

Behavioral Economics and Agency  
Problems: Empirical Studies

Inauguraldissertation  
zur  
Erlangung des Doktorgrades  
der  
Wirtschafts- und Sozialwissenschaftlichen Fakultät  
der  
Universität zu Köln

2009

vorgelegt  
von  
Diplom-Volkswirt Peter Werner  
aus  
Solingen

Referent: Prof. Dr. Axel Ockenfels

Korreferent: Prof. Dr. Dirk Sliwka

Korreferentin: Prof. Dr. Jeannette Brosig

Datum der Promotion: 18.12.2009

# Acknowledgements

First and foremost, I would like to thank my supervisor Axel Ockenfels for encouraging and inspiring me and teaching me how to conduct research. I have strongly benefited from his advice in our discussions about the projects in this thesis.

I am also very grateful for the support and advice of my other supervisors Jeannette Brosig and Dirk Sliwka.

I thank my co-authors Ben Greiner and Felix Lamouroux for their support and collaboration in our joint research projects.

Financial support from the DFG is gratefully acknowledged.

I thank my former and present colleagues Felix Ebeling, Georg Gebhardt, Veronika Grimm, Jens Großer, Burkhard Hehenkamp, Lyuba Ilieva, Imke Jungjohann, Felix Müsgens, Xavier del Pozo Somoza, Alexander Rajko, Matthias Sutter, Gabriele Thorn and Nadja Trhal for our academic discussions, non-academic get-togethers and the very nice working atmosphere.

I am indebted to the student assistants Pouya Azimi, Helge Brinkmann, Julian Conrads, René Cyranek, Christoph Feldhaus, Meike Fischer, Dietmar Ilsen, Johannes Mans, Julia Mohs, Marcel Oestreich, Martin Petzold, Andreas Pollak, Karin Ruëtz, Philipp Tillmann, Johannes Wahlig, Christian Wittneben and Thomas Wolfgarten for their valuable research support.

I am deeply grateful to my parents, Siegfried and Hiltrud Werner, for everything they did for me. I also thank the Hillenbrand family and my friends for their support and confidence. Finally, I thank my wife Petra for sharing her life with me and for all her encouragement, care and patience.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>An Experiment on Inequality and Trust</b>	<b>8</b>
2.1	Introduction . . . . .	8
2.2	Related Literature . . . . .	10
2.3	Experimental Design and Hypotheses . . . . .	12
2.4	Experimental Results . . . . .	15
2.4.1	Aggregate Data . . . . .	15
2.4.2	Individual Decisions . . . . .	20
2.5	Discussion and Conclusions . . . . .	22
2.6	Appendix . . . . .	24
2.6.1	Instructions . . . . .	24
2.6.2	Additional Results . . . . .	26
<b>3</b>	<b>Reference Points, Job Satisfaction and Performance</b>	<b>27</b>
3.1	Introduction . . . . .	27
3.2	The Data . . . . .	29
3.2.1	Survey, Compensation and Performance Data . . . . .	29
3.2.2	Characteristics of the Bonus System . . . . .	29
3.3	Results . . . . .	32
3.3.1	Reference Points and Job Satisfaction . . . . .	33
3.3.2	Reference Points and Performance . . . . .	39
3.4	Discussion and Conclusions . . . . .	43
3.5	Appendix . . . . .	45

<b>4</b>	<b>Social Comparisons in a Real-Effort Experiment</b>	<b>49</b>
4.1	Introduction . . . . .	49
4.2	Related Literature . . . . .	50
4.3	Experimental Design and Hypotheses . . . . .	52
4.4	Experimental Results . . . . .	55
4.5	Conclusions . . . . .	63
4.6	Appendix . . . . .	65
4.6.1	Instructions . . . . .	65
4.6.2	Additional Test Statistics . . . . .	68
<b>5</b>	<b>The Dynamics of Cooperation in a Microfinance Game</b>	<b>69</b>
5.1	Introduction . . . . .	69
5.2	Related Literature . . . . .	72
5.3	Decision Situation and Experimental Parameters . . . . .	74
5.4	Experimental Design and Hypotheses . . . . .	77
5.5	Experimental Results . . . . .	80
5.5.1	Aggregate Results . . . . .	80
5.5.2	Individual Decisions . . . . .	84
5.5.3	Selection into Contracts . . . . .	87
5.6	Conclusions . . . . .	88
5.7	Appendix . . . . .	90
5.7.1	Instructions . . . . .	90
5.7.2	Parametrization . . . . .	95
5.7.3	Additional Figures . . . . .	99
5.7.4	Additional Results . . . . .	100
<b>6</b>	<b>Conclusions</b>	<b>101</b>

# List of Figures

2.1	Average Send Rates over Rounds . . . . .	16
2.2	Average Return Rates over Rounds . . . . .	17
2.3	Average Total Economy Wealth over Rounds, in ET . . . . .	18
2.4	Observed and Simulated Gini Coefficients over Rounds . . . . .	19
3.1	Distribution of Average Bonus Payout Percentages . . . . .	31
3.2	Estimated Relation between Payout Percentage Intervals and Job Satisfaction relative to Managers in the [100%,105%] Interval . . . . .	38
4.1	Average Output per One-Minute Interval . . . . .	56
4.2	Relative Performance Increase per Treatment in % . . . . .	58
4.3	Input Mask . . . . .	66
5.1	Average Effort Run 1 per Treatment . . . . .	81
5.2	Average Effort Run 2 per Treatment (Individual Liability) . . . . .	82
5.3	Average Effort Run 2 per Treatment (Joint Liability) . . . . .	83
5.4	Expected Sum of Period Payoffs per Borrower Pair and Effort Level . . . . .	99

# List of Tables

2.1	Determinants of Individual Send and Return Rates . . . . .	21
2.2	Determinants of Individual Send and Return Rates (2-factorial Set of Independent Interaction Effects) . . . . .	26
3.1	Distribution of Performance Ratings . . . . .	30
3.2	Proportions of Answers to the Question “I am very satisfied with my job”. . . . .	33
3.3	Determinants of Job Satisfaction – Germany (Dependent Variable: Adjusted Satisfaction Score) . . . . .	35
3.4	Determinants of Job Satisfaction – US (Dependent Variable: Adjusted Satisfaction Score) . . . . .	36
3.5	Incentive Effects of Reference Point Violations (Dependent Variable: Performance Rating of Supervisor in Subsequent Year) . . . . .	41
3.6	Determinants of Job Satisfaction (Germany) – Robustness Checks . . . . .	45
3.7	Determinants of Job Satisfaction (US) – Robustness Checks . . . . .	46
3.8	Determinants of Job Satisfaction - Reference Point Effects . . . . .	47
3.9	Incentive Effects of Reference Point Violations – Robustness Checks . . . . .	48
4.1	Average Working Performance per Treatment . . . . .	57
4.2	Percentage of Incorrectly Completed Forms by Source of Error . . . . .	59
4.3	OLS Regressions of Individual Performance in Task 2 . . . . .	61
4.4	P-values of Two-Sided Wald Tests on Differences in Regression Coefficients . . . . .	62
4.5	P-values of Two-Sided Mann-Whitney-U Tests on Differences in Relative Performance Increases between Treatments . . . . .	68

4.6	P-values of Two-Sided Mann-Whitney-U Tests on Differences in Sources of Errors between Treatments . . . . .	68
5.1	Parameter Set . . . . .	77
5.2	Average Effort Levels per Treatment and Run . . . . .	80
5.3	Relative Repayment Rates per Treatment and Run in % . . . . .	83
5.4	Determinants of Effort Choices under Joint Liability . . . . .	85
5.5	Averages Run 1 per Contract Choice . . . . .	87
5.6	Determinants of Effort Decisions under Joint Liability – Robustness Checks . . . . .	100

# Chapter 1

## Introduction

Research in behavioral economics has made enormous progress in integrating insights from psychology and traditional models of economic decision-making. The concept of man in standard theory relies on the idea that each individual has stable preferences and – based on these preferences – maximizes the expected value of a well-defined utility function. However, a large number of experimental and empirical studies have provided evidence that this concept is in many cases not a good approximation of real behavior. In a survey article, DellaVigna (2009) identifies three fields of “behavioral anomalies”.<sup>1</sup> First, utility may be influenced by other factors than a person’s absolute payoff (non-standard preferences). Second, beliefs, i.e. subjective expectations of the probabilities of certain events, may be biased in a systematic way (non-standard beliefs). Finally, economic decision-makers may be guided by heuristics or be influenced heavily by framing or emotions (non-standard decision-making).

This thesis focuses on the application of the first group of behavioral phenomena – non-standard preferences – to the context of agency problems. More precisely, we investigate possible impacts of non-standard preferences on trust, efficiency and economic development (chapter 2), analyze their effects on the functioning of incentives in organizations (chapters 3 and 4) and show how they might influence the success of contract schemes in microfinance credit programs (chapter 5).

First, social preferences play a central role for the presented research studies. Theories of social preferences propose that subjects do not only care about their

---

<sup>1</sup>For other surveys of research in behavioral economics see Earl (1990), Rabin (1998), Kahneman (2003) and Camerer and Loewenstein (2004).

own material payoffs, but also about the payoffs of others. In the ‘Handbook of the Economics of Giving, Reciprocity and Altruism’, Fehr and Schmidt (2006) distinguish between three types of social preference theories. Besides models of altruism (Levine, 1998) and envy (Bolton, 1991), distributive preferences have been formalized in the models of inequality aversion by Bolton and Ockenfels (2000) and Fehr and Schmidt (1999). Here, utility depends not only on absolute payoffs but also on a subject’s relative payoff in a given reference group.<sup>2</sup> Preferences for relative payoffs are particularly important for our analysis of the impact of economic inequality on trust and trustworthiness and for the incentive effects of compensation schemes. In these contexts, preferences for social status, i.e. the notion that agents are affected by their relative rank in a given population, are also highly relevant (see Frank, 1984*a,b*).

Second, a group of theories postulates that subjects focus on the perceived intentions behind economic actions. Concerns for intentions have been modeled by theories of reciprocity (see Rabin, 1993, Dufwenberg and Kirchsteiger, 2004, and Falk and Fischbacher, 2006). If an action by another person is perceived as kind, a person will be more willing to choose a kind action as well, whereas any unkind action will provoke the desire to get even.

In repeated games, reciprocity is reflected also in the pattern of conditional cooperation. There is strong evidence that persons are more willing to contribute in a social dilemma if they believe that others will contribute as well.<sup>3</sup> Conditional cooperation will play an important role for our studies addressing the evolution of behavior, i.e. the dynamics of trust and trustworthiness (chapter 2) and cooperation within groups of borrowers who are jointly responsible for loan repayments (chapter 5).

Finally, the focus will be on loss aversion as proposed by Kahneman and Tversky (1979) in one chapter of this thesis (see also Tversky and Kahneman, 1991, and Kahneman et al., 1991). Loss aversion describes the phenomenon that subjects evaluate economic outcomes not in absolute terms but rather in terms of gains and losses from a given reference point. Furthermore, utility functions capturing loss aversion have a kinked form around the reference point which implies that losses relative to the reference point decrease utility stronger than same-sized gains increase

---

<sup>2</sup>There are also models that additionally incorporate preferences for social efficiency, see, for example, Charness and Rabin (2002).

<sup>3</sup>For a survey of related findings from laboratory and field experiments see Gächter (2006).

it. In our empirical analysis of a bonus system implemented in a multinational company (chapter 3), loss aversion among managers will have strong implications for the incentive and satisfaction effects created by the system.

The four research projects presented in this thesis follow an empirical approach and explore the relevance of the described behavioral regularities for the functioning of incomplete contracts. Our starting point is that the existence of non-standard preferences might crucially alter economic incentives in the respective environments and create inefficiencies. Thus, the first step is to identify where and under which circumstances such inefficiencies might occur. Second, the results of the studies are used to infer practical implications for the design of economic institutions in the fields of personnel and financial economics. In the following, a brief overview is given on the research questions and central results of the studies.

The second chapter “An Experiment on Inequality and Trust” (based on Greiner et al., 2007) explores behavioral impacts of inequality for economic development. The central research question is what the effects for the efficiency of transactions in a society are if agents care about relative positions. On a societal level, the existence of trust makes economic agents more willing to engage in interactions involving the risk of being deceived which in turn enhances efficiency and economic growth. However, data from survey studies show that a too strong dispersion of wealth is detrimental for the process of economic development, because higher social distance lowers peoples’ willingness to trust (Alesina and Ferrara, 2002; Gustavsson and Jordahl, 2008). Our experiment complements this literature and investigates the dynamic interplay of trust, efficiency and distribution in a controlled laboratory setting.

Our working-horse is a growth game, a modified dynamic version of the trust game introduced by Berg et al. (1995). In our dynamic game, we explore behavioral consequences of economic inequality on trust and trustworthiness by varying initial endowments of the agents and allowing them to accumulate wealth from trusting transactions over the rounds. We find that in experimental economies starting with equally distributed endowments, trust is relatively prevalent at the beginning but decreases over time. When endowments are unequal, trust is initially lower yet remains relatively stable. Furthermore, there is no evidence of deliberate redistribution from rich to poor agents in the treatment with heterogeneous endowments that would be in line with straightforward concerns for equality. However, trust in the

economies starting with equal endowments is conditioned on the investor's and the trustee's wealth. This difference in behavior is partly due to the fact that relative wealth has a different information value and source between the treatments.

Our findings are in line with related studies on the effect of the source of inequality on cooperation in dilemma games (see, for example Haile et al., 2008) and studies about the effect of procedural fairness for the acceptance of unfavorable outcomes (see Bolton et al., 2005). Thus, while our laboratory experiment captures only few of the characteristics of economic inequality in the field, it provides an indication that the process through which inequality emerges can influence the relation between income dispersion and development.

The following two chapters of this thesis are concerned with principal-agent relationships in working environments. Moral hazard problems arise when working output is insecure and the effort of an employee cannot be directly observed. The reason is that employer and employee have different objectives. Whereas the goal of the employer is to induce high effort exertion, because this implies higher success probabilities and higher expected profits, a worker has the incentive to choose the lowest possible effort level. In this situation the employer can implement high efforts through monetary rewards such as piece-rates and bonuses that are related to the observable output of a worker.<sup>4</sup> However, if workers are not only concerned about their own wages, monetary incentives may not have the desired impact.

If workers have concerns for relative wages, a central question for the design of compensation schemes is how to differentiate in payments. Frank (1984*a,b*) suggests that preferences for relative ranks lead to wage compression in firms. In his model, a worker's productivity is exogenous per assumption. However, concerns for relative positions might also influence worker productivity endogenously through effort choices. Yet the sign of this relationship is not clear per se. On the one hand, low relative wages might increase efforts in some cases to overcome the dispersion in income. On the other hand, low relative pay might crowd out a worker's motivation.

Anecdotal evidence for the latter effect is provided by Bewley (1999), who surveyed more than 300 US companies and concluded that "within a company, pay inequity offends (indeed, sometimes outrages) employees and destroys trust". With this result as the starting point, we investigate in two studies if and how reference-

---

<sup>4</sup>For general introductions into the economic modeling of moral hazard problems see, for example, the textbooks by Laffont and Martimort (2002) and Bolton and Dewatripont (2005).

dependent preferences influence the functionality of incentives in working relations.

The third chapter entitled “Reference Points, Job Satisfaction and Performance” (based on Ockenfels et al., 2009) applies the phenomenon of relative comparisons and loss aversion to the context of bonus schemes. We analyze the impact of reference points both on job satisfaction and performance with the example of a bonus system for the executives of a world-wide operating company.

A connection of survey, compensation and performance data on the individual level enables us to assess direct implications of incentive devices. In particular, the implemented system has the advantage that it provides a clear reference point for the managers – their yearly bonus budgets – to evaluate their bonus payments. This reference point has two important features: first, it is a personal reference point that determines the individual potential bonus payment in a given year. Second, it is also a social reference point, because it reveals information about the relative standing of a particular person with respect to her peer managers.

We find – in line with loss aversion and inequality aversion – that negative deviations from reference points have significant detrimental effects on reported job satisfaction while there is no significant impact of positive deviations. Furthermore, lack of transparency mitigates the importance of reference points, and spot bonuses have a stronger effect on satisfaction than regular bonus payments even though their economic relevance is much smaller. Furthermore, we find a detrimental effect of reference point violations on performance. The more managers a supervisor reduces below their personal budgets, the lower is her estimated performance in the subsequent year.

These findings have important implications for the design of incentive schemes in business practice. Although regular bonuses account for substantial shares of the managers’ yearly incomes, they are not related to job satisfaction in our sample. Instead, managers evaluate bonus payments based on their relative standings where the reference point is salient. Hence, due to the asymmetric effect of deviations from the reference point, the system handling creates inefficiencies where it fosters its emergence.

The fourth chapter “Social Comparisons in a Real-Effort Experiment” (based on Greiner et al., 2009) picks up the idea that wage transparency and relative comparisons between workers may affect incentives. In a controlled laboratory setting, we let participants perform simple working tasks to examine the net effects of discrimi-

natory wage increases and decreases. Our experimental data provides no evidence for a direct impact of wage changes on performance. However, in a treatment with piece-rates, we find an influence of social comparisons on effort exertion. Information about peer wages triggers performance differences between high-wage and low-wage subjects that are absent when participants do not know about the wage difference. This effect is mainly driven by negative responses to wage cuts. In a fixed wage environment under public wage information, performance differences between high- and low-paid workers are largely mitigated.

Similar to the study in chapter 3, this experiment provides evidence that transparency of wages might be counterproductive from a company's perspective. Relative comparisons lead to inefficiencies, because subjects under piece-rates increase working quantity to catch up with their highly paid co-workers (a pattern also in line with inequality aversion) which comes at the cost of lower quality. Contrary, positive output effects as a reaction to wage increases are not strong enough to yield a positive net effect for the employer.

Chapter 5 "The Dynamics of Cooperation in a Microfinance Game" (based on Werner, 2009) is related to the design of credit contracts in microfinance lending. The latter term subsumes alternative financial services to provide poor population groups the access to credit markets. Under the widely-used contract scheme of group or joint liability lending, several borrowers form groups that are jointly responsible for the repayment of the entire credit sum. While group lending provides a "social" collateral against involuntary default and mitigates agency problems between banks and borrowers, it has the drawback that moral hazard problems are potentially created within the group. As the responsibility for repayment is borne by others, borrowers have incentives to free-ride by shifting investment risks to group members.

Given that in practice borrower groups typically receive more than a single loan, the dynamics of behavior are particularly important and are therefore in the focus of our study. Moreover, we analyze whether different mechanisms are suited to maintain cooperation under joint liability after some participants have been converted to individual credit contracts. In our experiment, subjects decide about the effort invested into risky projects. Under group responsibility, the highest feasible effort maximizes joint payoffs while minimum effort is the individually optimal choice. However, preferences for equal outcomes as well as for reciprocity might motivate positive effort levels.

Indeed, despite the absence of direct sanctioning mechanisms, average effort remains high until the end of the game. Nevertheless, moral hazard is still persistent within groups, as subjects converted into individual liability increase their effort significantly. Moreover, we find a path dependency of behavior, as the willingness to exert effort declines with lower partner efforts and also with the frequency of partner repayments. The correlation of effort choices with partner effort is in line with the findings of conditional cooperation in dilemma games (see, for example, the study by Fischbacher and Gächter, forthcoming). With respect to conversion mechanisms, we find that after self-selection of participants into contract schemes, average effort levels persist on a high level under group lending. Contrary, when groups are converted on the basis of repayment performance, average effort levels among the remaining borrower groups tend to be lower than in the reference treatment.

To sum up, the experiment suggests that the functioning of group lending schemes might indeed be influenced by the dynamic interaction between group members. Hence, increased monitoring from outside might become necessary to mitigate within-group moral hazard.

In the following, all studies are presented in detail. The final chapter reviews their main results and briefly discusses their implications and remaining questions for further research.

# Chapter 2

## An Experiment on Inequality and Trust

### 2.1 Introduction

The degree of trust in an economy may positively influence economic growth and the distribution of wealth. Yet trust is difficult to identify and measure in the field, both on the microeconomic and the macroeconomic level. Survey data frequently discover individual attitudes towards trust, but cannot easily identify to what extent such self-reported attitudes reflect actual economic behavior, and how trust evolves in a dynamic environment. Furthermore, as has been argued by Durlauf (2002), the causal relationship between trust and economic variables is often ambiguous. He thus advocates the use of laboratory studies. In this paper we follow Durlauf's advice and complement the empirical and survey literature with laboratory experiments that systematically investigate the dynamic interplay of trust, efficiency and distribution.

The working horse of our experiment is a growth game, which embeds a variant of the trust game introduced by Berg et al. (1995) into a dynamic context. In this game, an investor can send an amount of money to an anonymous trustee. Before received by the trustee, the amount sent is multiplied by a factor greater than one, and thus yields efficiency gains. Subsequently, the trustee decides on how much of the amount received she wishes to send back to the investor. The amount sent can be interpreted as a measure of trust, while the amount returned measures

the degree of trustworthiness.<sup>1</sup> However, in our game, income from interactions is cumulated over time. Participants start with either an unequal or equal distribution of initial endowments within a group. In each of several rounds they play the trust game with a randomly matched anonymous partner. Before making decisions, both transaction partners are informed about the current wealth of their opponent. Round payoffs are added to endowments, and therefore determine the amount that can be exchanged in future rounds. That is, investments and repayments (i.e. trust and trustworthiness) jointly affect the current and potential future growth rates of the ‘laboratory economy’, as well as the evolution of economic inequality.<sup>2</sup>

We observe that initial investment levels are lower in the treatment starting with unequal endowments (IEQ) compared to the treatment with equal endowments (EQ). However, in IEQ trust depends less on wealth comparisons. Part of the reason is that the source of inequality plays a role in what can be inferred from wealth comparisons: while in EQ all wealth differences must be due to differences in trust and trustworthiness across subjects, in IEQ differences in behavior are concealed by differences in the initial wealth allocation. As a result, trust is triggered differently across treatments. Investment rates decrease steadily and strongly over time in EQ, yet they remain rather stable in IEQ. The wealth distributions in equal and unequal economies converge to each other.

In Section 2.2 we review the literature related to our experiment. Section 2.3 explains the details of our experimental design and procedures, and sketches hypotheses based on previous empirical results and economic models. Our experimental data and statistical analysis are presented in Section 2.4. We discuss our results and conclude in Section 2.5.

---

<sup>1</sup>The original game by Berg et al. (1995) is sometimes called ‘investment game’, and the amount sent is interpreted as a measure for investment in risky projects. In our setting, that interpretation fits as well.

<sup>2</sup>E.g., if all investments yield the same positive rate of return, the dynamic game allows initially rich subjects to increase their endowments much more than initially poor subjects.

## 2.2 Related Literature

There is a large body of empirical and theoretical economic literature on the relationship between inequality within a country and its level of growth and prosperity. The evidence is, however, not unambiguous.<sup>3</sup> Some authors have argued that trust is the key for understanding this relationship: Inequality decreases the level of trust and trustworthiness in an economy, which in turn negatively affects growth.<sup>4</sup> Empirical evidence is provided by Knack and Keefer (1997) and Zak and Knack (2001), who found that countries with higher income dispersion (measured by the Gini coefficient for income) exhibit significantly lower values for a trust measure derived from the World Values Survey (WVS).<sup>5</sup> Similarly, Alesina and Ferrara (2002) found a negative connection between social distance and trusting behavior in a study restricted to the United States. Gustavsson and Jordahl (2008) combine Swedish individual panel data with aggregate data on inequality and find that stronger disparities among people in the bottom half of the income distribution have a detrimental effect on trust. Furthermore, a number of empirical studies established a positive impact of generalized trust on economic development (Knack and Keefer, 1997; Zak and Knack, 2001; La Porta et al., 1997).

---

<sup>3</sup>The academic discussion started in the 1950s with the Kuznets-Curve (Kuznets, 1955), which proposed a relation between inequality and economic development in the form of an inverted U. Most of the more recent theoretical literature assumes a negative relationship, including the models of Galor and Zeira (1993) and Persson and Tabellini (1994). Bénabou (1996), Ros (2000) and Glaeser (2005) survey the differing strains of literature. The majority of early empirical studies of the relationship of inequality and trust find a negative link between income disparity and growth (see Bénabou, 1996). However, some of the more recent studies, employing panel data and advanced econometrics, yield either no effect (e.g. Barro, 2000) or even a positive relationship (e.g. Forbes, 2000; Castelló-Climent, 2004). Banerjee and Duflo (2003) argue that non-linearity of the relationship might be a reason for the ambiguous results. They find that any *change* in inequality – in each direction – affects growth detrimentally. In their meta-analysis of empirical studies, De Dominicis et al. (2008) show that estimation techniques, included independents, development status of countries, and length of considered growth period have a significant impact on the estimated size and direction of the effect of inequality on growth.

<sup>4</sup>See Jordahl (2008) for an overview of different mechanisms explaining the negative impact of inequality on generalized trust. Other authors see other forms of human capital, such as education (e.g. Castelló-Climent, 2004), or social preferences as possible links. Corneo and Grüner (2000) and Corneo and Jeanne (2001) discuss concerns for social status, as these might discourage both poor and rich subjects to accumulate income in an unequal society and lower the political will for redistribution.

<sup>5</sup>The World Values Survey consists of repeated interview studies with representative population samples on the changes in moral values and beliefs, conducted in 80 countries all over the world since 1981. One question is: “Generally speaking, would you say that most people can be trusted or that you can’t be too careful in dealing with people?”. The percentage of positive responses is used as a measure of generalized trust in a country.

Durlauf (2002), however, notes that there are various problems of causality and identification in many empirical studies on the relationship between social capital, trust and economic indices. He thus proposes the use of laboratory experiments to investigate the causal structure between these measures. Results from such economic experiments allow to build models of individual behavior to explain the relationship between social capital and economic measures on the aggregate (see, for example, the model by Glaeser et al., 2002).

There is experimental evidence on the relationship between cooperation and inequality in public goods games, which share a couple of features with the trust game studied here. This evidence is, however, mixed. In a survey on repeated public goods games with complete information, Ledyard (1995) comes to the conclusion that economic heterogeneity among subjects generally lowers cooperation levels. Chan et al. (1996) find that poor subjects contribute more to a public good than rich subjects. Buckley and Croson (2006) conduct a linear public good game with heterogeneous endowments of the subjects. In their study, rich and poor subjects contribute on average the same absolute amount to a public good. Thus, as poor subjects contribute a higher share of their respective endowments, economic inequality increases within the experimental groups.

Other studies are devoted to the relationship between social distance (measured on various scales) and investment behavior in the trust game introduced by Berg et al. (1995). Glaeser et al. (2000) combine questionnaires on social backgrounds and trust attitudes with an experimental trust game. In their experiment, subjects interacting face-to-face with a partner of a different race or nationality exhibit a lower level of trustworthiness. In addition, a higher social status of the sender seems to be positively related to the earnings of a trusting decision. Hence, the results of this study indicate detrimental effects of social distance. However, their survey measures of generalized trust are not correlated with actual trusting behavior.

Fershtman and Gneezy (2001) find significantly different degrees of trust towards different ethnical groups in the Israeli-Jewish society, although these groups do not differ concerning their trustworthiness. In a recent study, Haile et al. (2006) conduct a trust game experiment with South-African students. They find negative effects of socio-economic differences, as low-income subjects trust less when confronted with a high-income transaction partner from another ethnic group.

To our knowledge, there are only two experiments which specifically study the role

of payoff inequality in the trust game. Contrary to the studies discussed above, social distance is induced by the experimental design. Brülhart and Usunier (2007) vary endowments of the trustees, which however does not affect trust. Anderson et al. (2006) employ an equal as well as a symmetric and a skewed unequal distribution of show-up fees in a trust game. The distribution of show-ups is either private or public information. The authors observe only small and non-systematic effects of unequal endowments on trusting behavior.

## 2.3 Experimental Design and Hypotheses

In our study, we focus on the dynamic interaction of trust, trustworthiness and inequality. Therefore, we develop a growth game which embeds the essentials of trust game introduced by Berg et al. (1995) but puts them into a dynamic growth and distribution context. The growth game is played over 20 rounds. In each round, two randomly and anonymously matched subjects play a variant of the trust game. One of the subjects is randomly assigned the role of the investor, the other the role of a trustee. Before decisions are made, each subject is informed about his own and the opponent's wealth in the current round. Wealth is defined as the initial endowment plus any payoffs that have been accumulated in earlier rounds. A player's wealth limits the amount that he can invest or return in the current round of the growth game in the following way. The investor decides on an amount  $S$ , which is not allowed to exceed his current wealth, to be sent to the trustee. Any amount sent is multiplied by the factor 1.2, i.e. the trustee receives  $1.2S$ . Next, the trustee can decide on the amount  $R$  to be sent back to the investor. The minimum amount to be returned is  $0.9S$ , or 90% of the amount sent.<sup>6</sup> The upper limit is given by the sum of the current wealth of the trustee plus the received amount. Because payoffs are accumulated over the course of the repeated trust game, our laboratory economies could maximally grow by an expected factor of 6.7.<sup>7</sup>

As our main experimental parameter, we varied the distribution of the initial endowments across our two treatments. In the equality condition (EQ), all subjects

---

<sup>6</sup>These rules make the one-round interaction in our game equivalent to the original trust game interaction with a sent amount multiplier of three, with the exception that the amount that can be sent is restricted to 10% of the investor's wealth. Without such restrictions, the experiment could have gone beyond the scope of any reasonable financial budget.

<sup>7</sup>As in each round only half of the subjects in the economy are randomly assigned to the role of the investor, the expected maximum growth rate over 20 rounds with full investments corresponds to  $1.1^{20}$ .

were endowed with an amount of 500 ET (Experiment Talers) before the first round. In the inequality treatment (IEQ), half of the subjects in each matching group received 200 ET, and the other half received 800 ET. In order to investigate experience effects and to test robustness of behavior, we played two runs of 20 rounds; that is, after the first 20 rounds of the experiment we restarted the game for another 20 rounds. Subjects were told before the session that the experiment consisted of several runs, one of which would be randomly selected for payoff.

The experimental sessions took place in the Cologne Laboratory for Economic Research. We conducted four sessions, two for each of our treatments. Subjects were recruited using the Online Recruitment System by Greiner (2004). Altogether 128 student subjects participated. Each session consisted of 32 participants. Random matching per round was restricted to groups of 8 participants.<sup>8</sup> It was publicly known that two subjects would never interact with each other in consecutive rounds. Due to this procedure, we obtained observations on 8 statistically independent ‘economies’ for each treatment. Overall, we collected 2,560 choices for each player role.

The experiment was computerized using the z-Tree software (Fischbacher, 2007). After subjects arrived and were randomly assigned to a cubicle, instructions were distributed.<sup>9</sup> Questions were answered privately. At the end of the experiment subjects filled in a post-experimental questionnaire asking for demographical data and containing open questions for motivations of subjects’ decisions. Finally, either run 1 or run 2 was selected for payoff by publicly rolling a die. Participants were paid out privately and left the laboratory. The exchange rate was fixed at 150 ET = 1 Euro. The average payoff was 12.25 Euros (including a show-up fee of 2.50 Euros) with a standard deviation of 5.09 Euros. Each session lasted approximately one and a half hours.

In the rest of this chapter, we will motivate a number of competing hypotheses for the dynamic interplay of inequality and trust based on (simplified) theoretical reasoning and empirical findings. These hypotheses help organizing our analyzes and findings. At the same time, however, we wish to caution that our experiment is mainly designed to complement the empirical studies, and not as a test of any particular theory - if only because there is no theory yet that addresses the potentially

---

<sup>8</sup>Subjects were not informed that the matching procedure was restricted in such a way, conveying the impression that being matched with the same opponent more than once is very unlikely.

<sup>9</sup>Instructions are included in section 2.6.

complex dynamics we are interested in.

The standard game theoretic prediction is trivial. Because of the finiteness of the growth game, there is no trust and no trustworthiness among selfish and rational players if selfishness and rationality are common knowledge. However, starting with Berg et al. (1995), numerous experiments have shown that subjects are willing to invest and return non-trivial amounts of money in the trust game. For a survey of the trust game literature see, for example, Camerer (2003).

While the experimental one-shot version of the trust game is by now well-analyzed and -understood, the dynamic interplay of inequality and trust in the context of our growth game is not easily predicted. However, observe that both of our treatments start with identical *average* endowments. If inequality does not affect subjects' willingness to send and return money, relative to their endowments, the two treatments may be expected to yield equivalent results with respect to growth rates.<sup>10</sup> On the other hand, the empirical and experimental literature on social and economic heterogeneity cited in Section 2.2 suggests that we may observe a negative impact of inequality on trust in our setting. Dispersion of wealth could increase social distance between economic agents and, as a result, trust and trustworthiness may decrease. To the extent our experiment captures some of the underlying mechanisms assumed in this literature, we should expect less growth and lower efficiency in the IEQ treatment.

Finally, we note that theories of social preferences can organize some of the deviations from standard equilibrium behavior observed in the trust game. For instance, inequity aversion models (Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999) can in principle explain both trust and trustworthiness in the trust game.<sup>11</sup> However, these models do not yield unambiguous comparative static predictions across the two treatments of our growth game. To see why, observe for instance that a rather fair-minded investor who is matched with a relatively poor trustee may send money to equalize payoffs, while a rather selfish investor may not send money because he cannot expect to get anything back from a relatively poor opponent. Thus, the

---

<sup>10</sup>In the beginning of the first round, the average endowment of investors in treatment EQ is equal to 500, as it is in treatment IEQ. Thus, if the same share is sent and returned, expected overall invested amounts are the same, as well as the amounts returned. Therefore, the expected endowments of investors in round 2 are the same in both treatments. The same reasoning applies to all consecutive rounds of the game.

<sup>11</sup>See Bolton and Ockenfels (2000), page 187, for a detailed description of the mechanics of the fairness models in the context of the trust game by Berg et al. (1995).

predictions of inequity aversion models will depend on the distribution of preferences. It appears, though, that ‘myopic’, straightforward concerns for equal payoffs lead to more trust and trustworthiness in IEQ in the following sense: Even when an inequality averse subject assumes that everybody else behaves in a completely selfish manner, he still has reason to trust and to be trustworthy towards relatively poor opponents in the inequality treatment (where, in the beginning of round 1, the payoff distribution is unfair), but no such incentive exists in the equality treatment (where the payoff distribution is fair if everybody behaves selfishly).

## 2.4 Experimental Results

### 2.4.1 Aggregate Data

Figures 2.1 and 2.2 depict the evolution of average send and return rates over time. We define the ‘send rate’ in a particular round as the share of the investor’s wealth in this round that she invests in the transaction. The ‘return rate’ is defined as the amount returned minus the mandatory 90% ( $R - 0.9S$ ), divided by the amount received minus the mandatory 90% ( $1.2S - 0.9S$ ). For example, a return rate of  $1/3$  implies that the trustee returns exactly the amount invested by the investor. (The dashed horizontal line in Figure 2.2 indicates this ‘break-even line’.) For figures and non-parametrical tests the send rate averages are calculated by adding up all amounts sent in a matching group, and dividing the sum by the total wealth of the senders.<sup>12</sup>

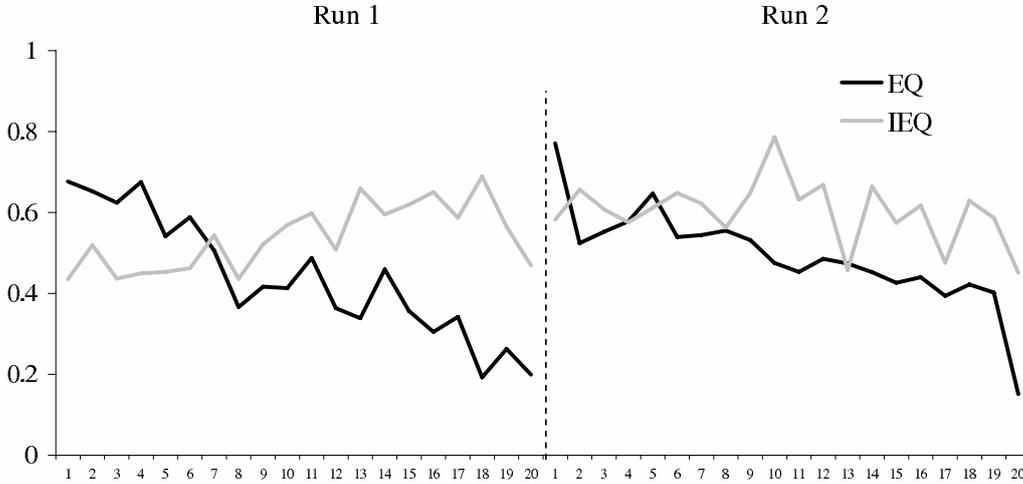
Figure 2.1 shows that the dynamics of trust differ markedly between the treatments. In the first round of the games, the equal distribution of wealth leads to higher trust levels (54% more, to be exact) than the unequal distribution. This observation is in line with previous empirical and theoretical work suggesting that inequality hampers efficiency. However, applying two-sided Mann-Whitney-U (MWU) tests to (statistically independent) individual send rates and to respective matching group data in round 1 shows that the difference is not significant ( $p = .136$  and  $p = .130$ , respectively).<sup>13</sup>

---

<sup>12</sup>This procedure seems appropriate since here our focus is on aggregate behavior and independent observations. However, our conclusions from statistical tests would not be different if we had used individual averages.

<sup>13</sup>The reason is the heterogeneity of subjects in treatment IEQ. More specifically, poor IEQ subjects send less in absolute terms than EQ subjects ( $p < .01$ ), but not in relative terms, while

FIGURE 2.1  
AVERAGE SEND RATES OVER ROUNDS



Send rates in treatment EQ strongly and steadily decrease over time from 68% in round 1 to 20% in round 20 in run 1, and from 77% to 15% in run 2, while send rates in IEQ increase slightly in run 1 and decrease slightly in run 2. Correspondingly, in EQ average send rates of the matching groups are significantly and negatively correlated to the number of rounds (Pearson-R=-.586,  $p < .001$  and Pearson-R=-.394,  $p < .001$  for run 1 and 2, respectively) while this is not (strongly) so in IEQ (Pearson-R=.061,  $p = .442$  and Pearson-R=-.154,  $p = .051$  for run 1 and 2, respectively).<sup>14</sup>

As shown in Figure 2.2, average return rates are about the break-even level that makes an investment profitable, with probably a small advantage for IEQ in run 1. In fact, differences in return rates across treatments and over time are much less pronounced than differences in send rates. In the aggregate data, we find that, in the first run, average return rates are 24% lower in treatment EQ than in treatment IEQ<sup>15</sup> - yet the effect disappears in run 2.

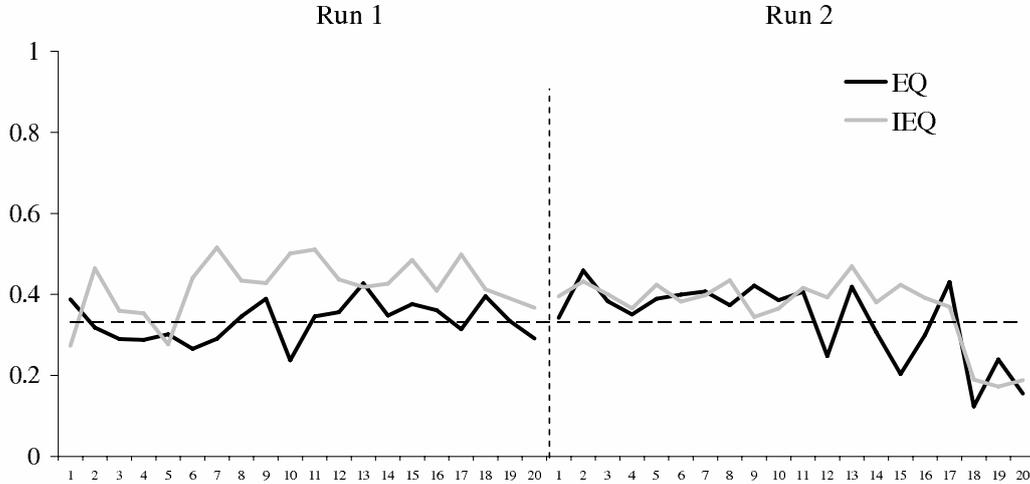
---

rich subjects send less in relative terms ( $p < .05$ ), but not in absolute terms. However, our analysis of individual behavior in the next subsection, where we control for these wealth effects, confirms the observation that initial inequality hampers efficiency on any standard significance level.

<sup>14</sup>Applying two-sided Wilcoxon Matched Pairs Signed Ranks (WMPSR) tests, a similar conclusion is reached when comparing matching group averages in the first and the second half of each run.

<sup>15</sup>A two-sided MWU test applied to independent matching group averages indicates significance with  $p = .021$ .

FIGURE 2.2  
AVERAGE RETURN RATES OVER ROUNDS



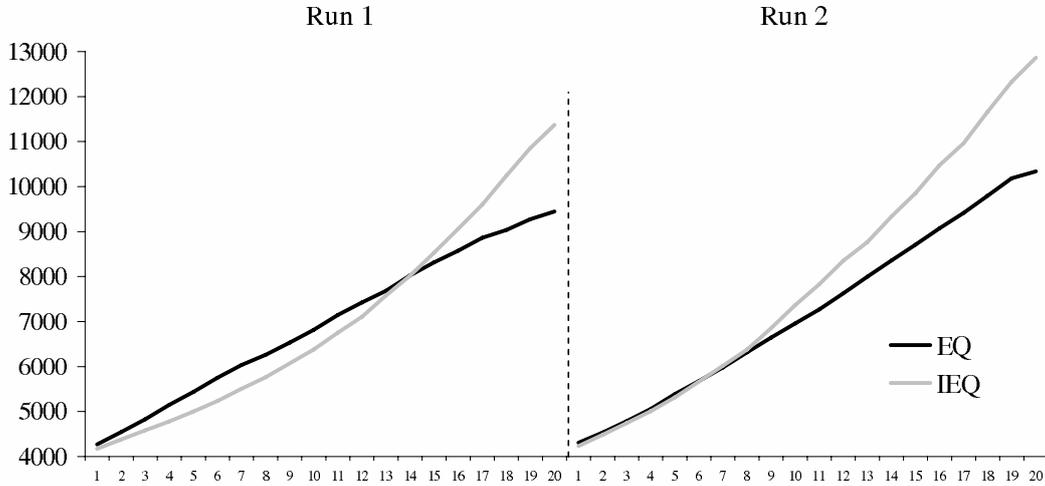
More investment directly expands overall wealth, because the latter is a cumulative measure of the former. So, the different dynamics in trust across laboratory economies are reflected in different growth rates of overall wealth. Figure 2.3 depicts average economy wealth over time. There are substantial efficiency gains in both treatments and runs, with total average wealth more than doubling in all runs of both treatments. Initially, wealth in treatment IEQ lags behind the one in treatment EQ. However, as average send rates remain on a relatively high level in treatment IEQ and significantly decrease in treatment EQ, the lag is eventually counterbalanced and reversed in the last few rounds.<sup>16</sup> In run 2 we do not observe large initial differences, and after the first few rounds treatment EQ lags behind indefinitely. Accumulated wealth in IEQ finally accounts for more than 300 % of initial endowments.

Not only efficiency gains but also the distributions of wealth in our laboratory economies evolve endogenously through sending and returning decisions. We use Gini coefficients to analyze the dispersion of individual wealth levels.<sup>17</sup> Figure 2.4

<sup>16</sup>Statistically, the wealth of IEQ economies in rounds 1 to 5 is weakly significantly lower than in EQ ( $p=.060$ , two-sided MWU test). Final wealth levels are not significantly different from each other.

<sup>17</sup>The Gini coefficient as a measure for disparity takes the value of zero if the income is equally distributed among the subjects and  $(n - 1)/n$  if all wealth is concentrated on only one subject. Here, the maximum value of the Gini coefficient is  $7/8$ , as the number of subjects per experimental matching group is  $n = 8$ ).

FIGURE 2.3  
AVERAGE TOTAL ECONOMY WEALTH OVER ROUNDS, IN ET



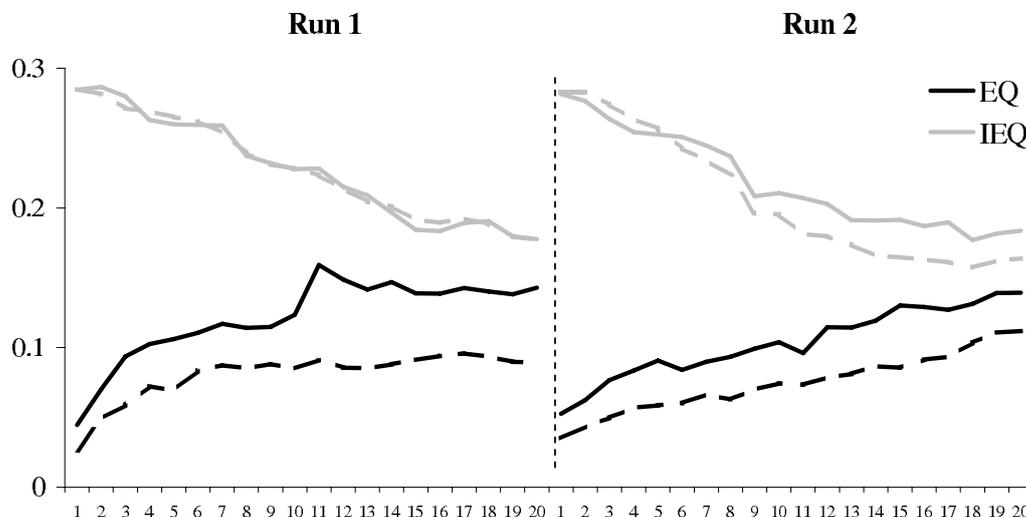
shows average matching group Gini coefficients in treatments EQ and IEQ (solid lines). We observe that Gini coefficients strongly and significantly decrease (increase) in treatment IEQ (EQ).<sup>18</sup> Furthermore, the values for the Gini coefficients tend to converge to each other towards the end of a run. In the last round of a run, IEQ and EQ Ginis are not significantly different (MWU,  $p = .505$  and  $p = .105$  for run 1 and 2, respectively).

Redistribution in our setting might have two different sources: on the one hand, it could be the result of random, homogenous interaction, in the sense that rates are not conditioned on individual wealth states or wealth comparisons. Because, for a given rate, richer subjects send more in absolute terms than poorer subjects, unconditional behavior moves the economy towards more equality when starting with unequal endowments. On the other hand, redistribution could be the result of send and return rates which systematically depend on own and probably others' wealth in the current state. Depending on the nature of conditional behavior (which will be analyzed in subsection 2.4.2) and the heterogeneity of the behavioral patterns, the resulting system behavior may increase or decrease equality relative to what can be expected from unconditional homogenous interaction.

In order to isolate the effects of these two mechanisms, we simulate Gini coef-

<sup>18</sup>Two-sided WMPST tests applied to matching group averages for rounds 1-10 and rounds 11-20 of each run yield  $p = .008$  for treatment IEQ (both runs) and  $p = .055$  and  $p = .008$  for treatment EQ (run 1 and run 2, respectively).

FIGURE 2.4  
OBSERVED AND SIMULATED GINI COEFFICIENTS OVER ROUNDS



ficients for unconditional, homogenous behavior. More specifically, simulations are based on the same role and group matchings as implemented in our experiment. Additionally, we assume that in every round all participants in a matching group behave identical – like the group average.<sup>19</sup> If actual behavior is unconditional with respect to wealth levels and differences, simulations and actual behavior cannot differ.

The average simulated Gini coefficients (see the dashed lines in Figure 2.4) follow the same general pattern as the observed ones. In treatment IEQ, simulated and observed Gini curves are nearly the same (run 1) or differ only slightly (run 2). WMPSR tests yield no significance comparing average observed and simulated Ginis for whole runs or 10-round intervals. In treatment EQ, observed Gini values are constantly higher than the simulated values, and the differences are significant at  $p < .016$  with two-sided WMPSR tests applied to whole runs or 10-round intervals.

Thus, on the aggregate level, we find little evidence for systematic and deliberate redistributive behavior from rich to poor in treatment IEQ. Contrary, inequality rises faster than expected in treatment EQ, suggesting that there are indeed heterogeneous behavioral patterns that systematically affect wealth distribution – as studied in the next subsection.

<sup>19</sup>This procedure yields the same economy growth rates in the simulation as in the experiment.

### 2.4.2 Individual Decisions

We regressed the individual send and return rates on a number of independent variables in order to investigate the determinants of individual behavior.<sup>20</sup> Besides the *Round* number (1-20), and two dummies for *Treatment* (0 for EQ, 1 for IEQ) and *Run* (0 for 1st, 1 for 2nd run), we include the relative wealth standing of the sender ( $W_{SDR}$ ) and the responder ( $W_{RSPR}$ ) prior to the current transaction. Both variables are derived by dividing the respective decision maker's wealth by the average wealth in her economy (matching group). For the regression of the return rate we also include the send rate of the counterpart. To account for group-specific as well as for subject-individual characteristics, we use Mixed Effects models. Due to the rather complex dynamic nature of our game, we cannot exclude any interaction effects between the independent variables, which poses a model selection problem that we addressed in the following way: in our main regressions, we start with the full factorial set of potential interaction effects. We then iteratively throw out insignificant effects. After two iterations we ended up with the models presented in Table 2.1. Note that, by construction, all included interaction effects are significant. As a second approach we ran regressions on the complete 2-factorial set of interaction factors. The results are presented in Table 2.2 in the Appendix and basically confirm the analysis discussed here.

The model for investor decisions, presented in the left column of Table 2.1, shows that wealth positions influence trusting behavior in treatment EQ, but are of only minor importance in treatment IEQ. In treatment EQ, the effect of both direct wealth variables is negative. That is, participants send less the richer they are and the richer the responder is. However, the positive interaction effect indicates that the more equal sender and responder are, the less pronounced are the wealth effects. The lowest send rates are found for poor senders towards rich responders, and vice versa. Contrary, in treatment IEQ all these three effects are mitigated (see the interaction effects of *Treatment* with  $W_{SDR}$ ,  $W_{RSPR}$ , and  $W_{SDR} * W_{RSPR}$ ). These observations are consistent with the simulation results of the Gini coefficient dynamics. While the trust decisions in EQ systematically affect the wealth distribution in the economy

---

<sup>20</sup>We had to exclude 6 and 314 observations in the models on the send rate and return rate, respectively, because the send rate is only defined for positive wealth of the investor, and the return rate is only defined for positive amounts sent. The reason for the 6 observations with a sender wealth of zero was that a participant erroneously returned her entire round wealth in one interaction.

TABLE 2.1  
DETERMINANTS OF INDIVIDUAL SEND AND RETURN RATES

Dependent Variable	Send Rate		Return Rate	
	Coefficients	(Std.Error)	Coefficients	(Std.Error)
Round	-0.017**	(0.001)	-0.005**	(0.001)
$W_{SDR}$	-0.950**	(0.152)	-0.080**	(0.023)
$W_{RSPR}$	-0.922**	(0.145)	0.029	(0.030)
$W_{SDR} * W_{RSPR}$	0.662**	(0.142)		
Treatment	-1.255**	(0.198)	0.075*	(0.033)
Treatment*Round	0.013**	(0.002)		
Treatment* $W_{SDR}$	0.929**	(0.160)		
Treatment* $W_{RSPR}$	0.892**	(0.151)		
Treatment* $W_{SDR} * W_{RSPR}$	-0.631**	(0.147)		
Run	0.036**	(0.011)	0.016	(0.033)
Run* $W_{SDR}$			-0.077*	(0.038)
Run* $W_{SDR} * W_{RSPR}$			0.063*	(0.027)
Send rate			0.154**	(0.015)
Constant	1.866**	(0.173)	0.275**	(0.050)
<i>Random Effects</i>				
Group Std.Dev.	0.186	(0.045)	0.0001	
Subject Std.Dev.	0.264	(0.019)	0.174	(0.013)
Residual Std.Dev.	0.272	(0.004)	0.246	(0.004)
No. of obs.	2554		2246	
Wald $\chi^2$	305.10		249.81	
Log-restricted likelihood	-529		-207	

Standard errors are given in parentheses. \* and \*\* denote significance on the 5% and 1%-level, respectively. Regression models are derived by starting with a full factorial set of interaction effects and iteratively throwing out insignificant effects.

beyond what can be expected from non-conditional, homogenous trust patterns, this is not the case in IEQ.

The effect of the repetition of the game (Run) is positive and corresponds to an increase of average send rates in the second run of the game, across treatments. With respect to the evolution of investments over time, we find a negative effect of the number of rounds for treatment EQ, whereas in IEQ the effect of time is somewhat mitigated. Finally, the coefficient of the treatment dummy is highly significant, pointing out a negative effect of *initial* inequality introduced by the variation of

endowments.<sup>21</sup>

The model for trustee decisions indicates that return rates are generally higher in treatment IEQ, and shrink over time. We find that participants reciprocate high investments, as the coefficient for the send rate is positive and significant: The more of his wealth the investor sends, the higher his expected profit margin. Rich senders can expect to earn less from their trusting decisions than poor senders, while there seems to be no clear effect of the wealth of the responder herself. Also, there are no significant interaction effects between treatment and the relative wealth indicators.

## 2.5 Discussion and Conclusions

We analyze the behavioral dynamics of economic inequality and trust. In our laboratory economies, participants start with either equal or unequal endowments. They then repeatedly play an investment game and, by accumulating their payoffs, endogenously create growth and wealth distributions. In each round, both transaction partners are informed about the current wealth of their opponent.

Initially, investments are higher in economies starting with equal endowments (EQ) compared to the economies starting with unequal endowments (IEQ). However, in EQ cooperation deteriorates over time, while trust remains stable in IEQ. As a result, EQ economies initially grow faster, but are ultimately outperformed by the IEQ economies in terms of efficiency. With respect to the distribution of wealth, IEQ economies become more equal, while EQ economies become more unequal, such that the distributions of wealth are converging to each other over time.

The different dynamics of EQ and IEQ at the aggregate level are mirrored by two differences at the individual level. First, there is initially less trust in IEQ. Second, trust in the EQ economies is conditioned on the investor's and the trustee's wealth, while no analogous effects can be identified in IEQ. Conditional trust appears to be the main reason for the downward trend in EQ.

We speculate that the differences are partly due to the fact that relative wealth has a different information value and source in EQ compared to IEQ economies. A

---

<sup>21</sup>Several robustness checks have been conducted with respect to the results of our regressions. First, Tobit Random Effects models controlling for censored send and return quotas yield the same results as described above. Second, our main result - the conditioning of trust on sender and responder wealth - is robust against inclusion of a 'personal experience' variable (the average return rate a sender experienced in previous rounds), and is also present when comparing send rates from rich/poor senders to rich/poor responders using non-parametric WMPSSR tests on the economy level, not controlling for other factors as in the regression above.

large relative wealth in EQ is a rather reliable signal for not having been trustworthy in the past: unfair agents become richer. A large relative wealth in IEQ, on the other hand, is not only the result of relatively selfish behavior but also of the exogenously imposed unequal endowments.<sup>22</sup> Both, models of strategic and social behavior, are then in line with the observed patterns of (conditional) trust. Because a higher wealth tends to suggest lower trustworthiness in EQ (but not in IEQ), richer people should be trusted less in EQ (but not in IEQ). Moreover, a number of studies (see for example, Bolton et al., 2005 and Frey et al., 2004) suggest that people are more tolerant towards inequitable outcomes if inequality is the result of a procedurally fair allocation mechanism. Finally, a study by Haile et al. (2008) shows that the willingness to cooperate in a dilemma game depends on the source of inequality.

Thus, to the extent that high wealth exogenously and randomly imposed in IEQ economies is perceived as fair while high wealth endogenously resulting from selfish behavior is perceived as unfair, inequality in EQ may invoke a different social response than inequality in IEQ. Modeling the strategic and social roots of the dynamic interaction of distribution and efficiency is left to future research.

---

<sup>22</sup>This reasoning is supported by the data. Results of Pearson correlations of average return rates and final wealth in half-runs of treatment EQ are  $R=-.358$ ,  $p=.004$  and  $R=-.479$ ,  $p<.001$  for rounds 1-10 and rounds 11-20 in run 1, and  $R=-.397$ ,  $p=.001$  and  $R=-.309$ ,  $p=.013$  for rounds 1-10 and rounds 11-20 in run 2, respectively. Contrary, in treatment IEQ half-run correlations between average return rates and wealth are low or insignificant;  $R=-.142$ ,  $p=.262$  and  $R=-.048$ ,  $p=.704$  for rounds 1-10 and rounds 11-20 in run 1, and  $R=-.240$ ,  $p=.056$  and  $R=-.152$ ,  $p=.231$  for rounds 1-10 and rounds 11-20 in run 2, respectively.

## 2.6 Appendix

### 2.6.1 Instructions

*Below we include the instructions used in the first run of treatment IEQ, translated from German. Instructions for the other runs and treatments were worded analogously.*

Welcome to this experiment! In this experiment you can earn money. How much money you earn depends on your decisions and the decisions of the other participants.

From now on, please do not communicate with other participants. If you have a question concerning the experiment, please raise your hand! We will come to your place and answer your question privately. If you do not comply with these rules, we will have to exclude you from the experiment and all payments.

In the experiment, we will use ET (“Experiment-Taler”) as the currency. At the end of the experiment, your payoff will be converted into Euros and will be paid out in cash. The exchange rate is  $150 \text{ ET} = 1 \text{ Euro}$ . In the experiment, all amounts in ET are rounded to whole numbers.

The experiment consists of several parts. The payoff of only one of these parts will be paid out at the end of the experiment. When the experiment is finished, a die will be used to determine which part will be used for payment. The following instructions refer to the first part of the experiment. After the first part is finished you will receive new instructions.

In this part all participants receive an initial endowment. Half of the participants receive an initial endowment of 800 ET, the other half receive an initial endowment of 200 ET. It will be determined by chance which participant receives which initial endowment.

This part consists of 20 rounds. In each round pairs are formed randomly, each pair consisting of participant A and participant B. It is guaranteed that you do not interact with the same participant in two consecutive rounds. The roles A and B within the pair are assigned randomly in every round. The identity of the participant you are interacting with is secret, and no other participant will be informed about your identity. In this sense, your decisions are anonymous.

Every round proceeds as follows:

- At the beginning of the round both participants are informed about their roles (A or B), the current round (1-20), their own current wealth and the current wealth of the other participant.
- Then participant A decides how much of his/her wealth he/she wants to send to participant B.
- The amount sent by participant A is multiplied by 1.2. This means participant B not only receives the amount sent, but 120 % of the amount sent ( $1.2 \times \text{amount sent}$ ).
- Then participant B decides how much he/she sends back to participant A. He/she must send back at least 90 % of the amount sent ( $0.9 \times \text{amount sent}$ ). The upper limit for the amount sent back is the wealth of participant B.

After that the round is over. Wealth at the end of the round is calculated as follows:

- Participant A: Wealth at the end of the round = wealth at the beginning of the round - amount sent + amount sent back (at least  $0.9 \times \text{amount sent}$ )
- Participant B: Wealth at the end of the round = wealth at the beginning of the round +  $1.2 \times \text{amount sent}$  - amount sent back (at least  $0.9 \times \text{amount sent}$ )

Wealth at the beginning of a new round is equal to wealth at the end of the preceding round. The payment for this part in case it is selected is given by the wealth at the end of the last round of this part.

## 2.6.2 Additional Results

TABLE 2.2  
DETERMINANTS OF INDIVIDUAL SEND AND RETURN RATES (2-FACTORIAL  
SET OF INDEPENDENT INTERACTION EFFECTS)

Dependent Variable	Send Rate		Return Rate	
	Coefficients	(Std.Error)	Coefficients	(Std.Error)
Round	-0.012*	(0.005)	-0.011*	(0.005)
$W_{SDR}$	-0.375**	(0.076)	-0.239**	(0.072)
$W_{SDR}$ *Round	0.000	(0.003)	0.006	(0.003)
$W_{RSPR}$	-0.243**	(0.065)	0.010	(0.082)
$W_{RSPR}$ *Round	-0.005	(0.003)	0.004	(0.003)
$W_{RSPR}$ * $W_{SDR}$	0.069*	(0.035)	0.067	(0.037)
Treatment	-0.629**	(0.130)	0.075	(0.099)
Treatment*Round	0.013**	(0.002)	0.000	(0.002)
Treatment* $W_{SDR}$	0.308**	(0.065)	0.063	(0.047)
Treatment* $W_{RSPR}$	0.253**	(0.043)	-0.058	(0.062)
Run	0.120*	(0.056)	0.065	(0.063)
Run*Round	-0.003	(0.002)	-0.006**	(0.002)
Run* $W_{SDR}$	0.025	(0.033)	-0.027	(0.033)
Run* $W_{RSPR}$	-0.082*	(0.032)	0.076*	(0.032)
Run*Treatment	0.014	(0.022)	-0.059**	(0.021)
Send rate			0.293**	(0.074)
Send rate*Round			-0.003	(0.002)
Send rate* $W_{SDR}$			-0.013	(0.040)
Send rate* $W_{RSPR}$			-0.095*	(0.046)
Send rate*Treatment			0.039	(0.031)
Send rate*Run			-0.038	(0.027)
Constant	1.190**	(0.123)	0.330**	(0.120)
<i>Random Effects</i>				
Group Std.Dev.	0.171	(0.040)	0.000	(0.000)
Subject Std.Dev.	0.265	(0.019)	0.173	(0.012)
Residual Std.Dev.	0.272	(0.004)	0.243	(0.017)
No. of obs.	2554		2246	
Wald $\chi^2$	301.43		299.70	
Log-restricted likelihood	-500		-156	

Standard errors are given in parentheses. \* and \*\* denote significance on the 5% and 1%-level, respectively.

# Chapter 3

## Reference Points, Job Satisfaction and Performance

### 3.1 Introduction

There is a broad and growing economic literature stressing that reference points strongly affect how individuals evaluate economic outcomes. But while there are many theoretical contributions (Kahneman and Tversky, 1979; Gul, 1991; Köszegi and Rabin, 2006) and some laboratory studies, there is surprisingly little field evidence on this issue.<sup>1</sup>

This lack of research is particularly problematic for the question of optimal incentive design in firms. The main reason is that bonus payments as the typical key component of incentive schemes are necessarily uncertain and that employees form expectations about the bonus payment they will receive. Hence, the design of a bonus scheme may possibly have a strong impact on the formation of reference points. In turn, implications for the optimal design of incentive schemes are altered when reference points matter for the behavior of employees (compare de Meza and Webb, 2007 or Herweg et al., 2008 for recent theoretical contributions).

In this paper we therefore study the causes and effects of reference point formation empirically in the context of bonus plans. We use a unique combination

---

<sup>1</sup>A notable exception is for instance the study by Mas (2006) showing that police performance is sensitive to pay rises compared to reference points set by final offer arbitrations. The number of crimes cleared by arrest decreased significantly after arbitrators in a compensation dispute decided in a way unfavorable to the demand of the police force.

of administrative personnel data and survey data on managerial employees from a large multinational company in Germany and the US.

A key feature of our data set is that for each managerial employee in the company a supervisor has to decide on the percentage of a target bonus paid out under the constraint that the sum of all bonus payments to his direct subordinates is equal to the sum of their target bonuses. Hence, the 100% payout percentage generates a clear reference point for the evaluation of bonus payments. Moreover, while the computational logic of the plan is exactly the same in the US and Germany, German employees learn the target bonus and the payout percentage while US employees only learn the dollar amount leading to a much higher saliency of the reference point in Germany. A connection of survey, compensation and performance data on the individual level now enables us to assess incentive and satisfaction effects of reference point violations in both countries.

We find that negative deviations from the 100%-reference point lead to a significantly lower reported job satisfaction while positive deviations are not associated with a significant increase in satisfaction. Hence, these results are well in line with loss aversion postulating that negative deviations from a given reference point decrease utility to a stronger extent than positive deviations of the same size (see Kahneman and Tversky, 1979, Tversky and Kahneman, 1991 and Kahneman et al., 1991).

Using panel data on performance assessments, bonus payments and subsequent supervisor performance, we also find that violations of employees' reference points negatively affect future performance in the German subsidiary where the reference point is very salient. This negative effect is robust against alternative specifications of reference points, evaluation behavior and team composition.

Our results also add to the literature on social comparisons as the payout percentage reveals information on the relative position of an employee within the relevant department.<sup>2</sup> Hence, they provide field evidence on the fair-wage effort hypothesis proposed by Akerlof and Yellen (1990).<sup>3</sup> Moreover, they are well in line with theories of inequality aversion (Bolton and Ockenfels, 2000 and Fehr and Schmidt, 1999)

---

<sup>2</sup>There is now broad evidence showing that subjective well-being is highly related to relative wealth positions (see for instance Frank and Sunstein, 2000 or Clark et al., 2007).

<sup>3</sup>While there are now numerous experimental studies on the relationship between fair wages and effort (see for instance Fehr et al., 1993 or Fehr et al., 1997), there is still not much field evidence as it is typically difficult to assess employee performance.

stating that utility does not depend solely on absolute payoffs, but also on relative payoff comparisons to a reference group.

The paper proceeds as follows: Section 3.2 describes the survey and the compensation data; section 3.3 presents the results for reported job satisfaction and performance. In section 3.4, we provide a brief discussion of our results and conclude.

## 3.2 The Data

### 3.2.1 Survey, Compensation and Performance Data

We use survey and compensation data from managers of a multinational company (>100,000 employees) operating in diverse business fields. In cooperation with the company, we conducted a survey among the managerial staff in Germany (autumn 2007) and the United States (summer 2008). While the system of variable compensation is the same for all managers in the sample, handling of the system differs in subtle details that are nevertheless decisive for the formation of reference points.

As the survey was part of a larger study, managerial employees were asked some 60 questions about job satisfaction and motivation, workplace characteristics, and personal preferences on incentive schemes. Together with the survey, we collected personal data about demographic characteristics, department affiliations, performance evaluations and compensation over the years 2004-2006 (Germany) and 2004-2007 (US). The technical environment allowed us to connect the background data with survey answers in a way that guaranteed anonymity of the participants. Altogether 4,997 executives took part in the survey (3,122 in Germany and 1,875 in the US).<sup>4</sup>

### 3.2.2 Characteristics of the Bonus System

At the end of every business year, individual bonus budgets are determined by linking company performance, divisional performance, hierarchical level and fixed salary of a manager. Each supervisor receives the sum of individual bonus budgets

---

<sup>4</sup>In the remainder of the paper, we use only data from individuals for which it was possible to connect survey answers with demographic and compensation data. Comparing participating and non-participating managers, we do not find strong differences in background or compensation variables that would limit the representativeness of the sample with respect to the survey answers.

TABLE 3.1  
DISTRIBUTION OF PERFORMANCE RATINGS

	GER 2006	US 2007	Recommended
‘Excellent’	0.6	2.0	$\leq 5\%$
‘Above Average’	22.4	33.9	$\leq 25\%$
‘Average’	74.1	62.7	$\sim 60\%$
‘Below Average’	2.8	1.4	$\leq 10\%^*$
‘Inadequate’	0.1	n/a	

\*Recommended share refers to the sum of ‘Below Average’ and ‘Inadequate’ employees.

to distribute among the managers in her department. Then, she allocates personal payout percentages (markups on or reduction below the 100% bonus budget).

Allocation decisions are constrained by the individual performance evaluation of a person. Each managerial employee is rated on a five point performance scale (either ‘Excellent’, ‘Above Average’, ‘Average’, ‘Below Average’, or ‘Inadequate’). For each grade, payout percentage intervals are defined. For example, a person rated ‘Average’ (the large majority of employees) should be assigned between 80% - 110% of her individual bonus budget. If a supervisor evaluates all her managers with ‘Average’ ratings, she is able to allocate every person the actual budget of 100%. However, the interrelation between performance ratings and payout percentages makes bonus assignments a constant-sum game. In the moment a supervisor differentiates in ratings, externalities are created by the budget: If one manager receives a bonus payment larger than 100% of her budget, another manager must necessarily receive a bonus below 100%.

Bonus payments of German and US managers are roughly comparable during the period of analysis. Average bonus payments of German managers account for some some 21,800 USD in 2006 while US managers received some 22,500 USD in 2007 (the years relevant for the survey). Furthermore, average bonuses remain largely constant over the years in the sample.<sup>5</sup> In particular, regular bonuses have substantial economic importance for the managers, as they account for some 20% of fixed salaries on average.

Given the connection between ratings and bonuses, an important aspect is the practice of performance evaluation. Table 3.1 lists the proportions of the perfor-

---

<sup>5</sup>It is important, however, that there is a substantial spread of bonus payments between managers on different hierarchy levels.

mance grades in the year relevant for our survey. The actual shares are contrasted to the recommended distribution provided by the company.<sup>6</sup>

We find the typical ratings biases often discussed in the literature on subjective performance evaluations. There is the general tendency to rate employees too positive (the so-called leniency biases) as well to compress ratings (centrality bias).<sup>7</sup> In Germany, differentiation is very weak, as the large majority of managers receives the medium rating ‘Average’. There is only a very small share of excellent ratings or grades ‘Below Average’ (in sum less than 5%). Supervisors in the US on average do not differentiate strongly either. But they assign the highest performance grades significantly more often and the lower-than-average ratings significantly less often than their German counterparts ( $p < .001$  and  $p = .001$ , two-sided  $\chi^2$ -tests).

FIGURE 3.1  
DISTRIBUTION OF AVERAGE BONUS PAYOUT PERCENTAGES

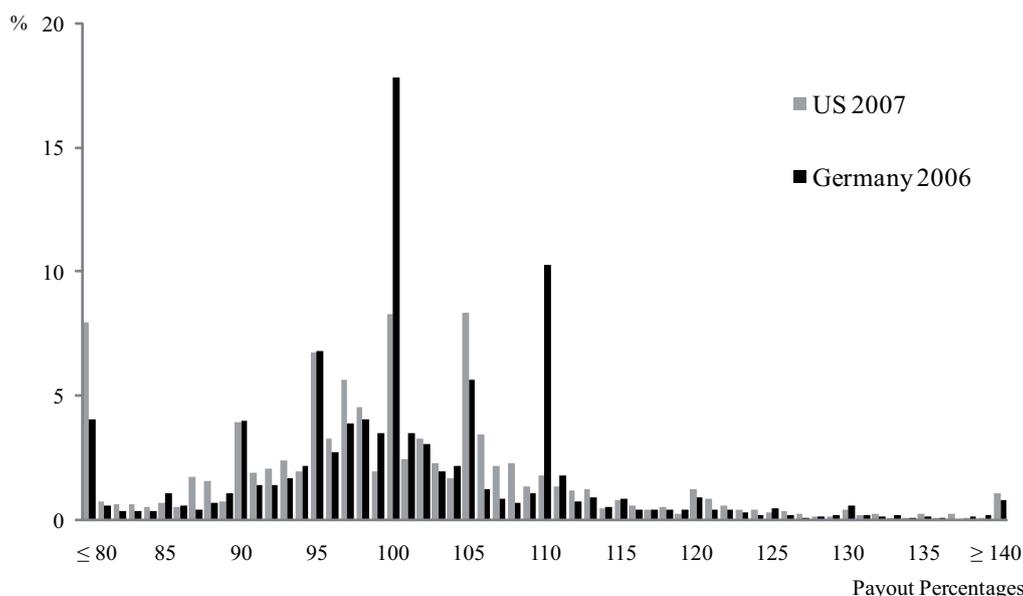


Figure 3.1 shows the distribution of bonus payout percentages in Germany and the US. Although the bonus system enables supervisors to differentiate strongly between employees (theoretically it is possible to assign between 0% and 160% of the bonus budget), payout percentages are compressed. The large majority of payout

<sup>6</sup>The company explicitly states that the grading distribution is not a forced distribution but should rather give supervisors a guideline for their appraisals.

<sup>7</sup>See Prendergast (1999) for a survey. There is also evidence suggesting that supervisor evaluation biases also influence subordinates’ subsequent performance (see the recent study by Bol, 2008).

percentages lies close to 100% of the budget. Moreover, the exact bonus budget is a much stronger focal point for German supervisors: Here, 17.8% of the managers receive exactly 100% payout compared to 8.3% in the US (significant with  $p < .001$ , two-sided  $\chi^2$ -test).

This observation can be related to a small but important difference in system handling, namely that the mode of communication differs across countries. When receiving their yearly bonus letters, managers in Germany explicitly learn their personal payout percentage together with the information about their bonus payments. In contrast, US managers only get to know the absolute amounts of their bonus payments without explicitly knowing their bonus budgets.<sup>8</sup> We expect the variation in system handling to have different effects on the emergence of the reference point, as we will explore in the next section.

Finally, in addition to regular bonus payments, supervisors can allocate individual spot bonuses to employees, for instance for special achievements or exceptional performance in particular projects. However, there exist no formal allocation rules, and the practice of spot bonus allocations differs strongly between supervisors. On average, supervisors assign spot bonus payments to 30.5% (25.5%) of their managers, however, the variation in shares is high in both countries. Compared to regular bonus payments, the economic relevance of these payments is low: The sum of spot bonuses equals some 3% of the total bonus budget; on average spot bonuses account for only 1.5 - 2.5% of the respective managers' yearly income.

### 3.3 Results

We first study the importance of reference point formation in Germany and in the US by investigating the effect of payout percentages on job satisfaction. In particular, we focus on the effects of deviations from the conjectured 100% reference point on individual job satisfaction. In the next step we then study how the frequency of violations of subordinates' reference points affects subsequent performance of supervisors using panel data on evaluations and compensation.

---

<sup>8</sup>In principal, US managers may infer their payout percentage. But as the bonus budget is computed based on a complex formula including divisional and company performance this requires substantial effort. Open survey answers of US managers convey that this lack of information is seen as a problematic aspect of the system. Example comments include the statements: "The letter that comes with the award should clearly spell out your rating", "At a minimum, the employee should be told what % was applied", "Most people have no idea if their award is correct".

### 3.3.1 Reference Points and Job Satisfaction

Our proxy for job satisfaction is the degree of approval to the statement: “I am very satisfied with my job”. Subjects could respond on a scale from 1 (totally disagree) to 7 (fully agree), so that increasing values of the score reflect higher satisfaction. The distributions of answers in the US and Germany are rather similar as table 3.2 shows. In all model specifications we use a unit normal transformation of the satisfaction score in line with Freeman (1978) by subtracting the sample mean from a person’s score and dividing the result by the standard deviation of the sample.

TABLE 3.2  
PROPORTIONS OF ANSWERS TO THE QUESTION “I AM VERY SATISFIED WITH MY JOB”.

	GER 2006	US 2007
1 - totally disagree	1.0	1.1
2	3.0	3.9
3	6.0	6.6
4	10.2	13.9
5	27.9	27.5
6	40.0	34.4
7 - fully agree	11.9	12.7
Mean	5.285	5.242
Std.Dev.	1.252	1.253

In the first step, we estimate ordered probit regressions with job satisfaction as the dependent variable for the years 2006 (Germany) and 2007 (US) in which the surveys took place. Tables 3.3 and 3.4 list alternative specifications for the satisfaction models. We control for age, gender, firm tenure, hierarchical level and division. Furthermore, in model 1 we also include dummies for each performance grade with the ‘Average’ category as a reference group.<sup>9</sup>

First, note that in neither country we find a significant positive impact of the absolute salary or the salary increase on self-reported satisfaction.<sup>10</sup>

Furthermore, – despite their substantial economic relevance – absolute bonus payments have no significant influence when we control for the performance grade

<sup>9</sup>The results reported in the following remain identical if controls for evaluation behavior of the respective supervisors are also included in the regression models.

<sup>10</sup>In some specifications for the US managers, the coefficient of the fixed salary becomes even negative and significant. Additional calculations show, however, that this result can be attributed to the positive correlation between hierarchy levels and job satisfaction. Once we control for the correlation between hierarchy levels and fixed salaries, the effect of the latter disappears.

of a manager.<sup>11</sup> This finding is in line with empirical studies showing little or no effect of absolute wages on job satisfaction (see Clark and Oswald, 1996 and Clark, 1999).

By contrast, the performance rating significantly and robustly influences job satisfaction. For both countries, regression coefficients have the expected signs: Better-than-average ratings are positively associated with satisfaction while the opposite is true for lower-than-average ratings. One reason for their high importance is that they not only influence short term bonus payments but also long term career opportunities. In the data set, we find a positive effect of the performance rating on the probability of a future promotion.

Unlike absolute bonus payments, spot bonuses are significantly associated with a higher satisfaction score in Germany.<sup>12</sup> In the US, the coefficient of spot bonuses is also positive but significant only in some specifications. A possible explanation is that in the US the spot bonus is more similar in nature to the regular bonus payment as employees typically do not infer the computational logic of the regular bonus.

In the next step, we focus on the role of payout percentages and the 100% budget threshold. In specification 2 we first include the bonus payout percentage.<sup>13</sup> Indeed, we find payout percentages to be positively and significantly correlated with reported satisfaction. However, following the literature on the effect of reference point violations on utility, we expect that the relation of payout percentages and satisfaction is non-linear. At this point, we exploit the fact that the system provides the clear reference point of 100% bonus budget for each individual. Therefore, in model 3 we split payout percentages into those above and those below 100% bonus budget and allow for the possibility that the effect of the bonus payout percentage has different slope above than below the reference point. This model is defined as

$$JobSatisfaction_i = \alpha + \beta \cdot X_i + \gamma \cdot (z_i - 100) \cdot I_{z_i > 100\%} + \delta \cdot (100 - z_i) \cdot I_{z_i < 100\%} + \varepsilon_i$$

with  $z_i$  being the payout percentage of individual  $i$  and  $I_{z_i > 100\%}$  ( $I_{z_i < 100\%}$ ) is a

---

<sup>11</sup>Please note that if performance ratings are excluded, absolute payments become significant in the German sample, see model 5, table 3.6 in the Appendix.

<sup>12</sup>This is in line with the results of Engellandt and Riphahn (2004) indicating a positive relation between surprise bonuses and subsequent overtime hours.

<sup>13</sup>As the performance grades determine the intervals from which the percentages can be selected we omit the performance grades in this specification.

TABLE 3.3  
DETERMINANTS OF JOB SATISFACTION – GERMANY (DEPENDENT VARIABLE:  
ADJUSTED SATISFACTION SCORE)

Country No. Model	GER 1 Ordered Probit	GER 2 Ordered Probit	GER 3 Ordered Probit	GER 4 Ordered Probit
Fixed Salary (000s)	0.005 [0.006]	0.003 [0.005]	0.004 [0.005]	0.002 [0.006]
Salary Increase (000s)	0.007 [0.005]	0.006 [0.005]	0.006 [0.005]	0.005 [0.006]
Spot Bonus (000s)	0.042** [0.022]	0.045** [0.022]	0.046** [0.022]	0.051* [0.028]
Performance Rating 'Excellent'	0.480* [0.276]			
Performance Rating 'Above Average'	0.279*** [0.064]			
Performance Rating 'Below Average'	-0.920*** [0.166]			
Performance Rating 'Inadequate'	-1.08 [0.732]			
Absolute Bonus Payout (000s)	-0.007 [0.010]			
Bonus Payout Percentage		0.013*** [0.002]		
Positive Deviation from Bonus Budget (=100%)			0.005 [0.003]	-0.002 [0.012]
Negative Deviation from Bonus Budget (=100%)			-0.023*** [0.004]	-0.022*** [0.006]
Sample Observations log-likelihood	All 2094 -3181	All 2099 -3198	All 2099 -3193	'Average' 1545 -2369

Standard errors are given in brackets. \*, \*\* and \*\*\* denote significance on the 10%, 5% and 1%-level, respectively. Demographic control variables include a manager's age and gender, total years of affiliation to the company and at a given hierarchy level, dummies for hierarchy levels and for promotion in the last year.

dummy variable taking the value 1 if the payout percentage is above 100% (below 100%). Hence,  $\gamma$  captures the effect of a positive deviation and  $\delta$  of a negative deviation from the 100% budget. If 100% indeed constitutes a reference point relative

TABLE 3.4  
DETERMINANTS OF JOB SATISFACTION – US (DEPENDENT VARIABLE:  
ADJUSTED SATISFACTION SCORE)

Country No. Model	US 1 Ordered Probit	US 2 Ordered Probit	US 3 Ordered Probit	US 4 Ordered Probit
Fixed Salary (000s)	-0.003 [0.003]	-0.003 [0.003]	-0.005* [0.003]	-0.005 [0.004]
Salary Increase (000s)	-0.019 [0.015]	-0.017 [0.015]	-0.02 [0.015]	0.007 [0.022]
Spot Bonus (000s)	0.033 [0.025]	0.041* [0.024]	0.050** [0.025]	0.051 [0.043]
Performance Rating 'Excellent'	0.551** [0.248]			
Performance Rating 'Above Average'	0.199*** [0.075]			
Performance Rating 'Below Average'	-1.213*** [0.295]			
Performance Rating 'Inadequate'				
Absolute Bonus Payout (000s)	-0.004 [0.006]			
Bonus Payout Percentage		0.006** [0.003]		
Positive Deviation from Bonus Budget (=100%)			0.006 [0.006]	0.003 [0.008]
Negative Deviation from Bonus Budget (=100%)			-0.021*** [0.006]	-0.002 [0.005]
Sample	All	All	All	'Average'
Observations	1008	1010	957	633
log-likelihood	-1553	-1569	-1481	-999

Standard errors are given in brackets. \*, \*\* and \*\*\* denote significance on the 10%, 5% and 1%-level, respectively. Demographic control variables include a manager's age and gender, total years of affiliation to the company and at a given hierarchy level, dummies for hierarchy levels and for promotion in the last year.

to which employees evaluate their bonus payment, we expect  $\gamma$  and  $\delta$  to be different from each other. In particular, loss aversion would imply that  $|\gamma| < |\delta|$ .

Indeed, model 3 shows that decreases below 100% have a significantly nega-

tive effect on satisfaction levels while increases above 100% do not increase self-reported satisfaction in both countries. However, given the strong negative impact of lower than average performance evaluations on satisfaction, we have to rule out the possibility that the presumed reference point effect is not in fact driven by ratings. We control for this possibility by estimating model 3 with a reduced sample including only employees who receive an ‘Average’ rating (model 4). In this specification, the reference point effect remains robust and highly significant for Germany. For the US, however, the effect disappears entirely. Recall that in Germany the employees directly learn the payout percentage and the 100% budget is therefore a very prominent reference point. Contrary, US managers do not orientate on their personal budgets, and negative deviations from the personal budget do not lower their satisfaction accordingly.

These conclusions remain identical in a number of robustness checks presented in tables 3.6 and 3.7 in the Appendix. First, we test if payout percentages influence job satisfaction per se when controlling for performance ratings (models 6 and 7). In model 6, we restrict our analysis again to managers with an ‘Average’ rating. In Germany, the effect of payout percentages is positive and significant whereas it is not significant in the US. Model 7 (calculated over the whole sample) captures the relation between ratings and payout percentages by including interaction terms.<sup>14</sup> Again, the positive relation between payout percentages and satisfaction remains significant in the German sample and disappears in the US sample.

Finally, the asymmetric effect of deviations from the 100% budget on job satisfaction remains equally robust in the German sample if dummies for ratings and the corresponding interaction effects are included (model 8).<sup>15</sup>

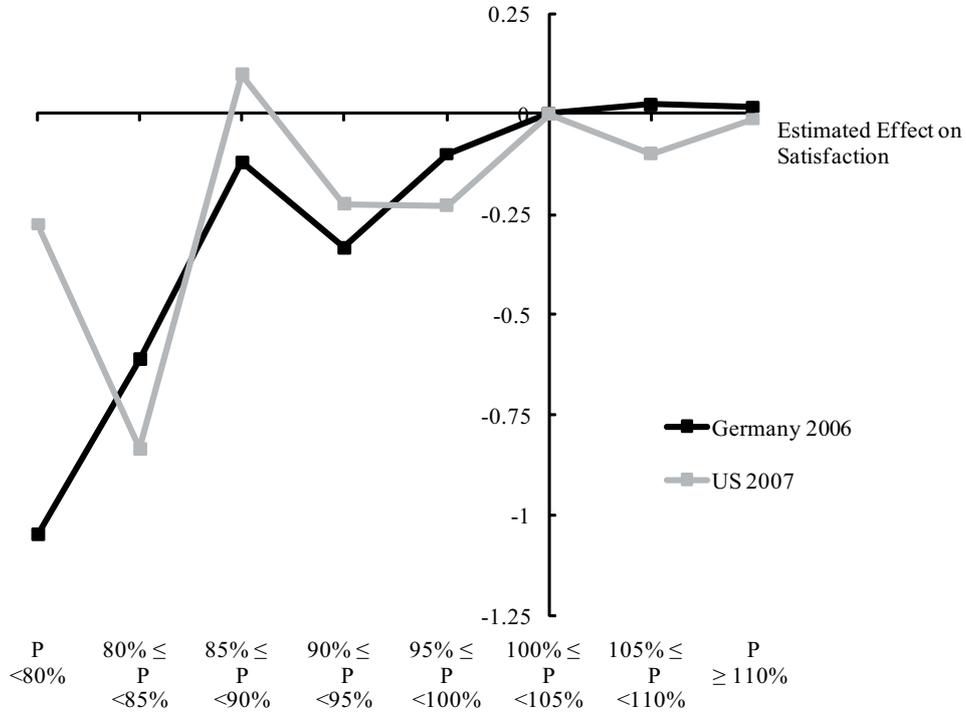
To further confirm and quantify the reference point violation effect, we calculate an additional OLS model with adjusted satisfaction as the dependent variable (see table 3.8 in the Appendix). In addition, we include fixed effects per supervisor to control for unobserved heterogeneity. Instead of payout percentage values we

---

<sup>14</sup>Average payout percentages per rating account for some 139%, 114%, 98%, 71% and 0% for the ratings ‘Excellent’, ‘Above Average’, ‘Average’, ‘Below Average’, and ‘Inadequate’ in the German sample. The corresponding values in the US sample are 127%, 109%, 95% and 60%. The lowest rating ‘Inadequate’ was not assigned by US supervisors.

<sup>15</sup>In model 8, the interaction variable Negative Deviation X ‘Above Average’ is included in the US sample. The reason is that for some managers with this rating, US supervisors discretionally assigned bonuses outside the relevant percentage intervals. However, less than 10 observations in our sample fall under this category.

FIGURE 3.2  
 ESTIMATED RELATION BETWEEN PAYOUT PERCENTAGE INTERVALS AND JOB SATISFACTION RELATIVE TO MANAGERS IN THE [100%,105%[ INTERVAL



The figure shows the estimated effect of payout percentage intervals on adjusted job satisfaction, controlling for compensation, performance rating, and demographic background of the managers (see table 3.8 in the Appendix). Mean predicted values per payout percentage interval are derived from a linear model with fixed effects for supervisors.

include dummies for percentage intervals. The reference group consists of managers with payout percentages from the [100%, 105%[ interval. In the German sample, all dummies below 100% have negative signs and are significant in most cases, indicating a lower satisfaction score compared to managers around the threshold. Dummies for intervals above 100% are insignificant and small in size. Corresponding to the expected low importance of the reference point (and the non-transparency of the payout percentage) there is no clear tendency in the US and most interval dummies are insignificant.

A graphical display of the results is shown in figure 3.2. For the graph, we compute the estimated adjusted satisfaction score derived from the regression model for each manager. We calculate for each percentage interval the average satisfaction

score and subtract the average score of the reference group, i.e. managers in the [100%, 105%[ interval. In the case of Germany, estimated scores exhibit roughly the kinked shape around the reference point implied by loss aversion: Managers with a higher percentage payout exhibit only a marginally higher satisfaction on average relative to the control group. In contrast, the further away payout percentage intervals are from the reference interval, the stronger is the detrimental effect on job satisfaction.<sup>16</sup> Furthermore, the described relation remains equally robust if the analysis is restricted to ‘Average’ managers. For the US managers, no clear tendency is observable.

### 3.3.2 Reference Points and Performance

So far, we only examined the effect of reference point violation on a subjective survey measure. But especially from an economic point of view it is important to study whether reference point violations have consequences for employee performance. To address incentive effects of the system handling, we use the performance ratings of a supervisor as a proxy measure for the performance of the respective team of her subordinates.

If reference point violations negatively affect performance we should expect that a higher frequency of employees with a payout percentage below 100% in one year lead to a reduced supervisor performance in the subsequent year. Due to the panel structure of our data, we can track the supervisors’ evaluation behavior and their own performance grades over several years. We therefore study the following specification

$$Rating_{it+1} = \alpha + \beta \cdot DevRef_{it} + \gamma \cdot X_{it} + a_i + \eta_t + \varepsilon_{it}$$

where  $X_{it}$  is a vector of individual background variables of a supervisor  $i$  in year  $t$ ,  $a_i$  are supervisor fixed effects and  $\eta_t$  are year dummies. We use a unit normal transformation of the performance rating as the dependent variable. Besides organizational affiliations and hierarchy levels, the background variables include compensation data (fixed salary and salary increase) and evaluation behavior in year  $t$ . For the latter, we use the shares of employees rated ‘Excellent’, ‘Above

---

<sup>16</sup>Managers in the [85%, 90%[ interval are outliers in the trend both for the German and the US sample. However, there is no systematic reason related to the compensation scheme or the practice of performance evaluation that could explain this observation.

Average’ and ‘Below Average’ by the supervisor.<sup>17</sup>

Our hypothesis is that the share of employees rated below the 100% reference point (measured by  $DevRef_{it}$ ) in year  $t$  by supervisor  $i$  has a negative effect on his performance in year  $t + 1$ . We control for unobserved supervisor heterogeneity by including fixed effects. In addition, it is important to account for the general evaluation behavior of the supervisor by including the relative shares of all performance grades he assigned to his subordinates. The reason for the latter is that this grading distribution determines the restrictions under which the supervisor can distribute the budget. By controlling for the distribution of grades we study the effect of variations in the choice to violate the 100% reference point for given budgeting restrictions. Table 3.5 reports the results for Germany and the United States.

In the models, we use the unit-normal transformation of a supervisor’s performance rating in  $t + 1$  as the dependent variable. Across countries we find results in line with the regressions on reported satisfaction. Controlling for the background and the evaluation behavior of the supervisor, reference point violations negatively affect supervisor performance in Germany. Thus, the more managers are reduced below 100% budget in a given year, the lower is estimated performance in the next year. This result suggests that reference point violation not only decreases job satisfaction, but also mitigates incentives. A given supervisor discourages her managers when violating reference points, and subsequently should take this into account when deciding about bonus assignments.

In the US data, we do not find a significant effect of reference point violations. As managers do not orientate strongly on the reference point of 100% budget, performance is not negatively affected. Supervisors can therefore achieve a stronger variation in payments among the managers in their departments. At the same time, the absence of the reference point enables supervisors to assign a higher share of ‘Above Average’ ratings to their subordinates, because budgetary pressure is lower. This corresponds to the observation that in the US some 11% more of the managers receive ‘Above Average’ than in Germany.<sup>18</sup>

Furthermore, the reported effect of reference point violation remains equally ro-

---

<sup>17</sup>We include no dummy for the rating ‘Inadequate’, because only a negligible share of supervisors used this rating throughout the years in our sample.

<sup>18</sup>Please note that managers rated ‘Average’ in the US receive about 3% less of their budgets than in Germany. Among supervisors who belong to the quartile with the strongest dispersion in performance ratings, this difference even increases to more than 7% on average.

TABLE 3.5  
INCENTIVE EFFECTS OF REFERENCE POINT VIOLATIONS (DEPENDENT  
VARIABLE: PERFORMANCE RATING OF SUPERVISOR IN SUBSEQUENT YEAR)

Dependent Variable	Adjusted Rating GER 1 Fixed Effects	Adjusted Rating GER 2 Fixed Effects	Adjusted Rating US 1 Fixed Effects
Share of Managers rated ‘Average’ and awarded below 100%	-0.825** [0.403]	-0.828** [0.368]	0.211 [0.278]
Fixed Salary	0.062* [0.035]	0.075 [0.157]	-0.031 [0.022]
Salary Increase	-0.009 [0.012]	0.089 [0.093]	0.025* [0.014]
Constant	3.990* [2.063]	-9.047 [14.795]	4.874 [3.314]
Sample	Full	Restricted	Full
Observations	367	86	297
Number of Supervisors	249	71	176
R-squared (within)	0.20	0.49	0.19

Standard errors are given in brackets. \*, \*\* and \*\*\* denote significance on the 10%, 5% and 1%-level, respectively. Demographic control variables include the age of the supervisor, total years of her affiliation to the company, correction for part-time employment and evaluation behavior measured by the share of respective performance ratings.

bust in Germany if we calculate model 1 with a reduced sample (model 2). Here, we include supervisors who assigned solely ‘Average’ ratings to their subordinates.<sup>19</sup> Due to the lack of differentiation in ratings, these supervisors are not restricted in bonus assignments by budgetary concerns. Thus, performance effects of reference point violations do not interact with evaluation behavior of the supervisor in this subsample. Again, the coefficient of the share of managers below 100% remains negative and highly significant.

<sup>19</sup>In the US sample, assigning only ‘Average’ ratings is much less frequent among supervisors. In fact, there are not enough observations in the sample to calculate model 2, as only 30 observations in the data set fall under this category in the respective years.

A final observation is that - in contrast to job satisfaction models - compensation variables have significant positive effects in some specifications. Here, the data suggests that higher fixed salaries (Germany) and stronger salary increases (United States) are correlated with higher performance in the subsequent year.

The effects of reference point violations in Germany (and the lack of an effect in the US) remain robust in alternative specifications. Table 3.9 in the Appendix lists models with alternative measures for reference point violation to validate our previous statements. First, we include the mean payout percentages per supervisor for ‘Average’ managers as an alternative measure for reference point violation (see model 3). The respective coefficient is positive and significant for Germany, indicating that a stronger reduction in payments of ‘Average’ managers decreases incentives accordingly. Second, our main result could be driven by general financial constraints of the supervisor rather than behavioral effects of reference point violation. Therefore, we control for absolute bonus budgets per supervisor (model 4). The less budget per capita a supervisor is able to distribute, the more managers have to be reduced below the threshold for a given ratings distribution. However, the inclusion of bonus budgets does not have a significant impact for both countries, and the effect of reference point violation remains robust for Germany.

Summarizing, our data provides evidence that reference point violation does not only affect employee perception but also performance. As in the case of satisfaction, our results highlight the impacts of differences in system handling: If orientation on the individual budget is fostered by the system, negative deviations from this reference point have detrimental effects. However, if the reference point is not salient, there is no negative effect on performance found if managers’ bonus payments are reduced below the individual budgets.

### 3.4 Discussion and Conclusions

We have conducted an empirical study in which we connect demographic, compensation and performance data with survey results to derive insights about the relevance of reference points for the design of bonus schemes. Comparing managerial employees in the United States and Germany, we find effects consistent with loss aversion when reference points are very salient. Negative deviations from a natural reference point induced by the bonus system significantly reduce self-reported job satisfaction levels, whereas positive deviations have little influence. Moreover, violations of subordinates' reference points also harm performance, as the supervisor's own ratings are reduced when they assign payout percentages below personal budgets to a higher share of their subordinates. In the US, system handling differs in a way that prevents the formation of reference points, as personal payout percentages are not explicitly communicated. Subsequently, the described detrimental effects of reference point violations are not observed.

It is important to note that the exhibited patterns are not only in line with loss aversion but also with theories of inequality aversion. In the present bonus system, the 100% reference point coincides with a 'social reference': When having a payout percentage below 100%, an employee knows that she receives a lower bonus payment than the average fellow employee in the same unit.

From our analysis, two main fields for further research can be derived in the context of bonus schemes. First, our results suggest that transparency may be a double-edged sword. On the one hand, transparency may increase procedural fairness perceptions of compensation schemes or wage changes (see, for example, the psychological studies by Schaubroeck et al., 2000 and Werner and Ones, 2000). But as we have shown, transparency can also be detrimental as it may trigger relative social comparisons and the formation of precise expectations which may eventually be disappointed.<sup>20</sup> At the same time, however, low transparency provides supervisors with more degrees of freedom when deciding about bonus payments to their employees. In our sample, US supervisors are able to differentiate more between their subordinates, apparently because they do not have to care so much about the violation of their subordinates' reference points.

Second, our study does not identify the reference point dependent concerns;

---

<sup>20</sup>This is also in line with experimental findings suggesting that a lack of transparency may dampen social comparison effects (see for instance Gehrig et al., 2007)

managers may compare their bonus relative to their individual bonus target or relative to their peer group, or to both. In fact, we believe that one reason for why we find strong reference point effects is that different reference points coincide: falling behind my individual target implies falling behind the peer group because the bonus not paid to me goes to my peers. However, it would be interesting to know the contribution of each of those effects for our results.

As our study shows, it is highly important to take behavioral effects into account when making recommendations for incentive plan design in practice. Small differences in handling may have large effects on the perception of the employees and the subsequent incentive effects. At the moment, the research on behavioral incentive design is only starting.

### 3.5 Appendix

TABLE 3.6  
DETERMINANTS OF JOB SATISFACTION (GERMANY) – ROBUSTNESS CHECKS

Country No. Model	GER 5 Ordered Probit	GER 6 Ordered Probit	GER 7 Ordered Probit	GER 8 Ordered Probit
Fixed Salary (000s)	-0.001 [0.006]	0.002 [0.006]	0.003 [0.005]	0.003 [0.005]
Salary Increase (000s)	0.009 [0.005]	0.006 [0.006]	0.006 [0.005]	0.006 [0.005]
Spot Bonus (000s)	0.058*** [0.021]	0.051* [0.028]	0.036* [0.022]	0.036* [0.022]
Performance Rating 'Excellent'			7.950* [4.127]	2.129* [1.214]
Performance Rating 'Above Average'			3.302*** [0.845]	0.399*** [0.106]
Performance Rating 'Below Average'			-0.454 [1.058]	-0.540 [0.426]
Performance Rating 'Inadequate'			0.472 [0.841]	
Absolute Bonus Payout (000s)	0.025*** [0.008]			
Bonus Payout Percentage		0.015*** [0.005]	0.015*** [0.005]	
Positive Deviation from Bonus Budget (=100%)				0.000 [0.012]
Negative Deviation from Bonus Budget (=100%)				-0.021*** [0.006]
'Excellent' X Percentage			-0.058** [0.029]	
'Above Average' X Percentage			-0.029*** [0.008]	
'Below Average' X Percentage			-0.001 [0.014]	
Positive Deviation X 'Excellent'				-0.043 [0.031]
Positive Deviation X 'Above Average'				-0.014 [0.013]
Negative Deviation X 'Below Average'				0.006 [0.015]
Negative Deviation X 'Inadequate'				0.010 [0.009]
Sample	All	'Average'	All	All
Observations	2099	1545	2094	2094
log-likelihood	-3213	-2370	-3172	-3171

Standard errors are given in brackets. \*, \*\* and \*\*\* denote significance on the 10%, 5% and 1%-level, respectively. Demographic control variables include a manager's age and gender, total years of affiliation to the company and at a given hierarchy level, dummies for hierarchy levels and for promotion in the last year.

TABLE 3.7  
DETERMINANTS OF JOB SATISFACTION (US) – ROBUSTNESS CHECKS

Country No. Model	US	US	US	US
	5	6	7	8
	Ordered Probit	Ordered Probit	Ordered Probit	Ordered Probit
Fixed Salary (000s)	-0.006*	-0.005	-0.006**	-0.006*
	[0.003]	[0.004]	[0.003]	[0.003]
Salary Increase (000s)	-0.016	0.007	-0.021	-0.020
	[0.015]	[0.022]	[0.015]	[0.015]
Spot Bonus (000s)	0.045*	0.065	0.033	0.039
	[0.024]	[0.044]	[0.025]	[0.025]
Performance Rating 'Excellent'			-2.882	0.455
			[2.061]	[0.809]
Performance Rating 'Above Average'			1.782**	0.171
			[0.694]	[0.133]
Performance Rating 'Below Average'			-2.114	-0.484
			[1.513]	[0.986]
Absolute Bonus Payout (000s)	0.007			
	[0.006]			
Bonus Payout Percentage		0.004	0.002	
		[0.008]	[0.004]	
Positive Deviation from Bonus Budget (=100%)				-0.003
				[0.025]
Negative Deviation from Bonus Budget (=100%)				-0.004
				[0.010]
'Excellent' X Percentage			0.027	
			[0.017]	
'Above Average' X Percentage			-0.015**	
			[0.007]	
'Below Average' X Percentage			0.016	
			[0.024]	
Positive Deviation X 'Excellent'				0.010
				[0.037]
Positive Deviation X 'Above Average'				-0.001
				[0.027]
Negative Deviation X 'Below Average'				-0.014
				[0.026]
Negative Deviation X 'Above Average'				0.058
				[0.082]
Sample	All	'Average'	All	All
Observations	1010	599	1008	957
log-likelihood	-1572	-947	-1548	-1475

Standard errors are given in brackets. \*, \*\* and \*\*\* denote significance on the 10%, 5% and 1%-level, respectively. Demographic control variables include a manager's age and gender, total years of affiliation to the company and at a given hierarchy level, dummies for hierarchy levels and for promotion in the last year.

TABLE 3.8  
DETERMINANTS OF JOB SATISFACTION - REFERENCE POINT EFFECTS

Country	GER	US
No.	9	9
Model	OLS	OLS
Fixed Salary (000s)	0.004 [0.006]	-0.009** [0.004]
Salary Increase (000s)	0.000 [0.006]	-0.031 [0.020]
Spot Bonus (000s)	0.029 [0.026]	-0.037 [0.039]
Performance Rating 'Excellent'	0.112 [0.323]	0.875** [0.422]
Performance Rating 'Above Average'	-0.041 [0.136]	0.236 [0.218]
Performance Rating 'Below Average'	-0.162 [0.411]	-1.327* [0.760]
Performance Rating 'Inadequate'	-0.24 [0.945]	
Payout <80%	-0.852* [0.466]	-0.030 [0.623]
80% ≤ Payout <85%	-0.574*** [0.173]	-0.724* [0.410]
85% ≤ Payout <90%	-0.087 [0.156]	0.128 [0.351]
90% ≤ Payout <95%	-0.304*** [0.104]	-0.172 [0.250]
95% ≤ Payout <100%	-0.087 [0.083]	-0.237 [0.197]
105% ≤ Payout <110%	0.025 [0.106]	-0.203 [0.216]
Payout ≥ 110%	0.069 [0.149]	-0.134 [0.263]
Observations	2093	957
Number of Supervisors	707	535
R-squared (within)	0.06	0.09

The models include fixed effects calculated on the level of supervisors. Standard errors are given in brackets. \*, \*\* and \*\*\* denote significance on the 10%, 5% and 1%-level, respectively. Demographic control variables include a manager's age and gender, total years of affiliation to the company and at a given hierarchy level as well as dummies for hierarchy levels and promotion in the last year.

TABLE 3.9  
INCENTIVE EFFECTS OF REFERENCE POINT VIOLATIONS – ROBUSTNESS  
CHECKS

Dependent Variable	Adjusted Rating GER 3 Fixed Effects	Adjusted Rating GER 4 Fixed Effects	Adjusted Rating US 3 Fixed Effects	Adjusted Rating US 4 Fixed Effects
Share of Managers rated 'Average' below 100%		-0.814** [0.404]		0.222 [0.280]
Fixed Salary	0.064* [0.035]	0.064* [0.036]	-0.031 [0.024]	-0.031 [0.023]
Salary Increase	-0.009 [0.012]	-0.009 [0.012]	0.032** [0.015]	0.024 [0.015]
Mean Payout Percentage for 'Average' Managers	0.056** [0.026]		-0.002 [0.006]	
Bonus Budget per Employee		-0.033 [0.049]		0.013 [0.026]
Constant	-7.987 [14.132]	4.369** [2.143]	3.570 [3.628]	4.687 [3.346]
Observations	365	367	270	297
Number of Supervisors	248	249	164	176
R-squared (within)	0.20	0.20	0.23	0.19

Standard errors are given in brackets. \*, \*\* and \*\*\* denote significance on the 10%, 5% and 1%-level, respectively. Demographic control variables include the age of the supervisor, total years of her affiliation to the company, correction for part-time employment and evaluation behavior measured by the share of respective performance ratings.

# Chapter 4

## Social Comparisons in a Real-Effort Experiment

### 4.1 Introduction

When designing compensation schemes, the natural question from a company's perspective is how to differentiate between wages of employees with different productivity. Standard theory implies that in equilibrium each worker should be paid her marginal product. Frank (1984*a*) shows that the distribution of wages observed in the real world is less dispersed – more compressed – than the distribution of worker's marginal products. In a formalization of this argument, Frank (1984*b*) shows that the existence of rank-dependent preferences may lead to more compressed wages.

However, in this model, a worker's productivity is fixed and given exogenously. It seems rather plausible to assume that workers' efforts and performance depend on their relative wage, although the sign of this relationship is not clear per se. On the one hand, low relative (piece-rate) wages might increase worker's efforts as a mean to overcome the dispersion of income. On the other hand, low relative pay might crowd out a worker's motivation.

In this experimental study, the focus is on the effect of social comparisons with respect to discriminatory wage increases and decreases. Our aim is to disentangle the direct performance effect of wage changes from the indirect effect induced by changes in relative positions.

Participants in our experiment are hired to carry out simple but laborious typing tasks. We allow for shirking by paying only for quantity, but not for (not imme-

diately observable) quality of work. While identical wages are paid for a first task, we introduce wage differentials in a similar second task. Our design tests both a fixed and in a piece-rate wage environment, and controls for the effect of social comparisons by varying information about co-workers' wages.

The importance of social comparisons is reflected in behavioral responses to wage changes in our treatment with piece-rates. Information about peer wages triggers performance differences between high wage and low wage subjects that are absent when wages are identical or peer information is not provided. This effect is mainly driven by negative responses of participants with lower wages who become more likely to cheat in the second task. Nevertheless, there is some evidence that wage increases trigger positive adjustments among high wage subjects, as errors attributable to inaccuracy significantly decrease in the second task. In general, effects in the fixed-wage environment – as compared to piece-rate wages – are qualitatively similar but largely mitigated and statistically insignificant.

Section 4.2 gives a short overview on the related empirical and experimental literature. In Section 4.3, we present the details of our experimental design and introduce hypotheses about expected behavior. Our results are reported in Section 4.4, and we briefly conclude in Section 4.5.

## 4.2 Related Literature

In the laboratory, agency problems of employment relationships are mostly studied in the framework of the ‘gift exchange’ game introduced by Fehr et al. (1993).<sup>1</sup> In this game, the employer sets a fixed wage, and the employee makes a costly ‘effort’ choice. While strict rationality implies minimum wages and efforts, a substantial share of employers pay higher-than-minimum wages in this kind of setting, and employees reciprocate by choosing higher-than-minimum effort levels.

Some variations of the gift exchange game are devoted to the effects of social comparisons between multiple workers – the results, however, are ambiguous. A number of studies suggest an influence of wage differentials and social comparison on effort exertion: In the study by Clark et al. (forthcoming), workers were informed about the wages offered to other workers prior to their effort choices. The authors found evidence for the impact of relative income, as effort choices increased with the

---

<sup>1</sup>See Gächter and Fehr (2001) for a survey of numerous variations.

rank of a worker's wage in the population. Rivas and Sutter (2008) conducted an experiment in which one principal was matched with four agents of differing productivity. Controlling for a worker's own wage, effort decreased with rising inequality among the agents. Finally, in the repeated game studied by Abeler et al. (forthcoming), effort had to be provided before wages were set. Here, efforts diminished over time when the employer had no possibility to discriminate in wages between workers choosing different effort levels.

Other gift exchange experiments found no robust correlation between peer payments and effort levels. Charness and Kuhn (2007) conducted a study in which a firm was matched with two workers of differing productivity. Here, the information about the co-worker's wage was only of minor importance. In a gift exchange game with one principal and two agents, Gächter et al. (2008) found that co-workers' wages had an influence only if subjects were additionally provided information about co-workers' corresponding effort choices.

There are a few experimental studies which try to bridge the gap between the laboratory and the real work environment and let participants perform *real* tasks rather than making theoretical "effort" choices. In a psychological real-effort field experiment, Greenberg (1987) showed that fair (performance-based) and unfair (room number-based) procedures to assign wages yield different willingness to accept a low wage. In a further study, Greenberg (1988) reported that workers assigned to offices associated with lower (higher) status decreased (increased) their performance. Burchett and Willoughby (2004) compared performance of experimental subjects under alternative initial reward systems (flat payment versus high and low piece rates). It turned out that different systems resulted in performance differences only if subjects were informed about the existence of these differences. Hennig-Schmidt et al. (forthcoming) measured the effects of pay increases of differing sizes for real effort tasks conducted in the lab and in the field and found no direct effect on performance. They found positive effects on efforts only when workers were informed about the surplus their activity created for the employer.

Several studies provide evidence for the impact of relative wage positions for survey measures of job satisfaction. While the influence of absolute wages seemed to be limited (Clark, 1999), income increases of peers in comparable positions were associated with significant declines in reported job satisfaction (Clark and Oswald, 1996).<sup>2</sup> In a survey study among US colleges and university faculties, Pfeffer and

---

<sup>2</sup>Please note that the reference group appears to have a decisive impact. In a recent study, Clark

Langton (1993) found – consistent with rank preferences – that satisfaction was positively correlated with relative positions, and average satisfaction increased with lower salary variation. Evidence for the incentive effect of relative wages is provided by Torgler et al. (2008) who showed that German soccer players’ individual performances were sensitive to their income positions within their teams.

Finally, a strain of literature investigates behavioral effects of wage changes. In a field experiment conducted by Fehr and Goette (2007), an exogenous wage increase had a positive impact on labor supply of bicycle messengers. Bellemare and Shearer (2007) found that a temporary wage increase had positive effects on productivity of workers in a tree-planting company. However, several laboratory and field experiments showed that positive responses to wage increases are weaker than negative responses to wage decreases (see Gneezy and List, 2006; Kube et al., 2007). Moreover, psychological experiments provide evidence that wage freezes or wage cuts result in negative reactions of employees, such as declining employee satisfaction (Schaubroeck et al., 2000) or an increase of employee theft (Greenberg, 1990).

### 4.3 Experimental Design and Hypotheses

The experiment consisted of two identical working tasks which both lasted 20 minutes.<sup>3</sup> Participants were provided with a staple of questionnaire forms filled in by participants in an unrelated class room experiment. The subjects were asked to type in the 15 decimal numbers found on each form into an input mask on the computer screen.<sup>4</sup> Participants were told that the data from these questionnaires were required in electronic form for research purposes. It was known that the correctness of entries was not checked prior to payments, giving rise to the possibility to shirk by filling in arbitrary values.<sup>5</sup> This specific working task was chosen for two reasons: First, it allowed the precise measurement of working performance with respect to both quantity and quality. Second, demands to subjects were not different

---

et al. (2009) found that co-worker wages as signals for future earnings were positively related to job satisfaction measures.

<sup>3</sup>Instructions translated from German can be found in Appendix 4.6.1.

<sup>4</sup>See Figure 4.3 in Appendix 4.6.1 for a screenshot.

<sup>5</sup>We limited the possibilities for shirking in two dimensions: First, only 50 forms were distributed in each task, therefore limiting achievable payments in the piece-rate treatment. In addition, the software required subjects to fill in a number into each field of the input mask to be able to continue to the next input mask.

to usual economic experiments: reading text and filling in a form at the computer screen.

At the beginning of each session, subjects were randomly assigned to either group A or group B, and received instructions in a red (A) or blue (B) envelope. For the first task all subjects were paid identical wages. In the second task, each participant in group A received a 60% wage increase, while the wages of participants in group B were cut by the same share.<sup>6</sup> All participants were assigned the same set of questionnaires.

To systematically study the effects of these wage changes, we implemented a 2x2 design. In the *private information* dimension, participants were only informed about their own wage. In the *public information* condition, A (B) participants were informed that they were assigned to a B (A) participant working on the exact same set of questionnaires, and were told the other participant's wage. This allows us to isolate the impact of social comparisons from direct effects of wage in- and decreases. In the second design dimension, participants either received a *fixed wage* for the task, or a *piece-rate wage*, i.e. a wage depending linearly on the number of forms filled in. Thus, while under a piece-rate wage participants could shirk in quality, but not in quantity, none of these two effort dimensions was contractible in the fixed wage conditions. Our analysis will include all three performance measures: quantity (number of completed questionnaires), quality (share of correct questionnaires) dimension, as well as net output, i.e. the number of correct questionnaires.

If effort is costly, and participants are egoistic and view the work tasks as a one-shot interaction with the experimenters, then we should observe no effort under fixed wages and minimal effort (full quantity with zero quality) under piece-rate wages. In any case, participants should not produce any usable output, and there should be no treatment effects (Null hypothesis).

However, if – as intended – participants understand the experiment as a real world working task, and they believe to be in a repeated game context and/or behave reciprocal to provided wages, we would expect them to produce some usable output (hypothesis 1).<sup>7</sup> To organize behavioral hypotheses with respect to wage

---

<sup>6</sup>It is important to note that wage changes were not related to any performance measure, and subjects had no possibility to assess their own performance relative to the others.

<sup>7</sup>In our setting, experimental subjects face a multi-tasking problem, because they have to allocate effort between working quantity (number of completed questionnaires) and working quality (avoidance of mistaken entries). For theoretical considerations on multi-tasking problems see Lazear (1986), Holmstrom and Milgrom (1991) and, for a survey of the literature, Prendergast

changes in this case, we distinguish between two possible effects of our treatment variations: the effect of wage changes under private wage information, and the effect of social comparison when information about peer wages is provided.

Under private information and in a fixed wage environment, where no work dimension is contractible, reciprocal reactions to a wage change would predict increased (decreased) effort and output after a wage increase (decrease) (hypothesis 2a). Under straightforward piece-rate wages, the same effect would be predicted by reciprocity or by a standard assumption of increasing marginal costs of effort (hypothesis 2b). However, the asymmetry in observability of quantity and quality complicates a prediction, as a wage change may additionally induce a shift of effort from the unobservable quality dimension to the observable quantity dimension, specifically for a wage cut (hypothesis 2c).

With information on peer wages, rank-dependent or otherwise distribution-based preferences might interact with the direct effects of wage changes. In general, the effects described above might be fortified (hypothesis 3a, e.g. crowding out of motivation after wage cuts) or mitigated (hypothesis 3b, e.g. additional incentives to improve relative income position after wage cut). However, given existing empirical evidence about the impacts of relative ranks, we would expect that specifically the negative effort effects of wage cuts are stronger when information about wages of other, higher paid workers is provided.

The experimental sessions took place in February and April 2006 in the Cologne Laboratory for Economic Research. We conducted four sessions, one for each cell of our 2x2 design. Altogether 128 subjects participated, most of them with a major in Economics, Business Administration or related fields.<sup>8</sup> Subjects were recruited using the online recruitment system ORSEE (Greiner, 2004). In the invitation e-mail, the experiment was announced as a ‘work experiment’, indicating that subjects were recruited for a research-related real effort task rather than a usual decision making experiment. The computerized input mask used in the two real effort tasks was implemented utilizing the z-Tree software (Fischbacher, 2007). At the end of the experiment, subjects filled in a post-experimental questionnaire. Each session lasted

---

(1999).

<sup>8</sup>Two out of 128 subjects had to be excluded from the data set: One person accidentally received wrong instructions; the other person did not understand the compensation system in the experiment. In the post-experimental questionnaire, the person stated that he realized not until the course of the second task that piece-rates would be paid without controlling the correctness of the working input.

approximately one hour. Participants were paid out their accumulated earnings for both tasks privately and left the laboratory. Piece-rate treatments were conducted first, where participants received 0.20 Euros per completed form in the first task, and 0.08/0.32 Euros in the second task. To ensure comparability of overall financial incentives, wages in the fixed rate treatments were set equal to the average payoff in the piece-rate conditions. Specifically, fixed wages were 6.80 Euros in the first task, and 2.70/10.90 Euros in the second task. The overall average payoff was 16.10 Euros, including a show-up fee of 2.50 Euros.

## 4.4 Experimental Results

In the remainder of this paper, we use the following notations for our four experimental treatment cells:

- VAR Public for piece-rate wages and public wage information,
- VAR Private for piece-rate wages and private wage information,
- FIX Public for flat wages and public wage information, and
- FIX Private for flat wages and private wage information.

Figure 4.1 illustrates the development of average output over 1-minute intervals for both wage regimes, averaged over information and payment conditions. Examination immediately leads to a rejection of the minimal-effort Null hypothesis and confirms hypothesis 1: participants invest considerable effort and produce usable output. Under both wage schemes, average performance increases throughout task 1, but shrinks again roughly in the second half of task 2.<sup>9</sup> Hence, in our real effort experiment the implemented incentive schemes (fixed wages and piece-rates) largely yield identical output despite the differences in economic incentives for workers.

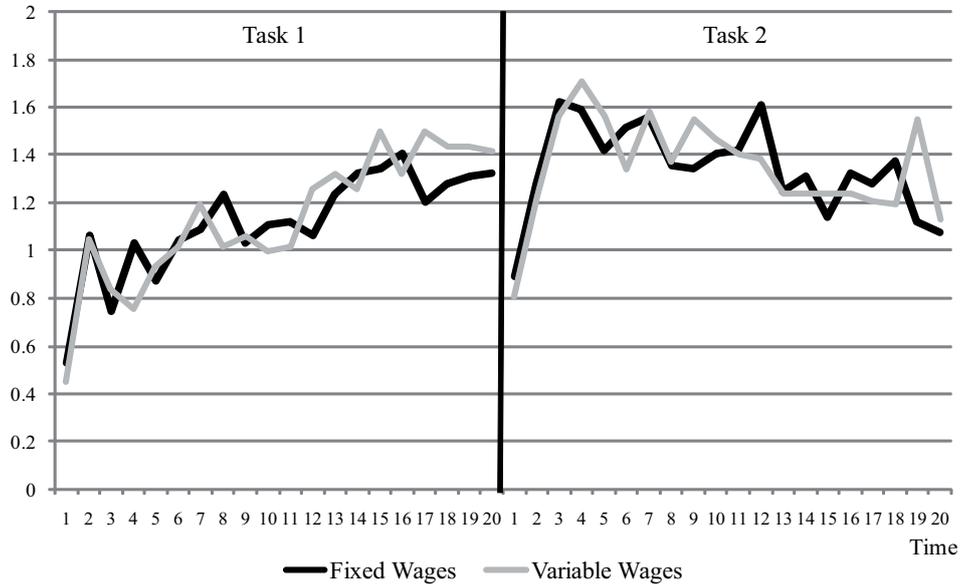
Table 4.1 lists mean and standard deviation values for the performance measures of all our experimental groups. On average, subjects complete between some 26 and 35 forms in the first task (between some 30 and 41 in the task 2).<sup>10</sup> Despite large

---

<sup>9</sup>The reason for the decrease is not that the majority of subjects reached the threshold of 50 forms. In fact, only 15 out of 126 subjects completed all forms in the second task.

<sup>10</sup>It is important to note that there is a strong variation between the subjects. The minimum and maximum values of correct forms accounted for 0 and 42 forms in task 1 (2 and 45 forms in task 2).

FIGURE 4.1  
AVERAGE OUTPUT PER ONE-MINUTE INTERVAL



incentives to cheat, output and working quality are generally high. Over treatment groups, the average share of correct questionnaires ranges between 66% and 88%. Moreover, there is evidence that subjects become remarkably better in solving the task, as average output increases in all experimental groups.<sup>11</sup> This performance increase is corroborated by non-parametric tests: comparing first- and second-task output on the individual level with two-sided Wilcoxon Matched Pairs Signed Ranks (WMPSR) tests yields  $p$ -values of  $p < .001$  in all experimental groups except low-paid subjects in VAR Public.<sup>12</sup>

Learning effects and the subsequent general increase in performance make a direct analysis of behavior after the wage change difficult, because skill differences and the speed of learning might possibly overlay the direct effects on motivation. Therefore, as a first step to isolate incentive effects, we compare not the absolute, but the relative performance increases. For all three performance measures we calculate subjects' individual performance changes in percent of their respective task 1 performance. Averages per treatment are depicted in Figure 4.2.

<sup>11</sup>In the post-experimental questionnaire, 66 out 126 subjects stated that their ability to fill in the values increased throughout the experiment.

<sup>12</sup>Comparing quantity and quality separately yields similar results. Whereas the increase in quantity is significant in 7 out of 8 treatment groups ( $p < .002$ , two-sided WMPSR tests), the quality increase is significant in 5 of those groups ( $p < .05$ , two-sided WMPSR tests).

TABLE 4.1  
AVERAGE WORKING PERFORMANCE PER TREATMENT

(a) Quantity		Task 1			Task 2	
		N	Mean	Std.Dev.	Mean	Std.Dev.
VAR Public	High	16	29.3	5.9	34.0	6.6
	Low	16	33.5	8.9	41.3	9.4
VAR Private	High	15	35.4	9.9	40.3	8.3
	Low	15	26.7	8.4	31.5	9.9
FIX Public	High	16	28.1	8.4	30.1	9.7
	Low	16	28.4	7.6	31.8	9.2
FIX Private	High	16	30.4	10.6	35.0	9.0
	Low	16	29.1	9.4	33.3	9.6

(b) Quality		Task 1			Task 2	
		N	Mean	Std.Dev.	Mean	Std.Dev.
VAR Public	High	16	74.3	14.4	82.5	18.7
	Low	16	75.3	13.2	69.5	17.5
VAR Private	High	15	69.5	26.3	71.2	28.8
	Low	15	67.5	19.8	74.9	23.4
FIX Public	High	16	75.7	11.2	88.3	6.1
	Low	16	74.6	9.4	79.5	16.0
FIX Private	High	16	75.3	11.3	81.2	11.6
	Low	16	77.1	10.5	81.3	10.2

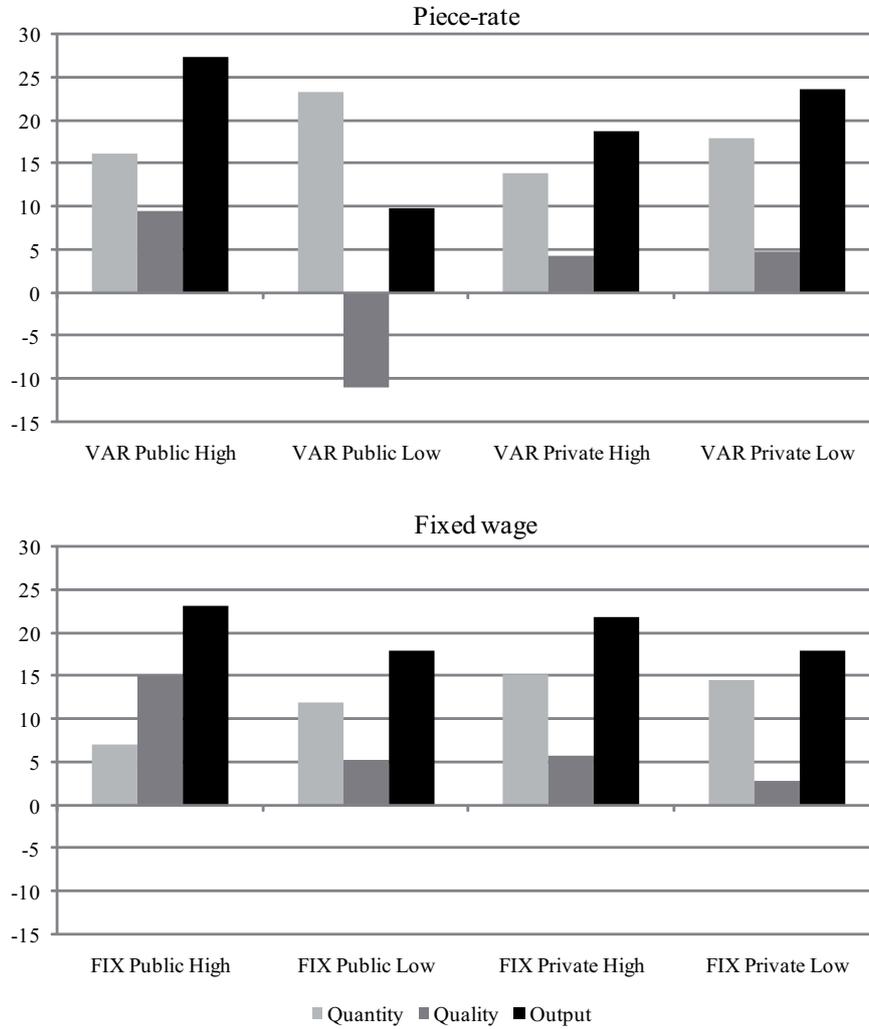
  

(c) Output		Task 1			Task 2	
		N	Mean	Std.Dev.	Mean	Std.Dev.
VAR Public	High	16	22.2	7.4	28.3	8.5
	Low	16	25.9	9.0	28.5	9.4
VAR Private	High	15	23.4	10.6	27.8	12.4
	Low	15	18.9	9.3	23.3	10.8
FIX Public	High	16	21.6	7.9	26.6	8.6
	Low	16	21.6	7.7	25.5	9.4
FIX Private	High	16	23.4	9.9	28.6	8.6
	Low	16	22.9	9.1	27.1	8.3

The first thing to note is that there are no distinct differences between high wage and low wage subjects under private information. Neither under piece-rates nor under fixed wages we find a significant difference if we compare the experimental groups statistically using two-sided Mann-Whitney-U (MWU) tests (all  $p$ -values are higher than  $p > .1$ ).<sup>13</sup> In particular, there is no evidence for hypotheses 2b and 2c

<sup>13</sup> $P$ -values of all two-sided MWU tests can be found in Table 4.5 in the Appendix.

FIGURE 4.2  
RELATIVE PERFORMANCE INCREASE PER TREATMENT IN %



that (dis)advantaged subjects increase (lower) output subsequently or shift effort from quality to quantity.

Contrary, high and low wage workers adjust performance differently if they are informed about co-worker wages under piece-rates. Here, low-wage subjects increase quantity stronger than high-wage subjects ( $p = .060$ , two-sided MWU test) while lacking behind with respect to quality increase ( $p = .019$ , two-sided MWU test). In fact, the group of low-paid subjects in VAR Public is the only group in which average quality becomes lower in the second task. This behavior is in line with incentives

for subjects in the inferior relative position and would also correspond to inequity aversion: Increasing quantity enables low-paid subjects to mitigate their payment disadvantage. As the working task requires substantial effort and concentration, this has a detrimental effect on working quality. Thus, we can accept hypothesis 3a in the sense that it is information about peer wages which eventually invokes the reciprocal effects expected in hypotheses 2b and 2c.

In the fixed wage treatments, adjustments are strongly mitigated under public information, which lets us statistically reject hypothesis 2a and any additional effects of peer wage information. Yet differences in relative performance changes between high and low wage subjects have the same signs as in the piece-rate treatment: Low-paid subjects increase quantity (quality) stronger (weaker) than high wage subjects. However, under fixed wages none of these differences is significant at conventional levels.

In our piece-rate scheme, subjects have the clear incentive to cheat by completing as much questionnaire forms as possible. The question is whether the wage change and social information affect shirking behavior systematically. To identify such behavior, we distinguish between ‘regular’ errors and deliberate false entries to increase quantity and payments. We classify all arbitrary entries (for example if only 0s were filled in) and deliberate omissions (for example if the participant did not fill in decimal places) as forms of shirking. All other incorrect forms are classified as unsystematic errors.

TABLE 4.2  
PERCENTAGE OF INCORRECTLY COMPLETED FORMS BY SOURCE OF ERROR

		<b>Shirking</b>		<b>Unsystematic Errors</b>	
		Task 1	Task 2	Task 1	Task 2
VAR Public	High	4.5	6.1	19.9	10.8
	Low	1.9	13.3	20.7	18.0
VAR Private	High	21.3	20.7	15.4	12.4
	Low	6.3	11.4	23.0	15.0
FIX Public	High	5.1	2.9	18.4	8.9
	Low	1.8	4.1	22.0	15.6
FIX Private	High	2.1	4.5	20.8	13.9
	Low	1.9	4.5	19.1	14.3

Table 4.2 displays the share of incorrect forms submitted for each treatment group, distinguished per source of errors. While there is variation with respect to

the level of cheating between the experimental groups in task 1, average behavior of the subjects is roughly in line with incentives, as the level of shirking is higher among groups under piece-rate compensation.<sup>14</sup> Furthermore, the above-mentioned learning effects are displayed in the data because the share of unsystematic errors decreases in all treatments in task 2. However, the experimental treatments differ with respect to the size of those adjustments.

First, increasing payments has no clear effect on shirking, as adjustments are only small in all experimental groups. Contrary, wage cuts result in increased shirking under piece-rates as low-paid subjects increase cheating by some 11% (5%) in VAR Public (VAR Private). Under fixed wages, the effects of wage cuts are negligible which seems plausible because subjects cannot influence their payoffs by completing more forms with false entries.

In line with the results on relative performance changes, we find no systematic differences between high- and low-wage subjects in the private information groups. Comparing the number of incorrect forms on the individual level between treatments with two-sided MWU tests yields  $p$ -values of  $p > .1$ .<sup>15</sup>

Contrary, the effect of peer information becomes visible again. Under piece-rates, the shares of mistakes due to shirking and due to lower accuracy are (weakly) significantly higher for low-paid subjects ( $p = .072$  and  $p = .001$ , two-sided MWU tests, respectively). The reason for this difference is that the relative share of shirking increases among low-paid subjects ( $p = .001$ , two-sided WMSPR test) whereas accuracy of high-paid subjects becomes higher ( $p = .008$ , two-sided WMSPR test). Thus, we find behavioral effects of the wage change in both directions: Whereas the public wage cut increases some subjects' willingness to deliberately choose actions that harm the principal, the wage increase obviously motivates the respective workers to increase effort into accuracy.

The latter effect can be observed also in our treatments with fixed wages. Under public information, there is a significant difference in the share of unsystematic errors between high and low wage earners ( $p = .013$ , two-sided MWU test), because the former increase the level of accuracy after the wage change ( $p < .001$ , two-sided WMSPR test).

---

<sup>14</sup>Please note that the relatively high value for shirking in VAR Private High is mainly driven by two subjects who entered only arbitrary values in both tasks.

<sup>15</sup>Table 4.6 in the Appendix displays the significance levels of all applied two-sided MWU tests.

TABLE 4.3  
OLS REGRESSIONS OF INDIVIDUAL  
PERFORMANCE IN TASK 2

Dependent Variable	Quantity in Task 2	Quality in Task 2	Output in Task 2
VAR Public High	7.815*** [1.795]	0.241*** [0.067]	8.222*** [1.776]
VAR Public Low	11.323*** [1.956]	0.103 [0.067]	5.086*** [1.919]
VAR Private High	8.643*** [2.052]	0.166** [0.064]	6.677*** [1.852]
VAR Private Low	7.594*** [1.726]	0.219*** [0.063]	6.303*** [1.693]
FIX Public High	4.947*** [1.754]	0.289*** [0.068]	7.098*** [1.753]
FIX Public Low	6.348*** [1.763]	0.209*** [0.067]	5.979*** [1.755]
FIX Private High	7.808*** [1.837]	0.221*** [0.067]	7.406*** [1.822]
FIX Private Low	7.295*** [1.788]	0.207*** [0.069]	6.357*** [1.803]
Quantity Task 1	0.895*** [0.048]		
Quality Task 1		0.785*** [0.078]	
Output Task 1			0.903*** [0.054]
N	126	126	126
R-squared	0.99	0.98	0.97

Standard errors are given in brackets. \*\* and \*\*\* denote significance on the 5% and 1%-level, respectively.

So far, our analysis has not controlled for individual abilities. However, as the experimental subjects are very heterogeneous with respect to performance, we have to assure that our results about the effects of social comparisons are robust against the inclusion of measures for individual ability.

Table 4.3 lists simple linear regression models with measures for individual performance in task 2 as dependent variable: quantity (model 1), quality (model 2), and output (model 3). To control for differences in ability among subjects, we include

performance in the task 1 as an independent variable. As expected, task 1 performance is a strong and highly significant predictor of performance in task 2 in all specifications. Additionally, we include dummies for each experimental treatment (the constant is suppressed). The dummy variables are all positive and significant in most cases. Therefore, to investigate if the strength of adjustments differs among the treatments, we compare the sizes of the regression coefficients using Wald tests. The resulting two-sided  $p$ -values are displayed in Table 4.4.

TABLE 4.4  
P-VALUES OF TWO-SIDED WALD TESTS ON  
DIFFERENCES IN REGRESSION COEFFICIENTS

Tests on	Quantity	Quality	Output
High versus low wages			
VAR Public	<b>0.032</b>	<b>0.003</b>	<b>0.095</b>
VAR Private	0.540	0.272	0.846
FIX Public	0.383	<b>0.087</b>	0.546
FIX Private	0.750	0.766	0.572
Public versus private information			
VAR High	0.617	0.113	0.670
VAR Low	<b>0.027</b>	<b>0.016</b>	0.527
FIX High	<b>0.077</b>	0.142	0.868
FIX Low	0.555	0.961	0.839

Note: Significant or weakly significant  $p$ -values are indicated by bold numbers.

In line with our former statements, we find no per se impact of wage changes, irrespective of the compensation scheme. Estimated coefficients for high wage and low wage workers do not differ significantly in any of the performance measures when subjects are not informed about their relative standing. However, there are significant differences between treatments which allow for social comparisons. Under piece-rates, low-paid subjects increase quantity stronger than high-paid subjects ( $p = .032$ ) and lack behind with respect to quality ( $p = .003$ ) which leads to a weakly significant lower output effect ( $p = .095$ ). In addition, comparing low-paid subjects with and without peer information under piece rates shows that the former group increases quantity (quality) stronger (weaker) than the latter ( $p = .027$  and  $p = .016$ , respectively). Under fixed wages those effects are again mitigated. We observe a weakly significant difference between high-paid and low-paid subjects with

respect to quality ( $p = .087$ ), whereas there is no quantity or output effect ( $p = .383$  and  $p = .546$ , respectively).

Comparing high-paid subjects with and without information does not yield significant effects, with the exception of the quantity effect which is weakly significantly higher under fixed wages and private information ( $p = .077$ ). Hence, there is no further evidence that information about the superior wage position triggers performance increases.

## 4.5 Conclusions

We have conducted a real-effort experiment in which subjects had to solve two simple working tasks. After identical wages for the first task, we introduce wage differentials in the second task. Whereas subjects generally improve performance due to learning processes, there are performance differences found if relative comparisons between subjects are possible due to peer wage information.

Unlike related studies, wage changes per se do not lead to significantly different performance adjustments among the subjects in our setting. Only if information about peer wages is provided, performance changes of low- and high-paid workers differ from each other. This effect is stronger under piece rates where subjects have the chance to influence the distribution of absolute payoffs. Moreover, the result is mainly driven by subjects in inferior positions who show an increased probability to shirk although there is some evidence that high-wage workers increase their working effort displayed by lower error rates due to inaccuracy.

Responses to wage changes are generally weak under flat wages. In particular, there is no evidence for negative adjustments of low-paid workers under public information as observed under piece-rates. A possible reason for this observation in our experiment could be income effects: Low-paid workers were already guaranteed a comparably high payoff before task 2 started (12.00 Euros including show-up fee for approximately one hour of presence in the laboratory) which is more than the usual hourly wage for students. Thus, the fact that low-paid subjects under fixed wages did not decrease their effort could be in line with the experimental results of Greenberg (1987) suggesting that the perceived unfairness of lower relative payoffs diminishes with rising absolute payoffs. Moreover, such behavior is consistent with the model of reference-dependent preferences proposed by Köszegi and Rabin (2006). Here, subjects are assumed to evaluate economic outcomes relative to a

reference point based on rational expectations. In our environment, this would suggest that subjects evaluate realized payments relative to expected payments. If in our case subjects used an average hourly wage (which is also a common base for the calculation of average participant payoffs in economic experiments) as the reference income, the wage cut in task 2 would not be interpreted as a loss and subsequently would not alter effort decisions.<sup>16</sup>

An alternative explanation is related to the character of a fixed wage contract. Sliwka (2007) introduced a model in which the principal can signal her positive expectation of employees' cooperativeness by offering a flat wage rather than a piece-rate wage contract. If a substantial share of employees orientate on behavior of their peers and believe them to deliver decent work, this signal can induce high effort exertion. A similar mechanism could lead to a lower sensitivity of our experimental subjects to the wage cut.<sup>17</sup> Moreover, this would explain why aggregate quantity and output in the fixed wage treatment are not substantially lower than in the piece-rate treatment.

Summarizing, our study provides new evidence that relative economic positions matter for effort and performance in working environments. However, it also shows that the impact of relative wages might depend on the form of compensation. For the design of incentive schemes it would be therefore important to gain further insights into this motivational interaction. While the research on social preferences and incentives has been very active in the last years, there is still a large potential for studies about the design of mechanisms that create incentives but do not discourage those who do not profit from the system.

---

<sup>16</sup>Naturally, a similar argument applies also to subjects under piece rate wages. The difference here is, however, that participants had the possibility to decrease the difference between expected and realized income by filling in more forms.

<sup>17</sup>However, the model differs to our setting in the sense that there is only one observable dimension of performance that the principal can fully control by setting piece-rates.

## 4.6 Appendix

### 4.6.1 Instructions

*Below we show the instructions translated from German for the first and second task in treatment VAR Public (High-wage workers). Instructions for the other treatments were formulated in a similar way.*

#### *Instructions for task 1*

Welcome! You can earn money in this experiment during the next hour. From now on, please do not communicate with the other participants. If you have a question about the experiment, please raise your hand! We will come over to you to answer your question. If you break this rule, we will have to exclude you from the experiment and all payments.

You will receive two working tasks today. The wages for both tasks will be paid out at the end of the experiment.

All subjects were assigned either to group A or to group B. The group assignment remains constant in both tasks. You belong to group A. One participant of group B will be assigned to you.

The following instructions refer to the first task. You will receive the instructions for the second task after the first task is over.

#### **Task 1**

You find a staple with copies of questionnaires on the desk in front of you. These questionnaires were filled in by the participants in a recent class-room experiment. The data from the questionnaires are required in electronic form for analysis at our institute. Your task is to fill in the data into the input mask on your screen. The data will be written automatically into an Excel-spreadsheet.

The participants in the classroom experiment had to answer 15 questions. These questions were labeled 1, 2, 3, 4, 5, 1.o, 1.u, 2.o, 2.u, 3.o, 3.u, 4.o, 4.u, 5.o and 5.u. For each question, the participants had to give estimates in percent. The following screen-shot shows the input mask for the data entry.

FIGURE 4.3  
INPUT MASK

The screenshot shows a software interface for data entry. At the top, it says 'Bearbeitete Formulare: 0'. Below this is a grid of input fields. The first column is labeled 'Nummer' and contains the value '12245'. The next five columns are labeled '1 oben', '2 oben', '3 oben', '4 oben', and '5 oben', with values '50', '60', '30', '40', and '20' respectively. The next five columns are labeled '1 o links', '1 u links', '2 o links', '2 u links', '3 o links', '3 u links', '4 o links', '4 u links', and '5 u links', with values '32', '29', '59', '50', '45', '33', '89', '90', and '100' respectively. An 'OK' button is located at the bottom right of the grid.

Nummer	1 oben	2 oben	3 oben	4 oben	5 oben	1 o links	1 u links	2 o links	2 u links	3 o links	3 u links	4 o links	4 u links	5 u links
12245	50	60	30	40	20	32	29	59	50	45	33	89	90	100

First, please copy **the number on the upper right side** on each questionnaire and write it into the field ‘Number’ on the screen. Then, fill in **only the left column** of the percentage values from the questionnaires, i.e. the numbers for the categories ‘up’ and ‘left’, into the corresponding fields on the screen. You do not need to fill in the %-sign. Please note that you cannot fill in commas into the fields on the screen. Instead, you have to use decimal points according to the English syntax. After you have filled in the values, press the OK-button. The data is then transferred and written into the Excel-spreadsheet. Then go on with the next questionnaire.

You have 20 minutes time for the data input. The remaining time is displayed on the screen. You receive an amount of **0.20 Euro** for each questionnaire completely filled in in task 1.

The participant of group B assigned to you works on the same set of questionnaires as you do, and receives the same amount of **0.20 Euro** for each questionnaire completely filled in. During later data validation entries are compared and, if necessary, corrected.

*Instructions for task 2*

**Task 2**

For task 2 you receive a new staple of questionnaires. These questionnaires contain data of other participants from the same classroom experiment. These data are also to be transferred in the same way as in task 1. Again, you have 20 minutes time for the data input.

You receive an amount of **0.32 Euro** for each completely filled in questionnaire in task 2.

The same participant of group B as in task 1 is assigned to you and works again on the same questionnaires as you. The participant of group B receives an amount of **0.08 Euro** for each questionnaire completely filled in in task 2.

#### 4.6.2 Additional Test Statistics

TABLE 4.5  
P-VALUES OF TWO-SIDED MANN-WHITNEY-U  
TESTS ON DIFFERENCES IN RELATIVE  
PERFORMANCE INCREASES BETWEEN TREATMENTS

	Quantity Increase	Quality Increase	Output Increase
High versus low wages			
VAR Public	<b>0.060</b>	<b>0.019</b>	0.171
VAR Private	0.560	0.102	0.137
FIX Public	0.933	0.224	0.283
FIX Private	0.450	0.485	0.462
Public versus private information			
VAR High	0.822	0.208	0.236
VAR Low	0.172	<b>0.008</b>	0.113
FIX High	0.365	0.146	0.609
FIX Low	0.462	0.361	0.963

Significant and weakly significant  $p$ -values are indicated by bold numbers.

TABLE 4.6  
P-VALUES OF TWO-SIDED MANN-WHITNEY-U TESTS ON  
DIFFERENCES IN SOURCES OF ERRORS BETWEEN  
TREATMENTS

	Shirking		Unsystematic Errors	
	Task 1	Task 2	Task 1	Task 2
High versus low wages				
VAR Public	0.333	<b>0.072</b>	0.105	<b>0.001</b>
VAR Private	0.360	0.267	0.325	0.967
FIX Public	0.615	0.608	<b>0.080</b>	<b>0.013</b>
FIX Private	0.749	0.612	0.266	0.955
Public versus private information				
VAR High	0.451	0.172	0.689	0.562
VAR Low	0.501	0.176	0.338	<b>0.015</b>
FIX High	0.708	0.129	0.099	<b>0.014</b>
FIX Low	0.863	0.179	0.206	0.970

Significant and weakly significant  $p$ -values are indicated by bold numbers.

# Chapter 5

## The Dynamics of Cooperation in a Microfinance Game

### 5.1 Introduction

Developed in the 1970s to alleviate poverty in the lowest-income countries, microfinance programs are now widespread and have also entered the western world. The non-profit organization MIX Market reports that in 2007, 890 surveyed microfinance institutions (MFI) served more than 30.1 million borrowers and managed a loan portfolio of more than 64 billions USD (see the MicroBanking Bulletin, 2008).

The goal of microfinance lending is to provide poor population groups the access to credit markets that would otherwise be blocked due to prohibitively high risks and transaction costs. Most MFI rely on non-standard contract schemes and practices, the most prominent being group responsibility or joint liability lending. Here, several borrowers form groups that are jointly responsible for the payback of the entire credit sum. If one borrower is not able to repay her credit, the other group members have to bail her out.

Joint liability contracts benefit both clients and MFI. First, they offer a ‘social’ collateral for borrowers who are not able to provide tangible assets. Second, moral hazard and adverse selection problems between MFI and borrowers are mitigated because costly screening or monitoring activities are carried out by the group.<sup>1</sup>

---

<sup>1</sup>Stiglitz (1990), Banerjee et al. (1994), and Armendàriz de Aghion (1999) show how monitoring and sanctioning among agents in a borrower group can mitigate moral hazard problems with respect to project choices. Furthermore, the formation of borrower into credit groups helps to overcome

Indeed, repayment rates in joint liability microfinance programs are usually very high given the large share of households subject to extremely high economic risks in the portfolios of MFI.

However, various factors can limit the functionality of joint liability lending. First of all, moral hazard problems may be created within a borrower group. If the responsibility for repayment is borne by the group, borrowers have the incentive to free-ride by shifting the repayment burden or investment risks to fellow group members. Therefore, if monitoring is incomplete – for example because social ties between group members are loose or non-existent – or if social sanctioning within groups is not possible, joint liability schemes may not function in the desired way. Moreover, the threat of moral hazard within groups may affect the composition of an institution’s credit portfolio, as joint liability programs become less attractive for good risks.<sup>2</sup>

In practice, several large MFI have refrained from group responsibility contracts or have started to offer individual liability contracts to their clients. However, there is little empirical evidence that helps to evaluate in which circumstances which contract scheme is preferable. Giné and Karlan (2007) state that, “despite being a question of first-order importance, empirical research on group versus individual liability lending has not provided policymakers and institutions the clean evidence needed to determine the relative merits of the two methodologies.”

One inherent problem of empirical studies about the impact of contract schemes is identification, because most MFI initially offer one type of contract, and estimated effects might potentially suffer for example from selection biases. Therefore, the goal of the present study is to complement existing empirical work by investigating behavior under joint and individual liability contracts in a controlled laboratory environment. In particular, the focus is on the dynamics of behavior, because the interaction among group members might crucially influence the functioning of contract schemes over time. So far, the evolution of interaction between borrowers under group lending has received relatively little attention. Moreover, we consider the mentioned case where no mechanism such as social sanctioning is available to

---

adverse selection problems (see Armendàriz de Aghion and Gollier, 2000, Ghatak, 1999 and van Tassel, 1999). Finally, peer pressure within borrower groups might reduce strategic defaults (see Besley and Coate, 1995). For an overview of the theoretical literature see Morduch (1999).

<sup>2</sup>Other potential disadvantages of joint liability are related to the evolution of credit demand. If group members are heterogeneous concerning the required amounts of funding, individual contracts may become optimal for the wealthier clients, see Madajewicz (2004).

directly enforce cooperative behavior. Finally, the study compares the performance of different conversion mechanisms in mitigating moral hazard problems among borrowers.

We let experimental subjects play a microfinance game repeatedly under joint liability (run 1) before partly being converted into individual liability (run 2). Subjects decide about the effort invested into a risky project for up to 12 periods in each run. Under joint liability, subjects face a dilemma situation, as minimum effort is the individually optimal choice while the maximum feasible effort maximizes joint payoffs. Furthermore, in line with common practice of MFI, we introduce dynamic incentives. Irrespective of liability schemes, credit provision in a given period is contingent on successful repayment in the previous period.<sup>3</sup>

Two conversion mechanisms from joint to individual liability are implemented. In a first treatment, subjects can choose between an individual and a joint liability contract after the first run of the experiment. In a second treatment, borrower groups with above-average repayment performance are subsequently offered individual contracts for run 2 while the other subjects remain under joint liability. Finally, we compare the performance of these mechanisms with a baseline setup in which subjects remain under joint liability for both runs of the game.

Despite the absence of direct sanctioning mechanisms, average effort levels remain high until the end of the game in all experimental treatments. Nevertheless, within-group moral hazard is still observable, as subjects converted into individual liability increase their effort significantly relative to the first run.

Moreover, individual effort decisions are positively related to partner effort and negatively related to the frequency of partner repayments. Thus, the results suggest a path dependency of behavior that may have a detrimental effect for the functioning of joint liability programs from a dynamic perspective.

In line with effort decisions, subjects become more likely to select individual contracts the lower the partner effort and the higher the frequency of partner repayments were. Regardless of this pattern, average effort provided by subjects opting for a joint liability contract persists on a high level in run 2. Contrary, in the treatment with performance based conversion, average effort levels among the subjects remaining under group lending tend to be lower than in the baseline treatment.

---

<sup>3</sup>Armendàriz de Aghion and Morduch (2000) and (2005) show formally that the threat of not receiving follow-up loans may prevent strategic default of borrowers.

In the following, we briefly review the relevant empirical and experimental literature (section 5.2). In the next step, we introduce our experimental decision situation (section 5.3), describe the experimental details and formulate hypotheses (section 5.4). We present aggregate and individual results in section 5.5. Finally, section 5.6 gives a short conclusion.

## 5.2 Related Literature

Experimental and empirical studies about the behavioral impacts of microfinance contracts are scarce up to now and yield mixed results. In a laboratory experiment where subjects had the option to default strategically, Abbink et al. (2006*a*) found that repayment performance increased under joint liability relative to individual liability. Giné et al. (2009) conducted experiments in urban Peru with a large sample of owners and employees of micro-enterprises. Among other things, the authors found that the introduction of joint liability lending increased the riskiness of project choices in a dynamical setting. Kono (2006) reported results of experiments in three districts of Ho-Chi-Minh City showing that strategic default increased if subjects were jointly responsible for repayment.<sup>4</sup>

There are few studies investigating empirically the effects of contract schemes and conversion into individual liability. Giné and Karlan (2007) carried out a field experiment in which a random sample of joint liability borrowers was converted into individual liability. The authors found no difference in repayment performance between the contractual schemes after the conversion. However, the data confirmed some inherent problems of joint liability lending, as individual liability programs were more likely to attract new customers. Cull et al. (2009) compared the performance of microfinance institutions and found no large differences in loan portfolio quality with respect to the implemented contract schemes. Finally, in a survey study conducted in cooperation with the group lending organization FINCA-Peru, Karlan (2007) showed that repayment performance of borrower groups increased with stronger social ties between the members.

Ahlin and Townsend (2007) report evidence from a rural microfinance institution

---

<sup>4</sup>Other microfinance experiments include the studies by Abbink et al. (2006*b*) analyzing the role of interest rates for repayment performance and Karlan (2005) connecting individual choices in trust and public goods games to performance in microfinance programs. Here, a positive relation between trustworthiness in investment games and repayment and savings behavior is established.

in Thailand that leaves clients the decision about the form of liability. The authors found evidence for adverse selection among borrowers, as the selection into joint liability contracts and the probability of a household's project success were negatively correlated. In a study of microfinance loans in Canada, Gomez and Santor (2003) found superior repayment performance of group lending schemes due to both selection effects and decreasing moral hazard problems within groups.

As moral hazard within borrower groups is modeled as a social dilemma in the present study, evidence from public goods games is also highly relevant. In this extensively studied experimental game, positive cooperation levels among participants are usually observed, and this effect becomes stronger for partner matchings (for an extensive survey of the related literature see Ledyard, 1995). Furthermore, for our treatment in which subjects can choose between individual and joint liability, the results of experiments investigating the effect of selection processes are particularly important. These studies show that self-selection of participants into groups can foster social norms of cooperation and may help to mitigate or eliminate dilemma problems (see Bohnet and Kübler, 2005, Page et al., 2005, Gürer et al., 2006 and Grimm and Mengel, 2009). Out-of-equilibrium payoffs or sanctioning possibilities may lead to a self-selection of subjects into highly-cooperative groups and in most cases to perpetuation of high cooperation levels. For our setting, these results would suggest that the selection into joint liability could be a signal for the willingness to cooperate.<sup>5</sup>

---

<sup>5</sup>At the same time, positive expectations about peer cooperation can make it optimal for selfish subjects to enter joint liability and exploit the fellow group members by not providing effort.

## 5.3 Decision Situation and Experimental Parameters

In our experimental setting, we model the decision situation of borrowers after being granted a credit amount of  $I$  by a bank or MFI. In every period in which a credit has been provided, events unfold as follows:

1. Each credit is automatically invested into a risky project that pays  $R > 0$  in case of success and 0 in case of failure.
2. Every borrower chooses the effort level  $e$  invested into the project from the interval  $[e_0, e_{max}]$ . Effort exertion positively influences the probability  $p$  of project success according to the function  $p(e)$ . Effort costs are captured by the cost function  $c(e)$  and have to be incurred irrespective of the project outcome.
3. Project payoffs are determined. In case of joint liability, payoffs and effort levels of all borrowers are public information, but there is no possibility of sanctioning. Loan repayments  $L$  for each credit are automatically deducted from borrowers' payoffs in case of sufficient funds. If the credit is fully repaid, a new credit is provided in the next period and the decision situation starts again. Contrary, if borrowers default, they receive no further credit in subsequent periods.

We pose a number of restrictions on the variables. First, a borrower is able to repay two loans in case of success under joint liability, because  $R \geq 2L$ .<sup>6</sup> Next, there is an exogenous success probability  $p_0 > 0$  regardless of effort provision so that total success probability  $p$  accounts for  $p = p_0 + p(e)$ . Moreover, we assume a constant marginal success probability  $p'(e) = \bar{p}$  and constant marginal effort costs  $c'(e) = \bar{c}$ . Hence, every unit of effort increases the probability of project success by a fixed factor and is associated with constant additional effort costs. Finally, we assume that providing the maximum feasible effort still bears a risk of failure, as  $p(e_0) + p(e_{max}) < 1$ .

---

<sup>6</sup>This assumption is necessary to avoid non-linearities of profit functions under joint liability contracts.

In the individual liability case a borrower's expected payoff (or utility under risk-neutrality) is

$$E[\pi_i(e)] = [p_0 + p_i(e_i)] \cdot (R - L) - c(e_i) \quad (5.1)$$

The first-order-condition (FOC) for the optimal effort level yields  $p' \cdot (R - L) = c'$ . As  $p'(e) = \bar{p}$  and  $c'(e) = \bar{c}$ , only corner solutions are possible: depending on the relative size of marginal gains and marginal costs of effort provision, it is either optimal to choose the maximum effort level  $e_{max}$  or the minimum  $e_0$  in a one-shot game.

We compare individual loan contracts to joint liability. In our simple case, we assume that two persons form a borrower pair and are jointly responsible for the repayment of the loan sum  $2L$ . As the partners have to bail out each other, the expected payoff of borrower  $i$  now accounts for:<sup>7</sup>

$$E[\pi_i(e_i, e_j)] = [p_0 + p_i(e_i)] \cdot [p_0 + p_j(e_j)] \cdot (R - L) + [p_0 + p_i(e_i)] \cdot [1 - p_0 - p_j(e_j)] \cdot (R - 2L) - c(e_i) \quad (5.2)$$

With probability  $[p_0 + p_i(e)] \cdot [p_0 + p_j(e)]$ , both partners succeed and repay their own loans. However, with probability  $[p_0 + p_i(e)] \cdot [1 - p_0 - p_j(e_j)]$  borrower  $i$ 's project succeeds while partner  $j$ 's project fails. In that case, the payoff of the successful borrower is reduced also by the other group member's obligation. For borrower  $i$ 's FOC follows  $p' \cdot [p_0 + p_j(e_j)] \cdot (R - L) + p' \cdot [1 - p_0 - p_j(e_j)] \cdot (R - 2L) = c'$ . Replacing  $p' = \bar{p}$  and  $c' = \bar{c}$  and rearranging, we get  $\bar{p} \cdot [R - L - (1 - p_0 - p_j(e_j)) \cdot L] \geq \bar{c}$ .

Comparing first-order conditions under individual and joint liability shows the moral hazard problem among borrowers in the latter contract. As the term  $1 - p_0 - p_j(e_j)$  is greater than zero per assumption, the provision of  $e_{max}$  becomes less likely under group lending. This is because the possibility that a borrower has to stand in for her partner makes own project success and effort exertion less attractive than in the individual case.<sup>8</sup>

---

<sup>7</sup>Success probabilities and project payoffs of the two borrowers are independent.

<sup>8</sup>Please note that the decision situation would be analogous if social sanctions existed but were not carried out in a group despite the possibility to observe the partner's effort. Armendàriz de Aghion and Morduch (2005) mention kinship among villagers or anonymous credit programs in cities as examples for situations in which social interaction between members of borrower groups might not be effective.

Optimal effort decisions in this setting depend on the precise set of parameters. Hence, we have to pose a number of restrictions to model the situation in the desired way.

First, it should be individually optimal to choose  $e = e_0$  under both contracts in the one-shot game. This restriction depicts the inherent moral hazard problem between MFI and borrower. Next, the borrower group should face a social dilemma. It follows that the maximum sum of payoffs for the borrowers is reached if both borrowers exert  $e = e_{max}$ . For success probabilities  $p_0$  and  $p(e_{max})$ , this is true as long as effort is not too costly:  $\bar{c} \cdot e_{max} \leq [p(e_{max}) - p_0] \cdot R - [p(e_{max}) \cdot (2 - p(e_{max})) - p_0 \cdot (2 - p_0)] \cdot L$ .<sup>9</sup>

Additional requirements have to be met in the repeated setting. The first requirement is related to individual liability contracts and the functioning of dynamic incentives: the threat of not receiving follow-up loans must provide a disciplining effect on agents' effort choices. Therefore, we assume that full effort exertion is optimal for the individual borrower  $i$  in the repeated setting, so that the following condition for expected payoffs holds:  $E[\pi_i(e_{i,t} = e_{max})] > E[\pi_i(e_{i,t} = e_0)] \forall t$ .

Furthermore, parameters should ensure that while the subgame perfect Nash equilibrium of the game consists of both borrowers choosing  $e = e_0$  in each period, the choice that maximizes joint payoffs is full effort exertion. Therefore,  $E[\pi_i(e_{i,t} = e_{max} | e_{j,t} = e_{max})] \geq E[\pi_i(e_{i,t} = e_0 | e_{j,t} = e_0)] \forall t$ .<sup>10</sup>

Finally, the insurance effect of joint liability lending has to be captured by  $E[\pi_{i,JL}(e_{i,t} = e_{max} | e_{j,t} = e_{max})] > E[\pi_{i,IL}(e_{i,t} = e_{max})] > E[\pi_{i,JL}(e_{i,t} = e_0 | e_{j,t} = e_0)] \forall t$ . Here,  $E[\pi_{i,JL}]$  and  $E[\pi_{i,IL}]$  denominate expected payoffs for borrower  $i$  in the repeated game under group and individual lending, respectively. According to this condition, expected payoffs from the full effort strategy under joint liability exceed maximum payoffs under individual contracts. However, individual contracts have to be superior in payoffs compared to a joint liability setting in which both players choose  $e = e_0$  in every period. This refers to the argument that borrowers

---

<sup>9</sup>Effort costs have to be lower than the expected gain for the borrower from switching to high effort. The term  $[p(e_{max}) - p_0] \cdot R$  displays the expected increase in project revenue. However, higher effort increases also the probability that the credit has to be repaid. Therefore, expected gains have to be corrected for the expected increase in loan repayments  $[p(e_{max}) \cdot (2 - p(e_{max})) - p_0 \cdot (2 - p_0)] \cdot L$ . This term is always positive as long as  $p(e_{max}) > p_0$ .

<sup>10</sup>We also show in the Appendix that the full effort strategy  $e_{i,t} = e_{j,t} = e_{max} \forall t$  yields higher expected payoffs than an alternating effort strategy in which group members chose in turn maximum and minimum effort levels.

are made worse off if low cooperation is established in the groups.

TABLE 5.1  
PARAMETER SET

Variable	Parameter Value
$R$	200
$L$	100
$I$	50
$e_0$	0
$e_{max}$	8
$p_0$	0.5
$p(e)$	$0.05 \cdot e$
$c(e)$	$6 \cdot e$

The parameters listed in table 5.1 meet the described conditions (for a more detailed proof of all described parameter restrictions see section 5.7.2 in the Appendix.)

## 5.4 Experimental Design and Hypotheses

We conducted 8 experimental sessions in which altogether 256 subjects took part (32 subjects per session), most of them with a major in Economics, Business Administration or related fields. In each session, participants were divided into matching groups of 8 persons (subjects were not aware of this procedure). The experiment consisted of two runs of up to 12 rounds of the described decision situation.

Subjects started run 1 under joint liability. Before the first period began, every person was randomly assigned an anonymous partner from her matching group with whom she formed a borrower pair for the entire first run. In the first period, both partners simultaneously decided about effort levels. As long the sum of loan repayments was covered in a given period, the borrower pair received two new loans in the subsequent period and decided again about effort levels. After each period, subjects were informed about effort levels and project revenues of both group members.

The treatment variation consisted of different conversion mechanisms from joint into individual liability that took place after run 1 was completed. We conducted four treatments denoted with SELECT, PERFORM (divided into treatments NOINFO and INFO) and BASE in the remainder of the paper. Our matching procedure yields 8 statistically independent observations for each treatment.

First, in treatment SELECT subjects could choose the contract in the second run. After the first 12 periods, people were handed instructions describing individual and joint liability contracts and were asked to choose one contract that would be applied for the next 12 periods. Subjects who had chosen joint liability contracts were paired with a new partner from their experimental matching group.<sup>11</sup>

Second, borrower pairs were converted into individual liability on the basis of repayment performance in the first run (treatment PERFORM). The two pairs per matching group with the highest repayment performance (i.e. the highest number of successfully repaid loans) were converted into individual liability, while the other two pairs remained in joint liability and were rematched with a new partner in run 2.<sup>12</sup> We conducted two specifications of this treatment: first, treatment INFO where subjects were informed about the conversion mechanism before the experiment started and second, NOINFO, in which the information was provided after run 1.

Finally, treatment BASE served as the benchmark to control for dynamic effects of repeated interaction in the borrower groups. Here, all subjects played the decision situation under joint liability in both runs. Similar to the other treatments, subjects were rematched after the first run.

Over the periods in which they received loans, subjects accumulated experimental payoffs. If period losses occurred, for example if a project failed and the subject had exerted positive effort levels, these losses were subtracted from accumulated payoffs. After the experiment, one run was randomly determined by the role of a die. Accumulated payoffs of this run were subsequently converted into Euro and paid out to the participants.

First of all, our focus is on the dynamics of moral hazard among the borrowers. In line with evidence from public goods games, we expect positive average effort levels. However, we hypothesize that moral hazard problems are still persistent and average effort deteriorates over time (hypothesis 1a). As a further indication for within-group moral hazard, we expect that effort levels of subjects receiving individual contracts increase relative to the first run (hypothesis 1b).

The aim of our treatment variations is to investigate if and how conversion mechanisms influence effort exertion and repayment performance. Both selection based

---

<sup>11</sup>If an uneven number of subjects per matching group had opted for joint liability, one person was randomly chosen and converted into individual liability.

<sup>12</sup>If two or more pairs had the same repayment performance and this was relevant for the assignment to contract schemes in the second run, one pair was randomly chosen for conversion.

and performance based conversion do not change incentives to exert minimum effort. However, the question is if behavioral adaptations of the subjects to the mechanisms have systematic effects on cooperation under group lending.

In line with the relevant studies about selection in public goods games, we hypothesize that self-selection into joint liability contracts induces higher cooperation levels among borrower pairs in the second run (hypothesis 2).<sup>13</sup>

Under performance based conversion, we expect lower effort exertion among the remaining borrower pairs in the second run of the game (hypothesis 3). If there is a correlation between repayment performance and effort exertion, the willingness to cooperate should be lower among borrower pairs in the second run, because subjects know that they are matched with a partner who chose relatively low effort levels in the first part of the experiment.<sup>14</sup>

Finally, as the joint liability offers an insurance device against involuntary default, it is important to note that risk preferences of the subjects might also influence both the effort levels and the contract choice. We will come back to this issue in the next section.

The experimental sessions took place in the period from April 2008 to February 2009 in the Cologne Laboratory for Economic Research. Subjects were recruited using Greiner's Online Recruitment System (Greiner, 2004). The experiment was computerized using the z-Tree software (Fischbacher, 2007). After subjects arrived and were randomly assigned to a cubicle, instructions were distributed.<sup>15</sup> Ques-

---

<sup>13</sup>This effect hinges on the assumption that players with a higher willingness to cooperate choose the group contract. For this choice, expectations about partner efforts play a central role. Under strict rationality, each subject would expect minimum effort exertion of the group member. Subsequently, expected payoffs are maximized by choosing individual liability and providing  $e_{i,t} = e_{max}$  in each period. However, if a subject expects higher-than-minimum effort levels  $e_{j,t}$  from her partner, it may become optimal to opt for joint liability if the expected effort level is high enough. With the present parametrization, expected partner effort has to take values of  $e_{j,t} \geq 7$  to make the choice of joint liability profitable (see 4. in the Appendix).

<sup>14</sup>In addition to behavioral changes in run 2, ex-ante information about performance related conversion might influence behavior already in the first run of the experiment (specification INFO in the PERFORM treatment). However, the provision of information has ambiguous effects in our setting. First, it could play a disciplining role for effort exertion, as potential gains from individual liability in the second run are high enough to compensate for the provision of non-minimum effort levels and losses from repayments of the partner's loans in run 1. On the other hand, presenting both contract schemes could lead to a higher saliency of incentive problems among the borrowers and subsequently to lower efforts. Please note also that in the SELECT treatment, ex-ante information about conversion would not alter effort decisions, as subjects have the possibility of opting for the loan contract.

<sup>15</sup>Instructions translated from German can be found in the Appendix.

tions were answered privately. Before payoffs were paid out, subjects filled out a post-experimental questionnaire. They were asked about demographical data, the underlying motivations for their decisions and in some sessions expectations about average behavior of the participants. Subjects were privately paid out their earnings for the selected run and left the laboratory. The average payoff was 14.79 Euros (including an increased show-up fee of 5.00 Euros to cover potential losses in the course of the game). Each session lasted approximately one and a half hours.

## 5.5 Experimental Results

In the following, we will present results on the aggregate level and compare the experimental treatments with respect to effort exertion and repayment performance. In the next step, we will analyze influence factors on individual behavior. Finally, we will briefly investigate motivations for contract choices in the SELECT treatment.

### 5.5.1 Aggregate Results

Table 5.2 lists average effort levels calculated over all active borrower pairs in both runs for each experimental treatment.

TABLE 5.2  
AVERAGE EFFORT LEVELS PER TREATMENT AND RUN

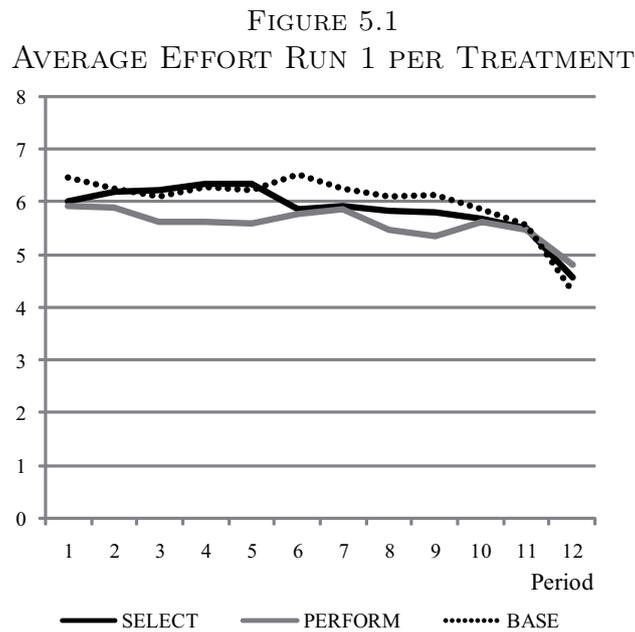
Treatment	Run 1		Run 2
	Joint Liability	Individual Liability	Joint Liability
SELECT	5.89	7.27	6.28
NOINFO	5.70	7.09	6.00
INFO	5.53	6.82	5.30
BASE	6.04	–	6.37

The first thing to notice is that high levels of cooperation are achieved in run 1, as average effort levels account for 69 to 75% of the maximum feasible effort. Furthermore, non-parametric tests reveal that experimental treatments do not differ from each other with respect to effort levels or repayment performance (two-sided Mann-Whitney-U (MWU) tests comparing matching group averages yield  $p > .1$ ).

For the further analysis, we pool the data of NOINFO and INFO in treatment PERFORM. Although average effort levels are somewhat lower in INFO, the dif-

ference is not significant at conventional levels.<sup>16</sup> Hence, ex-ante information about the conversion mechanism do not seem to alter behavior compared to the other treatments.

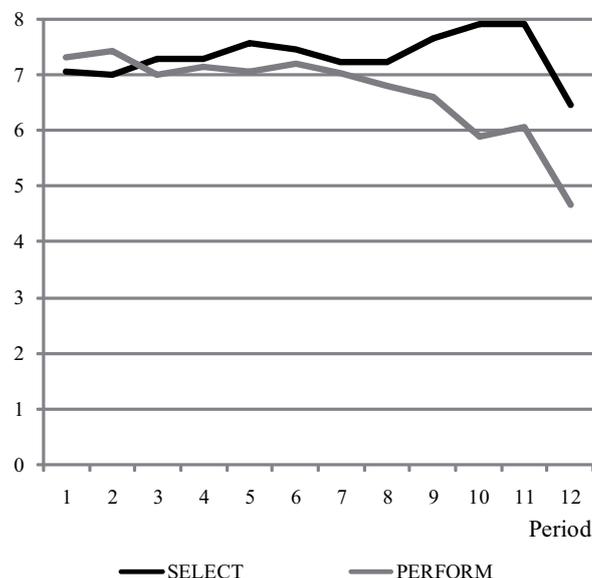
The development of average effort levels for all active borrowers is depicted in figures 5.1 to 5.3. First, figure 5.1 shows that effort levels undergo a small downward trend over time, but still account for more than 50% of the maximum level in the final period.



In line with incentives, subjects converted to individual liability in run 2 converge to full effort exertion in the repeated game (see figure 5.2). Here, average effort levels for the second run increase to 87% (PERFORM) and 91% (SELECT) of the maximum feasible level. This increase is corroborated by non-parametric tests: subjects converted to individual contracts exhibit significantly higher effort levels compared to run 1 ( $p = .016$  for SELECT and  $p < .001$  for PERFORM, two-sided Wilcoxon-Matched-Pairs-Signed-Ranks (WMPSR) tests). Hence, we find evidence for the hypothesized within-group moral hazard under group lending.

<sup>16</sup>The  $p$ -values of two-sided MWU tests for effort comparisons between NOINFO and INFO yield  $p = .721$  (run 1),  $p = .458$  (run 2, individual liability) and  $p = .368$  (run 2, joint liability). Furthermore, comparing NOINFO and INFO pairwise with the other treatments yields  $p > .1$  in all cases.

FIGURE 5.2  
 AVERAGE EFFORT RUN 2 PER TREATMENT (INDIVIDUAL LIABILITY)



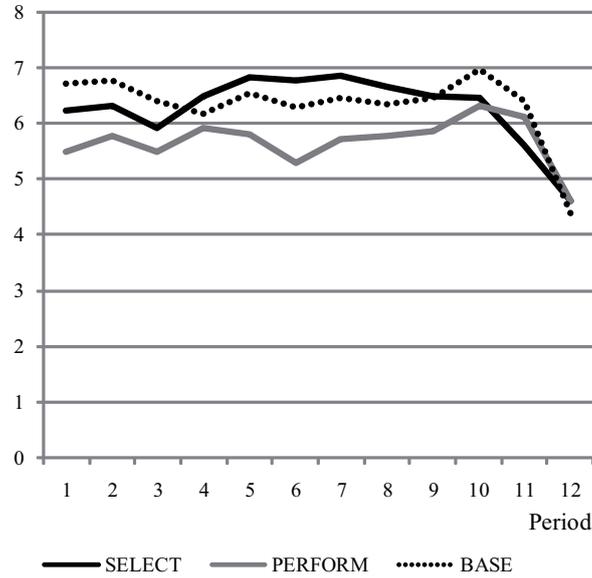
Next, the dynamics of effort exertion for joint liability pairs in the second run are plotted in figure 5.3. Effort levels remain on roughly the same level as in run 1. While there is a small increase of average effort in all treatments, this difference is weakly significant only in the BASE specification ( $p = .078$  for BASE,  $p = .844$  for SELECT and  $p = .706$  for PERFORM, respectively, two-sided WMPSR tests).

Comparing effort levels between the treatments, we find no difference between treatments BASE and SELECT ( $p = .574$ , two-sided MWU test). Thus, there is no evidence in line with hypothesis 2 that selection into group contracts systematically increases cooperation among the borrowers.

Subjects in treatment PERFORM provide lower average effort levels than their counterparts in the BASE treatment which corroborates hypothesis 3. Given that subjects are informed about the conversion mechanism, this result seems plausible. The prospect of being matched with a fellow borrower who had a low repayment performance in the previous run might decrease own willingness to cooperate by choosing high effort levels. However, the effect is only weakly significant ( $p = .068$ , two-sided MWU test).

Finally, to measure repayment performance, we use relative repayment rates which we define as the percentage share of loans repaid at total feasible loans per

FIGURE 5.3  
AVERAGE EFFORT RUN 2 PER TREATMENT (JOINT LIABILITY)



run. Table 5.3 lists relative repayment rates in all treatments for both runs.

TABLE 5.3  
RELATIVE REPAYMENT RATES PER TREATMENT AND RUN IN %

Treatment	Run 1		Run 2
	Joint Liability	Individual Liability	Joint Liability
SELECT	83.3	47.8	77.9
PERFORM	81.3	45.1	71.6
BASE	78.1	—	71.4

There are no significant differences concerning repayment performance found between the three treatments (all tests yield  $p$ -values  $> .1$  irrespective of the run).<sup>17</sup> Relative repayment rates on average decrease in run 2, but only significantly so among subjects under individual liability ( $p = .008$  for SELECT and  $p < .001$  for PERFORM, two-sided WMPSR tests). Although subjects increase the effort invested into their projects, their vulnerability against unintentional default increases.

<sup>17</sup>In general, our experimental data confirms that in treatment PERFORM, repayment performance on average is valid signal for exerted effort. Calculated over the whole first run, the correlation between repayment success and average effort levels is positive and highly significant (Spearman's  $\rho = .253$ ,  $p = .001$ ).

Hence, the conversion of subjects do not pay off from a bank’s perspective in our setting. The reason is that cooperation levels in the first run are so high that the insurance effect of joint liability contracts dominates the moral hazard effect.

### 5.5.2 Individual Decisions

To identify the drivers of individual behavior under joint liability, we calculate regression models with random effects on the level of the particular subject. Our dependent variable is the individual effort choice in a given period. As a robustness check, we calculate all models also with OLS and robust standard errors for each subject. Table 5.4 contrasts random effects (RE) and OLS-models.<sup>18</sup>

In the first specification (model 1) for run 1, we include the variables *Period* (number of period in respective run) and  $e_{j,t-1}$  (last period effort level of the partner) in the model. The coefficient for *Period* is negative and significant in both specifications indicating a downward trend of effort over time. More important, we find evidence for responsiveness of the subjects concerning partner effort, as the sign of  $e_{j,t-1}$  is positive and highly significant. Thus, cooperation in the group lending scheme is path-dependent: The lower the partner’s effort in the previous period, the lower is the effort choice of a subject in a given period. It follows that in a dynamic setting, the moral hazard problem of joint liability lending may increase over time.<sup>19</sup>

In specification 2, we additionally include the variable  $TimesPaid_{t-1}$  indicating the accumulated number of periods in which a subject had to repay the loan of her partner. Due to the correlation to the effort level  $e_{j,t-1}$ , we also include the interaction term  $e_{j,t-1} \times TimesPaid_{t-1}$ .<sup>20</sup> Model 2 shows that – controlling for the partner’s effort level – a subject’s willingness to exert effort declines significantly the more often she had to stand in for the partner’s loan. As the interaction term is positive and significant, the negative effect of  $TimesPaid_{t-1}$  becomes smaller with higher levels of partner effort. Yet its estimated net effect remains negative if

---

<sup>18</sup>Please note that conclusions remain similar if models with fixed effects per subjects are used.

<sup>19</sup>The dependency of effort choices on partner decisions is already indicated by simple correlation analysis. Under the joint liability contract, the Spearman  $\rho$ -values between the effort choice of a person and her partner’s effort choice in a given period of run 1 (run 2) account for .573 (.348) in case of treatment SELECT, .549 (.424) in treatment PERFORM and .471 (.524) in treatment BASE, respectively. All  $\rho$ -values are highly significant ( $p < .001$  in all cases).

<sup>20</sup>There is a significantly negative correlation between  $TimesPaid_{t-1}$  and  $e_{j,t-1}$  found in both runs: Spearman  $\rho$ -values calculated over all treatments account for  $\rho = -.188$  in run 1 ( $p < .001$ ) and  $\rho = -.186$  in run 2 ( $p < .001$ ).

TABLE 5.4  
DETERMINANTS OF EFFORT CHOICES UNDER JOINT LIABILITY

No.	1a	1b	2a	2b	3a	3b
Dependent Variable	$e_{i,t}$	$e_{i,t}$	$e_{i,t}$	$e_{i,t}$	$e_{i,t}$	$e_{i,t}$
Model	Run 1 RE	Run 1 OLS	Run 1 RE	Run 1 OLS	Run 2 RE	Run 2 OLS
<i>Period</i>	-0.152*** [0.010]	-0.175*** [0.016]	-0.046*** [0.014]	-0.023 [0.020]	-0.080*** [0.019]	-0.035 [0.033]
$e_{j,t-1}$	0.209*** [0.014]	0.433*** [0.023]	0.336*** [0.028]	0.571*** [0.034]	0.138*** [0.040]	0.403*** [0.083]
<i>TimesPaid</i> <sub><i>t-1</i></sub>			-0.577*** [0.099]	-0.568*** [0.129]	-0.742*** [0.138]	-0.422* [0.217]
$e_{j,t-1}X$ <i>TimesPaid</i> <sub><i>t-1</i></sub>			0.055*** [0.016]	0.063*** [0.019]	0.083*** [0.023]	0.046 [0.033]
SELECT					-0.267 [0.336]	-0.020 [0.291]
PERFORM					-0.519* [0.285]	-0.281 [0.253]
Constant	5.483*** [0.110]	4.536*** [0.159]	4.231*** [0.197]	2.722*** [0.261]	6.878*** [0.464]	4.425*** [0.743]
Observations	2594	2594	2338	2338	1336	1336
Subjects	256	.	252	.	160	.
Wald- $\chi^2$	347.6	.	581.1	.	171.5	.
R-squared	.	0.25	.	0.40	.	0.21

Random effects and clustered standard errors are calculated on the level of experimental subjects. Standard errors given in brackets. \*\*\*, \*\* and \* denote significance levels of  $p < 0.01$ ,  $p < 0.05$  and  $p < 0.1$ , respectively.

evaluated at the mean of  $e_{j,t-1} \approx 5.35$  in run 1 and  $e_{j,t-1} \approx 5.85$  in run 2. Hence, these results suggest the insurance effect of group lending becomes weaker over time: While group responsibility offers an insurance against involuntary default and increases the probability of further credits, the willingness to exert effort decreases

after agents had to stand in for each other.<sup>21</sup>

The described results remain robust for joint liability borrowers in run 2 (model 3). Here, we include also the dummy variables `PERFORM` and `SELECT` for the respective treatments. In line with non-parametric tests, the `PERFORM` dummy has a negative sign and is (weakly) significant in the random effects specification while the `SELECT` dummy is not significant in either case. The signs and significance levels of variables capturing the interaction within borrower pairs remain comparable to run 1. The exception is the interaction term  $e_{j,t-1} \times TimesPaid_{t-1}$  that is not significant in the OLS specification of model 3.<sup>22</sup>

A final aspect that could influence decisions in our setting is risk aversion. The threat of not receiving further credit might induce higher effort exertion among the more risk-averse subjects regardless of the partner choices. Therefore, as a further robustness check, we collected a survey measure for risk aversion in 5 of 8 experimental sessions. This measure is the answer to the question "Are you generally willing to take risks, or do you try to avoid risks?" taken from the German Socio-Economic Panel (SOEP).<sup>23</sup> People could answer on a scale from 0 (risk averse) to 10 (fully prepared to take risks). We rerun regression models 2 and 3 including the variable for risk preference. The results for individual decisions in run 1 and 2 are listed in table 5.6 in the Appendix. First, in all specifications the coefficient has the expected sign and is significant in most cases: The more risk averse a person is – displayed by lower values for the risk preference variable –, the higher is her estimated effort. Hence, the possibility of default and exclusion from further credits has a disciplinary effect among borrower pairs. However, the impacts of both variables concerning partner interaction ( $e_{j,t-1}$  and  $TimesPaid_{t-1}$ ) remain valid if risk aversion is included.<sup>24</sup>

---

<sup>21</sup>Conclusions are the same if we include repayments for the partner in the preceding round instead of accumulated repayments.

<sup>22</sup>In all described OLS regression models, we calculate clustered standard errors on the level of the experimental subjects. However, using clustered standard errors per matching group yields identical significance levels of the variables.

<sup>23</sup>For a description of the SOEP see Wagner et al. (2007).

<sup>24</sup>The only exception is the OLS specification of model 5, in which the coefficient of  $TimesPaid_{t-1}$  is still negative but not significant any more.

### 5.5.3 Selection into Contracts

In the SELECTION treatment, 25 out of 64 subjects (39.1%) decided to choose the individual contract. To distinguish between possible motivations for the contract choice, we compare subjects opting for individual liability to those choosing joint liability with respect to several measures calculated over run 1 (see table 5.5). Average variable values are compared on the level of experimental matching groups using two-sided WMPSR tests.

TABLE 5.5  
AVERAGES RUN 1 PER CONTRACT CHOICE

Contract Choice	Individual Liability	Joint Liability	<i>p</i> -value
Own Effort	5.54	6.10	0.383
Partner Effort	5.28	6.25	<b>0.039</b>
Partner paid (Share of Periods)	0.14	0.17	0.375
Paid for Partner (Share of Periods)	0.21	0.12	<b>0.016</b>
Risk Preference*	4.00	4.53	0.375

Significant differences are indicated by bold numbers. Mean values are compared using two-sided WMPSR tests.

\*A measure for risk aversion was collected only for 32 out of 64 subjects in the SELECT treatment.

The first variable we address is own average effort. Although subjects choosing individual liability exert lower effort on average, the difference is not significant. In contrast, average partner effort seems to be more important for the contract choice: Effort exerted by partners of subjects choosing the individual contract is some 18% lower than in case of joint liability subjects, and this difference is significant ( $p = .034$ , two-sided WMPSR test). Moreover, in line with individual effort decisions, the extent to which a subject had to stand in for the partner matters also for contract selection: Persons choosing individual liability on average had to repay the partner's loan in 21% of all active periods compared to 12% of persons choosing joint liability (the difference is significant with  $p = .016$ , two-sided WMPSR test). Contrary, the insurance effect of joint liability lending does not seem to influence contract choices strongly: Individual liability subjects on average relied less often on their partners (14% of periods in contrast to 17%), but the difference is not significant. Finally, subjects do not differ with respect to risk preferences  $p = .375$ , two-sided WMPSR test).

Summarizing, our results suggest that personal characteristics and behavior of the

borrower are only of lower relevance for the selection into contract schemes. Instead, experience with the partner seems to direct the choices: The lower a partner's average effort and the higher the frequency of partner repayments, the more likely is a person to prefer the individual credit contract.<sup>25</sup>

## 5.6 Conclusions

We have conducted a microfinance experiment in which subjects decide about effort invested into risky projects. In the first run of the experiment, all participants receive group lending contracts, in which two randomly matched subjects are jointly responsible for the repayment of the sum of both credits. While it is collectively optimal to exert the highest possible effort, it is individually rational to choose the minimum effort level. In the second run, participants are converted to individual contracts on the basis of self-selection (treatment SELECT) and repayment performance (treatment PERFORM). Both treatments are compared to a reference setting in which subjects interact under joint liability for both runs (treatment BASE).

In all treatments and both runs, high average effort levels are achieved. Thus, group lending contracts work well to foster cooperation among borrowers. Due to the high levels of cooperation under group lending, a conversion to individual contracts does not pay off from a lender's perspective in our setting. However, whereas the insurance effect of group lending dominates the disciplining effects of individual liability, cooperation within borrower groups remains below the optimal level and decreases over time.

One reason is that within-group moral hazard is still observable: in the second run, individual liability subjects choose significantly higher efforts than in the first run. Moreover, effort decisions under joint liability are highly path-dependent. The decision of a borrower to exert effort is strongly influenced by her partner's effort level. Hence, this result is in line with the literature about conditional cooperation in public goods games (see for example Fischbacher et al., 2001 and Fischbacher and Gächter, forthcoming). Furthermore, also the realized outcomes matter for effort choices. The more often a subject had to repay the partner's loan, the lower

---

<sup>25</sup>This interpretation is also confirmed by expected effort levels under joint liability in run 2 collected among half of the subjects in the post-experimental questionnaire. While persons choosing joint liability expected an average effort level of  $e_{JL,Run2} \approx 6.2$ , the value for persons opting for individual liability was  $e_{IL,Run2} \approx 5.6$ .

is her estimated effort in subsequent periods. All in all, these results suggest the importance of within-group dynamics for the functioning of joint liability lending.

A comparison of conversion mechanisms shows that removing the groups with a high repayment performance may have a negative impact on the remaining borrowers. The latter subjects tend to exert lower effort than their counterparts in the reference treatment. This suggests that the willingness to cooperate may decline after the more successful borrowers leave the program and borrowers are rematched.

Contrary, self-selection of subjects into contracts does not systematically change average effort or repayment performance in run 2. In particular, there is no sign for systematic selection of cooperative persons in and out of the joint liability scheme. Instead, partner behavior and the frequency of double repayments are more important drivers for selection into individual liability.

Our controlled laboratory experiment certainly captures only few of the aspects that determine the success of microfinance programs in the field. Yet the study gives an indication that it might be necessary for microfinance institutions to investigate within-group dynamics over time – given that in practice borrower groups typically receive more than one loan.

Furthermore, our study provided a first test of the effects of alternative conversion mechanisms on cooperation avoiding sample-selection problems emerging with field data. To connect field research and laboratory experiments might be a promising approach for further research in the microfinance context.

## 5.7 Appendix

### 5.7.1 Instructions

*Below we show the instructions for the first and second run in treatment SELECT. Instructions for the other groups and treatments were formulated in a similar way.*

#### **Instructions: General Information**

Welcome to the experiment! In this experiment you can earn money. How much depends on your decisions and the decisions of other participants.

From now on, please do not communicate with other participants. If you have a question concerning the experiment, please raise your hand! We will come to you and answer your question. If you violate these rules, we have to exclude you from the experiment and all payoffs.

In the experiment ECU is used as the currency. At the end of the experiment, your payoff in ECU is converted into Euro and paid out in cash. The exchange rate is 30 ECU = 1 Euro. In the experiment, all ECU amounts are rounded to whole numbers. The experiment consists of two runs. After the experiment, one of the two runs is randomly selected. The sum of period payoffs in this run is paid out. In addition, you receive an amount of 5.00 Euros for your participation in the experiment, which is paid out at the end regardless of the decisions. If you make a loss in the course of the experiment, this loss will be set off against the amount of 5.00 Euros.

#### **Instructions: First Run**

The following instructions refer to the first run. After the first run you will receive new instructions.

The first run consists of 12 periods. Previous to the first run, pairs of two participants are formed randomly. These pairs interact with each other during the whole first run. The identity of the participant you are interacting with is secret, and no other participant will be informed about your identity. Thus, your decisions are anonymous.

The following explanations apply to you and the other participant. The other participant faces exactly the same decision situation as you.

In the first period of the experiment you and the other participant receive a credit of 50 ECU each. The credit amounts automatically fund a risky investment project. Each participant has a personal investment project into which her credit is invested. Possible payoffs of the investment project are as follows:

- In case of success, each investment project achieves a payoff of 200 ECU.
- In case of failure, each investment project achieves a payoff of 0 ECU.

The payoff of your investment project and the payoff of the other participant's investment project are independent of each other.

Success probability of your project is influenced by your level of effort. You can choose every integer between  $A = 0$  (lowest effort) and  $A = 8$  (highest effort) as effort level (in the following abbreviated with  $A$ ).

- If you choose  $A = 0$ , the success probability of your project is 50%.
- For every unit of effort, the success probability of your project increases by 5%.
- Examples:
  - If you choose  $A = 1$ , the success probability of your project is 55%.
  - In case of  $A = 2$ , the success probability of your project is 60%.
  - ...
  - In case of the highest effort level  $A = 8$ , the success probability of your project is 90%.

For every unit of effort that you choose, you have costs of 6 ECU. (Example: If you choose  $A = 4$ , the costs of effort are  $4 \cdot 6 \text{ ECU} = 24 \text{ ECU}$ .) The costs for your chosen effort level incur regardless of the project success.

You find an overview of possible effort levels and the associated costs in the following table.

Effort level chosen	0	1	2	3	4	5	6	7	8
Success probability of the project in %	50%	55%	60%	65%	70%	75%	80%	85%	90%
Effort costs	0	6	12	18	24	30	36	42	48

After you and the other participant have chosen effort levels, payoffs of both investment projects are randomly determined. You are informed about:

- the payoff of your project (either 200 ECU or 0 ECU).
- the payoff of the other participant's project (either 200 ECU or 0 ECU).
- your chosen effort level and effort costs.
- the other participant's effort level and her effort costs.
- your period-payoff and the sum of period payoffs in this run.
- the other participant's period payoff and the sum of her period payoffs in this run.

The repayment amount for your credit is 100 ECU, the repayment amount for the other participant's credit is also 100 ECU. After payoffs of the investment projects are determined, credits have to be repaid. You and the other participant are jointly responsible for the sum of both repayment amounts ( $100 \text{ ECU} + 100 \text{ ECU} = 200 \text{ ECU}$ ). If your project was successful (your payoff = 200 ECU), the repayment for your credit (= 100 ECU) is deducted automatically from your payoff. If at the same time the other participant's project was not successful (payoff = 0 ECU), she is not able to repay her credit. In that case the other participant's repayment is also deducted automatically from your payoff.

If your project was not successful (your payoff = 0 ECU), you are not able to repay your credit. If at the same time the project of the other participant was successful (payoff = 200 ECU), your repayment is deducted automatically from the payoff of the other participant.

It is not possible to use payoffs from previous periods for credit repayments.

Your payoff in a period in which you have received a credit is determined as follows:

$$\begin{array}{r} + \text{ Payoffs of your project} \\ - \text{ Credit repayment} \\ - \text{ Effort costs} \\ \hline = \text{ Period payoff} \end{array}$$

If the sum of both repayment amounts (= 200 ECU) is repaid, you and the other participant receive a new credit of 50 ECU in the next period and the decision situation proceeds again as described.

If less than the sum of both repayments (= 200 ECU) is repaid in one period, the first run is over for you and the other participant. You and the other participant will not receive a new credit for the rest of this run and will not make any decisions.

### **Instructions: Second Run**

Welcome to the second run of the experiment! The second run of the experiment consists of 12 periods. For this run, all participants are divided into two groups, group 1 and group 2. Before the experiment starts, you can choose the group you want to belong to. If you have chosen one group, your choice is valid for the entire second run.

Participants choosing group 1 are responsible only for the repayment of their own credit during the entire second run. Otherwise the decision situation is identical to the decision situation in the first run. If a participant of group 1 is not able to repay her credit (= 100 ECU) in one period, the second run is over for this participant. In that case, the participant will not receive a new credit for the rest of the run and will not make any decisions.

Participants choosing group 2 are again – together with another participant – responsible for the sum of repayments of both credits in the second run. Analogous to the first run, pairs of two participants are formed randomly previous to run 2. These pairs will interact with each other during the whole second run.

The decision situation is identical to the decision situation in the first run of the experiment. If a pair of participants repays less than the sum of both repayment amounts (= 200 ECU) in one period, the second run is over for both participants.

In that case, both participants will not receive a new credit for the rest of this run and will not make any decisions.

With respect to the assignment of participants in group 2, it is guaranteed that no participants interact with each other that were matched in the first run of the experiment. The exception is that two participants that were matched in the first run are the only participants in group 2. If an odd number of participants chooses group 2, one participant is randomly selected and assigned to group 1.

### 5.7.2 Parametrization

In the following, it will be shown that the experimental parameters fulfil the requirements described in section 5.3. We focus on the following conditions:

1. The provision of minimum effort  $e = e_0$  is individually optimal in the one-shot game under both contract schemes. A social dilemma exists within the borrower group under joint liability:  $e_i = e_j = e_{max}$  is collectively optimal.

2. In the repeated game, full effort exertion is optimal under the individual contract.

$$E[\pi_i(e_{i,t} = e_{max})] > E[\pi_i(e_{i,t} = e_0)] \forall t.$$

3. In the repeated joint liability game, minimum effort by both players is the subgame perfect Nash equilibrium, but full effort exertion maximizes the sum of joint payoffs.

$$E[\pi_i(e_{i,t} = e_{max} | e_{j,t} = e_{max})] > E[\pi_i(e_{i,t} = e_0 | e_{j,t} = e_0)] \forall t.$$

4. An insurance effect of group lending exists as long as a sufficient level of cooperation can be maintained.

$$E[\pi_{i,JL}(e_{i,t} = e_{max} | e_{j,t} = e_{max})] > E[\pi_{i,IL}(e_{i,t} = e_{max})] > E[\pi_{i,JL}(e_{i,t} = e_0 | e_{j,t} = e_0)] \forall t.$$

Please note that in case of joint liability, we focus predominantly on symmetric strategies of the players ( $e = e_i = e_j$ ). Furthermore, we consider mainly time-invariant strategies in the repeated game.

Ad 1.

Expected payoffs for borrower  $i$  under the individual contract is denoted by  $E[\pi_i(e)] = [p_0 + p(e_i)] \cdot (R - L) - c(e_i)$ . Inserting the given parameters, we can compare expected payoffs from  $e_{max} = 8$  and  $e_0 = 0$ . Thus,

$$E[\pi_i(e_i = 8)] = (0.5 + 0.4) \cdot (200 - 100) - 6 \cdot 8 < 0.5 \cdot (200 - 100) = E[\pi_i(e_i = 0)].$$

As  $42 < 50$ , it is not optimal to exert effort in the one-shot game for the individual lender.

Under the joint liability contract, the expected payoff for borrower  $i$  accounts for

$$E[\pi_i(e_i, e_j)] = [p_0 + p_i(e_i)] \cdot [p_0 + p_j(e_j)] \cdot (R - L) + \\ [p_0 + p_i(e_i)] \cdot [1 - p_0 - p_j(e_j)] \cdot (R - 2L) - c(e_i)$$

As  $R = 2L$  in our case, the second term of the equation disappears. We calculate expected payoffs if both partners choose  $e_{max}$  and compare it to the case when both choose the minimum feasible level  $e_0$ :

$$E[\pi_i(e_i = e_{max} | e_j = e_{max})] \leq E[\pi_i(e_i = e_0 | e_j = e_0)]$$

Inserting the parameters yields  $0.9 \cdot 0.9 \cdot (200 - 100) - 6 \cdot 8 \leq 0.5 \cdot 0.5 \cdot (200 - 100) \implies 33 > 25$ . Hence, full effort exertion would be socially optimal, however, given the partner chooses  $e_{max}$ , it is optimal to switch to  $e_0$ , as  $0.5 \cdot 0.9 \cdot (200 - 100) - 0 = 45 > 33$ .

Finally, the sum of payoffs for the borrowers should not be maximal at an asymmetric effort combination  $e_0 < e^* < e_{max}$ . However, figure 5.4 shows that the sum of payoffs for the borrowers is maximized with  $e_i = e_j = 8$ .

Ad 2.

In the repeated game, expected payoffs from a given strategy crucially depend on the probability that further credit is granted in subsequent periods which is denoted by  $q$ . The expected period profit  $E[\pi_i(e_i = e)]$  has to be weighted by  $q$ , which in the individual contract is equal to the success probability of the investment project. Hence, the probability that a borrower receives a credit in the  $t$ -th period is denoted by  $q(e)^{t-1} = [p_0 + p(e_i)]^{t-1}$ . Expected payoffs of a given strategy  $e_i = e \forall t$  in which a constant effort level is chosen in all periods account for

$$E[\pi_i(e_i = e)] = \sum_{t=1}^T q^{t-1} \cdot E[\pi_{i,t}(e_{i,t} = e)].$$

Comparing  $E[\pi_i(e_i = e_{max})]$  and  $E[\pi_i(e_i = e_0)]$  yields  $\sum_{t=1}^{12} (0.9)^{t-1} \cdot 42 \leq \sum_{t=1}^{12} (0.5)^{t-1} \cdot 50 \implies 7.18 \cdot 42 > 1.99 \cdot 50 \implies 301.6 > 99.5$ .

Under the minimum effort, the probability deteriorates that the borrower receives loans in latter periods. Hence, it becomes optimal to bear the cost of full effort exertion to receive follow-up loans with a higher probability.

Ad 3.

It can be shown by backwards induction that the symmetric choice of  $e_i = e_j = e_0$  is the subgame perfect Nash equilibrium in the repeated joint liability game. Moreover, parameters have to ensure that both borrowers would improve their payoffs by switching to the maximum effort level. The argumentation is similar to 2. However, expected payoffs and the probability  $q$  of further credit change. Due to the mutual dependency of borrowers, the group is able to repay if at least one member is successful with her investment project. Assuming symmetric effort levels  $e = e_i = e_j$ , this occurs with probability  $q(e) = [p_0 + p(e)]^2 + 2 \cdot [p_0 + p(e)] \cdot [1 - p_0 - p(e)]$ . As in the individual liability case, the expected value of a strategy  $e_{i,t} = e \forall t$  for both players is

$$E[\pi_i(e_i = e)] = \sum_{t=1}^T q^{t-1} \cdot E[\pi_{i,t}(e_{i,t} = e)].$$

If full effort exertion maximizes the sum of joint payoffs, it has to be true for each borrower that  $E[\pi_i(e_{i,t} = e_{max}|e_{j,t} = e_{max})] > E[\pi_i(e_{i,t} = e_0|e_{j,t} = e_0)] \forall t$ . Inserting the present parameters confirms that this requirement is met as  $\sum_{t=1}^{12} (0.99)^{t-1} \cdot 33 \leq \sum_{t=1}^{12} (0.75)^{t-1} \cdot 25 \implies 11.36 \cdot 33 > 3.87 \cdot 25 \implies 374.9 > 96.8$ .

Furthermore, expected payoffs under full effort exertion are also higher than payoffs of an alternating strategy in which players in turn choose  $e_{i,t} = e_{max}$  and  $e_{j,t} = e_0$ . The expected probability of further credit under this strategy accounts for  $q(e) = [p_0 + p(e_{max})] \cdot [p_0 + p(e_0)] + [p_0 + p(e_{max})] \cdot [1 - p_0 - p(e_0)] + [1 - p_0 - p(e_{max})] \cdot [p_0 + p(e_0)]$ . Expected period payoff for the borrower providing full effort is  $E[\pi_i(e_i = e_{max}|e_j = e_0)] = -3$  whereas the expected period payoff for the partner with minimum effort is  $E[\pi_i(e_i = e_0|e_j = e_{max})] = 45$  (see above). This yields an average period payoff of 21, and the expected value per borrower over all periods accounts for  $\sum_{t=1}^{12} (0.95)^{t-1} \cdot 21$ , which is lower than expected payoffs under mutual full effort exertion.

Ad 4.

Under full cooperation, expected payoffs in the repeated game should be higher than under optimal behavior in individual liability. However, individual lending should be advantageous from the borrower's perspective only if low levels of cooperation can be established. Thus,

$$E[\pi_{i,JL}(e_{i,t} = e_{max}|e_{j,t} = e_{max})] > E[\pi_{i,IL}(e_{i,t} = e_{max})] > E[\pi_{i,JL}(e_{i,t} = e_0|e_{j,t} = e_0)] \forall t.$$

The chosen parameters meet this condition (see 2. and 3.), because  $11.36 \cdot 33 > 7.18 \cdot 42 > 3.87 \cdot 25$ . This reflects the insurance effect of joint liability lending: If both partners provide  $e = e_{max}$ , expected payoffs per period drop compared to individual contracts, however, vulnerability against default also becomes lower. Hence, the probability of further credits increases, and this effect dominates in our case.

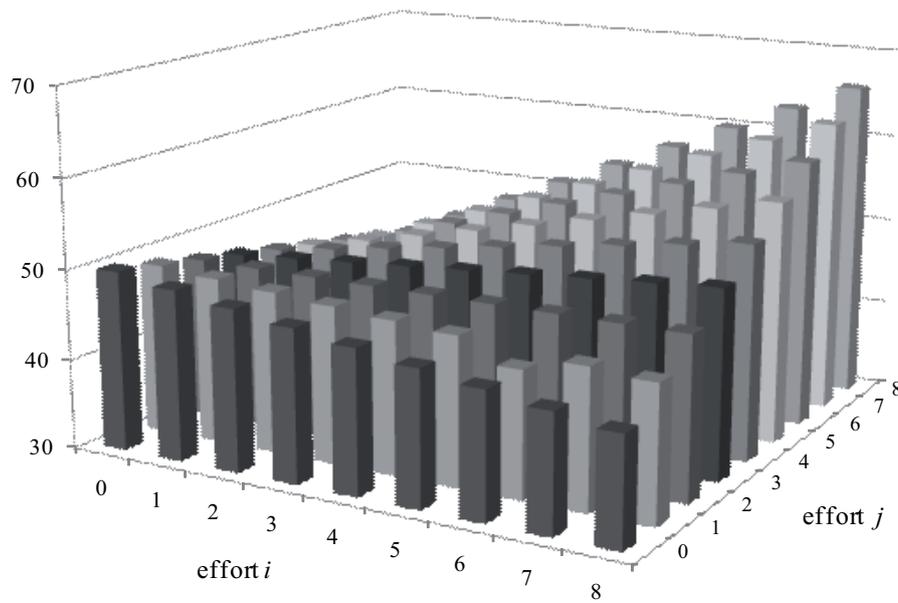
There exists a threshold for effort exertion under group responsibility below which the choice of individual liability in treatment SELECT is optimal. As was shown under 2., a subject under the individual contract can achieve expected payoffs of  $7.18 \cdot 42 = 301.6$ . A high value of cooperation would make it optimal to choose joint liability.<sup>26</sup> We calculate the threshold for the choice of the group contract by comparing  $E[\pi_{i,JL}(e_{i,t} = e_0|e_{j,t} > e_0)]$  for all feasible effort levels. For the given parameters, the joint liability contract becomes optimal only for effort levels  $e_{j,t} \geq 7$ . If the partner chooses  $e_{j,t} = 7$ , expected payoffs are  $E[\pi_{i,JL}(e_{i,t} = e_0|e_{j,t} = 7)] = \sum_{t=1}^{12} (0.93)^{t-1} \cdot 42.5 = 8.31 \cdot 42.5 = 353.2 > 301.6$ . However, a partner effort of  $e_{j,t} = 6$  makes individual liability optimal, as  $E[\pi_{i,JL}(e_{i,t} = e_0|e_{j,t} = 6)] = \sum_{t=1}^{12} (0.90)^{t-1} \cdot 40 = 7.18 \cdot 40 = 287.2 < 301.6$ . Furthermore, this threshold remains identical for all possible positive effort levels  $e_{i,t}$  of borrower  $i$ .

---

<sup>26</sup>However, after the subject has entered the contract, the social dilemma within the borrower group remains existent.

### 5.7.3 Additional Figures

FIGURE 5.4  
EXPECTED SUM OF PERIOD PAYOFFS PER BORROWER PAIR AND EFFORT  
LEVEL



Expected payoffs are calculated with parameter values  $R = 200$ ,  $L = 100$ ,  $I = 50$ ,  $p_0 = 0.5$ ,  $p(e) = 0.05 \cdot e$  and  $c(e) = 6 \cdot e$ .

### 5.7.4 Additional Results

TABLE 5.6  
DETERMINANTS OF EFFORT DECISIONS UNDER JOINT LIABILITY –  
ROBUSTNESS CHECKS

No.	4a	4b	5a	5b
Dependent Variable	$e_{i,t}$	$e_{i,t}$	$e_{i,t}$	$e_{i,t}$
Model	Run 1 RE	Run 1 OLS	Run 2 RE	Run 2 OLS
<i>Period</i>	-0.038** [0.016]	-0.021 [0.023]	-0.096*** [0.021]	-0.047 [0.033]
$e_{j,t-1}$	0.373*** [0.036]	0.547*** [0.044]	0.156*** [0.050]	0.412*** [0.111]
$TimesPaid_{t-1}$	-0.582*** [0.141]	-0.708*** [0.174]	-0.865*** [0.201]	-0.556 [0.439]
$e_{j,t-1} X TimesPaid_{t-1}$	0.057*** [0.021]	0.084*** [0.023]	0.114*** [0.031]	0.07 [0.060]
Risk Preference	-0.095** [0.037]	-0.071* [0.040]	-0.145** [0.069]	-0.09 [0.088]
SELECT			-0.192 [0.392]	0.054 [0.440]
PERFORM			-0.139 [0.310]	-0.014 [0.286]
Constant	4.635*** [0.318]	3.362*** [0.422]	7.735*** [0.629]	5.006*** [1.025]
Observations	1440	1440	968	968
Number of subjects	156	.	112	.
Wald- $\chi^2$	385	.	148.1	.
R-squared	.	0.39	.	0.21

Random effects and clustered standard errors are calculated on the level of experimental subjects. Standard errors given in brackets. \*\*\*, \*\* and \* denote significance levels of  $p < 0.01$ ,  $p < 0.05$  and  $p < 0.1$ , respectively.

# Chapter 6

## Conclusions

This thesis presents four studies that were designed to investigate the influence of non-standard preferences on agency problems in various economic settings.

Chapter 2 analyzes the role of economic inequality for trust and trustworthiness in the context of a dynamic growth game. In economies starting with equal endowments (EQ), cooperation deteriorates over time, while trust remains stable in economies starting with unequal endowments (IEQ). The reason is that during the course of the game, subjects start to condition their trusting decisions on relative wealth positions in EQ because wealth levels are interpreted as a signal for trustworthiness in previous rounds. Contrary, wealth levels in IEQ have no information value and are subsequently only of minor importance for sender decisions.

Our results suggest that the relationship between inequality and growth through the transmitter trust is not as linear and straightforward as suggested in the related empirical literature. Specifically, our results provide evidence that the source, timing and dynamics of inequality within a society may have explanatory power for its behavioral impact on prosperity and growth. Thus, new studies are required to gain insights into the interaction between changes of the income dispersion and their effects on the subsequent willingness of subjects to engage in economic transactions.

Chapters 3 and 4 focus on the role of relative comparisons for the functioning of wage schemes. In chapter 3, we use a combination of survey and compensation data from a multinational company. In line with loss aversion and inequality aversion, we find strong evidence for the importance of reference points with respect to bonus payments for both job satisfaction and performance. Negative deviations from a reference point – the managers' bonus budgets – lower job satisfaction signi-

ificantly and decrease incentives for performance. Contrary, positive deviations from the reference point have no significant effect.

An important question following from this observation is how transparent incentive systems should be. The effects of transparency can be ambiguous. On the one hand, transparency may increase procedural fairness perceptions of compensation schemes. Yet transparency can also mitigate incentives if it triggers relative social comparisons and the formation of precise expectations which may eventually be disappointed.

Moreover, in the case of the particular bonus system, a personal and a social reference point coincide. Falling short of her bonus budget yields a personal loss for the respective manager and also signals a low relative position among the other managers in her department. In our data set, it is not possible to distinguish between these two sources of discouragement. However, for the practical design of incentive systems it would be important to know the relative importance of each of those effects for our results.

In chapter 4, the impact of wage differentials on working performance is studied in a real-effort experiment. In a treatment with piece-rate wages and public wage information, we find a detrimental influence of social comparisons on effort exertion that is mainly driven by low-paid subjects. This effect is only existent if participants receive information about peer wages. Moreover, while our study provides further evidence that social comparisons matter for the incentive effect of wages, it also indicates that the form of compensation might be important: When wages are flat rather than piece-rate, performance effects under public wage information are largely mitigated.

The latter result would be consistent with the orientation on reference incomes (see Köszegi and Rabin, 2006, for a formalization of this idea). If subjects focused on a target income for the entire experiment rather than for the two separate tasks, the decrease in fixed wages would not be perceived as a loss and, subsequently, would not affect behavior. Therefore, for the choice between different forms of compensation, further research would be required to gain a better understanding about how wage expectations and reference incomes emerge in working environments.

Chapter 5 reports the results of a microfinance experiment in which subjects decide about effort invested into risky projects and are jointly responsible for credit repayment. Whereas high and relatively stable effort levels are achieved, within-

group moral hazard is still observable. Moreover, decisions are highly path-dependent because individual effort levels are related to partner efforts and partner repayments in previous rounds. A comparison of conversion schemes from group to individual lending reveals that self-selection into contracts do not decrease average effort in joint liability groups whereas after conversion based on repayment success, effort levels of borrower groups tend to be lower than in a reference treatment. The latter result suggests that the willingness to cooperate may decrease after the more successful borrowers leave the program and borrowers are rematched.

Our study points out that the dynamics of borrower behavior might influence the success of group lending contracts. Due to inherent moral hazard problems, it might become necessary for microfinance institutions to intensify monitoring of borrower groups. Yet, while our experiment allows to investigate the incentive effects of group and individual lending in a controlled environment, it captures only few of the aspects that determine the success of microfinance programs in the field. Thus, a connection of field data and laboratory experiments might be a promising approach for further research on the improvement of contracts and institutions in the microfinance context.

# Bibliography

- Abbink, K., Irlenbusch, B. and Renner, E. (2006a), ‘Group Size and Social Ties in Microfinance Institutions’, *Economic Inquiry* **44**(4), 614–628.
- Abbink, K., Irlenbusch, B. and Renner, E. (2006b), ‘Interest Rates in Group Lending: A Behavioural Investigation’, *Pacific Economic Review* **11**(2), 185–199.
- Abeler, J., Altmann, S., Kube, S. and Wibrals, M. (forthcoming), ‘Reciprocity and Payment Schemes: When Equality Is Unfair’, *Journal of the European Economic Association* .
- Ahlin, C. and Townsend, R. M. (2007), ‘Selection into and across Credit Contracts: Theory and Field Research’, *Journal of Econometrics* **136**(2), 665–698.
- Akerlof, G. A. and Yellen, J. L. (1990), ‘The Fair Wage-Effort Hypothesis and Unemployment’, *Quarterly Journal of Economics* **105**, 255–283.
- Alesina, A. and Ferrara, E. L. (2002), ‘Who trusts others?’, *Journal of Public Economics* **85**, 207–234.
- Anderson, L. R., Mellor, J. M. and Milyo, J. (2006), ‘Induced Heterogeneity in Trust Experiments’, *Experimental Economics* **9**, 223–235.
- Armendàriz de Aghion, B. (1999), ‘On the Design of a Credit Agreement with Peer Monitoring’, *Journal of Development Economics* **60**(1), 79–104.
- Armendàriz de Aghion, B. and Gollier, C. (2000), ‘Peer Group Formation in an Adverse Selection Model’, *Economic Journal* **110**, 632–643.
- Armendàriz de Aghion, B. and Morduch, J. (2000), ‘Microfinance beyond Group Lending’, *Economics of Transition* **8**(2), 401–420.

- Armendàriz de Aghion, B. and Morduch, J. (2005), *The Economics of Microfinance*, MIT Press, Cambridge.
- Banerjee, A. V., Besley, T. and Guinnane, T. W. (1994), ‘Thy Neighbor’s Keeper: The Design of a Credit Cooperative with Theory and a Test’, *Quarterly Journal of Economics* **109**(2), 491–515.
- Banerjee, A. V. and Duflo, E. (2003), ‘Inequality and Growth: What can the Data Say?’, *Journal of Economic Growth* **8**, 267–299.
- Barro, R. J. (2000), ‘Inequality and Growth in a Panel of Countries’, *Journal of Economic Growth* **5**, 5–32.
- Bellemare, C. and Shearer, B. (2007), ‘Gift Exchange within a Firm: Evidence from a Field Experiment’, CIRPÉE Working Paper 07-08.
- Bénabou, R. (1996), Inequality and Growth, in B. S. Bernanke and J. J. Rotemberg, eds, ‘NBER Macroeconomics Annual’, MIT press, Cambridge, pp. 11–74.
- Berg, J., Dickhaut, J. and McCabe, K. (1995), ‘Trust, Reciprocity, and Social History’, *Games and Economic Behavior* **10**, 122–142.
- Besley, T. and Coate, S. (1995), ‘Group Lending, Repayment Incentives and Social Collateral’, *Journal of Development Economics* **46**(1), 1–18.
- Bewley, T. F. (1999), *Why Wages Don’t Fall During a Recession*, Harvard University Press, Cambridge.
- Bohnet, I. and Kübler, D. (2005), ‘Compensating the Cooperators: Is Sorting in the Prisoner’s Dilemma Possible?’, *Journal of Economic Behavior and Organization* **56**(1), 61–76.
- Bol, J. C. (2008), ‘The Determinants and Performance Effects of Supervisor Bias’, Working Paper.
- Bolton, G. E. (1991), ‘A Comparative Model of Bargaining: Theory and Evidence’, *American Economic Review* **91**, 1096–1136.
- Bolton, G. E., Brandts, J. and Ockenfels, A. (2005), ‘Fair Procedures: Evidence from Games Involving Lotteries’, *Economic Journal* **115**, 1054–1076.

- Bolton, G. E. and Ockenfels, A. (2000), ‘ERC - A Theory of Equity, Reciprocity and Competition’, *American Economic Review* **90**, 166–193.
- Bolton, P. and Dewatripont, M. (2005), *Contract Theory*, MIT Press, Cambridge.
- Brühlhart, M. and Usunier, J.-C. (2007), ‘Verified Trust: Reciprocity, Altruism, and Randomness in Trust Games’, Working Paper.
- Buckley, E. and Croson, R. (2006), ‘Income and Wealth Heterogeneity in the Voluntary Provision of Linear Public Goods’, *Journal of Public Economics* **90**, 935–955.
- Burchett, R. and Willoughby, J. (2004), ‘Work Productivity when Knowledge of Different Reward Systems Varies: Report from an Economic Experiment’, *Journal of Economic Psychology* **25**(5), 591–600.
- Camerer, C. (2003), *Behavioral Game Theory*, Princeton University Press, Princeton.
- Camerer, C. and Loewenstein, G. (2004), Behavioral Economics: Past, Present, Future, *in* C. Camerer, G. Loewenstein and M. Rabin, eds, ‘Advances in Behavioral Economics’, Princeton University Press, Princeton and Oxford, pp. 3–52.
- Castelló-Climent, A. (2004), ‘A Reassessment Of The Relationship Between Inequality And Growth: What Human Capital Inequality Data Say?’, Working Paper 2004-15 of Instituto Valenciano de Investigaciones Económicas.
- Chan, K. S., Mestelman, S., Moir, R. and Muller, R. A. (1996), ‘The Voluntary Provision of Public Goods under Varying Income Distributions’, *Canadian Journal of Economics* **34**(1), 54–69.
- Charness, G. and Kuhn, P. (2007), ‘Does Pay Inequality Affect Worker Effort? Experimental Evidence’, *Journal of Labor Economics* **25**(4), 693–723.
- Charness, G. and Rabin, M. (2002), ‘Understanding Social Preferences with Simple Tests’, *Quarterly Journal of Economics* **117**, 817–869.
- Clark, A. E. (1999), ‘Are Wages Habit-forming? Evidence from Micro Data’, *Journal of Economic Behavior & Organization* **39**(2), 179–200.

- Clark, A. E., Frijters, P. and Shields, M. (2007), ‘Relative Income, Happiness and Utility: An Explanation for the Easterlin Paradox and Other Puzzles’, IZA Discussion Paper No. 2840.
- Clark, A. E., Kristensen, N. and Westergård-Nielsen, N. (2009), ‘Job Satisfaction and Co-worker Wages: Status or Signal?’, *Economic Journal* **119**, 430–447.
- Clark, A. E., Masclet, D. and Villeval, M.-C. (forthcoming), ‘Effort and Comparison Income: Experimental and Survey Evidence’, *Industrial and Labor Relations Review* .
- Clark, A. E. and Oswald, A. J. (1996), ‘Satisfaction and Comparison Income’, *Journal of Public Economics* **61**(3), 359–381.
- Corneo, G. and Grüner, H. P. (2000), ‘Social Limits to Redistribution’, *American Economic Review* **90**, 1491–1507.
- Corneo, G. and Jeanne, O. (2001), ‘Status, the Distribution of Wealth, and Growth’, *Scandinavian Journal of Economics* **103**(2), 283–293.
- Cull, R., Demirgüç-Kunt, A. and Morduch, J. (2009), ‘Microfinance Meets the Market’, *Journal of Economic Perspectives* **23**(1), 167–192.
- De Dominicis, L., Florax, R. J. and De Groot, H. L. F. (2008), ‘On the Empirics of Social Capital’, *Scottish Journal of Political Economy* **55**(5), 654–682.
- de Meza, D. and Webb, D. (2007), ‘Incentive Design under Loss Aversion’, *Journal of the European Economic Association* **5**(1), 66–92.
- DellaVigna, S. (2009), ‘Psychology and Economics: Evidence from The Field’, *Journal of Economic Literature* **47**, 315–372.
- Dufwenberg, M. and Kirchsteiger, G. (2004), ‘A Theory of Sequential Reciprocity’, *Games and Economic Behavior* **47**, 268–298.
- Durlauf, S. N. (2002), ‘On the Empirics of Social Capital’, *Economic Journal* **112**, F459–F479.
- Earl, P. E. (1990), ‘Economics and Psychology’, *Economic Journal* **100**, 718–755.

- Engellandt, A. and Riphahn, R. T. (2004), ‘Incentive Effects of Bonus Payments: Evidence from an International Company’, IZA Discussion Paper No. 1229.
- Falk, A. and Fischbacher, U. (2006), ‘A Theory of Reciprocity’, *Games and Economic Behavior* **54**, 293–315.
- Fehr, E., Gächter, S. and Kirchsteiger, G. (1997), ‘Reciprocity as a Contract Enforcement Device – Experimental Evidence’, *Econometrica* **64**, 833–860.
- Fehr, E. and Goette, L. (2007), ‘Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment’, *American Economic Review* **97**(1), 298–317.
- Fehr, E., Kirchsteiger, G. and Riedl, A. (1993), ‘Does Fairness Prevent Market Clearing? An Experimental Investigation’, *The Quarterly Journal of Economics* **108**(2), 437–59.
- Fehr, E. and Schmidt, K. (1999), ‘A Theory of Fairness, Competition, and Cooperation’, *Quarterly Journal of Economics* **114**, 817–868.
- Fehr, E. and Schmidt, K. (2006), The Economics of Fairness, Reciprocity and Altruism – Experimental Evidence and New Theories, in S.-C. Kolm and J. M. Ythier, eds, ‘Handbook of the Economics of Giving, Reciprocity and Altruism – Volume 1’, Elsevier, North-Holland, pp. 615–691.
- Fershtman, C. and Gneezy, U. (2001), ‘Discrimination in a Segmented Society: An Experimental Approach’, *Quarterly Journal of Economics* **116**(1), 351–377.
- Fischbacher, U. (2007), ‘z-Tree: Zurich Toolbox for Ready-made Economic Experiments’, *Experimental Economics* **10**(2), 171–178.
- Fischbacher, U. and Gächter, S. (forthcoming), ‘Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Good Experiments’, *American Economic Review* .
- Fischbacher, U., Gächter, S. and Fehr, E. (2001), ‘Are People Conditionally Cooperative? Evidence from a Public Goods Experiment’, *Economics Letters* **71**(3), 397–404.

- Forbes, K. (2000), ‘A Reassessment of the Relationship between Inequality and Growth’, *American Economic Review* **90**(4), 869–887.
- Frank, R. H. (1984a), ‘Are Workers Paid Their Marginal Products?’, *American Economic Review* **74**, 549–571.
- Frank, R. H. (1984b), ‘Interdependent Preferences and the Competitive Wage Structure’, *The Rand Journal of Economics* **15**, 510–520.
- Frank, R. H. and Sunstein, C. (2000), ‘Cost Benefit Analysis And Relative Position’, AEI-Brookings Joint Center For Regulatory Studies, Research Paper No. 00-5.
- Freeman, R. B. (1978), ‘Job Satisfaction as an Economic Variable’, *American Economic Review* **68**(2), 135–41.
- Frey, B. S., Benz, M. and Stutzer, A. (2004), ‘Introducing Procedural Utility: Not Only What, but Also How Matters’, *Journal of Institutional and Theoretical Economics* **160**, 377–401.
- Gächter, S. (2006), ‘Conditional Cooperation: Behavioral Regularities from the Lab and the Field and their Policy Implications’, CeDEx Discussion Paper No. 2006 – 03.
- Gächter, S. and Fehr, E. (2001), Fairness in the Labor Market – A Survey of Experimental Results, *in* F. Bolle and M. Lehmann-Waffenschmidt, eds, ‘Surveys in Experimental Economics. Bargaining, Cooperation and Election Stock Markets’, Physica Verlag, pp. 95–132.
- Gächter, S., Nosenzo, D. and Sefton, M. (2008), ‘The Impact of Social Comparisons on Reciprocity’, IZA Discussion Paper No. 3639.
- Galor, O. and Zeira, J. (1993), ‘Income Distribution and Macroeconomics’, *Review of Economic Studies* **60**(1), 35–52.
- Gehrig, T., Güth, W., Levati, V., Levinsky, R., Ockenfels, A., Uske, T. and Weiland, T. (2007), ‘Buying a Pig in a Poke: An Experimental Study of Unconditional Veto Power’, *Journal of Economic Psychology* **28**(6), 692–703.
- Ghatak, M. (1999), ‘Group Lending, Local Information and Peer Selection’, *Journal of Development Economics* **60**(1), 27–50.

- Giné, X., Jakiela, P., Karlan, D. and Morduch, J. (2009), ‘Microfinance Games’, Working Paper.
- Giné, X. and Karlan, D. (2007), ‘Group versus Individual Liability: A Field Experiment in the Philippines’, Working Paper.
- Glaeser, E. L. (2005), ‘Inequality’, Harvard Institute of Economic Research Discussion Paper No. 2078.
- Glaeser, E. L., Laibson, D. and Sacerdote, B. (2002), ‘An Economic Approach to Social Capital’, *Economic Journal* **112**, F437–F458.
- Glaeser, E. L., Laibson, D., Scheinkman, J. A. and Soutter, C. L. (2000), ‘Measuring Trust’, *Quarterly Journal of Economics* **115**(3), 811–846.
- Gneezy, U. and List, J. A. (2006), ‘Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments’, *Econometrica* **74**(5), 1365–1384.
- Gomez, R. and Santor, E. (2003), ‘Do Peer Group Members Outperform Individual Borrowers? A Test of Peer Group Lending Using Canadian Micro-Credit Data’, Bank of Canada Working Papers.
- Greenberg, J. (1987), ‘Reactions to Procedural Injustice in Payment Distributions: Do the Means justify the Ends?’, *Journal of Applied Psychology* **72**(1), 55–61.
- Greenberg, J. (1988), ‘Equity and Workplace Status: A Field Experiment’, *Journal of Applied Psychology* **73**(4), 606–613.
- Greenberg, J. (1990), ‘Employee Theft as a Reaction to Underpayment Inequity: The Hidden Cost of Pay Cuts’, *Journal of Applied Psychology* **75**(5), 561–568.
- Greiner, B. (2004), An Online Recruitment System for Economic Experiments, in K. Kremer and V. Macho, eds, ‘Forschung und wissenschaftliches Rechnen 2003, GWDG Bericht 63’, Ges. für Wiss. Datenverarbeitung, Göttingen, Germany, pp. 79–93.
- Greiner, B., Ockenfels, A. and Werner, P. (2007), ‘The Dynamic Interplay of Inequality and Trust – An Experimental Study’, CESifo Working Paper No. 2173.

- Greiner, B., Ockenfels, A. and Werner, P. (2009), ‘Social Wage Comparisons and Performance – A Real-Effort Experiment’, Mimeo.
- Grimm, V. and Mengel, F. (2009), ‘Cooperation in Viscous Populations – Experimental Evidence’, *Games and Economic Behavior* **66**(1), 202–220.
- Gul, F. (1991), ‘A Theory of Disappointment Aversion’, *Econometrica* **59**(3), 667–686.
- Güerker, O., Irlenbusch, B. and Rockenbach, B. (2006), ‘The Competitive Advantage of Sanctioning Institutions’, *Science* **312**, 108–112.
- Gustavsson, M. and Jordahl, H. (2008), ‘Inequality and Trust in Sweden: Some Inequalities are more harmful than others’, *Journal of Public Economics* **92**, 348–365.
- Haile, D., Sadrieh, A. and Verbon, H. A. A. (2006), ‘Cross-Racial Envy and Underinvestment in South Africa’, CESifo Working Paper No. 1657.
- Haile, D., Sadrieh, A. and Verbon, H. A. A. (2008), ‘Self-serving Dictators and Economic Growth’, *Journal of Economic Behavior and Organization* **67**, 573–586.
- Hennig-Schmidt, H., Rockenbach, B. and Sadrieh, A. (forthcoming), ‘In Search of Workers’ Real Effort Reciprocity – A Field and a Laboratory Experiment’, *Journal of the European Economic Association* .
- Herweg, F., Müller, D. and Weinschenk, P. (2008), ‘The Optimality of Simple Contracts: Moral Hazard and Loss Aversion’, BonnEcon Discussion Paper 17/2008.
- Holmstrom, B. and Milgrom, P. (1991), ‘Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design’, *Journal of Law, Economics, and Organization* **7**, 24–52.
- Jordahl, H. (2008), Inequality and Trust, in G. T. Svendsen and G. L. H. Svendsen, eds, ‘Handbook of Social Capital’, Edward Elgar, Cheltenham.
- Kahneman, D. (2003), ‘Maps of Bounded Rationality: Psychology for Behavioral Economics’, *American Economic Review* **93**(5), 1449–1475.

- Kahneman, D., Knetsch, J. L. and Thaler, R. H. (1991), ‘The Endowment Effect, Loss Aversion, and Status Quo Bias: Anomalies’, *Journal of Economic Perspectives* **5**(1), 193–206.
- Kahneman, D. and Tversky, A. (1979), ‘Prospect Theory: An Analysis of Decision under Risk’, *Econometrica* **47**(2), 263–91.
- Karlan, D. (2005), ‘Using Experimental Economics to Measure Social Capital and Predict Financial Decisions’, *American Economic Review* **95**(5), 1688–1699.
- Karlan, D. (2007), ‘Social Connections and Group Banking’, *Economic Journal* **117**, F52–F84.
- Knack, S. and Keefer, P. (1997), ‘Does Social Capital have an Economic Payoff? A Cross-Country Investigation’, *Quarterly Journal of Economics* **112**, 1252–1288.
- Kono, H. (2006), ‘Is Group Lending A Good Enforcement Scheme for Achieving High Repayment Rates? Evidence from Field Experiments in Vietnam’, Working Paper.
- Köszegi, B. and Rabin, M. (2006), ‘A Model of Reference-Dependent Preferences’, *Quarterly Journal of Economics* **121**(4), 1133–1165.
- Kube, S., Marèchal, M. A. and Puppe, C. (2007), ‘Do Wage Cuts Damage Work Morale? Evidence from a Natural Field Experiment’, Working Paper.
- Kuznets, S. (1955), ‘Economic Growth and Income Inequality’, *American Economic Review* **45**(1), 1–28.
- La Porta, R., Lopez-de-Silanes, F., Shleifer, A. and Vishny, R. W. (1997), ‘Trust in Large Organisations’, *American Economic Review* **87**(2), 333–338.
- Laffont, J.-J. and Martimort, D. (2002), *The Theory of Incentives – The Principal Agent Model*, Princeton University Press, Princeton and Oxford.
- Lazear, E. (1986), ‘Salaries and Piece Rates’, *Journal of Business* **59**(3), 405–431.
- Ledyard, J. O. (1995), Public Goods: A Survey of Experimental Research, in J. H. Kagel and A. E. Roth, eds, ‘The Handbook of Experimental Economics’, Princeton University Press, Princeton, pp. 111–194.

- Levine, D. (1998), ‘Modelling Altruism and Spitefulness in Experiments’, *Review of Economic Dynamics* **1**, 593–622.
- Madajewicz, M. (2004), ‘Joint Liability versus Individual Liability in Credit Contracts’, Columbia University Department of Economics Discussion Paper No. 0304-18.
- Mas, A. (2006), ‘Pay, Reference Points, and Police Performance’, *Quarterly Journal of Economics* **121**(3), 783–821.
- Microfinance Information Exchange (2008), ‘MicroBanking Bulletin’, Volume 17.
- Morduch, J. (1999), ‘The Microfinance Promise’, *Journal of Economic Literature* **37**(4), 1569–1614.
- Ockenfels, A., Sliwka, D. and Werner, P. (2009), ‘Bonus Payments and Reference Point Violations’, Mimeo.
- Page, T., Putterman, L. and Unel, B. (2005), ‘Voluntary Association in Public Goods Experiments: Reciprocity, Mimicry and Efficiency’, *Economic Journal* **115**, 1032–1053.
- Persson, T. and Tabellini, G. (1994), ‘Is Inequality Harmful for Growth?’, *American Economic Review* **84**(3), 600–621.
- Pfeffer, J. and Langton, N. (1993), ‘The Effect of Wage Dispersion on Satisfaction, Productivity, and Working Collaboratively: Evidence from College and University Faculty’, *Administrative Science Quarterly* **38**, 382–407.
- Prendergast, C. (1999), ‘The Provision of Incentives in Firms’, *Journal of Economic Literature* **37**(1), 7–63.
- Rabin, M. (1993), ‘Incorporating Fairness into Game Theory and Economics’, *American Economic Review* **83**, 1281–1302.
- Rabin, M. (1998), ‘Psychology and Economics’, *Journal of Economic Literature* **36**, 11–46.
- Rivas, M. F. and Sutter, M. (2008), ‘Wage Dispersion and Workers’ Effort’, University of Innsbruck, Faculty of Economics and Statistics Working Paper 15-2008.

- Ros, J. (2000), *Development Theory and the Economics of Growth*, The University of Michigan Press, Ann Arbor.
- Schaubroeck, J., May, D. R. and Brown, F. W. (2000), 'Procedural Justice Explanations and Employee Reactions to Economic Hardship: A Field Experiment', *Journal of Applied Psychology* **79**(3), 455–460.
- Sliwka, D. (2007), 'Trust as a Signal of a Social Norm and the Hidden Costs of Incentive Schemes', *American Economic Review* **97**(3), 999–1012.
- Stiglitz, J. E. (1990), 'Peer Monitoring and Credit Markets', *World Bank Economic Review* **4**(3), 351–366.
- Torgler, B., Schaffner, M., Schmidt, S. L. and Frey, B. S. (2008), 'The Power of Positional Concerns', CREMA Working Paper Series 2008-7.
- Tversky, A. and Kahneman, D. (1991), 'Loss Aversion in Riskless Choice: A Reference-Dependent Model', *The Quarterly Journal of Economics* **106**(4), 1039–61.
- van Tassel, E. (1999), 'Group Lending under Asymmetric Information', *Journal of Development Economics* **60**(1), 3–25.
- Wagner, G. G., Frick, J. R. and Schupp, J. (2007), 'The German Socio-Economic Panel Study (SOEP): Scope, Evolution and Enhancements', DIW Berlin, The German Socio-Economic Panel (SOEP), SOEPpapers No.1.
- Werner, P. (2009), 'The Dynamics of Cooperation in Group Lending – A Microfinance Experiment', Mimeo.
- Werner, S. and Ones, D. S. (2000), 'Determinants of Perceived Pay Inequities: The Effects of Comparison Other Characteristics and Pay-System Communication', *Journal of Applied Social Psychology* **30**(6), 1281–1309.
- Zak, P. J. and Knack, S. (2001), 'Trust and Growth', *Economic Journal* **111**, 295–321.