sanctions with higher increases in cooperation rates in the following round. In PRIVATE cooperation rates do not respond to punishment.

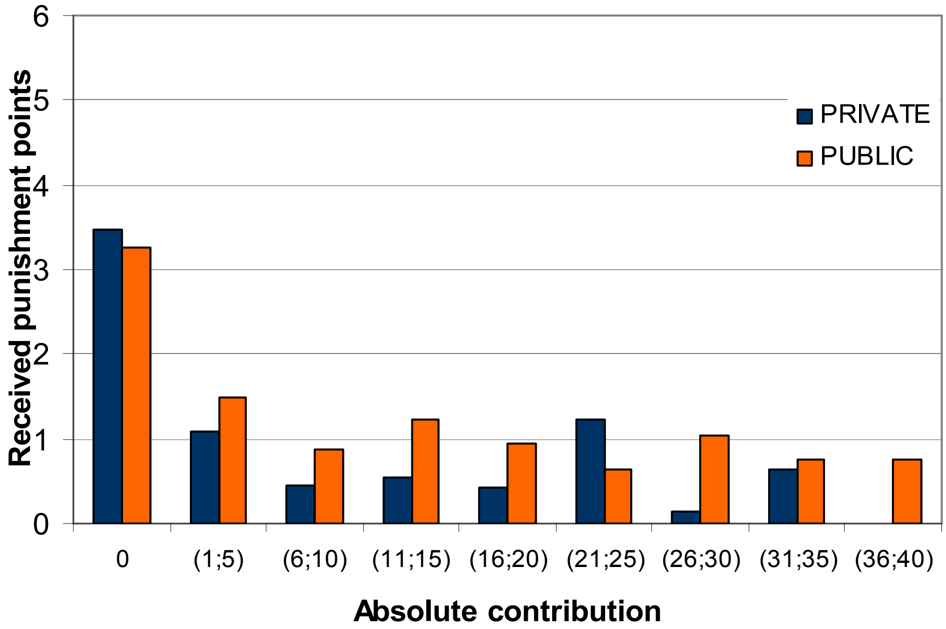
Why is punishment a rather blunt disciplining device in PRIVATE? A closer look at the punishment patterns in our two treatments provides an explanation. We first look at punishment with regard to absolute contributions. In both information treatments, subjects can observe the absolute contributions of others and react to this in the punishment stage. Figure 3 shows the relationship between received

---

[14] We use the following measure that expresses the reaction of a person as the change in her contribution rate in the following period relative to the maximum possible increase or decrease. We denote the absolute change in the contribution rate for subject *i* as $\Delta_i = (g_i^{t+1}/e_i^{t+1} - g_i^t/e_i^t)$. A decrease in contribution rate from period *t* to *t+1* implies that $\Delta_i < 0$ and we can denote the relative decrease by $\delta = \Delta_i/(g_i^t/e_i^t)$. For $\Delta_i > 0$ the relative increase in the contribution rate is denoted by $\delta = \Delta_i/(1 - g_i^t/e_i^t)$.

punishment points and absolute contributions for both treatments. Punishment patterns do not differ fundamentally across our two treatment conditions. Subjects who contribute nothing receive similar punishment in both treatments: on average 3.5 and 3.3 points in PRIVATE and PUBLIC respectively. For any positive contribution level punishment lies in the rather narrow range between 0 and 1.5 points, and is slightly higher in PUBLIC (with the exception of the 21-25 contribution level). Across all positive contributions punishment averages 0.7 points in PRIVATE and 1 point in PUBLIC.

This punishment pattern implies that in both treatments it clearly pays off to contribute at least one token (if possible) to avoid the high penalty associated with contributing nothing. However, since any positive contribution receives a slight punishment anyway, no matter how high it is (with the exception of 0 punishment for contributions higher than 35 in PRIVATE) there appears to be no monetary incentive to respond to punishment with increased contributions.[15]



.

Figure 3. Absolute contributions and received punishment points

[15] To show this more formally we ran linear regressions of contributions as the explanatory variable responsible for losses from being punished (ignoring observations where subjects contributed nothing in a round). The regressions yield y = -0.0664x + 2.8522, R^2 = 0,1723 for PRIVATE and y = -0.0693x + 4,3477, R^2 = 0,1635 for PUBLIC. The marginal benefits through reduced penalties are 0.0664 and 0.0693 respectively, which do not compensate for the marginal cost of 0.6 incurred from contributing one more token.

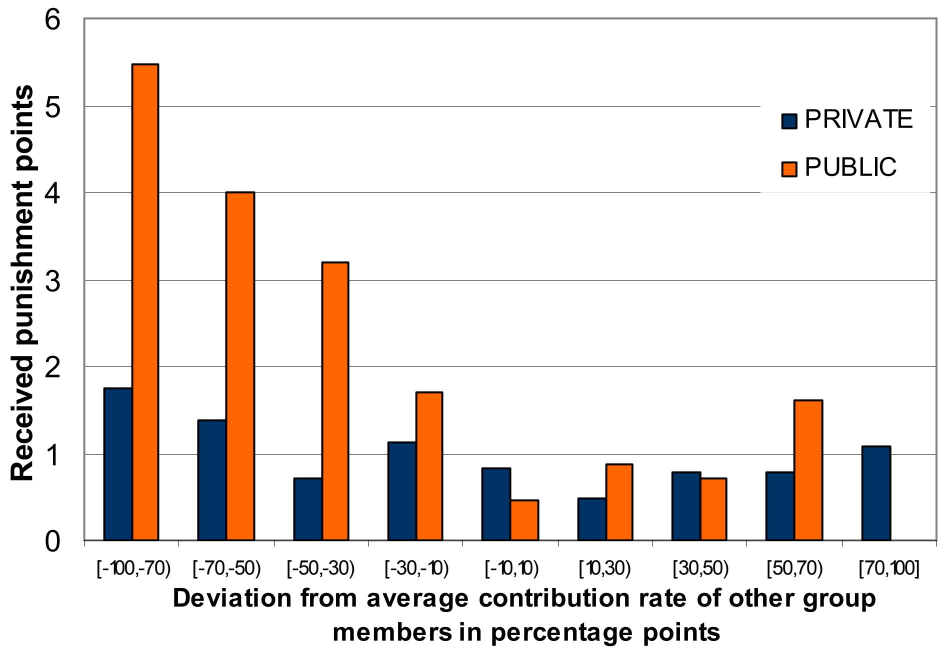


Figure 4. Percentage deviation from other group members' average contribution rate and received punishment points

This result is of particular importance for PRIVATE, where absolute contributions are observed, but contribution rates are not observed. In PUBLIC, however, this is only part of the story since subjects can also assess cooperation rates. Punishment according to contribution rates is displayed in figure 4, which shows the punishment received conditional on a subject's deviation from the other group members' average contribution rate.[16] In PUBLIC, we see a clear and pronounced punishment pattern: subjects are punished more heavily the more their contribution rate falls short of the rest of the group's average contribution rate. Punishment is lowest when the contribution rate is close to the average of the other group members and it is particularly high if someone under-contributes to a high degree. Spearman rank correlation coefficients between deviations from the group's average contribution rate and the average punishment points received provide further statistical evidence at the independent group level: in PUBLIC 15 coefficients are negative and 1 coefficient is positive (two-sided binomial test significant at p = 0.001; for details see table 2.d in Appendix 2.B).[17] This punishment pattern closely resembles the punishment behaviour observed in Fehr and Gächter (2000 and 2002) where absolute and relative contributions coincide due to uniform endowments. In the present study, where absolute and relative contributions are not perfectly correlated, punishment according to contribution rates appears more pronounced than punishment according to absolute contribution. This clearly suggests that cooperation rates are the dominant criteria for punishment decisions.

This has important implications for the incentives to respond to punishment. In PRIVATE there is still no incentive to react to punishment since contribution rates are unobserved. In PUBLIC, however, under-contributors can escape significant payoff reductions by closely matching the average cooperation rate in the group. Whether it pays off to do so depends on the size of the endowments. For higher endowments it is more costly to cooperate at a specific rate. For example, to match a

---

[16] It is clear that in PRIVATE, where cooperation rates are unobserved, figure 4 merely serves as an illustration of how punishment points, applied on the basis of absolute contributions, are distributed on over- and under- contributors in relative terms.

[17] In PRIVATE 8 groups have positive and 8 negative coefficients. We had to exclude 2 groups in PUBLIC and 3 groups in PRIVATE, because less than 3 punishment levels were observed in these groups.

contribution rate of 50% a person with an endowment of 40 tokens needs to contribute 20 tokens, whereas someone with an endowment of 20 only has to spend 10 tokens. Thus, the lower the endowment the more likely it is that the benefit of a lower penalty outweighs the costs of higher contributions. In order to demonstrate that the observed punishment is sufficiently strong to make it worthwhile to increase cooperation rates we conducted a simple cost-benefit analysis. For every decision we estimated the costs and benefit of a hypothetical increase in contributions. Using the average contribution rate of the other group members r as a reference point, we calculate for every under-contribution decision the loss of earnings in stage one that would be necessary to match this contribution rate (i.e. 0.6*(endowment*r - actual contribution)). Assuming a contribution rate of r would induce losses from punishment of 1.41 points (as observed in the data), we can calculate the hypothetical net impact of punishment consequences as the difference between the reduction associated with the actual decision and the reduction of 1.41. The results show that 55% of all under-contributors would have fared better if they had matched the contribution rate of the other group members. For all under-contributors the average net benefit of matching the rest of the groups' average rate is 1.83. We summarize these findings in our next result:

Result 6: In PUBLIC subjects punish other group members according to their degrees of cooperativeness. Subjects are punished more severely the lower their contribution rate relative to the average contribution rate of the other group members. Punishment is strong enough on average to create a monetary incentive to increase cooperation in order to match the groups' average rate. In PRIVATE, where punishment cannot be conditioned on contribution rates, there are no monetary incentives to increase cooperation.

In the psychology literature there is evidence that not only monetary concerns trigger reactions to punishment. Models of procedural justice, for example, emphasize that the perception of whether punishment is viewed as fair or unjustified matters for subsequent behaviour (e.g. Tyler, 1988). In PRIVATE punishment can be regarded as violating the fundamental principles "in dubio pro reo" and thus might be generally perceived as an unfair treatment. Similarly, it might matter whether punishment hits an "innocent" person with a high contribution rate or an actual free rider with a low contribution rate. To see whether the justification for punishment matters, we constructed figures 5a and 5b that show whether over- and under-contributors respond differently to punishment. Figures 5a and b show how relative (possible) increase or decrease in contribution rates from period *t* to *t+1* are related to received punishment points in period t. We again use the average contribution rate of the group as a reference point and discriminate between under- and over-contributors relative to this.[18] In PUBLIC subjects generally tend to increase contribution levels. This reaction is particularly strong for those who under-contributed. Cooperation rates only decrease (slightly) for those who over-contributed and received 0 or more than 3 punishment points. In PRIVATE the reactions of the two types clearly diverge: under-contributors tend to increase their contribution rates (although this is less pronounced than in PUBLIC). Over-contributors decrease their contributions and this reaction appears to be stronger the more they get punished.

[18] We chose this reference point for comparability with figure 4. Different reference points such as the average contribution rate over all rounds, or 50% contribution, give a very similar picture. Of course, in PRIVATE subjects cannot infer whether their contribution rate was higher or lower than other group members.

Figure 5a. Reaction to Punishment: PUBLIC

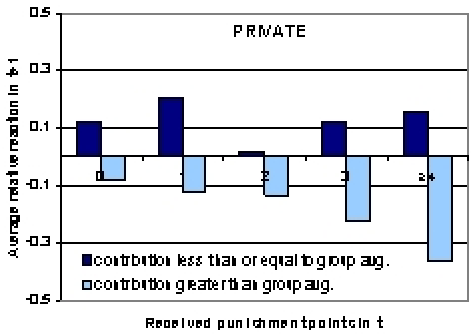


Figure 5b. Reaction to Punishment: PRIVATE

Overall 62% of the over-contributors in PRIVATE contribute at a lower rate after having received punishment points in the previous period, whereas only 28% of the under-contributors react negatively. In PUBLIC the percentages of over- and under-contributors who reduce contribution rates in response to being punished are only 23% and 12% respectively. Thus, the justification for punishment apparently matters for subsequent behaviour, and unjustified punishment seems to crowd out cooperation.

3.4 Welfare implications

Punishment is costly and the overall welfare implications are positive only if punishment triggers an increase in cooperation rates that is sufficiently high to outweigh the loss through sanctioning. The fact that punishment only has a moderate impact on cooperation rates in PRIVATE already suggests that punishment does more harm than good in this treatment. Figure 6 compares the welfare effects in terms of realized efficiency gains from cooperation of all treatments. In the PRIVATE punishment condition groups realize 46% of the gains of full cooperation after the first stage. However, punishment activities in stage two destroy a considerable part of these gains, which finally results in only 17% realized efficiency. Compared to the condition without punishment where 40% efficiency is reached, the net effect of punishment is clearly negative. In PUBLIC the result is similar: although the positive impact on cooperation is much higher, it is not sufficient to compensate for the losses from punishing and being punished. The realized gains are 78% of the full cooperation benchmark after the first stage. After punishment has taken place realized gains are reduced to 40%, which is clearly lower than the realized gains of 58% in the no punishment condition.

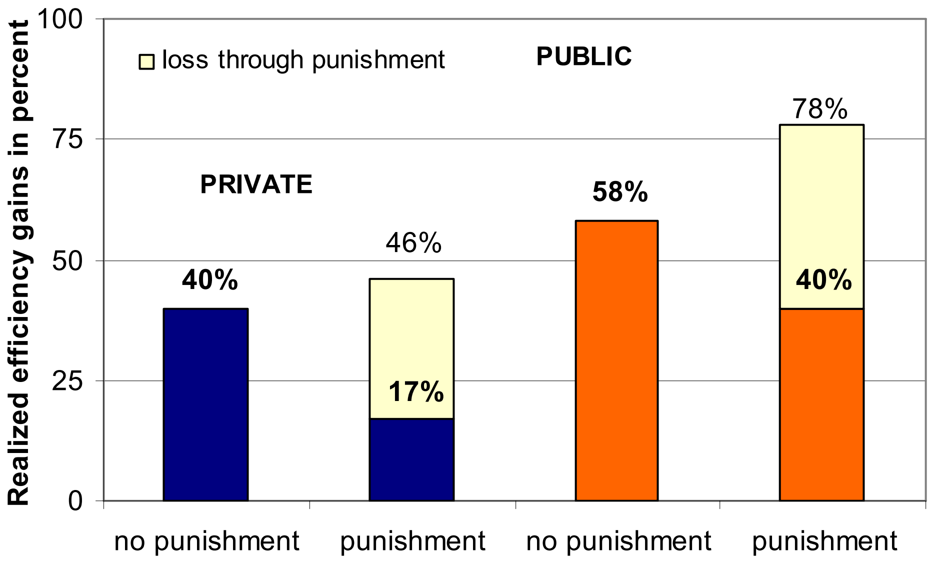


Figure 6. Realized efficiency gains

3.5 Conclusions

In this paper we have studied the role of heterogeneous endowments and incomplete information for cooperation and punishment decisions in the context of a public good. In each period, group members had different endowments. We also varied the information condition: in PUBLIC group members were fully informed about all endowments; in PRIVATE they only knew their own endowment.

We believe that this environment is interesting for practical and theoretical reasons: The practical relevance comes from the fact that in reality group members often differ in their capabilities, and therefore in their potential to contribute to their group. The theoretical relevance stems from the fact that punishment is considered to be a mechanism that can mitigate the free rider problem even in anonymous one-shot games (for theoretical arguments see e.g., Fehr and Gächter 2002; Gintis, Bowles, Boyd and Fehr 2003; Boyd, Bowles, Gintis and Richerson 2003). Yet, all evidence that supports this mechanism comes from experiments in which group members were homogeneous with respect to their endowments. Our experiments therefore provide a robustness check of the previous results.

Our experiment conveys two main insights: First, heterogeneity per se does not compromise voluntary cooperation when punishment is possible. In PUBLIC people punish other group members dependent on their contribution relative to their endowment. As a consequence, contributions increase to high levels, as they do in experiments in which group members are homogeneous with respect to their endowment. Our second main finding is that information about others' endowment is crucial. Even without punishment opportunities contributions in PRIVATE are lower than in PUBLIC. Thus, the possibility to "hide" when there is incomplete information leads to more free riding behaviour.[19] Moreover, punishment does not enhance cooperation if only the cooperation levels of other group members are observable but endowments are not (i.e., in PRIVATE). Surprisingly, this is not due to a reluctance to punish. Punishment is still widespread and therefore bound to involve many "type II errors" where people with low endowments but high contribution rates get punished. Thus, a practical implication is that establishing information about others' capabilities is crucial to avoid "type II errors".

[19] This finding is consistent with evidence from ultimatum games (e.g., Mitzkewitz and Nagel 1993; Güth, Huck and Ockenfels 1996) or from gift exchange experiments (e.g., Hennig-Schmidt, Rockenbach and Sadrieh 2003; Irlenbusch and Sliwka 2005), where incomplete information also led to more selfish behaviour.

This is consistent with what Knez and Simester (2001, p.767) report about the behaviour of employees who get partly paid according to a team compensation system: “Employees […] began contacting colleagues who had called in sick, to ask whether any assistance was needed. According to employees these calls enabled them to monitor whether the absences were due to valid illnesses.” Thus, establishing transparency is crucial to avoid detrimental effects of punishment.

3.6 References

Boyd, Robert, Herbert Gintis, Samuel Bowles, and Peter J. Richerson (2003): The evolution of altruistic punishment, Proceedings of the National Academy of Science 100(6), 3531–3535.

Carpenter, Jeffrey (2003): The demand for punishment, Mimeo.

Fehr, Ernst and Simon Gächter (2000): Cooperation and Punishment in Public Goods Experiments, American Economic Review, 90(4), 980-94.

Fehr, Ernst and Simon Gächter (2002): Altruistic punishment in humans, Nature, Vol 415, pp. 137-40.

Fischbacher, Urs (1999): ztree. Zürich Toolbox for Readymade Economic Experiments, University of Zürich, Institute for Empirical Research in Economics, Working Paper No. 21.

Gintis, Herbert, Samuel Bowles, Robert Boyd, and Fehr Ernst (2003): Explaining altruistic behavior in humans, Evolution and Human Behavior 24, 153-172.

Güth,Werner, Huck, Steffen, Ockenfels, Peter (1996): Two-level ultimatum bargaining with incomplete information. Economic Journal, 106, 593–604.

Hennig-Schmidt, Heike, Bettina Rockenbach, and Abdolkarim Sadrieh (2003): Incomplete and Asymmetric Surplus Information in Labor Relations. Discussion Paper No.: 2003-012E, University of Erfurt.

Bernd Irlenbusch and Dirk Sliwka (2005): Transparency and reciprocal behaviour in employment relations, Journal of Economic Behavior and Organization 56, 383-403.

Knez, Marc and Duncan Simester (2001): Firm-wide incentives and mutual monitoring at Continental Airlines, Journal of Labor Economics, 19 (4), 743-772.

Masclet, David, Charles N. Noussair, Steven Tucker, and Marie-Claire Villeval (2003): Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism, American Economic Review, 2003, 93(1), 366-380

Mitzkewitz, Michael and Rosemarie Nagel (1993): Experimental results on ultimatum games with incomplete information, International Journal of Game Theory, 22, 171–198.

Ostrom, Elinor, James Walker and Roy Gardener (1992): Covenants with and without a sword: Self-Governance is possible, American Political Science Review, 86, 404-417.

Page, T. and L. Putterman (2000): Cheap talk and punishment in voluntary contribution experiments, mimeo.

Romer, John E. (1998): Equality of opportunity, Harvard University Press, Cambridge Massachusetts

Sefton, Martin; Shupp, Robert and Walker, James (2002): The Effect of Rewards and Sanctions in Provision of Public Goods, Working Paper Nottingham University

Tyler, Tom (1988): What is procedural justice? Criteria used by citizens to assess the fairness and legal procedures, Law Society Review, 22, 301-355

4 Group size and social ties in microfinance institutions[20]

4.1 Introduction

In recent years microfinance institutions (MFIs) have become one of the most important instruments in development policy. The idea of microfinance arose in the mid-70s when Mohammad Yunus started a pilot scheme lending small amounts of money to villagers in Bangladesh who, due to a lack of collateral, had no access to conventional loans. Encouraged by high repayment rates, he founded the Grameen Bank to run such schemes on a larger scale. Today the Grameen Bank lends to about 2.4 million people. Since Grameen's early successes, the concept of microcredits has spread throughout the world and a plethora of organisations providing small loans to the poor have come into being.[21] Microfinance institutions are most widespread in less developed countries, although they are by no means confined to them. Microlending programmes have recently been introduced in transition economies like Bosnia and Russia and even in western economies like Canada and the United States.[22] There are more than 5 million households served by microcredit schemes in the world today.

Prior to the microfinance revolution poor people's opportunities to take up loans had been severely limited. First, with few substantial possessions poor households cannot offer collateral to back up their loans. Second, the potential addressees of small loans in less developed countries often live in remote rural villages beyond the reach of the traditional banking system. Third, although loans needed for individual projects are small, their myriad nature makes monitoring and enforcement costs prohibitively high. Poor villagers' only access to credit had been through non-commercial development programmes which provided subsidised credit. However, since these schemes faced the same monitoring difficulties as traditional banks they often suffered from poor repayment rates and high costs and were typically doomed to failure for that reason.

Microfinance institutions use innovative means to overcome these problems. Though the single schemes differ vastly in their concrete implementations most of them share some main characteristics, the most prominent of which is that of group lending.[23] In a typical microfinance scheme borrowers with individual risky projects form groups which apply for loans together. The whole group is liable if one or more group members default. Thus, joint liability provides an insurance against individual risks. Even if an individual project fails and some of the borrowers are unable to repay, the group as a whole might still be able to do so. In this sense joint liability serves as a substitute for collateral. Unless the individual risks are perfectly correlated, the overall risk of involuntary non-repayment can be substantially lower than with individual borrowing.

Compared with traditional credit programmes in less developed countries, microcredit schemes have proved to be a great success. Repayment rates leaped to levels previously unseen in less developed regions. Grameen reports repayment rates of more than 90%; other programmes replicated such figures. However, the story is not without blemish: while many were successful, numerous MFI programmes have failed to live up to their promise.[24] Furthermore, the ultimate goal of establishing

---

[20] This chapter is based on the working paper "Group size and sozial ties in Microfinance institutions", joint work with Klaus Abbink and Bernd Irlenbusch.

[21] Examples include BancoSol and PRODEM in Bolivia, Banka Rakyat and Badan Kredit in Indonesia, BRAC in Bangladesh, Pride Africa in Kenya, Tanzania and Uganda, Malawi and Uganda, FINADEV in Benin.

[22] See CONLIN (1999), or ARMENDÁRIZ DE AGHION and MORDUCH (2000) for an overview.

[23] Therefore, microfinance institutions are often referred to as *joint liability lending institutions* (JLLIs), though some institutions also give small loans to individuals with good reputation (Armendáriz de Aghion and Morduch (2000)).

[24] See e.g. Conlin (1999).

sustainable credit schemes for the poor has not been reached, and most programmes still rely on subsidies and donations.

To improve the performance of microlending it is vital to improve the design of these schemes. Among practitioners as well as academic scholars there is a heated debate on the appropriate design of their key features. Lending to groups involves a fundamental dilemma: It may insure the credit against involuntary defaults, but individual borrower's reliance on fellow borrowers to repay the loan gives the former an incentive to free-ride. Indeed, if the success of an individual project is not sufficiently verifiable by other group members the dominant strategy for each individual is to shirk and hold others liable for own default. Being aware of this peril, MFI schemes have usually incorporated a number of safeguards, the most prominent of which is that borrower groups be self-selected. This is the case in many programmes, the expectation being that close social ties enhance peer pressure and group solidarity. In a theoretical study, Besley and Coate (1995) show that the possibility of inflicting social sanctions on peers helps improve repayment.[25] However, the effectiveness of self-selection is not undisputed; Wydick (1999) empirically investigates MFIs in Guatemala and finds no evidence that groups made up of acquaintances have higher repayment rates than those consisting of strangers. Social ties may even hamper repayment discipline if they lead to more "forgivingness" towards defaulters.[26]

Free-riding incentives may depend crucially on the size of the borrowing groups. In practice, it is unclear how far group size affects repayment rates. FINCA, the organisation which pioneered the village banking concept, lends to large borrower groups of between 10 and 50 members, and boasts repayment rates of 96%.[27] On the other hand, Grameen prefers smaller groups with typically only five members, in order to keep free-riding and in-group co-ordination problems under control. In the academic literature, both positions have their advocates. Ghatak and Guinnane (1999) argue that despite the insurance effect of larger groups, smaller groups are to be preferred for their better in-group co-ordination and reduced level of free-riding.[28] On the other hand, Buckley (1996) empirically finds that groups with ten or more members still can work effectively.

While much of the literature focuses on group size and social ties, the importance of dynamic incentives is acknowledged to a much lesser extent. In general MFIs aim at forming long term relationships with their client groups. Follow-up loans are frequently made subject to whether precious loans have been repaid. These two features are intended to encourage compliance with repayment obligations. It can be argued that this aspect of microcredit schemes is at least as important for generating repayment discipline as peer pressure between the group members. Creating dynamic incentives may become vital if microcredit schemes are to be applied to other economies. In the urban contexts of transition economies, for instance, it may be more difficult to form self-selected borrowing groups than in closer-knit rural communities. For this reason Armendáriz de Aghion and Murdoch (2000) argue that in such economies the focus on group-lending should be abandoned and suitable dynamic incentive schemes should be sought.

The empirical evidence on how the various design features of microcredit schemes affect their success is still limited and controversial. Murdoch (1999) promotes the need for well-designed experiments to identify the impact of MFI design features on their performance. However, controlled experiments in which single properties of institutions are systematically varied are difficult to carry out

[25] Conlin (1999) argues that replicating the MFI concept in western cities requires that schemes rely less on social ties. See also Wenner (1995).

[26] Guinnane (1994) conjectures that such an effect may have contributed to the failure of Irish credit co-operatives in the nineteenth century.

[27] FINCA's scheme differs from others with respect to the internal organization of borrower groups. Villages in FINCA projects form self-governed groups to which the loan is given. The distribution of credit is left largely to the group members themselves.

[28] This argument is supported in empirical investigations by Mosley and Dahal (1987) and Devereux and Fishe (1993). Theoretical investigations into the tension between positive insurance and negative free-riding effects are provided by Impavido (1998) and Armendáriz de Aghion (1999).

in the field, due to problems of data accessibility and comparability.[29] Furthermore, many relevant variables, e.g. the individual project risk, are unobservable. Therefore, we introduce an alternative approach to the empirical analysis of microfinance institutions. In an interactive laboratory experiment, we can control for specific parameters and observe behaviour in simulated MFI scenarios directly. Furthermore, we can identify which factors influence behaviour by changing particular variables of the experimental environment, holding all other aspects unchanged.

As a starting point for our research agenda, we construct a stylised MFI scenario. To study free-riding behaviour connected to group lending we model a situation in which repayment depends on group solidarity alone. To implement dynamic incentives, follow-up loans are subject to full repayment in the past. In our experiment each member of a group of n players invests in an individual risky project. Whether the project succeeds is known only to the individual investor. Subjects decide individually whether or not to contribute to the group repayment. However, only those with successful projects are able to contribute. The experiment ends if too few contribute, that is, if the group as a whole cannot fulfil its repayment obligation. We focus on three instrumental variables identified as crucial for MFI success: (1) The group size, which we set to $n = 2$, $n = 4$ and $n = 8$ in three conditions, (2) the dynamic incentive structure, and (3) the intensity of social ties between group members. In a "group recruitment treatment" subjects already had to enrol as a group to capture the influence of social ties.

We observe a high and robust performance of group lending institutions in all our treatments. In fact repayment rates are generally higher than those achievable by individual lending. While individual contribution rates decrease slightly with larger groups, the impact of free-riding is alleviated by the greater dispersion of risks. We clearly identify the importance of dynamic incentives. Towards the end of the experiment repayment rates decrease substantially. Furthermore, we find that social ties only have a relatively moderate effect on repayment rates. Closer-knit groups have higher, but less stable repayment rates than those composed of strangers. We also observe gender effects: In our experiment women tend to contribute more than men.

4.2 Related experimental studies

To our knowledge laboratory experiments on microcredit institutions have not yet emerged in the literature. However, since microcredit institutions allow group members to free-ride at the cost of the group, valuable insights may be gained from the literature on public good games. In such games, each subject of a group of n persons can decide to invest an amount x (up to a certain limit y) in a public good. Everybody in the group of n individuals receives a return of cx, where $c < 1$, but $nc > 1$. Thus, it is a dominant strategy for rational players not to invest, but the Pareto efficient solution is realised if everybody co-operates by investing the maximum amount. In experimental public good games subjects typically contribute considerable amounts but fail to reach the social optimum.

Inspired by the microfinance theme, Barr and Kinsey (2002) conduct an experiment on such a public good game in Zimbabwean villages. Though they do not aim at modelling a microfinance scenario, their research question is closely linked to common MFI practices. Many microfinance institutions target women as their clients, partly because they consider women’s empowerment as a goal as such, but also because women are often seen as more reliable borrowers.[30] The authors test this conjecture by analysing women’s and men’s behaviour in the standard public goods game. The differences they find are small, but qualitatively supportive of the MFI practice: Women tend to contribute more to the public good than men.

The study most akin to our experimental set-up is the public good experiment by Suleiman, Weiss, and Bornstein (2002). In their study, the players’ endowments y are stochastic and private information. Thus, as in our design, players cannot identify whether other players’ failure to contribute is due to bad luck or shirking. The authors observe less free-riding in this environment than in a standard public

---

[29] These difficulties are discussed in Bolnick (1988) and Hulme (2000).

[30] There is some evidence that women’s repayment discipline is higher due to a better sense of responsibility (Ledgerwood 1999:38). Morduch (1999:1583) reports studies finding that male lending groups at Grameen have higher default rates than female groups. For critical accounts of the women-focused policy see Goetz and Sen Gupta (1996), and Kabeer (2001).

good setting with fixed endowments. Thus, players are less tempted to free-ride themselves when they cannot identify others' free-riding.

Several authors have examined the role of social factors in experimental public good games. All these studies, however, deal with symmetric situations in which endowments are the same for all and known to all players. Gächter and Fehr (1999) investigate whether social approval incentives reduce free-riding behaviour in a repeated public good game. Subjects have the opportunity for social approval towards their group members after the experiment. It turned out that social approval alone could not enhance co-operation. However, if in addition subjects could familiarise themselves with each other before the experiment co-operation increases significantly.[31] This suggests that ex-ante familiarity may be important in establishing co-operation. In another study Van Dijk, Sonnemans, and van Winden (2002), and Brandts, Riedl, and van Winden (2002) investigate the development of social connectedness as a result of repeated interaction in a public good setting and find that social attachment becomes stronger after successful co-operation.

The effect of group size has been studied first by Isaac, Walker, and Williams (1994) in a public good experiment with 4, 10, 40, and 100 participants. They find that contrary to the common conjecture contributions even increase with very large groups. A similar result is obtained by Carpenter (2002), who compares groups of 5 and 10 subjects. However, in both studies marginal social benefits increase hugely as the group size increases, which may account for this effect. Unless there are strong synergies between the individual projects within an MFI borrowing group this is typically not a characteristic of microfinance institutions.

Unlike in most public good games mentioned above repayment in a microfinance scenario does not solely depend on the willingness but also on the ability to repay. Through no fault of their own, individuals whose project fail cannot contribute to repayment and hence they rely on the solidarity of fellow group members. Individuals' sense of solidarity is assessed in the solidarity game experimented by Selten and Ockenfels (1998). Three players each roll a die to determine whether they win a prize. Winners can transfer money to losers. Contrary to the game theoretic prediction the great majority make substantial transfers where females show more solidarity than males.

None of these studies has been carried out with a microfinance background in mind. Consequently, there is no study examining group solidarity in a dynamic environment with follow-up benefits conditional on compliant behaviour. The incentive structure in the existing studies is quite different from a typical microfinance environment. Therefore, the findings of these studies cannot be immediately transferred to the microfinance institutions in question. Hence the need for a new experiment.

4.3 Model and experimental design

We consider a very simple experimental set-up capturing essential features of group-lending. Since microfinance institutions come in many different forms, we were forced to make design choices. In the present study, we focus on the conflict between free-riding and group solidarity in the borrower group, thus, we assume that individual repayment cannot be enforced. To focus on the group solidarity aspect, we assume that individuals have no means of verifying the success or failure of their fellow borrowers' individual projects.[32] Furthermore, by assuming that the success of the individual projects is uncorrelated, we abstract from complications that arise if risks are connected, as are, e.g., if their success depends on seasonal conditions. Finally, we consider a symmetric situation in which there are no differences in the individuals' strategic situation.

---

[31] This is in line with earlier findings by Dawes, McTavish, and Shaklee (1977), Isaac, McCue, and Plott (1985), Isaac and Walker (1988, 1991), Bohnet and Frey (1999). Although they examine different questions, they all find that pre-play communication increases co-operation.

[32] This is in line with the assumptions made by Besley and Coate (1995) in a theoretical study. The issue of peer-screening and peer-monitoring in group lending is addressed by Stiglitz and Weiss (1981), Stiglitz (1990), Varian (1990), Ghatak and Guinanne (1999), Armendáriz de Aghion and Gollier (2000), and Ghatak (2000).

4.3.1 The model

A group of n individuals receives a loan, for the repayment of which all group members are jointly liable. The loan enables each group member to invest in an individual risky project. All projects are of the same type, and the probability of success of any given project is 5/6. In case of success, the investor receives a project payoff of 420 talers (the fictitious experimental currency). If the project fails, however (the probability of which being 1/6), the subject receives a project payoff of zero.

After all projects have been carried out, the group loan has to be repaid. For repayment to ensue, we assume that each individual is supposed to repay 210 talers, and hence the group is liable to repay a total amount of 210n talers (for example, if we assume a loan of 175 talers per individual and an interest rate of 20%). Those individuals whose projects failed cannot contribute to the repayment; to ensure that this condition pertains we assume that no investor receives income from another source and that none possesses savings which could be used to repay the loan in the eventuality that the project fails.

Individuals whose project succeeds decide whether to contribute to the group repayment. As mentioned above, information on the project's success or failure is private; no other group member can ascertain whether an individual's default is strategic or due to the failure of the project. Hence, we model an idealised scenario in which the repayment of loans must ensue in the absence of means to enforce repayment.

To model joint liability in a simple and straightforward way, the debt of 210n talers is split evenly among those individuals who are able and willing to contribute. Thus, the less individuals contribute, the higher is the burden for the single contributor. Since contributions can only be financed from the current round's project payoffs, full repayment is only possible if at least half of the group members contribute.

Only if the group fulfils its repayment obligation, does the game continue into a further round, which proceeds in the same way with the same group members (A maximum of ten rounds can be played). If more than half the group members default (regardless of whether the default is strategic or due to project failure), then the group cannot repay the full amount, in which case no further rounds are played and the subjects of the group obtain no further payoffs during the experiment. With this feature, we model the practice of many MFIs which make follow-up loans conditional on the full repayment of previous loans.

After each round, players are informed about the number of contributors in the respective round (but not their identities), their own project payoff, and their round payoff (comprised of one's own project payoff minus the players' share of the repayment burden).

The game theoretic prediction, assuming that players maximise their own income, is that no contributions at all are made thus bringing play to an end after the first round. The intuition is straightforward. In the last round, it is obvious that no player would ever contribute. Now consider the penultimate round. table 4.1 shows the payoff distribution a player would get given the number of other players willing and able to play, for eight-player groups (n = 8).[33] The entries in italics mark the cases in which the number of players contributing suffices for the game to reach the final round. The payoffs are composed of the (sure) payoffs in the penultimate round plus the lottery that the player gets in the final round. It can easily be seen that the payoff distribution a player gets from not contributing stochastically dominates the one he or she would obtain by contributing. This is true regardless of the actual number of other players willing and able to repay. Therefore, it can never be profitable to contribute in the penultimate round, and hence one would predict that no game enters round 10. By induction, it follows that in a subgame perfect equilibrium, no contributions are made in any round, and play will end after the first round, in which no contributions are made.

[33] The argument is completely analogous for *n* = 2 and *n* = 4. It can easily be generalised for a wide range of other parameter constellations. Notice that the argument does not require assumptions about the subject's attitude to risk.

Table 4.1. Payoff with and without own contribution in the penultimate round (n = 8)

Decision	number of other players willing and able to repay							
	0	1	2	3	4	5	6	7
cont.	0	0	0	420; $^{5}/_{6}$ 0; $^{1}/_{6}$	504; $^{5}/_{6}$ 84; $^{1}/_{6}$	560; $^{5}/_{6}$ 140; $^{1}/_{6}$	600; $^{5}/_{6}$ 180; $^{1}/_{6}$	630; $^{5}/_{6}$ 210; $^{1}/_{6}$
not cont.	420	420	420	420	840; $^{5}/_{6}$ 420; $^{1}/_{6}$	840; $^{5}/_{6}$ 420; $^{1}/_{6}$	840; $^{5}/_{6}$ 420; $^{1}/_{6}$	840; $^{5}/_{6}$ 420; $^{1}/_{6}$

Notes: The entries in italics are lotteries, where each line represents one outcome. Within each line the semicolon separates the payoff and the probability of winning this payoff.

Through the unambiguous game theoretic property of the subgame perfect equilibrium the game draws an idealised picture of a microfinance situation in which repayment cannot be established by players' own payoff maximisation alone. Therefore, the game allows us to study the impact of joint liability in a microfinance relationship, since repayment out of group solidarity stands in sharp contrast to equilibrium behaviour.

It is worthwhile to look at a hypothetical benchmark of individual lending. A single individual can repay the loan if the project is successful. This occurs with a probability of 5/6. Contrary to the case of group lending, an individual would prefer to repay in all rounds bar the last, since the benefits of future credit outweigh the short-term profits of shirking. However, since projects may fail, expected repayment rates cannot exceed 5/6. Thus, while group-lending creates free-riding incentives (according the game-theoretic prediction there would be no repayment at all), the dispersion of risks makes it possible to generate higher repayment rates and more profitable loans for the lender. This, however, requires group solidarity, and one of our research questions is precisely whether group-lending mechanisms are able to outperform the benchmark of individual expected repayment rates of 5/6.

4.3.2 Treatments

We designed our experiment to examine two major issues in the design of microfinance schemes. The first issue concerns the effect of different group sizes on repayment performance. In absence of strategic default, larger group sizes provide some insurance against uncorrelated individual risks. However, it is unclear how group size behaviourally affects the tendency to free-ride. If free-riding is more pronounced in larger groups, this might counteract the insurance effect of larger groups. To test for the effect of group sizes, we conducted experiments with group sizes of n = 2, n = 4, and n = 8.[34]

The second issue we address is the effect of social ties on behaviour in a microcredit group. Typically, MFI borrowers are self-selected groups whose members have known each other for some time. Thus, we induce different levels of social ties by applying two distinct recruitment techniques: In the individual recruitment (IR) treatments, subjects register individually for the experiment, thus minimising social ties between participants interacting with each other; In the group recruitment (GR) treatment, potential participants are required to register for the experiment in groups of four. The latter method, ensures that groups are self-selected, since subjects need to form groups in order to register themselves for the experiment. This method resembles the self-selection process required by real-life microlenders. As in their procedures borrower groups are formed before they enter the microfinance scheme.[35] To control for the level of acquaintance between the group members, we requested the subjects to indicate the intensity of their contact to the other group members on a scale from 1 (no contact) to 7 (frequent contact), separately for each of the other group members. We distinguish between professional and private contacts. It seems plausible that social ties are more pronounced if

[34] This roughly covers the range of most typical MFIs. Only few lend to groups with more than eight borrowers.

[35] When they register for our experiment, however, subjects did not know the task that they would perform. This was explained to them only in the actual session. This ensured that the state of information about the task was the same for all subjects at the outset of a session.

the contacts are intensive and private. All statements were made anonymously.[36]

We conducted our GR treatment with a group size of n = 4, given our concern that a group size of two would allow couples (with a common household budget) to register as a "group". If the two players effectively act as one then the social conflict is removed and the incentive structure is reversed. In fact contributing in all rounds except the last one is socially as well as individually optimal. On the other hand we did not conduct experiments with n = 8 for practical reasons: it would have proved too difficult to find self-selected groups of eight subjects. Table 4.2 indicates the factorial design of our experiment, and, in brackets, the number of subjects in the corresponding treatment.

Table 4.2. The treatments of our experiment

Group size	Recruitment method Group	Individual
n = 2		IR2 (16)
n = 4	GR4 (32)	IR4 (28)
n = 8		IR8 (64)

Notes: In brackets are the numbers of subjects.

4.3.3 Experimental procedures

The experiment was conducted in the Erfurter Laboratorium für experimentelle Wirtschaftsforschung (eLab) at the University of Erfurt, Germany. Most subjects were students from various disciplines, where students of economics, law, and sociology constituted the largest fractions. To minimise presentation effects, we designed our experiment in a completely context free fashion. We presented the microfinance situation to the experimental subjects without connecting it to a microfinance story. We opted for a neutral presentation to avoid the uncontrolled effects of possible connotations raised by hypothetical stories, and to ensure best possible comparability with other experimental results.

Each session began with an introductory talk after which the written instructions were handed out to the subjects (Translations are provided in Appendix 3.A). The instructions were read aloud and explained in detail. After the introduction, the subjects were seated in cubicles, visually separated from one another by curtains. In the IR treatments, the cubicle numbers were randomly attributed to the subjects, and no subject was informed about the identity of the other players in his group. In the GR treatment, the subjects who had registered together formed a group. This was known to the subjects and further emphasised by decision sheets in different colours for the different groups in a session.

In each round, the success or failure of a project was determined through independent random draws for each subject, with a probability of one sixth for the failure of the project. Each subject rolled a die to determine his or her project's success or failure for each round. In order to overcome the difficulty of monitoring the veracity of subject's reports about their draw, we asked subjects at the very beginning of the session (before they knew the rules of the game) to roll a die ten times and to enter the outcomes in the first column of their decision sheet. Later in the experiment in each round the "losing number" which was the same for all subjects was determined by a randomly selected participant rolling a die. This number was publicly announced. All subjects whose individual number drawn in advance for this round matched the "losing number" met with the project failure and thus a project payoff of zero.

Each round proceeded as follows. The subjects were each handed one decision sheet, on which the complete history of play was presented on one page. The subjects indicated their repayment decision by ticking "Yes" or "No" boxes for the current round. After all subjects had completed their decisions, the experimenter collected the sheets. The "losing number" for the current round was then drawn. By

[36] For different research questions other techniques have been applied to study the impact of social ties. Gächter and Fehr (1999) and Bohnet and Frey (1999) invite strangers, but allow subjects in one condition to get acquainted with each other in a pre-play communication stage. Van Dijk, Sonnemans, and van Winden (2002) and Brandts, Riedl, and van Winden (2002) study the evolvement of social ties through interaction and assess social ties using psychological tests before and after play.

letting the subjects make their repayment decisions before the "losing number" was drawn, we gathered decisions also in the case that an individual's project has failed.[37] The experimenters computed the results of the rounds and distributed the decision sheets containing the results.

All groups in the session[38] completed this procedure ten times, even when their group's play had actually ended as a result of repayment default. This ensured, first, that the constitution of the groups were not revealed by diverging duration of play (this would have distorted both anonymity and the statistical independence of the groups), and second, that a preference for a short playing time could not counteract the monetary incentives.

Immediately after the session, the subjects were paid anonymously in cash. The exchange rate was set to DM 0.02 per taler. The total earnings in the session ranged from DM 7.80 to DM 65.20 with an average of DM 25.50 for 1½ hours, considerably more than a student's normal hourly wage in Erfurt. One DM is equivalent to € 0.51. At the time of the experiment, the exchange rate to the US dollar was approximately DM/$ 2.10.

We aimed at gathering eight independent observations per treatment. Since a few subjects in one session did not show up, one session in the IR4 treatment could only be conducted with 12 rather than 16 subjects, such that in this treatment, only seven independent observations are available. Since subjects in different groups within a session did not interact with one another, each group can be treated as a statistically independent observation.

4.4 Results

The design of our experiment focuses on the impact of group size, social ties, and the dynamics of play. Most behavioural effects in our data express themselves in the overall number of contributions we observe in the treatment. However, before we conducted the experiment, we decided to also look at the contribution rates for the rounds 1-9 only. Since the game is certain to end after the tenth round, we may expect that behaviour in that round to be substantially different. To test for treatment differences, we chose to apply Fisher's two-sample permutation test to all pairwise comparisons of the contribution rates in the independent subject groups.

Further, we also analyse the first round behaviour. Statistical tests on first round behaviour can be advantageous since at the very beginning of play each individual decision is a statistically independent observation.

Repayment performance

Table 4.3 shows how many participants agreed to contribute (chose yes on the decision sheet) on average in the four treatments of our experiment. Additionally it reports average contribution rates in the first round as well as averages over the first nine rounds. In the first round, between 82.3% (treatment IR8) and 98.9% (GR4) of the subjects decide to contribute if their project turns out to be successful. The contribution rates over all rounds of the experiment are lower but still considerably high. Notice that according to the selfish equilibrium prediction we should observe no contributions at all.

Table 4.3. Contribution decisions and average duration of play

Treatment	IR2	IR4	IR8	GR4
yes-choices in round 1	93.4%	85.7%	82.3%	98.9%
yes-choices all rounds	79.8%	76.0%	72.3%	74.1%
yes-choices rounds 1-9	81.3%	79.4%	72.3%	75.9%
average number of rounds played	7.5	7.0	5.1	7.0

[37] Brandts and Charness (2000) examine behaviour in different interactive games comparing spontaneous play and complete strategy elicitation, but find results to be unaffected. They suggest that both procedures be equivalent for low-complexity tasks.

[38] We conducted each session with at least two groups to ensure that subjects (in the IR treatment) could not identify which of the other subjects belong to their group.

From the lender's point of view the most interesting question regarding different MFI designs is how they affect the repayment rate. To address this question in our experimental framework one could look at the actual repayment rates realised in the sessions. However, these are highly influenced by the realisations of the random draws. Thus, we rather look at expected repayment rates, which we define as the repayment rate that the lender could expect if we take observed default rates as a proxy for the probability that a borrower defaults strategically. Denote by φ the probability that a borrower is willing to repay. Then, actual repayment probability is 5/6φ, since he can only repay when his project is successful. Given this probability, the number of group members actually contributing is binomially distributed with a single event probability of 5/6φ. Thus, the expected repayment rate ERR can be computed as

$$ERR(\varphi) = \frac{\sum_{k=0}^{n/2-1} 420 \cdot k \cdot B(n,k,\tfrac{5}{6}\varphi) + \sum_{k=n/2}^{n} 210 \cdot n \cdot B(n,k,\tfrac{5}{6}\varphi)}{210 \cdot n},$$

where B denotes the non-cumulated binomial distribution. The repayment (in talers) in the case that less than n/2 group members actually repay is given by the first sum of the numerator. In this case, the loan is only partially repaid, but those who pay must give their entire project payoff of 420. The second sum computes the expected repayment in case that the whole loan is repaid.[39]

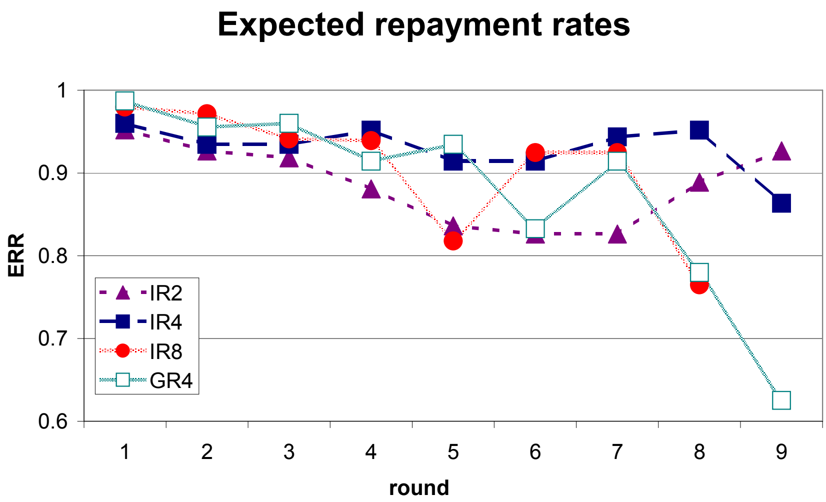


Figure 1. Expected repayment rates

Figure 1 depicts the expected repayment rates for the four treatments of the experiment, using for φ the observed rates of contribution decisions in the corresponding round and treatment.

As mentioned earlier, individual lending can at maximum generate a repayment rate of 5/6 = 83.3%. Recall that according to the theoretical predictions, individual borrowers would always, groups never repay. In our experiment, however, lenders would prefer group lending to individual lending. In all treatments, expected repayment rates are above this theoretical benchmark most of the time, and are substantially higher in earlier rounds. Thus, group lending outperforms individual lending. Even without

[39] The insurance effect of larger groups was principally understood by the subjects. In a post-experimental questionnaire, we asked subjects to estimate the probability to reach the final round for different group sizes ($n \in \{2, 4, 8\}$) and round numbers ($n \in \{5, 10, 20\}$) given that all group members are willing to contribute. On average the influence of the two dimensions was assessed qualitatively correct, though overall probabilities were quantitatively underestimated. The latter result replicates previous findings by Gneezy (1996) and Abbink, Irlenbusch, and Renner (2002) in the sense that dynamic effects are not sufficiently taken into account.

any monitoring opportunities, trust and group solidarity can establish substantial repayment rates making group-lending advantageous to the lender.

4.4.1 The effect of the group size

While large groups allow for a greater dispersion of risks one could expect that they are also prone to more free-riding. Indeed, table 4.3 suggests that contribution levels tend to decrease with the size of the group but only to a moderate extent. The statistical analysis provides only weak support. The non-parametric Jonckheere test, applied to the rates of yes-choices in rounds 1-9 in the single independent groups, rejects the null hypothesis of equal rates in favour of the hypothesis of decreasing rates at a weak significance level of $\alpha = 0.10$ (one-sided).[40] Of all pairwise comparisons, we can reject the null hypothesis of equal contribution rates only for the comparison of the contribution rates in rounds 1-9 for IR8 versus IR4. The rates tend to be weakly significantly ($\alpha = 0.10$, one-sided, Fisher's two-sample permutation test) lower in the IR8 condition. All other differences are not statistically significant.

Table 4.3 also indicates a decreasing average duration of play. However, the Jonckheere test does not reject the null hypothesis of an equal number of played rounds. Notice that the continuation of play is not only determined by the subjects' decisions, but also by the chance moves determining success and failure of the individual projects.

4.4.2 Dynamics of play

Since we have a finite number of rounds the dynamic incentives become weaker over time, as there are fewer rounds left in which profits can be made. Thus, we should expect contribution rates to decrease towards the end of play. Figure 2 shows that this is indeed the case. Though still more than two thirds of subjects contribute in round 7, late round contribution rates fall substantially. However, they do not reach zero, even in the last round of play. It seems surprising that even in the last round a considerable number of subjects contribute, though no further rounds can be expected. [41]

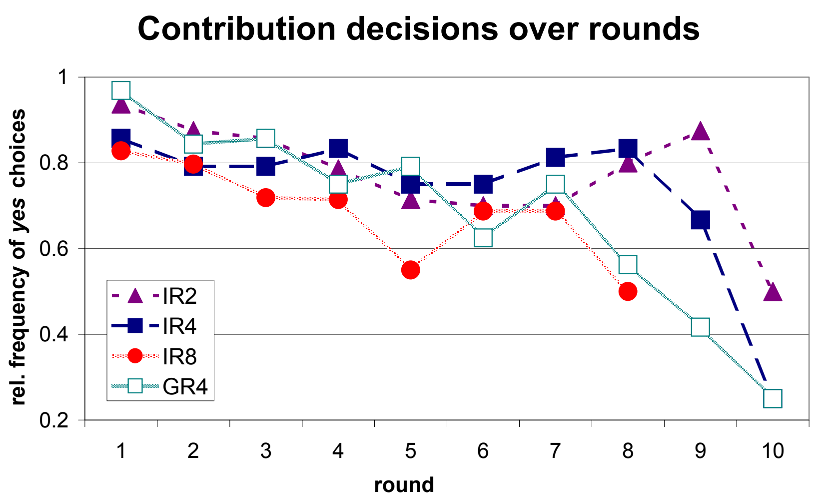


Figure 2. Relative frequency of yes choices over rounds

The decreasing trend of the contribution decisions can be observed not only in the aggregate data, but also in the single groups. We compute, for each session separately, non-parametric Spearman rank correlation coefficients between the number of yes-choices and rounds. Table 4.4 reports these rank correlation coefficients. Using these as summary statistics, the binomial test rejects the null hypothesis

---

[40] The analysis of group size effects is based on the three treatments with individual recruitment.

[41] Possibly a solidarity motive accounts for this type of behaviour. If a subject expects other subjects to contribute, then own contribution can ease the fellow players' burden of repayment.

that positive and negative correlation coefficients are equally likely at a significance level of at least α = 0.05 (one-sided) for all four treatments.[42]

Table 4.4. Spearman rank correlation coefficients of yes-choices over the played rounds (excluding round 10)

Treatment group	IR2	IR4	IR8	GR4
1	–	0	–0.77	–0.61*
2	–0.51	–0.52	–0.82	–0.88*
3	0	–0.62	–0.67	–0.63*
4	–0.27	–	–0.89	–0.77
5	–0.73*	–0.42	–	–0.95
6	0	–0.57	–0.98*	–0.65
7	–0.67	–0.75	–0.32	–
8	–0.87		–0.77	–0.75*

"*" significant at α = 0.05 (one-sided)

"–" number of played rounds lower than three

4.4.3 The effect of social ties

In the GR treatment, the groups who registered together can be expected to be a self-selected group in which social ties are stronger than in the anonymously matched groups of the IR treatment. The question arises whether these stronger social ties result in higher repayment rates due to a higher impact of group solidarity in self-selected groups. Our data provide mixed evidence with this respect. While the first round contribution rate rises from 85.7% in IR4 to 96.9% in GR4 (insignificant, p = 0.12 according to the Fisher exact test), we obtain an overall contribution rate which is even slightly lower in GR4 than in IR4 (74.1% versus 76.1%). Thus, the comparison of the contribution rate over the two treatments does not provide strong evidence for an effect of social ties on repayment rates. These results are in line with those of the survey study by Wydick (1999), who finds a "surprisingly small degree to which social ties within borrowing groups affect group performance".

However, further exploration of the GR4 data indicates some support for an effect of social ties. As figure 2 indicates, the rate of yes-choices starts at a higher level in GR4, but we can observe a sharp decrease in later rounds. In the IR4 treatment, the rates of yes-choices is more stable towards the end of the experiment. If we compare the Spearman rank correlation coefficients of the yes-choices in the single groups of these two treatments, as reported in table 4.4, we find that the coefficients in GR4 are greater (in absolute values) than in IR4. Fisher's two-sample randomisation test rejects the null hypothesis of equal rank correlation coefficients at a significance level of α = 0.01 (one-sided). Stronger social ties seem to induce a high, but less stable willingness to repay. Possibly subjects tend to be less "tolerant" towards their peers' suspected shirking when the fellow group members are their friends.[43]

Within the GR4 treatment, we find some evidence that the extent to which the group members are socially tied matters. As mentioned earlier, we asked for statements about the level of acquaintance between the group members. When we correlate the level of private social contacts to the overall contribution rate in the groups, we obtain a Spearman rank correlation coefficient of r_S = 0.73, which is significant at α = 0.05 (one-sided). Thus, stronger private contacts between the group members seem to have a positive effect on repayment decisions. The same analysis with the level of professional contact does not yield a significant result.

---

[42] Notice that the exclusion of round 10 is conservative to our findings.

[43] It is well-known that subjects in public good experiments cease co-operation when they observe free-riding by other group members (see, e.g. Keser and van Winden (2000) and Falk, Fischbacher, and Gächter (2002)). Conditional co-operation is much harder to achieve in our game since other players' shirking is not unambiguously observable. However, subjects may abandon co-operation if they feel that observed total contribution levels are suspiciously low.

4.4.4 Gender effects

Though the study of gender effects has not been one of our core issues at the outset of the experiment, we kept record of the gender of each participant. Table 4.5 shows first round and overall rates of yes-choices, for women and men separately. It can be seen that women's rate of yes-choices is higher than that of men in all treatments with individual recruiting. Fisher's exact test applied to the number of yes-choices in round 1 yields a one-sided significance level of $\alpha = 0.01$, if we pool the data from all three IR treatments.[44]

Though the frequencies in table 4.5 show sizeable differences in the contribution rates of females and males, the results should be interpreted with caution. First, the relatively small number of independent observations does not allow to comprehensively control for all possibly relevant socio-economic variables, such that we cannot detect compound effects with other background variables. Second, we can statistically detect the effect only for the pooled data of the first round data of the individual recruitment conditions, such that the evidence is relatively weak (men's contributions in the group recruitment treatment are even slightly higher). Finally, interpretations with respect to the current MFI practice to target women seem difficult, as the cultural background of our subject pool regarding gender relations might be considerably different from that in the target societies of microfinance schemes.

Table 4.5. Contribution decisions of women and men

Treatment	IR2	IR4	IR8	GR4
yes-choices of women, first round	100.0%	94.7%	90.0%	96.2%
yes-choices of men, first round	50.0%	66.7%	70.8%	100.0%
yes-choices of women, overall	78.4%	80.5%	78.8%	77.0%
yes-choices of men, overall	55.0%	55.0%	54.9%	81.3%

4.5 Summary and discussion

We introduce an experimental microfinance game to separate the impact of essential characteristics of group lending contracts on repayment performance. Small loans are given to groups who are jointly liable for repayment. Incentives to repay are provided through the prospect of follow-up loans. We report an experiment to investigate the influence of those features on strategic default of group members. Treatments involve different group sizes and a condition in which self-selected subject groups register for the experiment together.

We observe high willingness to repay in all treatments, though game theory would recommend free-riding. Through this high degree of solidarity, the experimental lending groups reach high repayment rates and are able to sustain the flow of further credits for several periods. The willingness to contribute declines as the experiment proceeds, but it remains remarkably high, even in the later periods, where the incentives to keep up the borrowing-relationship diminish.

We also examine the question of group size which is lively discussed in MFI policy. A dilemma arises when the advantage of larger groups through the insurance effect is counteracted by less solidarity and more free-riding. Our results show that the performance of the experimental microcredit groups is surprisingly robust with respect to group size. While the larger groups indeed manifest a higher tendency towards shirking, their superior dispersion of risk makes them perform at least as well as smaller groups in our parameter constellation.

Our results are also robust against variations of social ties between the members of the experimental borrower groups. Overall performance is not significantly worse with strangers than with good acquaintances. Self-selected groups exhibit a higher willingness to contribute in the beginning of the experiment, but their behaviour is less stable, possibly because friends are less willing to tolerate supposed free-riding by others.

[44] Notice that in the first round, the individual subjects can be treated as statistically independent observations.

Our data exhibit a gender effect. Females show a significantly higher willingness to contribute than males which is in line with previous findings in related experimental games (Selten and Ockenfels (1998)). This is also interesting as it seems to support the common practice of MFI lenders to give loans predominantly to women. However, as gender relations are different in different parts of the world, this finding should not be generalised prior to future research.

We believe that the experimental method is especially well suited to gaining a deeper understanding of how and why group lending schemes succeed or fail in practice. The present study provides a framework in which the effects of central MFI features can be disentangled. Of course, the present study should be seen as a starting point rather than a comprehensive exploration. To keep things simple, we have developed a very basic model that naturally lacks many of the complexities of real life group lending contracts. Furthermore, our experiment has been conducted in our laboratory using our student subject pool. Microfinance institutions are implemented all over the world within a great variety of economic and cultural backgrounds, such that the replication of our findings in different societies with different subject pools seems a promising research agenda for the future.

4.6 References

Abbink, Klaus, Bernd Irlenbusch, and Elke Renner (2002): An Experimental Bribery Game, Journal of Law, Economics, and Organization, 18, 428-454.

Armendáriz de Aghion, Beatriz (1999): On the design of a credit agreement with per monitoring, Journal of Development Economics, 60, 79-104.

Armendáriz de Aghion, Beatriz and Christian Gollier (2000): Peer group formation in an adverse selection model, The Economic Journal, 110, 632-643.

Armendáriz de Aghion, Beatriz and Jonathan Morduch (2000): Microfinance beyond group lending, Economics of Transition, 8, 401-420.

Barr, Abigail, and Bill Kinsey (2002): Do Men Really have no Shame? Working paper, University of Oxford.

Besley, Timothy and Stephan Coate (1995): Group lending, repayment incentives and social collateral, Journal of Development Economics, 46, 1-18.

Bohnet, Iris and Bruno S. Frey (1999): The sound of silence in prisoner's dilemma and dictator games, Journal of Economic Behavior and organization, 38, 43-57.

Bolnik, Bruce R. (1988): Evaluating Loan Collection Performance: An Indonesian Example, World Development, 16, 501-510.

Brandts, Jordi and Gary Charness (2000): Hot vs. Cold: Sequential Responses and Preference Stability in Experimental Games, Experimental Economics, 2, 227-238.

Brandts, Jordi, Arno Riedl and Frans van Winden (2002): Competitive Rivalry and Social Ties. An Experimental Study, mimeo. Institut d'Anàlisi Econòmica (CSIC), Barcelona.

Buckley, Graeme (1996): Rural and agricultural credit in Malawi: A study of the Malawi Mudzi Fund and the smallholder agricultural credit administration, in: David Hulme and Paul Mosley (Eds): Finance Against Poverty, London: Routledge.

Carpenter, Jeffrey P. (2002): Punishing Free-Riders: how group size affects mutual monitoring and the provision of public goods, Middlebury College Working Paper Series No. 0206.

Conlin, Michael (1999): Peer group micro-lending programs in Canada and the United States, Journal of Development Economics, 60, 249-269.

Dawes, Robyn M., Jeanne McTavish and Harriet Shaklee (1977): Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation, Journal of Personality and Social Psychology, 35, 1-11.

Devereux, John and, Rauymond P.H. Fishe (1993): An Economic Analysis of Group Lending Programs in Developing, Developing Economies, 31, 102-121.

Falk, Armin, Urs Fischbacher, and Simon Gächter (2002): Isolating Social Interaction Effects – An Experimental Investigation, mimeo, University of St. Gallen.

Fehr, Ernst and Simon Gächter (2000): Cooperation and Punishment in Public Goods Experiments, American Economic Review 90, 980-994.

Gächter, Simon and Ernst Fehr (1999): Collective action as a social exchange, Journal of Economic Behavior and organization, 39, 341-369.

Ghatak, Maitreesh (2000): Screening by the company you keep: Joint liability lending and the peer selection effect. The Economic Journal, 110, 601-631.

Ghatak, Maitreesh and Timothy W. Guinnane (1999): The economics of lending with joint liability: theory and practice, Journal of Development Economics, 60, 195-228.

Gneezy, Uri (1996): Probability judgments in multi-stage problems: Experimental evidence of systematic biases, Acta Psychologica, 93, 59-68.

Goetz, Anne Marie and Rina Sen Gupta (1996): Who Takes The Credit? Gender, Power and Control over Loan Use in Rural Credit Programmes in Bangladesh, World Development, 24, 45-63.

Guinnane, Timothy W. (1994): A failed institutional transplant: Raiffeisen credit cooperatives in Ireland 1894-1914, Explorations in Economic History, 31, 38-61.

Hulme, David (2000): Impact Assessment Methodologies for Microfinance: Theory, Experience and better Practice, World Development, 28, 79-98.

Impavido, Gregorio (1998): Credit rationing, group lending and optimal group size, Annals of Public and Cooperative Economics, 69, 243-260.

Isaac, R. Mark and James M. Walker (1988): Communication and free riding behavior: the voluntary contribution mechanism, Economic Inquiry, 26, 585-608.

Isaac, R. Mark and James M. Walker (1991): Costly communication: An experiment in nested public goods problem, in: Laboratory Research in Political Economy, Thomas Palfrey (ed.), Ann Arbor University of Michigan Press, 269-286.

Isaac, R. Mark, Kenneth F. McCue and Charles R. Plott (1985): Public goods provision in an experimental environment, Journal of Public Economics, 26, 51-74.

Isaac, R. Mark, James M. Walker and Arlington W. Williams (1994): Group Size and the voluntary provision of public goods, Journal of Public Economics, 54, 1-36.

Kabeer, Naila (2001): Conflicts Over Credit: Re-Evaluationg the Empowerment Potential of Loans to Women in Rural Bangladesh, World Development, 29, 63-84.

Keser, Claudia and Frans van Winden (2000): Conditional Cooperation and Voluntary Contributions to Public Goods, Scandinavian Journal of Economics 102, 23-29.

Ledgerwood, Joanna (1999): Microfinance Handbook. World Bank, Washington D.C.

Morduch, Jonathan (1999): The microfinance promise, Journal of Economics Literature, 37, 1569-1614.

Mosely, Paul and Rudra Prasad Dahal (1987): Credit for the Rural Poor: A Comparison of Policy Experiments in Nepal and Bangladesh, Manchester Papers on Development, 3, 45-59.

Selten, Reinhard and Axel Ockenfels (1998): An experimental solidarity game, Journal of Economic Behavior and Organization, 34, 517-539.

Stiglitz, Joseph E. and Andrew Weiss (1981): Credit Rationing in Markets with Imperfect Information, The American Economic Review, 71, 393-410.

Stiglitz, Joseph E. (1990): Peer monitoring and credit markets. World Bank Economic Review, 4, 351-366.

Suleiman, Ramzi , Taly Weiss and Gary Bornstein (2002): The Enhancing Effect of Uncertainty about Individual Endowments on the Provision of Public Goods, mimeo, Hebrew University Jerusalem

Varian, Hal R. (1990): Monitoring Agents with Other Agents, Journal of Institutional and Theoretical Economics, 146, 153-174.

Van Dijk, Frans, Joep Sonnemans and Frans van Winden (2002): Social Ties in a Public good Experiment. Journal of Public Economics, 85, 275-299.

Wenner, M. (1995): Group credit. A means to improve information transfer and loan repayment performance. Journal of Development Studies, 32, 263-281.

Wydick, Bruce (1999): Can social Cohesion be harnessed to repair market failures? Evidence from group lending in Guatemala, The Economic Journal, 109, 463-475.

5 Price rigidity in customer markets[45]

5.1 Introduction

The intuition that implicit contracts between buyers and sellers are an important source of price rigidity has been eloquently developed by Arthur Okun (1981), who coined the term “customer markets” for markets with long-term relationships between buyers and sellers. In customer markets, he argued, sellers may be reluctant to increase prices in response to temporary cost shocks because changes in the terms of the implicit contract may antagonize customers and disrupt the customer relationship. Okun also speculated that customers may be less antagonized if price increases can be justified by increases in sellers’ costs. Hence, according to this intuition, long-term relationships in customer markets are a source of price rigidity, but price rigidity is less pronounced if price increases can be justified by cost increases.

Despite the importance of price rigidity in macroeconomics and industrial organization, relatively little was known about the empirical validity of Okun’s intuitions until recently (see Carlton 1989 or Wynne 1995 for surveys). This lack of empirical knowledge is surprising since customer markets are of utmost economic importance. According to Blinder et al. (1998, p.302), “about 85 percent of all the goods and services in the U.S. nonfarm business sector are sold to ‘regular customers‘ with whom sellers have an ongoing relationship.” The reason why relatively little was known is that most theories of price stickiness rely on variables that are either unobservable by conventional methods or unobserved in practice. Hence, “the abject failure of the standard research methodology to make headway on this critical issue in the microfoundations of macroeconomics” (Blinder et al., p. 3) motivated researchers in recent years to use “unorthodox approaches” to investigate the issue of price rigidity in customer markets.

Empirical evidence on the relevance of Okun’s intuitions comes from questionnaire and experimental studies. In a much-cited questionnaire study, Kahneman et al. (1986) have investigated fairness perceptions of price increases in the general public, and find that people perceive price increases to be more acceptable if they can be justified by cost increases. However, their findings also suggest that price rigidity does not depend on whether buyers and sellers have a one-time or a long-term relationship. In a compelling interview study Blinder et al. evaluate motives for pricing decisions from the perspective of the sellers. Systematic interviews with managers who are responsible for pricing decisions show that long-term relationships are common and important, and that managers are careful not to antagonize their customers by their pricing decisions. The authors find support for the notion that managers believe customers to be less antagonized when price increases can be justified by cost increases. Much to their surprise, they do not find that price rigidity is related to long-term relationships. In a similar interview study with Swedish firms, Apel et al. (2001) find that implicit contracts are the most important explanation for price rigidity.

Experimental evidence usefully complements findings from questionnaire studies because market interaction can be observed in controlled environments. Cason and Friedman (2002) study price rigidity in an opaque market in which buyers develop customer relationships to save on search costs. The authors show that higher search costs tend to induce customers to remain with their seller, which in turn increases the sellers' market power and induces considerable price rigidity. However, the authors do not investigate the hypothesis that price rigidity is less pronounced if price increases can be justified by cost increases. The relevance of this hypothesis is investigated in a series of three closely related papers (Kachelmeier et al. 1991a, 1991b, Franciosi et al. 1995). These papers find weak and non-persistent price rigidity when price increases cannot be justified by cost increases. However, these studies do not analyze customer markets because long-term relationships are excluded by design in these experiments. It is important to note that all experimental studies cited

[45] This chapter is based on the paper “Price rigidity in customer markets”, Journal of Economic Behavior & Organization, Vol. 55, 2004, 575-593, joint work with Jean-Robert Tyran.

above analyze markets for “inspection goods.” In such markets, no quality uncertainty prevails.

Our experimental study is the first to investigate long-term relationships on markets for experience goods as an explanation of price rigidity. For experience goods, quality uncertainty prevails in the sense that buyers learn the quality of the good only after the purchase. We analyze markets for experience goods because our focus is on how trust in customer relationships generates price rigidity. We believe that trust is an important factor in shaping customer markets in addition to saving on search costs in an opaque market. In fact, Okun (p. 141) noted that “the extent to which firms are likely to enjoy repeat patronage depends both on the satisfaction of customers with previous purchases and their confidence that the supplier will maintain good performance.” Hence, long-term customer relationships may be upheld if buyers trust sellers to provide high quality and if sellers are trustworthy.

We compare behaviour in two experimental markets that differ with respect to whether buyers and sellers can develop long-term relationships. Both markets are opaque in the sense that market outcomes cannot be publicly observed and in both markets quality uncertainty prevails. High quality is more valuable to buyers, but its provision is more costly to sellers. In the “customer market,” buyers can trade repeatedly with the same seller and can thus develop long-term trading relationships. In contrast, this is not possible on the “anonymous market” where buyers and sellers do not know with whom they trade. In this market, economic theory predicts the “lemon outcome” (Akerlof 1970, see Cason and Gangadharan 2002 for experimental evidence). That is, low quality is provided at low prices because sellers cannot build a relation-specific reputation by providing high quality. In contrast, in the “customer market,” upholding customer relationships by providing high quality at a reasonable price may also be profitable to sellers if it induces customers to return to this seller.

To analyze whether customer relationships are upheld and whether they partially resolve the “lemons” problem, we compare the anonymous market and the customer market. To analyze whether the customer market causes price rigidity, we compare price adjustment in response to a cost shock in the two markets. To analyze the “cost justification” hypothesis of price rigidity, we compare price adjustment in two customer markets that exclusively differ with respect to the information about the cost shock. In one treatment, buyers are informed about the cost shock (i.e., information is public), and in the other treatment they are not (i.e., information is private to sellers). Hence, with public information a price increase is “justified” by a cost increase, while with private information it is not.

Our main results are as follows. First, we find that long-term customer relationships are frequently upheld and that average quality is about three times higher when customer relationships are possible. Higher quality on the customer market comes at a higher price, but upholding the relationship is profitable for both sellers and buyers on average. Hence, we find that customer relationships partially resolve the lemons problem. Second, we show that customers tend to penalize sellers for price increases and, in particular, for providing low quality by terminating the customer relationship. Third, we show that price rigidity is more pronounced if the price increase cannot be justified by a cost increase. In this case, sellers increase prices much less and bear considerable losses to avoid antagonizing customers. In all, we show that customer markets mitigate the “lemons” problem but are prone to price stickiness. Since long-term customer relationships are upheld in our customer market because buyers trust sellers to provide high quality, this paper also demonstrates how trust can shape market outcomes.

We proceed as follows. Section 2 explains the advantages of an experimental approach to price rigidity in customer markets and relates this study to the literature. Section 3 explains the experimental design, section 4 presents the experimental results, and section 5 concludes.

5.2 An experimental approach to price rigidity in customer markets

We test three hypotheses about to how trust shapes customer markets. Our first hypothesis H1 is that customer markets resolve the “lemons” problem. This hypothesis has already casually been suggested in the seminal article by Akerlof (see also Arrow 1973). That is, we test whether buyers trust sellers to deliver higher quality in a customer market than in an anonymous market.[46] We expect

[46] Note that the mechanism that leads to the "lemons outcome" of low quality at low prices in our paper as well as in the studies cited in this section is of the moral hazard type, while the mechanism described by Aero is of the adverse selection type.

this hypothesis to hold because two previous experimental studies have found that long-term relationships endogenously form in markets with quality uncertainty and that they improve market performance. Kollock (1994) obtains these findings by comparing markets with and without quality uncertainty and private vs. public offers. Brown et al. (2004) investigate a gift-exchange labor market in which buyers (firms) choose prices and sellers (workers) choose the quality (effort). They compare two markets that differ with respect to the anonymity of market participants and find that long-term relationships are upheld in non-anonymous markets and that these relationships are mutually beneficial. However, both studies analyze stable environments and do, therefore, not investigate whether long-term relationships cause price rigidity.

To test the “lemons” hypothesis, we compare market outcomes on the customer market and on the “anonymous market.” Note that our anonymous market does not correspond to an “auction market” as characterized by Okun. On auction markets homogenous goods are traded, and trading is organized as in exchanges for stocks or raw materials. Such markets are most closely approximated by experimental double auction markets. In contrast, we compare two posted-offer markets in which sellers post prices and in which goods are non-homogenous. We do not compare our customer market to a double auction market because in order to isolate the role of long-term relationships, we need to compare two settings that are otherwise identical (ceteris paribus variation).

The second hypothesis we test is the “invisible handshake” hypothesis H2. We analyze whether price increases on the customer market tend to undermine trust and disrupt the customer relationship. Because we investigate markets with quality uncertainty, both increasing one’s price and providing low quality may disrupt the customer relationship. For example, Okun (p. 154) suggested that “once the firm draws a clientele with attractive implicit contracts, any deviation unfavourable to customers is seen as a violation of these contracts.” Hence, our design also serves to investigate the relative importance of price and quality changes.

The third proposition we test is the “cost justification” hypothesis H3. We investigate whether customers are more willing to tolerate price increases if these are justified by cost increases. To this end, we compare price adjustment on the customer market when buyers are informed about the cost shock with a situation when they are not informed about the shock. Hence, we do not compare a cost and a demand shock to investigate the validity of the “cost justification” hypothesis. Okun suggested that prices are more flexible after a cost shock than after an increase in demand because price increases are perceived as more fair by customers in the former case. However, comparing cost and demand shocks involves a loss of experimental control because cost and demand shocks may differ in several dimensions. Instead, our approach of implementing a controlled variation of information conditions tests hypothesis H3 more directly.

Supportive evidence for H3 is provided by Blinder et al. (p. 157). The authors ask managers: “Does the understanding that prices should remain fixed hold when your costs increase, or do customers see price increases as justified when costs increase?” The responses (n = 111) were as follows: “When costs increase, our customers normally a) still want us to hold our prices (15.3%), b) attitudes are mixed (13.5%), c) tolerate price increases (71.2%).” Note that these findings reflect managers’ beliefs about customers’ price tolerance. However, we cannot infer from these responses how managers’ beliefs translate into actual pricing decisions or firm profits. For example, are managers willing to antagonize 15 or maybe 30 percent of their customers to increase profits on the remaining customers? While questionnaire studies are particularly useful to obtain information about decision-makers’ beliefs and attitudes, they are inapt to study market interaction and, ultimately, price rigidity. In contrast, our experiment analyzes market interaction under controlled conditions, but motives of market participants must be inferred from observed behaviour.

5.3 Experimental design

Section 5.3.1 provides a general description of the design, and section 5.3.2 explains the procedures and parameters in detail. Section 5.3.3 presents the predictions.

5.3.1 General description of the experimental design

In our experiment, two markets operate simultaneously: a customer market (CM) and an anonymous market (AM). In both markets, an experience good is traded. That is, buyers only learn the quality of the good after the purchase. The essential difference between the two markets consists in the possibility to become a repeat customer. In the CM, buyers and sellers can trade with the same partner repeatedly. Buyers receive an offer from a particular seller and decide whether to uphold their relationship with this specific seller. In contrast, the AM is anonymous. In the AM, buyers face offers from a number of unknown sellers, and due to the anonymity of the market, they cannot establish a customer relationship with a specific seller.

In the CM one buyer and one seller are matched at the beginning of a trading phase. In each period the seller posts a price to his buyer and chooses the quality he provides in case the buyer accepts. If the buyer accepts his seller's price offer, he learns the quality of the good and is re-matched with the same seller in the next period. Thus, the customer relationship is upheld as long as the buyer continues to accept his seller's offers. However, if the buyer rejects an offer from his seller, the customer relationship is terminated. As a consequence, both the buyer and the seller participate in the AM from this period on. According to this procedure, the customer relationship starts with an exogenous match but is endogenously upheld by customers. The reason why we start the relationship with an exogenous match is that the focus of our study is on trust within the relationship, not on consumer search.

In the AM all sellers simultaneously post prices to all buyers. Buyers see a list of anonymous price offers and can choose from these offers in random order. If a buyer accepts an offer, neither the seller nor the buyer can identify the trading partner. Hence, trading in the AM is completely anonymous. All buyers and sellers who are in the AM in a particular period remain there for the rest of the trading phase. A consequence of this procedure is that market participants can only switch from the CM to the AM, but not from the AM to the CM. This procedure implements a "grim trigger" strategy by customers and enforces an "unforgiving" behaviour that is empirically plausible (Engle-Warnick and Slonim (2003) show that this strategy is chosen by about 80 percent of the first movers in infinitely repeated trust games). This grim trigger strategy provides sellers in our experiment with strong incentives not to antagonize customers since customers cannot return to the seller at a later point.

To investigate the effect of information about the cost shock, we implement two treatment conditions (see table 5.1). In the treatment labelled PRIVATE, only the sellers are informed about the cost shock, while in the treatment called PUBLIC both sellers and buyers are informed about it.

Table 5.1: Overview over experimental conditions

	Information about cost shock is	
	PRIVATE	**PUBLIC**
Customer market (CM)	CMPrivate	CMPublic
Anonymous market (AM)	AMPrivate	AMPublic

5.3.2 Procedures, parameters and information conditions

Each experimental session has 5 trading phases with 10 periods each. In all sessions, 8 sellers and 4 buyers participate. In each period a seller can at most sell one unit of the good and a buyer can buy at most one unit. At the beginning of each trading phase 3 sellers and 3 buyers are randomly matched as trading pairs starting on the CM. All other participants start on the AM. As a consequence, there are at least five sellers and one buyer on the AM plus all trading pairs who at some point have terminated the customer relationship in the CM. This constellation implies that there is always an excess supply on the AM of 4 units of the good.

Participants can earn points by trading. These points are converted into money at the end of the experiment at a commonly known exchange rate. Participants' payoffs are Price – Seller Costs for a seller and Buyer Value – Price for a buyer if a trade is concluded, and zero otherwise.

Tabelle 5.2: Buyer values and seller costs

	buyer values	seller costs	surplus
High quality	200	80 (t = 3: 120)	120 (80)
Low quality	100	40 (t = 3: 60)	60 (40)

Table 5.2 shows the parameters of buyer values and seller costs. High quality is twice as valuable for buyers, but providing high quality is also twice as costly to the sellers as providing low quality. Therefore, the total surplus of a high-quality trade is twice the surplus of a low-quality trade. Since quality differences directly translate into efficiency differences, we will concentrate on reporting quality differences below. In period t = 3 of each trading phase, a temporary cost shock of 50 percent occurs. Hence, costs increase from period 2 to 3 from 80 to 120 for a high-quality good, and from 40 to 60 for a low-quality good. Note that costs fall back to the previous level from period 4 on.

The information conditions are as follows: In both treatments the payoff calculation, the number of trading phases and periods, and the buyer values are common information. The subjects do not know exactly how many sellers and buyers there are on the AM and the CM, but they are informed that there are always more sellers than buyers in the AM. The exact seller costs given in table 5.2 are only known to the sellers, but the buyers know that seller costs are lower for delivering low quality than for high quality and that the costs are the same for all sellers.

The two treatments differ exclusively with respect to the information about the cost shock. In both conditions, sellers are informed about the cost shock at the beginning of period 3. Buyers are informed about the cost shock in PUBLIC but are not informed about the shock in PRIVATE, and this condition is known to sellers. In particular, the information sellers obtain in PRIVATE has three aspects. Sellers know a) that costs increase by 50 percent, b) that this cost shock is temporary (they know that costs will fall back to the previous level in period 4), and c) that buyers are not informed about a) and b). In PUBLIC, a) and b) hold for both buyers and sellers, and sellers know that buyers are informed about a) and b). Note that our observations from period 3 are responses to the cost shock because participants are informed about it at the beginning of period 3.

At the beginning of each session the participants are randomly assigned to one of the booths in the computerized laboratory (we use the software z-tree, Fischbacher 1999). Written instructions which explain the experimental procedures in detail are handed out.[47] All subjects read instructions for both roles. There are no control questions, but subjects can ask questions before the experiment starts. Roles are randomly assigned in the first period and kept throughout the entire session. The written instructions contain no information on the sellers’ cost parameters, which are only displayed on sellers' computer screens from period 1 on. The information on the cost shock is shown on a separate screen that appears at the beginning of period 3 (see Appendix 5.B). In PRIVATE, the screen explicitly states that the information about the shock is shown to sellers only and that buyers do not have any information about the cost shock. In PUBLIC, the screen is shown to both buyers and sellers, and the information about the cost shock is announced aloud by the experimenter.

At the end of each period, subjects in both treatment conditions only get information about the outcomes (i.e., price, quality, and own profit) of their own transaction. That is, neither participants in the CM nor in the AM know about market outcomes (such as average quality, transaction prices, etc.) in the AM. Hence, both markets are opaque in the sense that no information about overall market outcomes is provided to market participants.

5.3.3 Predictions

In the AM, the theory of competitive markets predicts the “lemons outcome,” according to which low quality is provided at a low price. In particular, because of the anonymity of market interaction in the AM, a rational and self-interested seller always provides low quality. Because of the excess supply (of 4 units in each period) in the AM, prices should be driven down to the cost of providing low quality, for example, to 40 units (see table 5.2). However, we expect convergence to be slow and imperfect in the AM for three reasons. First, the AM is opaque in the sense that market participants only get information about their own transaction, but not about market outcomes. This lack of market feedback information is expected to lead to slow convergence (see Cason and Friedman for an extended discussion). Second, sellers in our experiment know buyer values, and fairness considerations may prevent prices from falling to costs. In particular, sellers may be reluctant to undercut each other knowing that the entire surplus goes to the buyers if prices equal costs. Third, buyers may trust sellers despite the anonymity of market interaction. It is known from many experimental studies on trust

[47] Instructions are available from the authors upon request.

games that people have a tendency to trust others even in perfectly anonymous one-shot interactions (e.g., Berg et al. 1995).

Two features of the CM contribute to the persistence of long-term customer relationships. First, markets are opaque. Participants in the CM only get feedback about their own market transaction, but do not know qualities and prices provided on the AM. Therefore, buyers have insufficient information to determine rationally whether to remain with their seller or to terminate the customer relationship to "experiment" (i.e., to uncover prices and qualities on the AM). However, a particular rule of our experiment may induce customers to remain with their seller: the termination of a relationship cannot be revoked. This rule makes "experimentation" potentially more expensive for buyers because they cannot return to the CM upon finding out that they earn less on the AM than on the CM. However, buyers know that the experiment has 5 phases in which participants can gather experience with both markets. Hence, we would expect customer relationships to be terminated at a higher rate in early phases and in late periods.

The second feature contributing to the persistence of long-term relationships is that buyers may trust sellers to provide high quality. Suppose a buyer trusts a seller to provide high quality and, upon receiving high quality at a "fair price", continues to trust the seller. Suppose further that the buyer terminates the relationship upon receiving bad quality or upon experiencing a price increase. Then, the seller is tempted in each period to reap a one-time gain by providing low quality, but has to trade off this one-time gain against the cost of losing his customer. While this cost is not known exactly to the seller because the market is opaque, it may be profitable for the seller to provide high quality if the relationship can be maintained for sufficiently many periods with sufficient probability (see Brown et al. for a more detailed account). A similar reasoning applies to price choices after the cost shock. Accordingly, a seller may find it profitable to refrain from increasing his price after the shock to avert antagonizing and losing his customer.

The remaining potential length of the customer relationship may be an important determinant of termination for two reasons. First, antagonizing the customer on the CM is more costly for the seller in early periods because the risk of not being able to trade on the AM is higher in early periods. Second, if the customer terminates the relationship in an early period, more periods of profitable interaction on the CM are foregone (see James 2002 for a more detailed analysis). These two factors lead us to expect that customer relationships tend to be terminated at a higher rate in late periods. However, since termination incentives are opposite for buyers and sellers, we have no clear prediction for the rate at which relationships end.

5.4 Results

We conducted 12 runs in each treatment at the University of Erfurt. In total, 288 undergraduate students from various disciplines participated in our experiment. Subjects earned € 15.3 (US$ 15, approx.) on average within about 120 minutes.

Table 5.3 summarizes important differences between the CM and the AM. The table shows the quality and prices averaged over all subjects, phases and periods. As can be seen in the table, high quality is provided about three times as often in the CM as in the AM. In particular, high quality is provided in 59 percent of the transactions in the CM, but only in 20 percent in the AM. We henceforth code high quality as 1 and low quality as 0, allowing us to speak of the percentage of high quality provided as "average quality." Higher quality in the CM comes at a higher price. In fact, the average price is 45 percent higher in the CM (115.1) than in the AM (79.9).

Since trading high quality fully translates into higher gains from trade (see table 5.2), the customer market yields considerable efficiency gains. In addition, both buyers and sellers earn more on the CM. In particular, average buyer earnings are 43.9 on the CM and 39.7 on the AM, and average seller profits are 47.8 on the CM, but only 12.2 on the AM. Hence, upholding the customer relationship is mutually profitable on average.

Table 5.3: Overview (averages, all periods, all phases, both treatments)

	Customer Market (CM)	**Anonymous Market (AM)**
Traded Quality	0.59	0.20
Transaction Price	115.1	79.9
Buyer Earnings	43.9	39.7
Seller Profits	47.8	12.2

Statistical testing for differences between the CM and the AM is problematic. The reason is that observations from the two markets are not strictly independent. For example, prices in the AM might be affected by the experience of subjects who were exogenously allocated to the CM but failed to uphold the relationship. As a consequence, selection problems prevail. Thus, we test for differences between the CM and the AM by considering the strictly independent observations of the first period in the first phase. We test transaction prices of those who were exogenously allocated to the CM (average 103.8) against those who were exogenously allocated to the AM (average 85.7). A Mann-Whitney test reveals that transaction prices are indeed higher in the CM ($p = 0.010$, one-tailed test). Comparing quality choices analogously across the CM (0.44) and the AM (0.22) yields a weakly significant difference ($p < 0.100$, Chi-square test, $\chi^2 = 3.24$, $df = 1$). Seller profits are significantly higher ($p = 0.000$, Mann-Whitney test) in the first period in the CM (45.9) than in the AM (9.2). Buyer earnings are also higher in the CM (40.9) than in the AM (32.3) in period 1 ($p = 0.000$, Mann-Whitney test).

We conclude that while the AM did not fully converge to the lemons outcome, the CM produced much better market outcomes. Average quality on the CM is much higher, and trading on the CM is mutually profitable for buyers and sellers. Hence, long-term relationships in the CM at least partially resolved the lemons problem, as suggested by hypothesis H1.

A) Pricing in customer markets

We now discuss how trust in customer markets affects prices, and, in particular, how price responses differ across treatments. Subsection B) considers quality, C) profits, and D) analyzes the determinants of customer relationships.

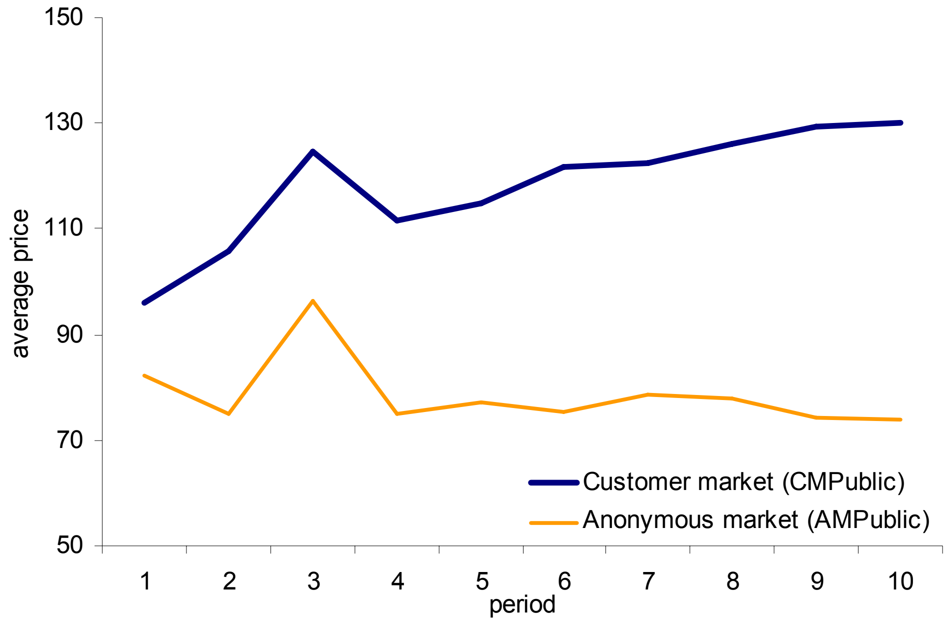


Figure 1: Average transaction prices with public information (PUBLIC)

Figure 1 shows average transaction prices per period in the two markets with public information (PUBLIC). To facilitate the exposition, we aggregate prices over all 5 phases (see Appendix 5.C for average prices in each phase). Several facts are evident from inspection of figure 1. First, prices are higher in CMPublic than in AMPublic. Second, prices do respond to the cost shock in both AMPublic and in CMPublic. Third, average prices are relatively stable in AMPublic, but tend to increase in CMPublic. Increasing prices in CMPublic result from a selection bias. The buyers with the lowest "tolerance" for high prices (i.e., those with the lowest degree of trust that the seller will provide high quality) tend to drop out of the CM in early periods. In addition, higher quality is strongly associated with higher prices, and providing low quality tends to disrupt the customer relationship (see section D). As a consequence, customer relationships exhibiting low prices and low quality tend to be disrupted first.

Figure 2 shows average transaction prices in the two markets when information is private (PRIVATE). As with public information, average prices are higher in CMPrivate than in AMPrivate, and prices have a positive trend in CMPrivate that is absent in AMPrivate. We again observe a pronounced reaction to the cost shock in period 3 in AMPrivate, but average prices do not seem to respond to the cost shock in CMPrivate. However, the price responses are difficult to assess from mere inspection of the figures because price responses are obscured by the upward trend in CM that results from the selection bias.

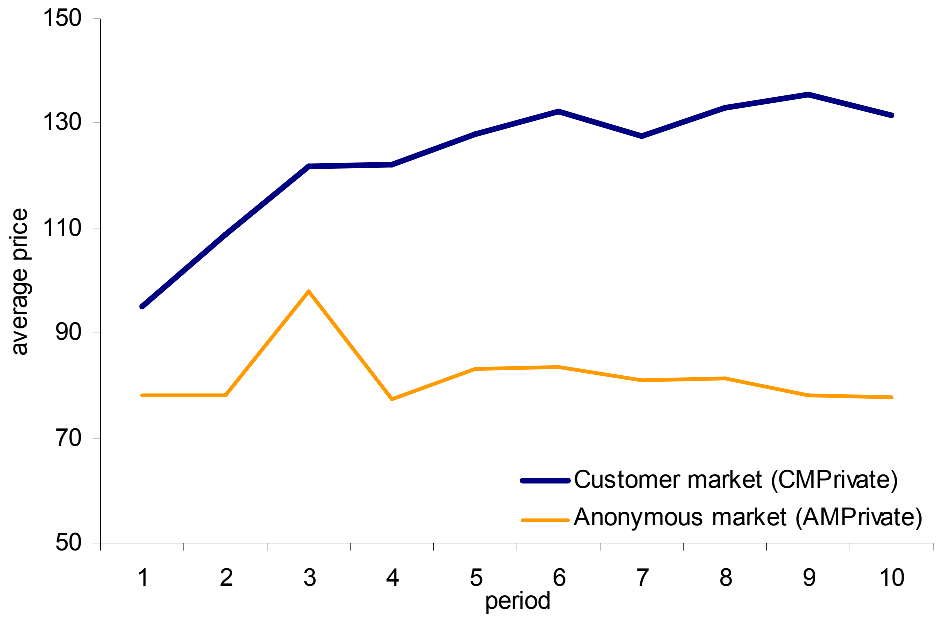


Figure 2: Average transaction prices with private information (PRIVATE)

The simplest way to correct for this selection bias is to consider price changes. We calculate how individual sellers' transaction prices responded to the cost shock (i.e., changed from period 2 to 3). Note that we correct for the bias by considering only transaction prices of those sellers in period 2 who also transacted in period 3.

Table 5.4: Price adjustment in response to cost shock

	Customer market			Anonymous market		
	(1)	(2)	(3)	(4)	(5)	(6)
	CMPrivate	CMPublic	Effect of information on CM	AMPrivate	AMPublic	Effect of information on AM
all phases	4.15	16.40	0.000	20.72	19.61	0.954
phase 1	4.40	23.65	0.013	23.03	29.21	0.507
phase 2	-0.79	12.39	0.028	16.75	31.31	0.225
phase 3	0.71	13.19	0.043	23.78	13.94	0.149
phase 4	10.63	14.56	0.032	17.00	17.56	0.773
phase 5	4.92	17.87	0.002	14.54	14.87	0.954

Notes: Columns (1), (2), (4) and (5) show the average change of transaction prices from period 2 to period 3. Columns (3) and (6) test for the effect of information on price adjustment. They show p-values of Mann-Whitney tests (one-tailed for CM, two-tailed for AM). All CM values are changes of transaction prices for a given customer relationship.

To evaluate the "cost justification" hypothesis H3, we compare price responses in the customer markets across information conditions. At the individual level, we find that 79.6 percent of sellers increase their prices in CMPublic, while only 40.3 percent of sellers do in CMPrivate. When aggregating over all 5 phases, we find that the average price response to the cost shock is almost 4 times more vigorous with public information (16.4, see column 2 of table 5.4) than with private information (4.2) in the CM. Mann-Whitney tests show that prices respond significantly more in CMPublic than in CMPrivate. This is true when aggregating over all 5 phases ($p = 0.000$), and it also holds at the 5 percent level for each individual phase (see column 3). Note that in contrast to the customer market, information about the cost shock has no effect on the anonymous market (see column 6).

If we analyze the responses of posted prices (rather than accepted prices as above) across information conditions in the customer market, we find that price offers on average increase by 7.5 in CMPrivate and by 17.1 in CMPublic. This difference across information conditions is highly significant (p = 0.000, Mann-Whitney), suggesting that sellers anticipated being turned down by customers if they increase prices too much without justification (i.e., in CMPrivate).

In all, we conclude that customer markets cause pronounced price rigidity if price increases cannot be justified by cost increases, but that price rigidity is much less pronounced if the price increase is justified by a cost increase, as suggested by hypothesis H3.

B) Quality

Figure 3 shows the percentage of transactions in which high quality is provided ("average quality") for both markets and information conditions. Average quality is higher in the CM than in the AM from period 1 on. While quality is relatively stable at low values in the AM, quality further increases in the CM to its peak in period 9, and then sharply falls in period 10. The increase in quality until period 9 reflects that higher quality tends to be associated with higher prices in the CM. Again, this increasing trend can be explained by a selection bias. Since providing low quality tends to disrupt the customer relationship, customer relationships with high quality tend to be upheld at a higher rate (see section D for details).

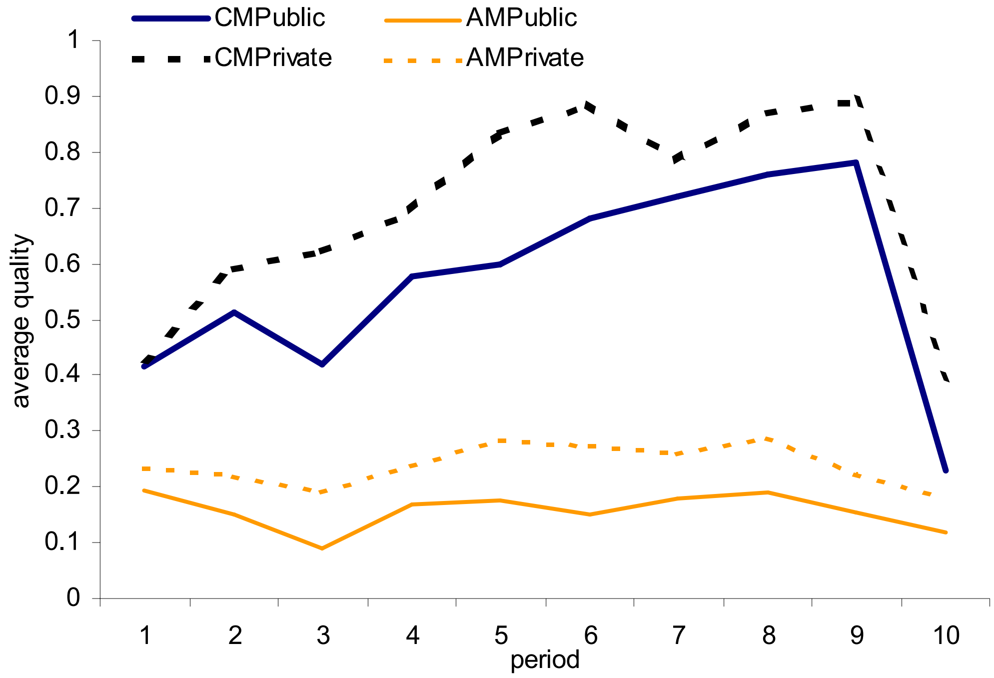


Figure 3: Average quality

The sharp drop in period 10 is an “endgame” effect that is frequently observed in finite trust (and public goods) games (e.g., Selten and Stoecker 1986). This sharp drop suggests that sellers provide high quality in the CM for strategic reasons. Sellers seem to have understood that providing low quality tends to disrupt the customer relationship and that upholding the relationship is profitable (see table 5.3). However, providing high quality ceases to be profitable for sellers in the last period since it cannot induce customer loyalty (i.e., induce a buyer to uphold the relationship by definition). It is remarkable that finite long-term relationships are apt to induce systematically higher quality in the CM than in the AM up to period 9. Remember that all participants knew that the customer relationship can at most last for 10 periods. Accordingly, depending on players' rationality and experience, one would expect quality to fall at some point before the end of the game (see Ch. 5 in Camerer 2003). However, the results of Engle-Warnick and Slonim suggest that players behave relatively similarly in a simple finite trust game (5 period) and an equivalent infinite trust game. The authors find that a lot of experience is needed (more than 20 repetitions) to yield clear behavioural differences between the finite and the infinite game. We only repeat the game 5 times (phases) and find no significant differences in buyer trust (breakup) or seller trustworthiness (quality) over phases.

Figure 3 seems to suggest that the cost shock depresses quality more with public than with private information. However, the decline in quality is not significantly different in the AM across information conditions (Mann-Whitney test, $p = 0.323$). If we control for the selection bias in the CM analogously as for prices (i.e., by considering how quality within a customer relationship changes from period 2 to 3), we find the following. Average quality falls by 12.5 percentage points (from 75% to 62.5%) in CMPrivate and by 17.4 percentage points (from 59.1% to 41.8%) in CMPublic. However, these changes are not significantly different ($p > 0.3$, Chi-square test, $\chi^2 = 1.84$, $df = 2$). We conclude that the information about the cost shock had no significant effect on quality.

C) Profits and Earnings

Upholding the customer relationship is profitable for both buyers and sellers on average (see table 5.3). However, the gains from trusting on the CM are unequally distributed. Sellers benefit much more from upholding the relationship than buyers. In particular, sellers earn almost four times as much in the CM than in the AM (47.8 vs. 12.2), but buyers on average only earn 10.2 percent more in CM than in AM (43.9 vs. 39.8). This implies that buyers can sanction sellers by terminating the relationship and that sellers have a strong incentive not to antagonize customers by increasing prices or by providing low quality. The average “sanction” of terminating the relationship for a seller is 35.6 per period (= 47.8 - 12.2) and results from two sources. First, trading on the AM is less profitable for sellers because prices are lower on the AM than on the CM. Second, and more importantly, there is a high risk of not being able to trade on the AM, which means that seller earnings are zero.

An interesting question is who bears the burden of the cost shock in the customer market. Of course, the cost shock depresses overall earnings since the gains from trade fall by a third, ceteris paribus (see table 5.2). Since the relationship is profitable for sellers, they should be willing to bear the burden of the cost shock if doing so contributes to uphold the relationship. According to hypothesis H3, the extent to which it does depends on the information available to customers. Hence, we would expect that sellers bear more of the burden when information about the shock is private than when it is public. This is indeed the case. With private information, the sellers bear the brunt of the cost increase because they are reluctant to increase prices, and they reduce quality only a little. In particular, seller losses in period 3 (as a percentage of period 2 profits) are more than twice as large as buyer losses (46.3 vs. 18.9 percent) on the CM when information is private. In contrast, with public information, buyers on the CM suffer losses that are about five times as large as seller losses. In particular, with public information buyers lose 62.3 percent of their earnings, while sellers only lose 12.6 percent of their profits (again relative to period 2). Hence, sellers are able to shift the burden of the cost shock to customers with public information but not with private information.

Interestingly, there are no pronounced long-term effects of the cost shock on seller profits or customer earnings in the CM. If we test seller profits across treatments by period (Mann-Whitney tests), we only find a significant difference in period 3 ($p = 0.002$); all other periods are insignificant ($p > 0.487$). Buyer earnings are only different in periods 3 ($p = 0.010$) and 6 ($p = 0.011$) at the 5 percent level.

D) Customer relationships

In total, 360 trading pairs start on the CM (= 12 runs x 3 pairs x 2 treatments x 5 phases). These pairs can uphold their relationship for at most 10 periods. The average breakup rate is similar in both treatments (16.4% in CMPrivate, 17.2% in CMPublic). The difference between the two treatments is at a maximum in the shock period (34.5% in CMPrivate, 20.3% in CMPublic). In the last period, the breakup rate is remarkably high (24.3% in CMPrivate, 36.6% in CMPublic). As a consequence, still 47.2 percent of all possible customer relationships are upheld in period 3, while this percentage drops to 15.0 in period 10.

To analyze the determinants of customer relationships, we run a maximum likelihood probit regression (see table 5.5) so that the estimates for dummy variables can be interpreted as the change in probability of a breakup. In particular, DF/dx is the marginal effect for a discrete change of the dummy variable from 0 to 1.

Table 5.5: Probit regression

	dependent Variable: probability of breakup of customer relationship	d*F*/d*x*	*P* > l *z*l
(1)	Treatment (dummy, 1 = PUBLIC)	.0079	0.731
(2)	Phase	.0093	0.218
(3)	Period	-.0121*	0.024
(4)	Last period (dummy)	.3477***	0.000
(5)	Shockperiod (dummy)	.1556***	0.000
(6)	Shockperiod * public info (dummy)	-.1177**	0.001
(7)	This period's posted price	.0030***	0.000
(8)	Price increase from last period (dummy)	.0702*	0.020
(9)	Extent of price change from last period	-.0000	0.998
(10)	Large (> 30 units) price increase (dummy)	-.1104*	0.010
(11)	Large (> 30 units) price cut (dummy)	.0624	0.403
(12)	Last period bad quality (dummy)	.3986***	0.000

Notes: *n* = 1346, Prob > Chi^2 = 0.0000, Log likelihood = -540.32, Pseudo R^2 = 0.1675;
*** = significant at the 0.1 percent level, ** = 1 percent, * = 5 percent.

The insignificant estimate for (1) Treatment shows that the probability to terminate the relationship is not different with public or with private information, which is what we expect in all non-shock periods. Phase (2) is insignificant, and Period (3) is weakly negative. This suggests that relationships tend to break-off early in a phase, but the tendency to uphold relationships is not different across phases. Both of these findings are surprising. Given our discussion of search in opaque markets in section 3.3, we would have expected customer relationships to be more prevalent in late phases when participants have gained experience with the market institutions. The (weakly significant) negative estimate for (3) is surprising since firms should be more hesitant to provide bad quality in early periods than in late periods (where the cost of terminating the relationship is lower). However, this seems only to hold for the very last period. The estimate for Last period (4) indicates that the probability that a relationship is terminated is 34.77 points higher in the last period than in other periods, suggesting that some customers anticipated that sellers are not going to be trustworthy and will deliver low quality in period 10 (as most of them do, see figure 3). Hence, many customers seem to understand that sellers can be more reliably trusted if they have an incentive to be trustworthy, but when this incentive is altogether absent in the last period, a considerable portion of customers ceases to trust.

Estimates (5) and (6) capture the effect of the shock period in the two treatments. The estimate for Shockperiod (5) indicates that the probability of separation is 15.56 points higher in period 3, and (6) suggests that the probability of breakup is 11.77 points lower with public information. Even though prices increase much less in CMPrivate than in CMPublic in response to the cost shock (see table 5.4), the breakup rate is higher in CMPrivate, indicating that sellers only partly anticipated that customers are antagonized by unjustified price increases.

The estimates (7) to (11) jointly capture the effect of pricing on the termination of customer relationships. Estimate (7) shows that higher levels of prices contribute to disruption, and (8) shows that if prices are increased, the probability of disruption is 7.0 points higher. Customers react to the extent of price increases in a rather surprising manner. Larger price increases (9) apparently do not additionally contribute to termination. Surprisingly, very large (by more than 30 units) price increases (10) reduce the probability of breakup. Apparently, very large price increases seem to be interpreted by customers as a signal that the seller is going provide high quality. In contrast, very large price cuts

(11) do not reduce the probability of breakup. Finally, the quality provided by the seller in the previous period has the quantitatively strongest effect on termination. The highly significant estimate (12) shows that providing low quality increases the probability of termination of the customer relationship by 39.9 points in the subsequent period.

5.5 Concluding remarks

This paper demonstrates how trust can shape market outcomes. In particular, we provide the first experimental investigation of how long-term customer relationships cause price rigidity in markets for experience goods. While we believe that our study provides interesting insights, we also feel the need to explore price rigidity in customer markets further. Fortunately, our experimental design lends itself to investigate some additional aspects of price rigidity in customer markets. For example, the design can easily be adapted to analyze asymmetries in price rigidity. To do so, a negative cost shock could be implemented in period 3, and the price adjustment could be compared with the price response to the positive shock. It would also be interesting to investigate further the implications of strategic trust for price rigidity. We speculate that many sellers do not respond to the cost shock by increasing prices and are willing to bear the corresponding costs in period 3 because they hope to reap the gains from trading with loyal customers in the remaining 7 periods of the game. To test for the importance of this trade-off, one could implement a cost shock in, say, period 7 and compare this price response with the one in period 3. If strategic trust is important, price rigidity should be less pronounced after the shock in period 7 because the gains from trading with loyal customers can now only be reaped in fewer periods.

Four remarks concerning the interpretation of our results seem appropriate.

First, we show that finite long-term relationships cause price rigidity in customer markets. In particular, it was known to all participants in our experiment that the customer relationship can last for 10 periods at most. In contrast, standard theory must assume infinite horizons to demonstrate possible effects of customer relationships with perfectly rational and self-interested agents (e.g., Klein and Leffler 1981). Fairness motives are likely to provide an explanation for this discrepancy (see Brown et al. 2004 for a theoretical discussion).

Second, while the intuition that price increases may antagonize customers motivated our experiment, our findings also draw attention to the importance of quality reductions. In particular, our results show that providing low quality is even more disruptive to customer relationships than price increases. Sellers anticipating this reaction have an incentive to leave prices and qualities unchanged after temporary cost shocks. That customer relationships may not only induce price rigidity but also “quality rigidity” is illustrated in the extreme example of the “Nickel Coke.” Levy and Young (2002) argue that because of an implicit contract with its customers, Coca-Cola not only held the nominal price of a bottle constant over a period of 70 years (from 1886 to 1959), but also held quality constant over this period despite important cost and demand shocks. However, firms may be forced to choose between price and quality adjustments at some point. Whether price or quality rigidity is likely to be more pronounced may importantly depend on the relative “visibility” of price and quality changes. For example, the journal Consumer Reports (1994, 59(10), p. 618) reports that “the typical can of dog food once held 14 ounces. Now it holds 13.2 ounces. The can itself hasn’t changed - and neither has the price.” Rotemberg (2002) provides further examples of how almost imperceptible quality reductions can be used to mask price increases.

Third, our study demonstrated that long-term relationships cause price rigidity in a setting in which reputations for providing high quality could only prevail within the relationship. That is, seller reputations were relationship-specific in our experiment, and reputations could not be transmitted to other potential buyers. However, if sellers can develop general reputations through markets (e.g., by advertising) or social networks (see Huck and Tyran 2003), incentives for upholding personal long-term relationships can be very different than in the absence of these transmission channels.

Fourth, we find that long-term relationships are profitable for sellers in our experiment. However, while long-term relationships clearly have benefits, they may also have drawbacks in more complex environments. For example, Grayson and Ambler (1999) suggest that long-term customers may hold higher expectations on service quality than one-time customers. Since higher expectations increase the potential for dissatisfaction with a given service quality, firms may have to incur higher costs to avert termination of the customer relationship.

5.6 References

Akerlof, G.A. (1970): The market for “lemons”: quality uncertainty and the market mechanism, Quarterly Journal of Economics, 84, 488-500.

Apel, M., Friberg, R., Hallsten, K, (2001): Micro foundations of macroeconomic price adjustment: Survey evidence from Swedish firms, Working paper Sveriges Riksbank no. 128.

Arrow, K.J. (1973): Social responsibility and economic efficiency, Public Policy , 21, 303-17.

Berg, J., Dickhaut, J., McCabe, K.(1995): Trust, reciprocity, and social history. Games and Economic Behavior 10, 122-42.

Blinder, A.S., Canetti, E.D., Lebow, D.E., Rudd, J.B. (1998):. Asking about Prices. A New Approach to Understanding Price Stickiness, Russell Sage Foundation, New York.

Brown, M., Falk, A., Fehr, E. (2004): Rational Contracts and the nature of market interaction. Econometrica, 72 (3), 747-780.

Camerer, C. (2003). Behavioral Game Theory. Experiments in Strategic Interaction. Princeton University Press, Princeton.

Carlton, D.W. (1989): The theory and facts about how markets clear: Is industrial organization valuable for understanding macroeconomics?, in: R. Schmalensee and R.D. Willig (Eds.), Handbook of Industrial Organization, Vol. I. North Holland, Amsterdam, 909-46.

Cason, T.N., Friedman, D., (2002): A laboratory study of customer markets. Advances in Economic Analysis and Policy 2, Article 1.

Cason, T.N., Gangadharan, L. (2002): Environmental labeling and incomplete consumer information in laboratory markets, Journal of Environmental Economics and Management,43, 113-34.

Engle-Warnick, J., Slonim, R.L. (2004): The evolution of strategies in a repeated trust game, Journal of Economic Behavior & Organization, Vol. 55, 575-593.

Fischbacher, U. (1999): z-tree. Zürich Toolbox for Readymade Economic Experiments. University of Zürich, Working Paper No. 21.

Franciosi, R., Kujal, P., Michelitsch, R., Smith, V.L., Deng, G. (1995): Fairness: effect on temporary and equilibrium prices in posted-offer markets, Economic Journal, 105, 938-50.

Grayson, K., Ambler, T. (1999): The dark side of long-term relationships in marketing services, Journal of Marketing Research, 36, 132-41.

Huck, S., Tyran, J.-R. (2003): Social ties and reciprocity in markets for experience goods. Mimeo.

James, H.S. (2002): On the reliability of trusting, Rationality and Society, 14, 231-58.

Kachelmeier, S.J., Limberg, S.T., Schadewald, M.S. (1991a): A laboratory market examination of the consumer price response to information about producers’ cost and profits. Accounting Review, 66, 694-717.

Kachelmeier, S.J., Limberg, S.T., Schadewald, M.S. (1991b): Fairness in markets: a laboratory investigation, Journal of Economic Psychology, 12, 447-64.

Kahneman, D., Knetsch, J.L., Thaler, R., 1986. Fairness as a constraint on profit seeking: entitlements in the market. American Economic Review 76, 728-41.

Klein, B., Leffler, K.B. (1981): The role of market forces in assuring contractual performance, Journal of Political Economy, 89, 615-41.

Kollock, P. (1994): The emergence of exchange structures: an experimental study of uncertainty, commitment, and trust, American Journal of Sociology, 100, 313-45.

Levy, D., Young, A.T. (2002): “The real thing”: nominal price rigidity of the nickel Coke, 1886-1959, Working paper Emory University.

Okun, A.M. (1981): Prices and Quantities: A Macroeconomic Analysis, The Brookings Institution, Washington, D.C.

Rotemberg, J. (2002): Customer anger at price increases, time variation in the frequency of price changes and monetary policy, NBER working paper 9320.

Selten, R., Stoecker, R. (1986): End behavior in sequences of finite prisoner's dilemma supergames, Journal of Economic Behavior and Organization, 7, 47-70.

Wynne, M.A. (1995): Sticky prices: What is the evidence? Economic Review, Federal Reserve Bank of Dallas, 1st quarter, 1-12.

Appendix

Appendix 1.A: Instructions

The instructions were originally written in German. We document the instructions of the treatment with a leader in period 1-10 and without a leader in period 11-20. The instructions in the other treatments were adapted accordingly. The original instructions are available from the author upon request.

General information on the experiment

You are taking part in an experiment on decision-making. If you read the following instructions carefully, you can, depending on your decisions, earn a considerable amount of money. It is therefore very important that you understand the following instructions.

Earnings

In the experiment you earn points. The points you have earned will be converted in Euros at the rate

1 point = 0.40 €.

At the end of the experiment your total income in Euros will be paid to you in cash.

Group membership and anonymity

During the experiment you are member of a group of four participants, i.e. in your group will be three more group members. The composition of the group is the same during the whole experiment. Thus you form a group with the same participants throughout the experiment. It will be randomly determined with whom of the other participants you form a group.

All participants decide anonymously, i.e. no participant will ever learn the identity of the other members of his group. To ensure anonymity it is imperative that all participants observe the following rule:

During the experiment all communication is prohibited, i.e. you are not allowed to speak or otherwise express yourself.

Should you have any questions please ask the experimenter.

The experimental procedure

The experiment consists of several periods.

The decision situation in period 1-10

The procedure is the same for each of these periods 1-10.

Every participant receives 20 tokens at the beginning of each period. Your task is to decide how you use your endowment. You have to decide how many of the 20 tokens you want to contribute to a project and how many of them to keep for yourself.

The calculation of your income

Your income consists of two parts:

1. The points you keep for yourself = Your endowment – your contribution to the project
2. The income from the project. The income from the project is calculated as follows:
   Your income from the project = 0.4 * (total contribution of all four group members to the project)

Your income in points is therefore:

(20 – your contribution to the project) + 0.4*(total contributions to the project)

The income of each group member from the project is calculated in the same way, this means that each group member receives the same income from the project.

Examples:

Suppose the sum of the contributions of all group members is 60 tokens. In this case each member of the group receives an income from the project of: 0.4*60 = 24 points.

If the total contribution to the project is 9 tokens, then each member of the group receives an income of 0.4*9 = 3.6 points from the project.

For each token, which you keep for yourself you earn an income of 1 point. Supposing you contributed this token to the project instead, then the total contribution to the project would rise by one point. Your income from the project would rise by 0.4*1=0.4 points. However the income of the other group members would also rise by 0.4 points each, so that the total income of the group from the project would rise by 1.6 points. Your contribution to the project therefore also raises the income of the other group members. On the other hand you earn an income for each point contributed by the other members to the project. For each point contributed by any member you earn 0.4*1=0.4 points.

The decision

One member in each group decides first on his/her contribution. When this member has made his/her decision, the other members of the group are informed of the amount of tokens he/she has contributed to the project. Thereafter the other group members decide.

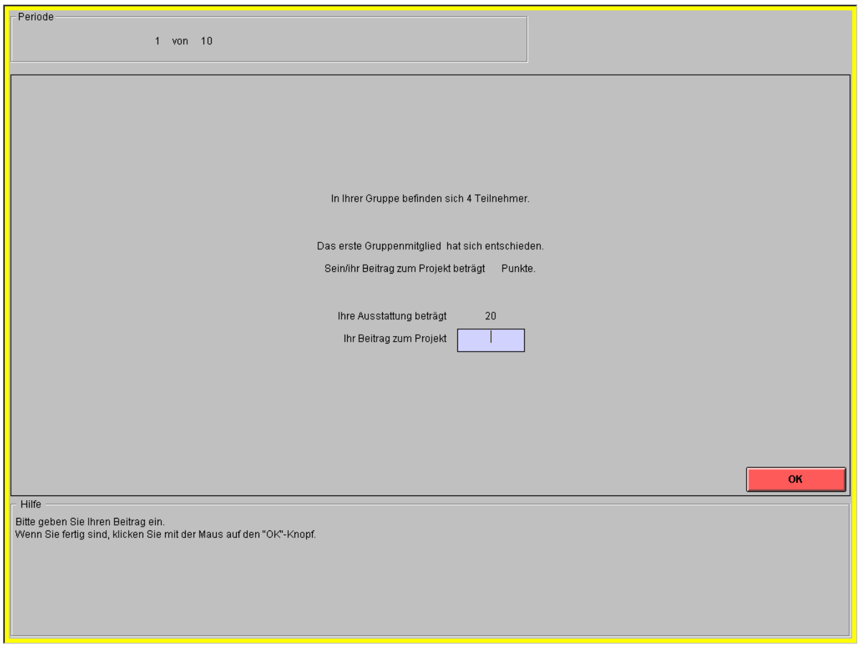
The person, who decides first in a group, remains the same for all 10 periods.

The person who decides first in a group will see the following input-screen:

The period number appears in the top left of the screen. In the middle of the screen you will find the following information: Your group consists of 4 members. You are the person who decides first. Your contribution will be displayed to the other members of your group. Then they will decide on their contribution. Your endowment is 20 tokens.

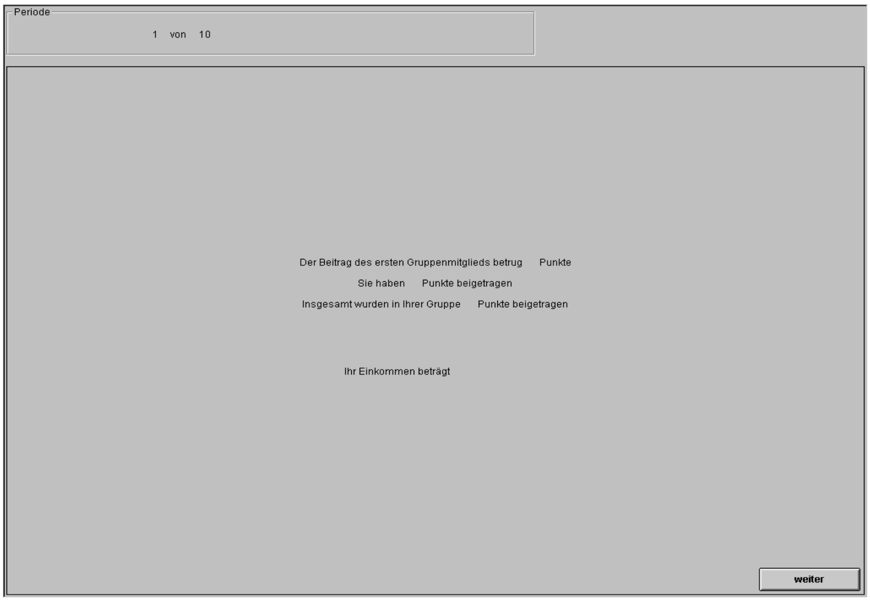
If this screen appears on your monitor you are the member of your group who decides first. You take your decision by typing a number between 0 and 20 in the input field. This field can be reached by clicking it with the mouse. As soon as you have decided how many points to contribute to the project, you have also decided how many points you keep for your self: This is (20 – your contribution) tokens. After entering your contribution you must press the O.K. button (either with the mouse, or by pressing the Enter-key). Once you have done this, your decision can no longer be revised.

As soon as the decision is made the following input-screen will appear for the other three members of the group:



The contribution of the group member, who has made his/her contribution decision first is displayed. You take your decision by typing a number between 0 and 20 in the input field and confirm by pressing the O.K.-button.

After all three have taken their decision the following income-screen is displayed for all members of the group:



Your own contribution and the contribution of the member who has made his/her decision first in your group are displayed. Furthermore, you are informed of the total contribution of all group members (including your contribution) and your income in the current period.

Do you have any questions?

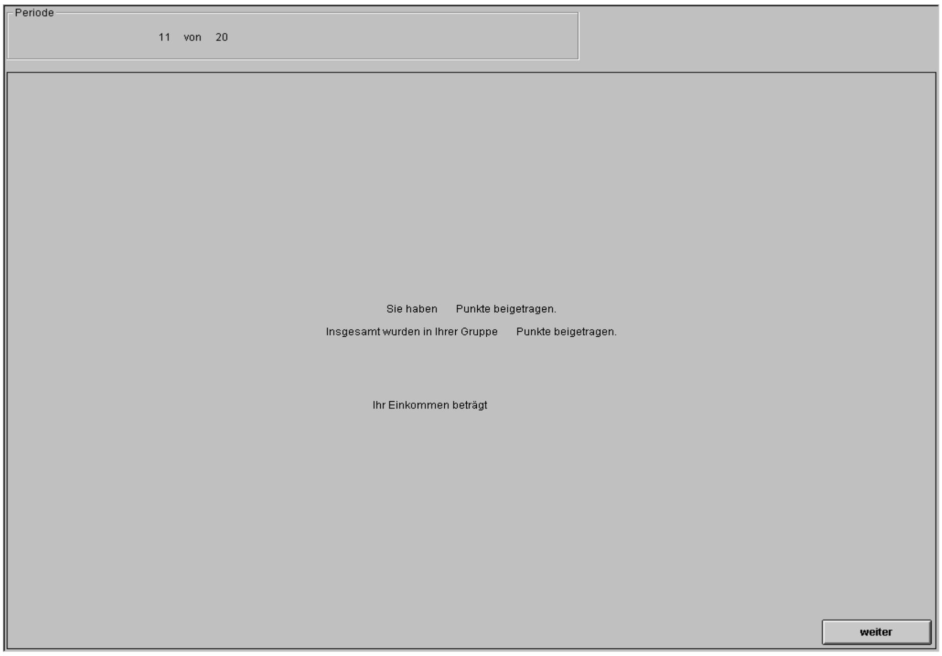
The experimental procedure in period 11-20

The experiment will continue for 10 more periods (period 11-20). The procedure is the same for each of these periods.

The decision situation in period 11-20

In the following periods the composition of your group remains unchanged, i.e. your group consists of the same persons as in the previous periods. The decision situation is basically the same as before. All values like the endowment of 20 token per period and the calculation of your income remain unchanged. The only difference in the following periods is that there will be no one in your group who decides first anymore. This means that every participant takes his decision without knowing the contribution of any other member of the group. The following input screen will appear for all participants:

On the income-screen your contribution, the total contribution of all members of your group (including your contribution) and your income displayed.



Appendix 1.B: Tables 1.a and 1.b

Table 1.a Spearman rank correlation coefficients of a leader's contribution and his average followers' contribution for each independent group with $\alpha = 0.4$

Leader in period 1-10		Leader in period 11-20			
Group No.	Random	Group No.	Free rider	Group No.	Pro-social
1	.877*	13	.848**	22	.134
2	.012	14	.839**	23	.984**
3	.745*	15	.706*	24	.692*
4	.834**	16	.752*	25	.471
5	.459	17	.742*	26	.887**
6	.757*	18	.873**	27	.806**
7	.218	19	.794**	28	.199
8	.696*	20	.904**	29	.486
9	.194	21	.228	30	.908**
10	.617				
11	.500				
12	.425				

"**" significant at $\alpha = 0.01$ (two-sided) "*" significant at $\alpha = 0.05$ (two-sided)

Table 1.b. Spearman rank correlation coefficients of a leader's contribution and his average followers' contribution for each independent group with α =.6

Leader in period 1-10		**Leader in period 11-20**			
Group No.	**Random**	**Group No.**	**Free rider**	**Group No.**	**Pro-social**
1	.56	10	.395	15	.412
2	.711*	11	.982**	16	.346
3	.799*	12	.859**	17	.781*
4	.304	13	.826**	18	.390
5	.521	14	.854**	19	.567
6	0				
7	.787**				
8	-.020				
9	.475				

"**" significant at α = 0.01 (two-sided) "*" significant at α = 0.05 (two-sided)

Appendix 2.A: Instructions

The original instructions were in German. We document the instructions of the public information treatment without punishment opportunity in the first sequence and with the opportunity to punish in the second sequence. The instructions in the other treatments were adapted accordingly. The original instructions are available from the author upon request.

General information on the experiment

You are taking part in an experiment on decision making. At the end of the experiment you will be paid according to the decisions you make. Please read these instructions carefully.

Earnings

During the experiment you earn points. The total income in points your earn during the experiment will be converted in Euros at the rate

1 point = 0,025 €.

At the end of the experiment your total income in Euros will be paid to you in cash.

At the beginning of the experiment every participant receives 100 points.

Group membership and anonymity

During the experiment you are member of a group of four participants, i.e. in your group will be three more group members. The composition of the group is the same during the whole experiment. Thus you form a group with the same participants throughout the experiment. It will be randomly determined with whom of the other participants you form a group.

All participants decide anonymously, i.e. no participant will ever learn the identity of the other members of his group. Therefore it is imperative that all participants observe the following rule:

During the experiment all communication is prohibited, i.e. you are not allowed to speak or express otherwise yourself.

Should you have any questions please ask the experimenter.

The experimental procedure

The experiment consists of ten periods. The procedure of each of these periods is the same. Your task is to decide, how many points you contribute to your groups project.

The decision situation in a period

Your endowment

At the beginning of each period each participant has an endowment of points. The amount of points can be between 0 and 40. How many points each participant has will be randomly determined for every single participant. Every number between 0 and 40 is drawn with the same probability. Every member in a group can therefore have a different endowment between 0 and 40 points. The endowment is newly determined at the beginning of each period for every single participant.

The decision

You have to decide how many points of your endowment you want to keep for yourself and how many you want to contribute to the project of your group.

The calculation of your income

Your income consists of two parts:

1. The points you keep for yourself = Your endowment – your contribution to the project
2. The income from the project. The income from the project is calculated as follows:
   Your income from the project = 0.4 * (total contribution of all four group members to the project)

The income of each group member from the project is calculated in the same way, this means that each group member receives the same income from the project.

Your income in a period is therefore:

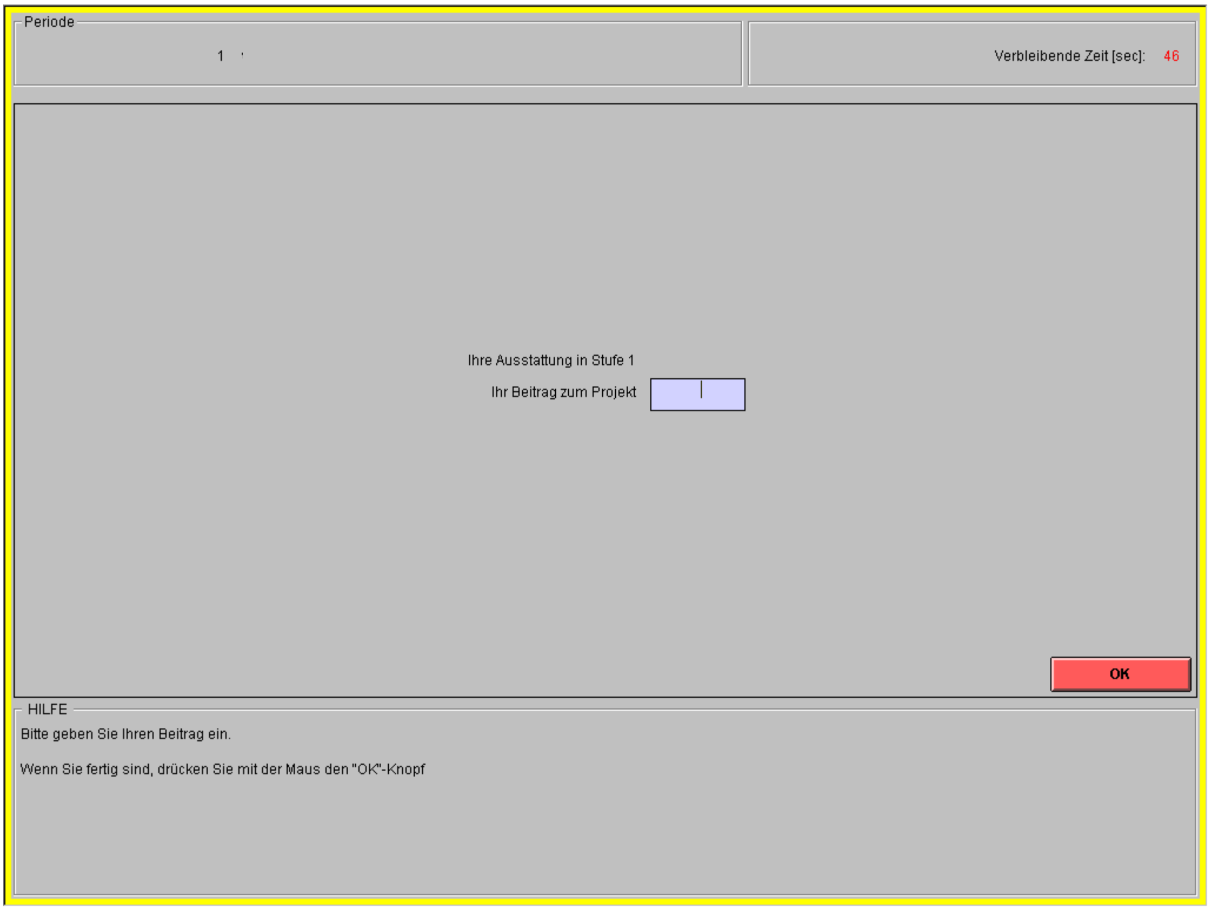
(Your endowment – your contribution to the project) + 0.4*(total contribution of all four group members to the project).

Examples: Suppose the sum of the contributions of all group members is 60 tokens. In this case each member of the group receives an income from the project of: 0.4*60 = 24 points. If the total contribution to the project is 9 tokens, then each member of the group receives an income of 0.4*9 = 3.6 points from the project.

For each token, which you keep for yourself you earn an income of 1 point. Supposing you contributed this token to the project instead, then the total contribution to the project would rise by one point. Your income from the project would rise by 0.4*1=0.4 points. However the income of the other group members would also rise by 0.4 points each, so that the total income of the group from the project would rise by 1.6 points. Your contribution to the project therefore also raises the income of the other group members. On the other hand you earn an income for each point contributed by the other members to the project. For each point contributed by any member you earn 0.4*1=0.4 points.

How to take your decision

At the beginning of each period the following input screen appears:



On this screen the amount of your endowment in the current period is displayed. You take the decision on your contribution to the project by typing the corresponding number in the input field and pressing the "O.K.-button". You can at maximum contribute your whole endowment. If you do not want to contribute, please enter 0.

As soon as all members of your group have made their decision the following income screen appears:

There you see your own contribution to the project and the total contribution of all group members (including your contribution). Furthermore it is displayed how many points you kept for yourself, the income from the project and your total income in this period, which is the sum of the points you kept and the income from the project.

Another screen will appear on which you can see for each of the four members of your group separately the endowment, how many points were contributed to the project and how many percent of the endowment this is. Your own contribution is always displayed in the first column in blue, the contributions of the other three group members in column two, three and four in random order. Please note: the order of the other group members will be new in each period.

Do you have any questions?

Information for the following 10 periods

The experiment will continue for 10 more periods. Like before the procedure is the same in each period.

In the following period the composition of your groups remains the same, i.e. you form a group with the same participants as before.

The decision situation in the following 10 periods

The decision task now has two stages.

First stage

The decision in the first stage is the same as in the periods before: You decide again, how many points of your endowment you contribute to the project. The procedure and the calculation of your income are the same as before and the endowments will be randomly determined for each participant in each period.

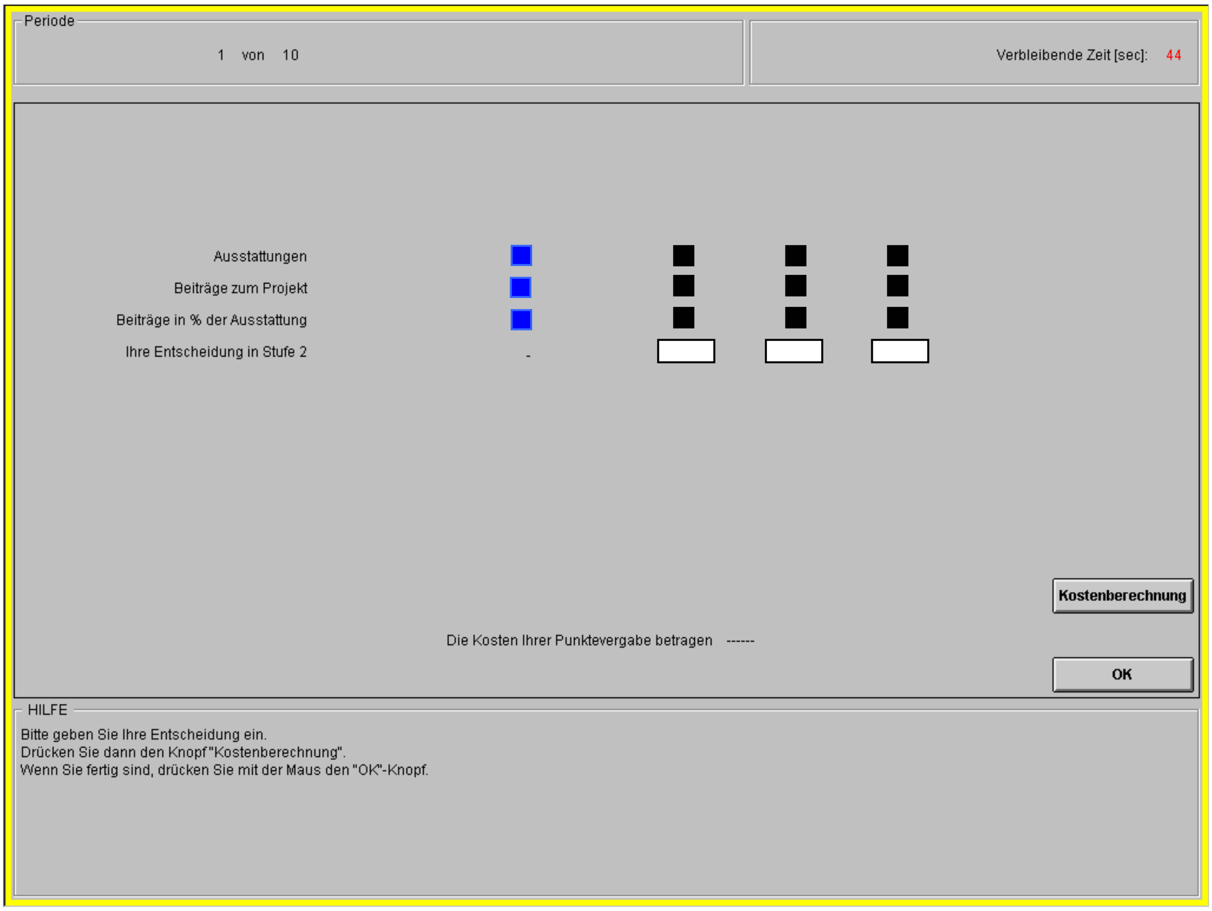
When all members of the group have taken their decision the income screen again appears, which will inform you on your income (in the first stage). Thereafter the second stage follows.

Second stage

In the second stage you decide whether and to what extend you reduce the income of the other members of your group.

The decision in the second stage

You will see again the screen (see bellow) on which the endowments, the contribution to the project and the contribution in percent of the endowment are displayed for all four group members. Your own contribution again is displayed in the first row in blue, the contributions of the other three members of your group in row two, three and four in random order. Please note: This order is renewed in each period.



You now must decide how many points you want to give to each other member of your group.

Each point you give to a specific member of your group reduces this member's income by three points.

If you choose 0 points, the income of this group member is not reduced. You can distribute at maximum 10 points to each of the other group members.

The sum of all points you give to other group members are reduced from your income.

After you have decided how many points you distribute to all group members please press the „O.K."-button.

After all participants have taken their decision in second stage the following income screen appears:

Your income from stage 1 and the points you have distributed is displayed. You are also informed of how many negative you received from other group members and how much your income is reduced by these points.

Your final income is calculated as:

Your income from the first stage

minus costs for distributing points

Minus reduction through received points (=3 * received points)

= Your Income in this period

Please note: Your income from the first stage is at maximum reduced to zero through the points you received from other group members. However you always bear the costs for the points you have distributed, such that your income can be negative in a period.

Do you have any questions?

Appendix 2.B: Tables 2.a – 2.d

Table 2.a. Spearman rank correlation coefficients of contribution rates and endowments

Treatment	PRIVATE		PUBLIC	
group no.	without punishment opportunity	with punishment opportunity	without punishment opportunity	with punishment opportunity
1	-.833*	-.476	-.714*	-.220
2	-.810*	-1.000**	-.738*	.060
3	-.881**	-.690	.119	-.548
4	-.095	-.357	-.686	-.195
5	-.667	-881**	-.667	-.038
6	-.333	-.119	-.333	-.405
7	-.156	-.143	-.452	-.611
8	-.857**	-.905**	-.143	-.643
9	-.167	-.905**	-.786*	-.805*
10	-.857**	-.738*	-.548	.143
11	.690	-.786*	.429	-.690
12	-.857**	-.857**	-.714*	-.805*
13	-.595	-.810*	-.810*	-.357
14	-.619	-.048	-.180	.330
15	-.095	.452	.262	.452
16	-.619	-.667	-.881**	-.333
17	-.738*	-.952**	-.714*	-.790*
18	-.333	-.048	.286	-.467
19	-.265	-.190		

"*" significant at α = 0.05 (two sided)

"**" significant at α = 0.01 (two sided)

Table 2.b. Spearman rank correlation coefficients of contribution rates over periods

Treatment	PRIVATE		PUBLIC	
group no.	without punishment opportunity	with punishment opportunity	without punishment opportunity	with punishment opportunity
1	-.236	.770**	-.267	.874**
2	.406	.588	.600	-0.28
3	-.152	.455	.876**	.939**
4	.345	-.345	.073	.813**
5	-.648*	.758*	.333	.500
6	-.745*	.079	-.612	.865**
7	-.176	-.345	.176	.865**
8	.564	.115	.212	.442
9	-.612	.855**	.304	.957**
10	.358	.042	.164	.588
11	-.952**	-.467	-.721*	.937**
12	.139	.709*	.782**	.915**
13	-.661*	-.164	.273	.127
14	-.648*	-.358	.091	-.074
15	.127	-.445	.867*	.738*
16	-.571	-.442	.224	-.091
17	-.091	.636*	.673*	.608
18	-.273	.742*	-.006	-.612
19	-.321	-.394		

"*" significant at α = 0.05 (two sided)

"**" significant at α = 0.01 (two sided)

Table2.c. Spearman rank correlation coefficients of received punishment points in period t and average relative change in contributions from period t to t+1

Treatment **group no.**	**PRIVATE with punishment opportunity**	**PUBLIC with punishment opportunity**
1	-	.649
2	-.200	.321
3	.500	.100
4	-.500	.700
5	-.400	.543
6	-.874**	.238
7	-.857*	.800
8	-1.000**	-.486
9	.200	.900*
10	.800	.109
11	-.486	1.000**
12	.143	-
13	.500	.410
14	-	.781*
15	-.248	.371
16	-.044	.319
17	-.800	.829*
18	-	-.500
19		

“*” significant at p = 0.05 (two sided)

“**” significant at p = 0.01 (two sided)

Table 2.d. Spearman rank correlation coefficients of deviation from groups average relative contribution (in percentage points) and received punishment points

Treatment group no.	PRIVATE with punishment opportunity	PUBLIC with punishment opportunity
1	-1.000	-.893**
2	-.200	-.679
3	.624	-.900*
4	-1.000**	-.700
5	.391	-.829*
6	.333	-.286
7	.179	-1.000**
8	.500	.029
9	.100	-1.000**
10	-.800	-.427
11	.486	-1.000**
12	-.257	-
13	-.500	-.700
14	-	-.810*
15	-.321	-.943**
16	-.122	-929**
17	.400	-.357
18	-	-
19	-	

"*" significant at α = 0.05 (two sided)

"**" significant at α = 0.01 (two sided)

Appendix 3.A: Instructions

The original instructions were written in German. The following instructions are those for the treatment IR8. The instructions for the treatments IR2 and IR4 differ only in the corresponding numbers. The instructions for treatment GR4 contain one different sentence which is indicated below. The original instructions are available from the author upon request.

Experimental Instructions

During the experiment you belong to a group consisting of 8 randomly chosen members.

{*GR4 treatment*: During the experiment you belong to a group consisting of the 4 members with whom you have registered}.

The composition of each group does not change throughout the experiment.

The experiment starts with the first round. Whether there will be a following round depends on the group's result of the previous round. A maximum of 10 rounds are played.

Contribution to the group payment?

In every round, each member of a group decides whether to contribute to a group payment or whether not to contribute.

In each round, the group has to raise a total amount of 1680 talers in order to reach a further round.

Who is able to pay and who pays how much?

For each group member it is randomly determined whether he receives an amount of 420 Taler. For this, in every round the experimenter asks a randomly chosen participant to roll a die. If the number is identical to the number for the respective round, which a participant has written down in the first column before the beginning of the experiment, this participant receives **0 talers**. If the numbers do not match, the participant receives **420 talers**. Thus, the decision whether a participant receives 420 talers takes place at random and independently ofall other participants.

A group member can only contribute to the group payment if he has received 420 talers in this round. A group member who has 0 talers cannot contribute. His payoff is 0 talers.

The amount of 1680 talers that must be raised by the group will be divided equally between those group members who decided to contribute to the group payment and who in addition are able to contribute, i.e. have received 420 talers. The more members of a group are willing and able to contribute, the less is the amount each member has to pay. This amount will be subtracted from the 420 talers received by each paying member.

The contributions and round payoffs of a paying member result as follows:

Number of paying group members	Contribution per paying member	Group raises 1680 talers?	Round payoff for a paying member
0	-	No → no further round	
1	420	No → no further round	**0**
2	420	No → no further round	**0**
3	420	No → no further round	**0**
4	420	Yes → further round	**0**
5	336	Yes → further round	**84**
6	280	Yes → further round	**140**
7	240	Yes → further round	**180**
8	210	Yes → further round	**210**

All members who have received 420 talers and have decided not to contribute to the group payment obtain this 420 talers as the payoff for this round.

Structure of the experiment

The structure of all rounds is identical. Each group member decides whether he wants to contribute to the group payment by ticking „yes“ or „no“ in column 2 of the decision sheet. **The decision is taken only for the current round.** Thus, please tick only the box which corresponds with the current round, but not those for the following rounds.

The experimenters collect all decision sheets and a randomly chosen participant rolls a die. This number is publicly announced and recorded in column 3 by the experimenter. The provisional round payoff results as explained above (column 4): if the numbers of column 1 and 3 match, the group member receives 0 for this round; if the numbers differ he receives a provisional round payoff of 420. The experimenter determines the number of paying group members (column 5) and writes down the final round payment for each member in column 6. Column 7 informs whether the group succeeded in raising the amount of 1680 talers, and the experiment continues with the next round.

For each participant the total payoff is determined (in talers) at the end of the experiment and is converted into DM at an exchange rate of 2.00 DM per 100 talers. Additionally each participant receives an amount of 5.00 DM.

The total amount is paid in cash to each participant at the end of the experiment.

Please note:

All decisions are made anonymously, i.e. at the end of a round each member is only informed about his own decision, his own payoff and the number of paying group members. During the experiment all participants sit in a cubicle. No conclusions can be drawn regarding the identity of the other group members, their decisions or payoffs.

To make sure that anonymity is guaranteed, it is necessary to adhere to the following rules:

During the experiment it is forbidden to speak or to communicate in any way.

All members are asked to stay in their cubicles till the end of the experiment.

Even if in a round a group does not succeed in raising the amount of 1680 talers, which means that no further rounds are taking place for this group, all participants have to stay in their cubicles until the end of the experiment. The experimenters continue to hand out and to collect the decision papers also for this group. Of course, those group members do not have to take decisions anymore

Appendix 4.A: Instructions

Original instructions were in German. The instructions were the same in both treatments. In the private information treatment the information on the temporary cost shock was shown on sellers' screens (see Appendix 5B) only at the beginning of period 3. In the public information treatment the information about the cost shock was shown on all participants' screens.). The original instructions are available from the author upon request.

General information on the experiment

You are taking part in a market experiment. At the end of the experiment you will be paid according to the decisions you make. Please read these instructions carefully.

Earnings. During the experiment you earn points. The total income in points you earn during the experiment will be converted in Euros at the rate 1 point = 0.012 € and paid to you in cash.

Please note: During the experiment all participants decide independently and anonymously, i.e., no participant will ever learn the identity of the persons with whom he interacts. Therefore it is imperative that all participants observe the following rule: **During the experiment all communication is prohibited, i.e. you are not allowed to speak or express otherwise yourself.** Should you have any questions please ask the experimenter.

Overview of the experimental procedures

In the experiment there are buyers and sellers. Whether you are a buyer or a seller will be randomly determined at the beginning of the experiment and displayed on your computer screen. You will keep your role throughout the whole experiment. The experiment is divided into several trading periods.

In every trading period a seller can sell one unit of a good and a buyer can buy one unit of the good. The seller can produce the good either in high or in low quality. The quality determines the production costs the seller incurs and the value of the good to the buyer.

Sellers and buyers can earn points by concluding a trade. A seller earns points if he sells the good at a price which exceeds his costs of production, a buyer earns points if he buys a good whose value exceeds the purchase price.

In every period a seller's task is to determine the price at which he sells the good. Moreover, he determines whether he will deliver high or low quality. This determines the production costs he incurs and the value of the good to the buyer. At the time of purchase a buyer knows only the price but not the quality of the good and hence he does not know the value of the good.

Calculation of your income in a period

Income of a buyer. For a buyer, a unit of a good has the following values: low quality: 100, high quality: 200. These values are the same for all buyers. The income of a buyer is calculated as follows: income = value of the delivered good – price.

If a buyer decides not buy a good in a period, his income is 0. If he buys at a price that exceeds the value of the good delivered, he makes a loss. Losses will be subtracted from your income.

Income of a seller. A seller has low production costs if he produces a low quality good and high production costs if he produces a high quality good. The exact production costs **are only known to the sellers and will be displayed on the sellers' computer screen. The production costs are the same for all sellers.**

If a seller does not sell his good in a particular period, his income is 0 in this period. If he sells his good, his income is calculated as follows:

Income = price – production costs subject to quality.

Every participant receives an additional 150 points at the beginning of the experiment.

How to trade

There are two markets: **market I** and **market II**. Trading rules differ across markets.

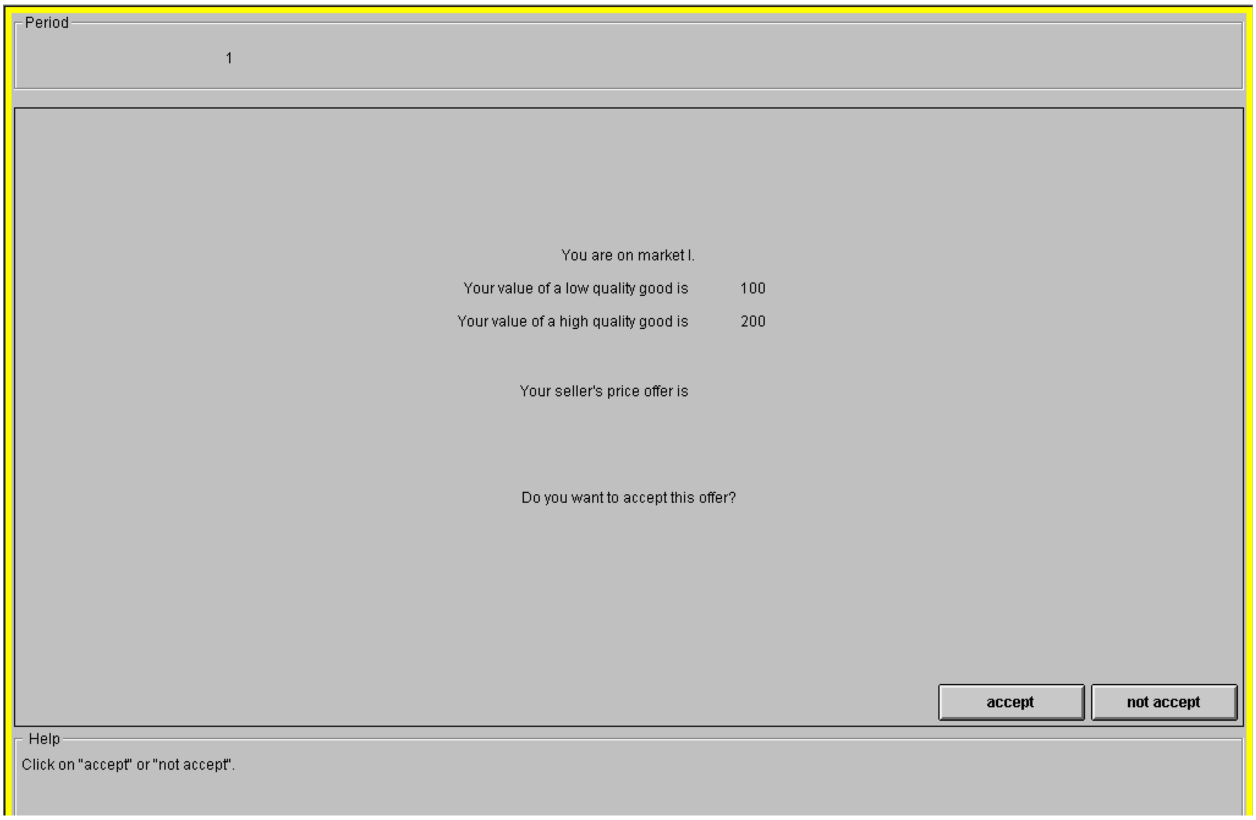
Trading rules for market I

On market I, a specific seller trades with a specific buyer, i.e. a seller and a buyer form a fixed matched trading pair as long as they are trading on market I. This means that you deal with **the same person** as long as you stay on market **I.**

A seller makes an offer by determining a price and the quality he will supply. This also determines his production costs and the value of the good for the buyer in case he accepts. On market I a seller sees the following input screen:

The production costs of high and low quality will be displayed. The seller makes his offer by entering a price and clicking on the quality he delivers if the buyer accepts. By clicking on the o.k.-button he submits his offer to the buyer. After this, the offer can not be revised.

Please note: The buyer will only see the price but not the quality the good. Only upon acceptance does he learn the quality and hence whether the good is of high or low value to him and which income he has earned.



A buyer will see the following screen:

The buyer decides, whether to accept or reject the offer by clicking the corresponding button.

- **The buyer accepts the offer**

The buyer will see an income screen with the following information: price; the delivered quality of the good (low or high); the value of the good (100 or 200) and his income for this period (= value – price)

The seller receives an income screen with: the price; his production costs, subject to the quality he has produced; his income in this period (= price – production costs subject to the produced quality). Only if the buyer has accepted the offer, will both stay on market **I**, i.e., the seller will submit an offer to the **same** buyer in the following period.

- **The buyer rejects the offer**

After a rejection on market **I**, both buyer and seller trade on **market II.** The seller submits a new offer for the current period and the buyer can buy one of the available offers on market **II**.

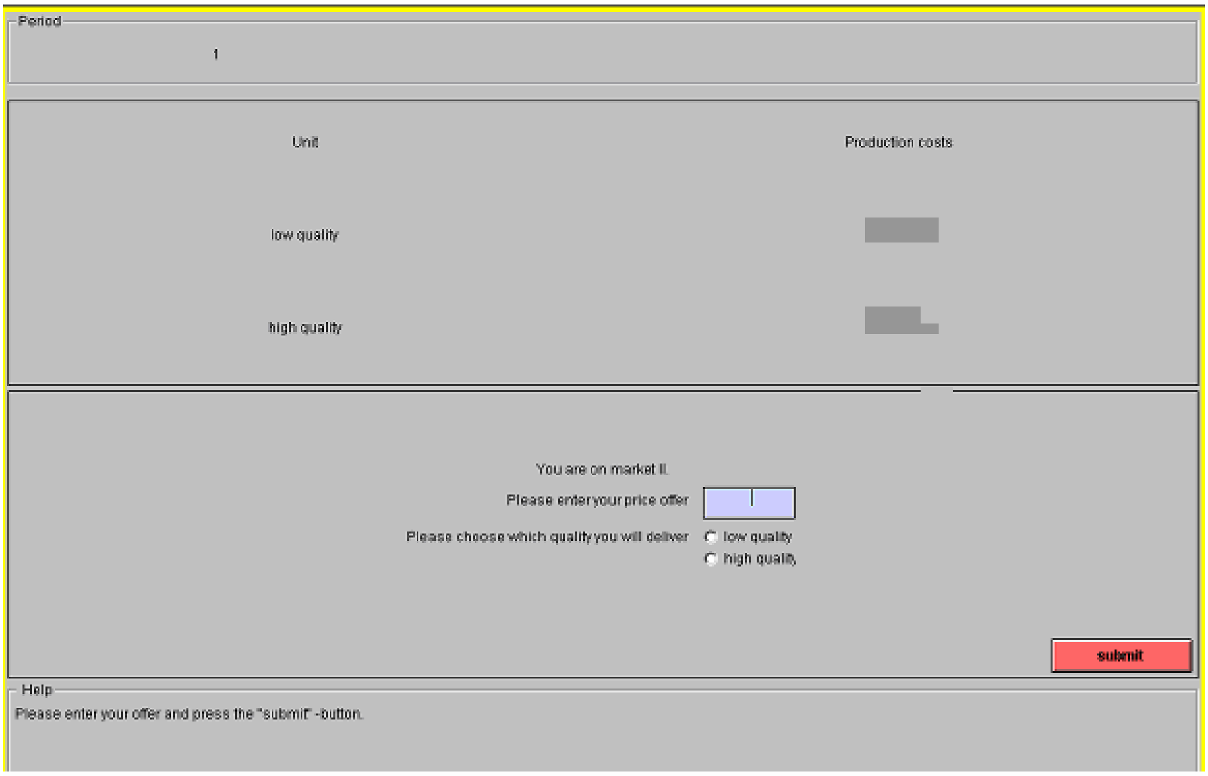
Trading rules on market II

On market II there are several sellers and several buyers. There are always more sellers than buyers, i.e., even if all buyers accept an offer there will be some sellers, who will not sell their good. Contrary to market I an offer cannot be submitted to a specific buyer and a buyer cannot identify which seller has made a particular offer.

How to submit an offer

Every seller has to submit an offer in every period by determining a price and the quality of his good.

A seller will see the following input-screen:

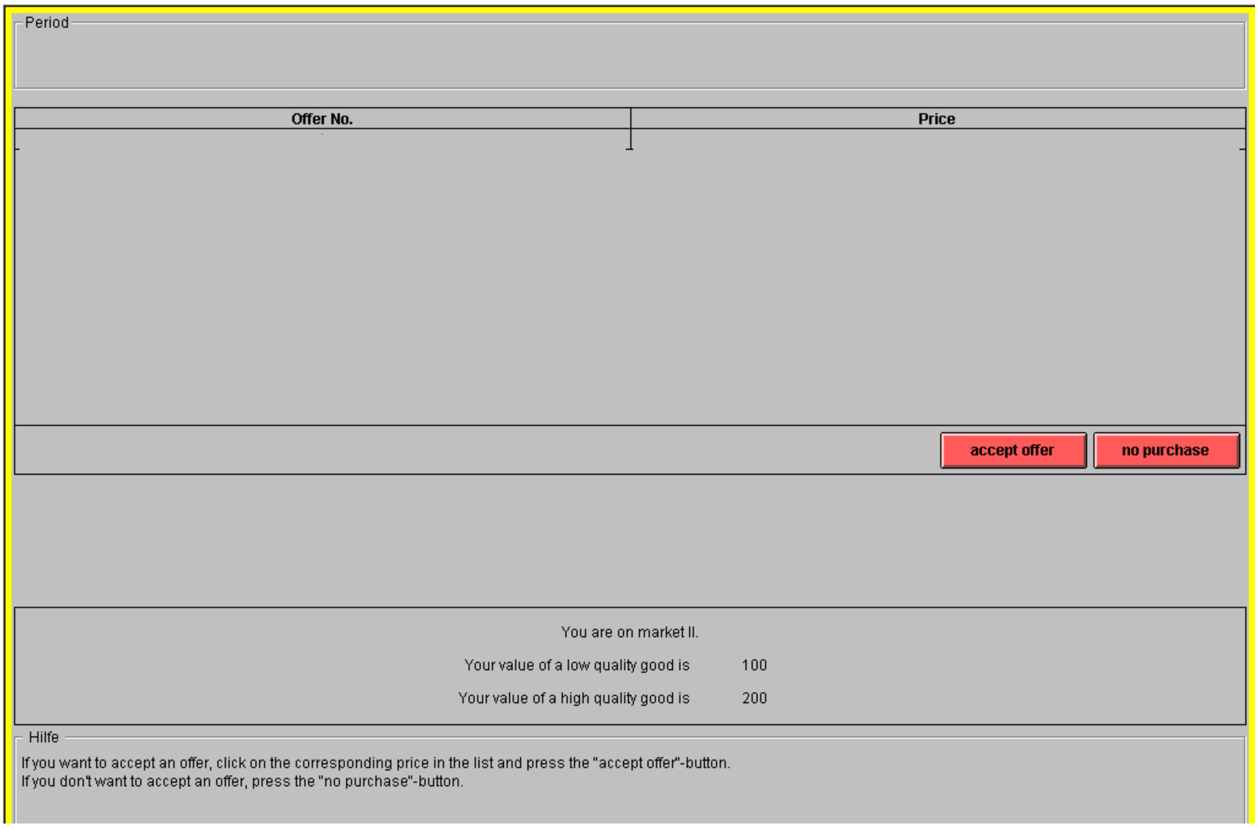


The production costs of a high and low quality good will be displayed. The seller makes his offer by entering a price and clicking on the quality he will produce if a buyer accepts his offer. By clicking on the "submit"-button with the mouse he submits his offer. Once an offer is submitted it cannot be revised.

The decision of the buyers

A list of all offers on market II in decreasing order of price will be displayed to buyers on that market. It is not possible to infer which seller has made a specific offer.

Please note: Buyers can only see the prices offered but not, which quality will be delivered. Only on acceptance does a buyer learn the quality of the good he bought and thus whether it has a high or a low value for him and which income he has earned in this trade.



Buyers will see the following input-screen:

The buyers on market II are asked, one after another, whether they want to accept one of the available offers. A first buyer is randomly chosen. This buyer will see all available offers on market II. If he wants to accept an offer, he clicks on it with the mouse. With another click on the “accept offer”-button the trade is concluded and the offer is removed from the list. If he does not want to buy any offer he clicks on the “no purchase”-button. As soon as the first buyer has finished his decision, the next buyer is randomly chosen. He will see all offers that are still available on market II and decide whether to accept an offer.

Once all buyers have decided they will be informed via an income screen:

whether they have concluded a trade; at which price they have purchased; whether the good is of low or high quality; the value of the good (100 or 200) and the income in this period (= value – price).

Sellers, in turn, are informed via their income screen: whether they have sold their good; their income in this period (= price – production costs subject to the produced quality).

Therewith the trading period is over and a new one starts in which new trades can be concluded.

Trading phases and access to market I and market II

There will be several trading phases. Every trading phase consists of 10 trading periods.

When a new phase starts the period counter at the top left of the screen will start again with period 1. In every trading phase some sellers and buyers start on market I. A seller and a buyer stay on market I as a fixed pair, as long as the buyer accepts the seller’s offers. As soon as the buyer rejects an offer they both join market II and stay there until the trading phase ends after 10 periods. It is not possible to return to market I. The sellers and buyers who start on market I are randomly chosen and randomly matched.

At the beginning of a new trading phase sellers and buyers are again randomly chosen and matched such that nobody will trade with the same person when starting again on market I.

All sellers and buyers who do not start on market I are on market II for the whole 10 periods of a trading phase.

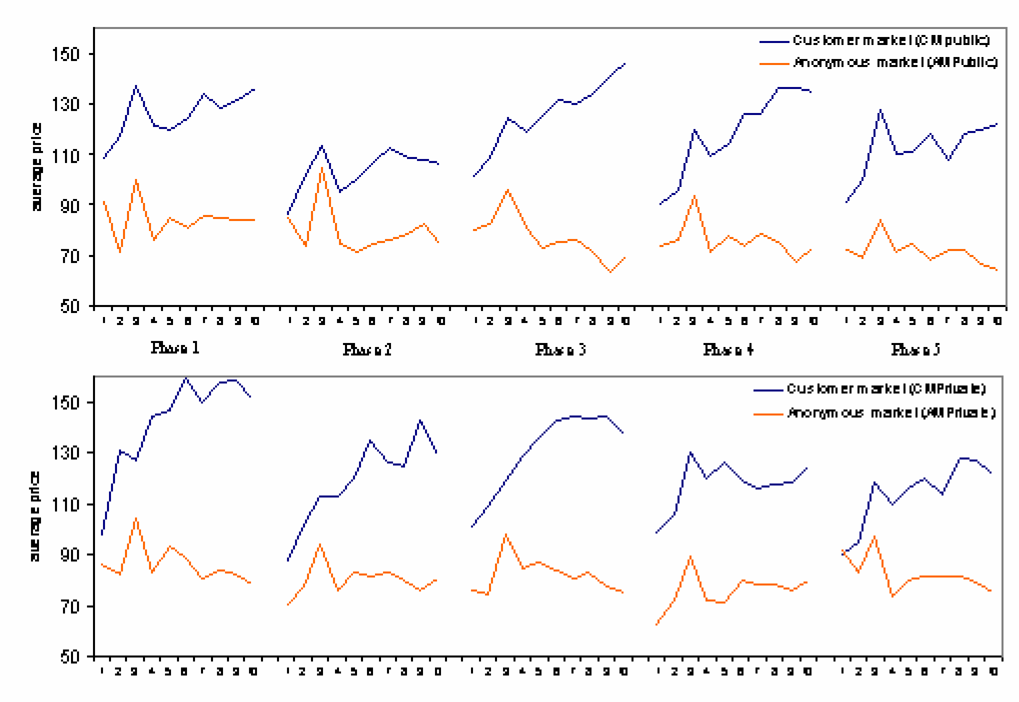
Appendix 4.B: Screens announcing the cost shock in period 3

At the beginning of period 3 the information on the temporary cost increase was shown on the screens below. In the private information treatment (PRIVATE) the following screen was shown to sellers:

In the public information treatment the screen below was shown to both sellers and buyer.

Appendix 4.C: Figure 4.a

Average transaction prices with public information (upper panel) and private information (lower panel) in phase 1-5



Curiculum Vitae

Elke Renner

Address University of Nottingham
School of Economics
University Park
Nottingham NG7 2RD
United Kingdom
Tel: ++44 (0)115 951 5399
Fax: ++44 (0)115 951 4159
e-mail: elke.renner@nottingham.ac.uk

Current Position

since 12/2003 Lecturer at the School of Economics, University of Nottingham

Education

2000-2003	Junior researcher ("wissenschaftlicher Mitarbeiter") and Ph.D. candidate at the University of Erfurt
1997- 2000	Junior researcher at the Wissenschaftliche Hochschule für Unternehmensführung (WHU) – Otto Beisheim Graduate School of Management in Vallendar and junior researcher at the Department for Organization and Management
1989-1997	Economics at the University of Bonn, "Diplom Volkswirtin" Title of diploma thesis: "Experimente zu einem Reziprozitätsspiel mit Bestrafungsmöglichkeit" (Experiments on Reciprocity Games with punishment opportunity), supervised by Professor Reinhard Selten

Research interests

Behavioural economics, microeconomics, game theory, organization theory