The Impacts of Social Preferences and Biases on Individual and Group Decision Making: Four Essays in Experimental Economics

Inaugural-Dissertation
zur Erlangung des akademischen Grades eines Doktors
der Wirtschafts- und Sozialwissenschaften
(Dr. rer. pol.)

der Heinrich-Heine-Universität Düsseldorf

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

vorgelegt von: Dipl.-Volksw. Holger Andreas Rau
aus: Heilbronn

Erstgutachter:	Prof. Dr. Hans-Theo Normann
Zweitgutachter:	Prof. Dr. Justus Haucap

Datum der letzten Prüfung: 20.12.2012

Preface

The research for this thesis has been conducted at the Goethe University Frankfurt, the Heinrich-Heine University Düsseldorf, and the Düsseldorf Institute for Competition Economics (DICE). The thesis has strongly been influenced by and benefited from discussions with professors and colleagues, as well as presentations at various national and international conferences, seminars, and workshops. I am very grateful to my colleagues: Volker Benndorf, Georg Clemens, Beatrice Pagel, Matthias Heinz, Steffen Juranek and to all who supported my work.

In particular, I would like to thank my supervisor Hans- Theo Normann who encouraged and supported me, for his excellent advice and patience when writing this thesis. I also want to thank my second supervisor Justus Haucap.

Finally I would like to particularly thank my parents and my sister for patiently supporting me during the whole time when writing this thesis. This thesis is dedicated to my deceased mother.

Contents

Chapter 1

Introduction

This thesis analyzes the impacts of social preferences and biases on decision making in different settings of individual and group decision making. During the last 30 years the analysis of biases in subjects' decision-making processes and the investigation of so-called "other-regarding preferences" became more and more important in Economics. In Experimental Economics there is overwhelming evidence of these cases where people deviate from standard neoclassical theory.

On the one hand, a bulk of papers report that in the presence of interaction partners, subjects' care about social preferences. That is, they do not only focus on their own monetary outcome and rather incorporate fairness issues in their utility functions which relate to other market participants. For instance a bulk of papers document that subjects in experiments focus on payoff differences between other subjects and themselves (e.g. Fehr and Schmidt (1999); Bolton and Ockenfels (2000)). Further papers emphasize that decision makers show reciprocal behavior and reward or punish their interaction partners (see Berg et al. (1995); Fehr et al. (1998)). There is also evidence that people are affected by "warm glow" (i.e., they receive a positive utility from behaving altruistically) and voluntarily contribute to public goods and charities (see Kahneman et al. (1986); Isaac and Walker (1988); Fehr and Gächter (2000)).

On the other hand there are papers which report that people use heuristics when deciding under uncertainty which often leads to decision biases (e.g. Tversky and Kahneman

(1974)). That is, subjects focus on reference points and their risk attitudes are not stable in investment decisions. They also show a pronounced degree of loss aversion and often feel regret when observing losses in their investment decisions (Kahneman and Tversky (1979)). There is also evidence which demonstrates that these preferences also differ for subjects' gender (for a survey see Croson and Gneezy (2009)). Hence, the presence of subjects' deviating from standard predictions have crucial impact on the efficiency of different market outcomes.

The first part of the thesis therefore focuses on the impacts of social preferences on the outcomes of individual decision making, whereas the second part deals with its consequences in group decision making.

Chapter 2 entitled **"Do Women Behave More Reciprocally than Men - Gender Differences in Real Effort Dictator Games"** (co-authored with Matthias Heinz and Steffen Juranek) analyzes gender differences in reciprocal behavior in an individual-decision-making environment. In our experiment we apply a modified dictator game where recipients have to earn the pot to be divided with a real-effort task. As the recipients move before the dictators, their effort decisions resemble the first move in a trust game. Depending on the recipients' performance, the size of the pot is either high or low. We compare this real-effort treatment to a baseline treatment where the pot is a windfall gain and where a lottery determines the pot size. In the baseline treatment, reciprocity cannot play a role. This chapter documents that female dictators show reciprocity and significantly decrease their taking-rates in the real-effort treatment. The treatment effect is larger when female dictators make a decision on recipients who successfully generated a large pot compared to the case where the recipients performed poorly. By contrast the chapter concludes, there is no treatment effect with male dictators, who generally exhibit more selfish behavior.

Chapter 3 entitled **"Step-Level Public Goods: Experimental Evidence"** (co-authored with Hans- Theo Normann) analyzes coordination and the efficiency of step-level public goods in the presence of social preferences. The chapter is an experiment to investigate how the order of moves (simultaneous vs. sequential) and the number of step levels (one vs. two) affects public-good provision in a two-player game. We find that the sequential order of moves significantly improves public-good provision and payoffs, even though second movers often punish first movers who give less than half of the threshold contribution. The additional second step level which is not feasible in standard Nash equilibrium leads to higher contributions but does not improve public-good provision and lowers payoffs. Based on an experimental data set in the literature, we calibrate Fehr and Schmidt's (1999) model of inequality aversion to make quantitative predictions. We find that actual behavior fits remarkably well with several predictions in a quantitative sense.

The second part of the thesis incorporates two chapters studying the effects of competition and decision-making biases in group-decision-making environments.

Chapter 4 entitled **"Competition in the Workplace: An Experimental Investigation"** (co-authored with Volker Benndorf) analyzes competitional effects and reciprocal behavior in a gift-exchange experiment where two workers are hired by the same employer. In the competitional treatment two employees simultaneously choose their effort. Whereas in the baseline treatment competition cannot occur since there is only one employee per employer. We find that in the competitional treatment employers implicitly set "tournament incentives" by rewarding employees who chose higher effort levels than their co-workers. The chapter concludes that employees' effort levels increase significantly faster which can be explained by Imitation learning. We find that employers decrease their wage payments per unit of effort exerted over time when employing two workers.

Chapter 5 entitled **"The Disposition Effect in Individual and Team Investments: Experimental Evidence"** analyzes biases in group decision making when jointly deciding about a portfolio. In this regard the chapter analyzes the disposition effect comparing single investors to teams of two investors. The baseline treatment is a replication of Weber and Camerer's (1998) laboratory experiment where subjects decide about investing in six risky assets. In the group-decision-making treatment two subjects decide about a joint portfolio. The baseline data replicates Weber and Camerer's findings: single investors are prone to a disposition effect. The magnitude of the disposition effect is higher for teams, who realize a significant higher proportion of capital gains than single investors.

At the same time teams realize a smaller proportion of stocks in the loss domain. The chapter concludes that these findings can be explained by a stronger impact of regret on the realization of capital losses for team members.

Part I

Individual Decision Making

Chapter 2

Do Women Behave more Reciprocally than Men? Gender Differences in Real Effort Dictator Games*

2.1 Introduction

Our study analyzes gender differences in reciprocal behavior. We apply a modified dictator game based on Cherry et al.'s (2002) study where dictators had to do a real-effort task before deciding on the money amount to be dictated to recipients. In contrast to most standard dictator games (e.g. Kahneman et al. (1986), Eckel and Grossman (1996) and Dana et al. (2006)), Cherry et al. (2002) report that 95% of their dictators behaved in line with the standard neoclassical prediction of selfish maximization of their own monetary income.[1]

Our study modifies the Cherry et al. (2002) setup in that we make the recipients (rather than the dictators) conduct the real-effort task. Since recipients move before dictators, their effort decisions resemble the first move in a trust game (Berg et al. (1995)). That is:

*The research of this chapter is part of a joint project with Matthias Heinz and Steffen Juranek.

[1]Further, some dictator games report that dictators decrease their taking-rates due to increased anonymity (e.g. Hoffman et al. (1996), Koch and Normann (2008)).

if the recipients do not trust dictators, they should not invest effort. Note here a crucial difference to the trust game is that our game does not include an outside option, i.e. all first-movers have to send their money to the second-movers.[2] According to our setup's resemblance to a trust game, we expect gender differences in terms of dictators' trustworthiness in our setup. We therefore hypothesize that female dictators will show increased reciprocity to recipients who have worked. We expect them to reciprocate a good working performance by strongly reducing their taking-rates if the recipients performed well. This is motivated by the trust game literature on gender differences, which reports that female second-movers more often reciprocate first-movers' offers by sending back positive amounts (e.g. Croson and Buchan (1999), Chaudhuri and Gangadharan (2003), Snijders and Keren (2004), and Buchan et al. (2008)).[3]

Our setup is most closely related to Ruffle's (1998) study on "tipping" behavior in the ultimatum game and in the dictator game. Here, the recipients also endogenized the money to be divided by doing a real-effort task, before dictators decided about the split. The main difference in our study is that we focus on gender differences in reciprocal behavior.[4] Related is also Oxoby and Spraggon (2008), who demonstrate that dictators make significantly lower offers when recipients did a real-effort task. However, there are two crucial differences: The authors used a standard dictator game as a control treatment. In contrast, we use a baseline treatment, where dictators can take money from recipients in a windfall environment and, furthermore, we focus on gender differences. Our results not only successfully replicate Ruffle's (1998) and Oxoby and Spraggon's (2008) findings; we also establish that female dictators are affected by reciprocity and show different magnitudes of reciprocal behavior depending on recipients' working performance.

[2] That setup ensures that dictators have to decide about both: successful and low performing recipients.

[3] There also exist papers which report gender effects in standard dictator games (e.g. Eckel and Grossman (1998), Bolton and Katok (1995), Andreoni and Vesterlund (2001) and Dufwenberg and Muren (2006)). For a complete survey see Croson and Gneezy (2009).

[4] Carlsson et al. (2010) also study gender effects in a dictator game with a real-effort task. However in this setup the dictators do the real-effort task. The authors do not find a significant gender difference.

2.2 Experimental Design

We use a modified dictator game with two differences to the standard game. The first difference is that the size of the pot which the dictators decide about is not constant. The second difference is that there are two stages in both the *Windfall* and the *Real Effort* treatment. As in Cherry et al. (2002), there is a first stage (*money-generation stage*) in which the size of the pot is determined. Dictators only decide in the second stage (*allocation stage*) about the allocation of the money.

In the *Real Effort* treatment the participants were randomly assigned to two groups of equal size and split between separate rooms, rooms A and B. In the *money-generation* stage, subjects in room A (the recipients) had the opportunity to take a quiz which consisted of 20 questions taken from the *Graduate Record Examination* test.[5] Depending on their results, we allocated money to the subjects such that subjects who answered at least 13 questions correctly[6] were given 10€, otherwise they received 5€. Subjects knew that they had 20 minutes to complete the quiz. As we corrected the tests, dictators in room B had to wait for approximately 30 minutes and we provided them with coffee and cake. The *allocation* stage of the *Real Effort* treatment randomly matched subjects in room A with those in room B. Neither subjects in room A nor subjects in room B knew the identities of their partners. Furthermore subjects in room A were informed about the amount of money we allocated, which depended on their results in the quiz. Individuals in room B were not told about the exact result of their interaction partner. They only learned whether the recipient generated 5€ or 10€.[7] Every subject in room B dictated in a one-shot dictator game a split of the wealth to the recipients in room A. Subjects in room A were informed by the experimenter about the allocation decision and the final earnings which they received according to the dictator's decision. Afterwards both the dictators and the recipients had to complete a short survey.[8] Finally subjects A and B were paid out at the

[5]These questions are based on basic arithmetic concepts (e.g. algebra, geometry and data analysis).

[6]The threshold of 13 correct answers was calibrated based on a pilot session of the *GRE* test among the undergraduate students of a seminar at Frankfurt University.

[7]Before making their decision dictators were also told that the recipients knew (before they started the real-effort task) that a dictator will decide on the allocation.

[8]Statistical analysis of this data revealed that only *gender* and *age* were significant variables, i.e. older people

end of the experiment.

Altogether we conducted five sessions of the *Real Effort* treatment. Four of these sessions had the structure explained above. In the remaining session we checked whether dictators are sensitive to overconfidence. We therefore provided them with a copy of the exam questions taken by those in group A (which may have induced positive or negative reciprocity, depending on the dictator's self-assessment of his or her own ability). Dictators were given the chance to have a look at the exam questions for 10 minutes, before they were asked to estimate the number of questions they would have solved correctly. After making their allocation decision we asked them to do the test, to check whether they had overestimated their own performance. However, we did not find significant differences in dictator-takings, if we compare this session with the other four *Real Effort* sessions (Mann-Whitney test, p-value = 0.980). Thus we pooled the data from this session with the data from the other four *Real Effort* sessions.

The *Windfall* treatment was identical to the *Real Effort* treatment except that the recipients did not have the opportunity to take the quiz. Instead, the pot size was determined randomly. Subjects had to draw a lottery ticket worth either 5€ or 10€, and they had a 50% chance of winning either a low or high stake size. In order to keep both treatments comparable, dictators also had to wait for 30 minutes and we provided them with coffee and cake.

We used ORSEE (Greiner, 2004) to recruit the subjects among the undergraduate students at Frankfurt. A total of 352 subjects attended the experiment. We ran five sessions of each treatment. A session lasted about 75 minutes and on average subjects earned 8.75€ including a 5€ show-up fee. To maintain transparency, all subjects were informed about the whole procedure of the experiment.

and women are more likely to take lower amounts. In contrast cultural differences (e.g. people's religion) were not significant at all.

2.3 Hypotheses

Our *Real Effort* treatment resembles a trust game, thus dictators benefit from recipients investments in effort. We hypothesize that this triggers reciprocal behavior. Since Croson and Buchan (1999) report that female second-movers are highly sensitive to reciprocal behavior, we expect female dictators in *Real Effort* to take lower amounts compared to the *Windfall* treatment. The trust game literature only reports increased reciprocity for female second-movers, i.e. we do not expect that male dictators are as sensitive to reciprocal behavior as female dictators.

Hypothesis 1. *Female dictator-taking will be lower in the Real Effort treatment compared to the Windfall treatment. In contrast, we do not expect male dictators to decrease their taking-rates by the same amount.*

However, there exists a second level at which reciprocity may play a role. Depending on recipients' performance, the size of the pot is either high or low in the *Real Effort* treatment, whereas, in the baseline treatment, the size of the pot is randomly determined. Dictators know in the *Real Effort* treatment that successful recipients will receive 10€ whereas less well-performing recipients get only 5€. Since the gender literature emphasizes reciprocity for women, we expect that female dictators especially reward successful recipients, i.e. if recipients generate 10€, dictators will take lower amounts.

Hypothesis 2. *In contrast to male dictators, female dictators in the Real Effort treatment take less from recipients who generate the large pot, i.e. in Real Effort there will be a stake size effect.*

From Hypothesis 1 and 2 it follows that female dictators care about recipients' performance. Thus positive reciprocity should additionally matter, when recipients worked and successfully generated a large pot. Furthermore, we expect that this will lead to a greater difference between female dictators' taking-rates when they decide about high-pot-size recipients in the *Real Effort* compared to the *Windfall* treatment. In contrast, negative reciprocity might reduce their generosity if recipients fail to get a high pot. That is: the difference in taking-rates between female dictators (in *Windfall* and *Real Effort*) will be larger when deciding about 10€ recipients compared to the case of 5€ recipients. Based on the evidence from gender trust games, men care much less about reciprocity. Thus we do not expect that men will show different magnitudes of reciprocity.

Hypothesis 3. *The difference in female dictators' taking-rates between the Real-Effort and the Windfall treatments will be larger if recipients generated 10€ compared to the case where only 5€ was achieved. This effect should hold only for female dictators.*

2.4 Results

This section starts with a brief outline of the average results. Afterwards, we analyze the gender differences and test our hypotheses. As our experiments are one-shot interactions, we count each participant as one observation in the statistical analysis. We report two-sided p-values and non-parametric Mann-Whitney tests throughout. Section 4.1 briefly summarizes dictator-taking in the *Windfall* and the *Real Effort* treatment (here we do not yet distinguish gender).

2.4.1 Dictator-taking: Average results

Table 2.1 presents the means of dictators' taking-rates in our two treatments separated into groups of 5€ and 10€ pot sizes. The standard deviations are in parentheses.

Pot size	Windfall	obs.	Real Effort	obs.	Avg.	obs.
5 €	71.30 (22.42)	43	69.13 (24.90)	45	70.19 (23.61)	88
10 €	76.75 (21.50)	40	70.17 (24.20)	48	73.16 (23.12)	88
Avg.	73.93 (22.02)	83	69.67 (24.41)	93	71.68 (23.35)	176

Table 2.1: Mean of taken amounts (*Windfall* and *Real Effort* treatment)

On average, the real-effort task marginally triggers dictators' reciprocity, i.e. dictators in the *Windfall* treatment take 73.93% compared to dictators in the *Real Effort* treatment who only take 69.67%. Nonetheless this small difference is statistically not significant (p-value = 0.223). There is also no treatment effect, if we focus on the 5€-pot (p-value = 0.709). Though if we concentrate on the 10€-pot, we find that dictators in *Real Effort* take 70.17% compared to dictators in the *Windfall* treatment who take 76.75% (one-sided p-value = 0.078). We therefore confirm the results of Ruffle (1998) and Oxoby and Spraggon (2008). Furthermore in the *Windfall* treatment we do not find a stake size effect (p-value = 0.238), this confirms Forsythe et al.'s (1994) findings. The same is true for the *Real Effort* treatment (p-value = 0.694). The brief analysis showed that dictators were prone to reciprocal behavior when recipients worked and successfully generated the large pot.

2.4.2 Dictator-taking: Gender effects

We now analyze dictator-taking, and separate the choices by gender in order to test Hypotheses 1-3. Table 2.2 presents male and female dictator taking-rates in the *Windfall* and *Real Effort* treatment.

Gender	Stake size	Windfall	obs.	Real Effort	obs.	Avg.	obs.
males	5€	68.73 (24.40)	15	74.21 (27.73)	24	72.10 (26.31)	39
males	10€	77.27 (21.62)	22	76.52 (25.65)	25	76.87 (23.38)	47
Avg.	-	73.81 (22.56)	37	75.39 (26.44)	49	74.71 (24.72)	86
females	5€	72.68 (21.62)	28	63.33 (20.33)	21	68.67 (21.38)	49
females	10€	76.11 (22.59)	18	63.26 (20.92)	23	68.90 (22.35)	41
Avg.	-	74.02 (21.82)	46	63.30 (20.40)	44	68.68 (21.70)	90

Table 2.2: Mean of taken amounts in our two treatments, split by gender and both stake sizes

In order to test Hypothesis 1, we compare females' taking-rates in the *Windfall* and *Real Effort* treatment. Since reciprocity plays an important role in the *Real Effort* treatment, Hypothesis 1 predicts that female dictators will be strongly affected by the fact that recipients have worked.

Testing Hypothesis 1 we find a significant treatment effect for female dictators, i.e. the average taking-rate of female dictators is 63.30% in *Real Effort* compared to 74.02% in *Windfall* (p-value $= 0.021$). Therefore we reject the null hypothesis that female dictators take the same amount from recipients who did a real-effort task compared to the case where recipients received a windfall gain. We thus find strong support for Hypothesis 1. If we analyze male decisions, we do not find that the real-effort task stimulates reciprocal behavior, i.e. male dictators take 73.81% in *Windfall* and 75.39% in *Real Effort* (p-value = 0.720). Interestingly male dictator-taking is very stable. They always take around 75%. Furthermore this effect holds for all of our sessions.[9] This also emphasized by Figure 2.1 which presents diagrams comparing the cumulative distribution functions (CDF) of male and female dictators in our two treatments. The left diagram shows the *Windfall* treatment and the right diagram the *Real Effort* treatment.

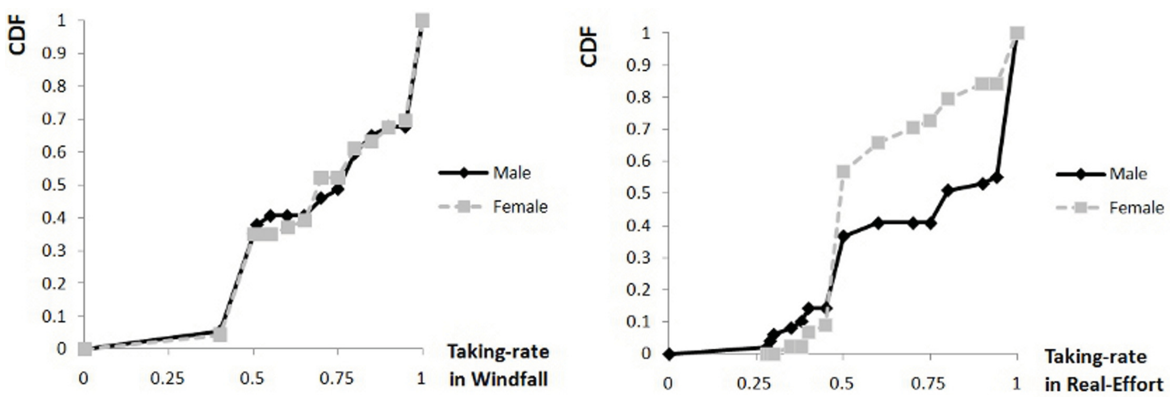

Figure 2.1: Gender effects in the *Windfall* (left diagram) and *Real Effort* treatment (right diagram)

First: male and female CDFs do not differ at all when dictators decide about windfall money (KS-Test, Max. D = 0.062, p-value = 1.000). However, there exist crucial gender differences in the CDFs when dictators decide about money which has been generated

[9] We also ran three sessions (with 96 subjects) of the *Real Effort* treatment where dictators did not have to wait 30 minutes, but decided immediately. This was done in order to control for possible time effects. Even though male dictators decreased their takings by 10 percent points (note there is great variance in the data: Some male dictators took 0% and others took 100%) we find no statistical support for a difference (p-value = 0.249). Furthermore our results show that female dictators take exactly the same amounts in both variants of the *Real Effort* treatment (p-value = 0.891). We thank an anonymous referee for pointing out this issue.

by a *Real Effort* task (KS-Test, Max. D = 0.326, p-value = 0.011). It is remarkable that 57% of female dictators choose the equal split decision as opposed to only 37% of male dictators. Furthermore only 16% of female dictators take the whole pot from the recipients. This stands in strong contrast to 45% of male dictators who choose the 100% taking-rate.

Result 1. *Comparing the Windfall with the Real Effort treatment we find significant gender differences, i.e. in Real Effort, female dictators take considerably smaller amounts compared to Windfall. In contrast, male dictators do not reduce their taking-rates at all.*

Table 2.2 shows that in the *Real Effort* treatment female dictators do not care about recipients' performance: They take 63.33% from 5€ recipients and 63.26% from 10€ recipients (p-value = 0.852). Thus we cannot reject the null hypothesis which postulates that female dictators do not care about recipients' performance. We therefore have to discard Hypothesis 2. Focusing on male dictator-taking in the *Real Effort* treatment, it appears that they also do not care about recipients' performance, i.e. they take 74.21% from recipients who generated a small pot and 76.52% from recipients who generated the large pot (p-value = 0.833).

Result 2. *In the Real Effort treatment, female as well as male dictators do not take smaller amounts from successful recipients who generated the large pot.*

In order to test Hypothesis 3 we now compare the treatment effect generated for the 5€ pot with the treatment effect for the 10€ pot. Focusing on female dictator-taking for the 5€ pot, we find that they take 72.68% in the *Windfall* treatment and 63.33% in the *Real Effort* treatment. However, this difference is not significant (p-value = 0.154).[10] If we focus on the large pot, we find that female dictators in the *Windfall* treatment take 76.11% compared to 63.26% in the *Real Effort* treatment. Thus the difference in taking-rates is larger when deciding about recipients who generated a large pot. In contrast

[10]Note there exists weak significance for a one sided p-value, i.e. p-value = 0.077.

to the 5€ pot, this difference is weakly significant (p-value = 0.069).[11] We therefore reject the null hypothesis that the difference in female dictators' taking-rates between 5€ recipients and 10€ recipients is not different. Thus we find support for Hypothesis 3. Interestingly females take nearly the same amounts from 5€ and 10€ recipients in *Real Effort*. Therefore, it cannot be that they were influenced by negative reciprocity in the case of 5€ recipients. Nevertheless, the difference in female taking-rates is larger if recipients generated a large pot. This is due to the fact that female dictators in *Windfall* take a larger amount from 10€ recipients compared to 5€ recipients.[12] Male dictators do not show different magnitudes of reciprocity. That is: taking-rates from 5€-recipients in *Windfall* and *Real Effort* are not different (p-value = 0.540). The same is true for the treatment difference in taking-rates for the 10€ recipients (p-value = 0.947).

Result 3. *The difference in female taking-rates, caused by the real-effort task is higher if recipients generated 10€ compared to the case where only 5€ was achieved. In contrast, male dictators do not show this behavior.*

2.5 Discussion

Do women behave more reciprocally than men? The answer is yes.

We analyzed a modified dictator game with a real-effort task (based on Cherry et al. (2002)) where dictators were asked to dictate a money amount which was generated by the recipients. Our results show that women significantly decrease taking-rates when the recipients generated the money (to be divided) by a real-effort task instead of a lottery task. Furthermore, female dictators decreased taking-rates more strongly if recipients generated a large pot compared to the opposed case, where recipients only generated a small pot. In contrast, male dictators did not show reciprocal behavior at all, i.e. they did

[11] Note there exists a significant difference for a one sided p-value, i.e. p-value = 0.035.

[12] Probably they do not grant a large pot to recipients because the money was endogenized by pure chance.

not lower their taking-rates in the environment of the real-effort task.

If we do not focus on gender, the general results show that dictators are sensitive to the real-effort task, however, the result depends on whether the recipient generated a large pot. That is: dictators only significantly lower taking-rates when the recipient was successful and generated 10€. Thus we confirm the results of Oxoby and Spraggon (2008) who argue that dictators are influenced by property rights legitimization.[13] Further, our results are in line with Ruffle (1998) who points out that dictators treat low performing recipients as if they did not work at all. Our study emphasizes real-effort's impacts on gender differences due to reciprocal behavior. That is: we extend Ruffle's (1998) and Oxoby and Spraggon's (2008) studies and give an explanation for their findings.

It is interesting that we only find a significant effect for female dictators. Thus our study sheds new light on gender differences in reciprocal behavior driven by a real-effort task. Our paper therefore provides important new insights as an explanation for Ruffle's (1998) and Oxoby and Spraggon's (2008) results. These findings seem to provide valuable new insights in terms of other-regarding preferences induced by a real-effort task.[14] For the future, it seems to be promising to uncover further gender differences in setups with real effort tasks.

[13] However we extend their framework by applying a treatment where dictators decide about windfall money which was won by recipients in a lottery. This enables us to emphasize the role of reciprocity induced by a real-effort task. Note there also exist other studies which emphasize the importance of endowments' origins (e.g. Mittone and Ploner (2006), Cherry and Shogren (2008)

[14] Fehr and Schmidt (1999), Bolton and Ockenfels (2000), and Falk and Fischbacher (2006) highlight the importance of other regarding preferences in their models.

References

[1] Andreoni, J., Vesterlund, L., (2001). “Which is the Fair Sex? Gender Differences in Altruism.” The Quarterly Journal of Economics 116, 293-312.

[2] Berg, J., Dickhaut, J., McCabe, K. (1995). “Trust, Reciprocity and Social History.” Games and Economic Behavior 10, 122-142.

[3] Bolton, G., E., Katok, E., (1995). “An Experimental Test for Gender Differences in Beneficial Behavior.” Economics Letters 48, 287-292.

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

[5] Buchan, R., Croson, R., T., A., Solnick, S., (2008). “Trust and gender: An examination of behavior and beliefs in the Investment Game.” Journal of Economic Behavior & Organization 68, 466-467.

[6] Carlsson, F., Haoran, H., Martinsson, P., (2010). “Windfall vs. Earned Money in the Laboratory: Do They Affect the Behavior of Men and Women Differently?” University of Gothenburg - Working Paper Series 478.

[7] Chaudhuri, A., Gangadharan, L., (2003). “Gender Differences in Trust and Reciprocity.” Department of Economics - Working Paper Series 875, The University of Melbourne.

[8] Cherry, T., L., Frykblom, P., Shogren, J., F., (2002). “Hardnose the Dictator.” American Economic Review 92, 1218-1221.

[9] Cherry, T., L., Shogren, J., F., (2008). “Self-interest, sympathy and the origin of endowments.” Economics Letters 101, 69-72.

[10] Croson R., Buchan, N., (1999). “Gender and Culture: International Experimental Evidence from Trust Games.” American Economic Review 89, 386-391.

[11] Croson R., Gneezy, U., (2009). “Gender Differences in Preferences.” Journal of Economic Literature 47, 448-474.

[12] Dana, J., Cain, D., M., Dawes, R., M., (2006). “What you don’t know won’t hurt me: Costly (but quiet) exit in dictator games.” Organizational Behavior and Human Decision Processes 100, 192-201.

[13] Dufwenberg, M., Muren, A., (2006). “Gender composition in teams.” Journal of Economic Behavior & Organization 61, 50-54.

[14] Eckel, C., Grossman, P., J., (1996). “Altruism in Anonymous Dictator Games.” Games and Economic Behavior 16, 181-191.

[15] Eckel, C., Grossman, P., J., (1998). “Are Women less Selfish than Men? Evidence from Dictator Experiments.” The Economic Journal 108, 726-735.

[16] Falk, A., Fischbacher, U., (2006). “A theory of reciprocity.” Games and Economic Behavior 54, 293-315.

[17] Fehr, E., Schmidt, K., M., (1999). “A Theory of Fairness, Competition, and Cooperation.” The Quarterly Journal of Economics, 114, 817-868.

[18] Forsythe, R., Horowitz, J., L., Savin, N., E., Sefton, M., (1994). “Fairness in Simple Bargaining Experiments.” Games and Economic Behavior 6, 347-369.

[19] 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, 79-93.

[20] Hoffman, E., McCabe, K., Smith, V., (1996). “Social Distance and Other-Regarding Behavior in Dictator Games.” American Economic Review 86, 653-660.

[21] Kahneman, D., Knetsch, J., L., Thaler, R., (1986). “Fairness and the Assumptions of Economics.” The Journal of Business 59, 285-300.

[22] Koch A., Normann, H., T., (2008). “Giving in Dictator Games: Regard for Others or Regard by Others?” Southern Economic Journal 75, 223-231.

[23] Mittone., L., Ploner, M., (2006). “Is It Just Legitimacy of Endowments? - An Experimental Analysis of Unilateral Giving.” CEEL Working Paper 02/2006.

[24] Oxoby R., J., Spraggon, J., (2008). “Mine and yours: Property rights in dictator games.” Journal of Economic Behavior & Organization 65, 703-713.

[25] Ruffle, J., B., (1998). “More is Better, But Fair is Fair: Tipping in Dictator and Ultimatum Games.” Games and Economic Behavior 23, 247-265.

[26] Snijders C., Keren, G., (2004). “Do You Trust? Whom Do You Trust? When Do You Trust?” Working Paper, Utrecht University.

Appendix

Experimental Instructions: Real Effort

You are participating in a scientific experiment, thus read the whole instructions carefully.

For attending the experiment you will be paid a show-up fee of 5€. However, you will have the possibility to earn more during the experiment. You will be paid out only after having the whole experiment completed.

- It is vital for the course of the experiment that you do not communicate with other participants
- Please switch off your cell phone. If you have any questions at any time, please raise your hand and we will come over to you

Your participation in the experiment is absolutely anonymous. During the experiment and afterwards, no identities will be revealed. In order to guarantee this, you will be allocated a randomly generated participant's Id. This Id consists of a letter (A or B) and a number (e.g.: "13"). The letter establishes whether you are type A or B in the experiment. The number identifies you as an individual. Make sure to keep your Id in mind during the whole experiment.

Course of the Experiment

During the experiment you will be allocated a participant of the opposite type. Each participant of type A has to allocate a sum of money between him and another participant B. The amount to be shared is determined by the results of a test that the participant of type B has previously taken.

Instructions for participants of type A

- Please go to room 4.201 and sit down at one of the places prepared for you.
- Please be a little patient
- Soon you will receive an envelope. It will contain a slip of paper with a sum of money, either 5€ or 10€. This amount depends on the results of a test that your corresponding participant of type B has taken (More about this later)

Please decide about the split of this amount between you and your corresponding participant in the experiment. Write down your decision on the slip of paper. Please note that the split can be any amount which adds up to the generated total amount. That is, it may involve not only rounded euros (e.g. 2€ and 3€). For instance, "Cent"-Amounts (e.g. 3.95€ and 6.05€) are also possible. Here are some examples of possible splits:

(i) Example - Total amount: 10€

- I keep 4€
- I give B 6€

(ii) Example - Total amount: 10€

- I keep 10€
- I give B 0€

(iii) Example - Total amount: 5€

- I keep 2.50€
- I give B 2.50€

(iv) Example - Total amount: 10€

- I keep 0.01€
- I give B 4.99€

Please note that these money splits will be actually paid, together with your show-up fee, at the end of the experiment. The amount of money that can be split depends directly on the results of a test taken by the participant of type B. This test consisted of questions that examine the mathematical and analytic skills of the participant. The participant had 20 minutes to answer the questions.

The maximum score was 20 points.

- For 0- 12 points 5€ was allocated.
- For 13- 20 points 10€ was allocated.

Subsequently, we will give you a questionnaire. Please complete it. Be patient until everyone has finished. We will then accompany you to the payment station to get your money.

Instructions for participants of type B

Please go to room 3.202 and sit down at one of the places prepared for you.

Soon your corresponding participant of type A will receive an envelope. It will contain a slip of paper with a sum of money, either 5€ or 10€. This amount depends on the results of a test that you will have taken (more about this later). This amount must be split between your corresponding participant and you. Please note that these money splits will be actually paid, together with your show-up fee, at the end of the experiment. The amount of money that can be split depends directly on the results of the test taken by you.

This test contains questions that examine your mathematical and analytic skills. You have **20 minutes** to answer the questions.

The maximum that could be scored was 20 points.

- For 0- 12 points 5€ was allocated.
- For 13- 20 points 10€ was allocated.

After your corresponding participant of type A has decided how to split the money we will tell you how you performed in the test and what the of money was. Subsequently, we will give you a questionnaire. Please complete it. Please be patient until everyone has finished. We will then accompany you to the payment station to get your money.

Experimental Instructions: Windfall

Instructions for participants of type A

The amount of money that you have to split between yourself and your corresponding participant of type B is determined by a lottery. Your corresponding participant draws a ball containing a folded slip of paper. On the slip, either “5€” or “10€” is written on it. This establishes the amount of money that you have to allocate.

After you have decided on the allocation between you and your corresponding participant we will give you a questionnaire. Please complete it. Please be patient until everyone has finished. We will then accompany you to the payment station to get your money.

Instructions for participants of type B

Please go to room 3.202 and sit down at one of the places prepared for you.

Soon your corresponding participant of type A will receive an envelope. It will contain a slip of paper with a sum of money, either 5€ or 10€. This amount is determined by a lottery. And it was previously drawn by you (more about this later). This amount must be split between your corresponding participant and you. Please note that this sum of money will be actually paid, together with your show-up fee, at the end of the experiment. The amount of money that can be split depends on the results of a draw that has previously been carried out by you. Subsequently we will give you a questionnaire. Please complete it. Please be patient until everyone has finished. We will then accompany you to the payment station to get your money.

Chapter 3

Step-Level Public Goods: Experimental Evidence*

3.1 Introduction

Public goods often have a step-level character, that is, the public good is provided only if some minimum threshold of contributions (or provision point) is met. Examples include the building of a bridge or a dike. Also, charities have properties of step-level public goods if the underlying production of the public good is subject to non-convexities (see Andreoni, 1998).

Our paper makes two contributions to the literature on public goods with step levels. First, we analyze whether sequential contributions as opposed to simultaneous decisions improve public good provision. Second, we analyze if an additional threshold, which is not feasible in standard Nash equilibrium and where the public good is provided at a higher level, improves public good provision.

*The research of this chapter is part of a joint project with Hans-Theo Normann.

The issue of sequential vs. simultaneous decisions is the subject of a substantial and growing literature. Following the theory contributions by Andreoni (1998), Hermalin (1998), Vesterlund (2003) and Andreoni (2006), researchers have analyzed *leading by example* in experiments. If a first mover gives an example that is mimicked by the followers, sequential contributions to the public good may be superior to simultaneous decisions. This will particularly be the case when a first mover is better informed about the return to contributions allocated to the common endeavor (Hermalin, 1998; Vesterlund, 2003) or about the quality of a charity (Andreoni, 2006).[1]

We study sequential vs. simultaneous decisions in a step-level game with two players and with complete information. For such a setting, one would at first expect a sequential-move game to seem superior to a simultaneous-move setting. A threshold public-good game is foremost a coordination game. With simultaneous moves there are multiple equilibria; coordination failures may occur and, moreover, the public good is not provided in all equilibria. With sequential moves, there is a unique subgame perfect equilibrium in which the public good is provided. Hence, coordination and therefore public good provision should be more frequent with sequential moves. There is, however, an aspect of sequential decision making that may reduce its alleged superiority. In the unique subgame perfect equilibrium with selfish players, the first mover contributes such that a best responding follower merely breaks even by meeting the threshold with her contribution. In other words, the first mover actually gives a *bad* example by contributing *less* than the followers. In an experiment, this may reduce the alleged efficiency of the sequential-move setting: players who try to exploit this first-mover advantage risk being punished by second movers who do not best respond but contribute zero to the public good. If such behavior occurs frequently, the higher efficacy of the sequential-move game will not materialize. Based on a calibrated model (see below), we hypothesize that the efficiency enhancing effect dominates so that sequential moves improve public good provision.

[1]The experimental literature on these issues (which we review in detail below) includes Erev and Rapoport (1990), Potters et al. (2005, 2007), Güth et al. (2007), Gächter et al. (2010a, 2010b), Figuières et al. (2010).

Now consider our second extension, the introduction of a second threshold. The general logic of multiple threshold public goods is that no return is obtained unless contributions meet the first level; and after this level, no additional return is earned until the second provision point is met. Multiple step levels have rarely been analyzed before (see Chewning et al., 2001; and Hashim et al., 2011, which we discuss below), but they seem realistic in many circumstances. For example, a charity may raise donations for building a hospital in a deprived area, which will typically require a minimum level of contributions. Adding a further specialized unit to the hospital may be subject to a minimum threshold just as building the main unit and, accordingly, the charity may announce two thresholds. Further examples include a public radio or TV station which may transmit more than one program, corresponding to multiple thresholds. Public bridges or highways may be built with one, two or more lanes. Finally, any kind of public good may be provided at various quality levels and the production of these quality levels may be subject to non-convexities, suggesting multiple thresholds.

The interaction of the two thresholds and the order of moves can be hypothesized as follows. In our experiments, first movers in the sequential-move game may aim for the second threshold since this yields higher payoffs—provided the threshold is met. Since such first-mover contributions must be higher than those required to meet the first threshold, second movers do not feel exploited and therefore do not punish first movers by making zero contributions. However, the second threshold is *not* a Nash equilibrium with selfish players. Given one player aims at the second threshold by contributing a high amount, the best response of a second player is to contribute low such that the first level only is met. Thus, with standard preferences the second level is not an equilibrium (with both simultaneous and sequential moves). However, when players have Fehr and Schmidt (1999) preferences, the second threshold is a Nash equilibrium—and meeting the second threshold is of course efficient. In any event, even if some second movers exploit those opponents who aim for the second level, the public good is still provided at least at the first level and so no efficiency loss occurs. In other words, behaviorally, the existence of a second threshold might make it more likely that the *first* threshold will be met. We thus hypothesize that the second step level improves public good provision.

Our main findings regarding the two treatment variables are as follows. Sequential contribution decisions significantly improve public good provision, even though second movers regularly punish first movers who contribute too little. This is in contrast to Gächter et al. (2010a) who find the opposite result, however, in an entirely different setting (see below). Coordination rates and payoffs are higher whereas contributions are not higher with sequential moves. The existence of a second threshold causes significantly higher contributions but this does not result in higher public good provision. To the contrary, payoffs are significantly lower when there are two step levels.

Our paper also makes *quantitative* predictions for our experiment based on a fully calibrated Fehr and Schmidt (1999) model. Whereas Fehr and Schmidt's (1999) model of inequality aversion has been used frequently in the literature, the predictions are almost always of a qualitative nature ("if players are sufficiently inequality averse, abc is an equilibrium"). We will calibrate Fehr and Schmidt's (1999) model on a (joint) distribution of the inequality parameters which we take from an experiment in the literature, and we will make exact quantitative predictions ("w percent of the first movers will contribute x"; or "given a first-mover contribution of y, the public good will be provided in z percent of the cases").[2]

We find that the calibrated Fehr and Schmidt (1999) model makes remarkably accurate quantitative predictions, but it also fails in two cases. The calibrated Fehr and Schmidt (1999) model predicts second-mover behavior (given first-mover behavior) in the sequential variant extremely well. Specifically, it accurately predicts the frequency of second-mover decisions (contribute such that the step level is met vs. punish first movers by contributing zero). The prediction regarding the first movers fails. First movers should

[2]Fehr, Kremhelmer and Schmidt (2008) also provide an analysis based on a calibration of Fehr and Schmidt's (1999) model. Their calibration is based on a two-type categorization (40 percent fair players and 60 percent standard players). See below.

anticipate (or learn) that second movers punish low contributions and thus always make the payoff-equalizing contribution. However, only slightly more than one-third of them do so. First movers behaving too greedily, as has been observed in previous experiments (e.g., Huck et al., 2001). The calibrated Fehr and Schmidt (1999) model also predicts well in the case with simultaneous-move contributions where some players contribute whereas others do not. Finally, the model rather precisely predicts the share of first movers who trust second movers by making a high contribution in the sequential two-threshold case. Here, the prediction regarding the second movers fails, as they exploit first movers significantly more frequently than predicted.

3.2 Literature Review

There are two major strands of the literature pertinent to our paper. The first literature is about simultaneous vs. sequential order of moves in public-good games. The second literature concerns public-good experiments with step-level character in general and, specifically, the small literature on experiments with more than one threshold.[3]

As mentioned in the introduction, several researchers have analyzed leading by example theoretically. Andreoni (1998) examines the efficiency of leadership giving. The paper provides an explanation of how seed money, from a group of "leadership givers," generates additional donations. In Hermalin (1998), a first mover may be better informed about the return to contributions allocated to the common endeavor. Therefore, she may plausibly give an example to followers who rationally mimic the first mover's behavior. Vesterlund (2003) shows that in the presence of imperfect information on a charity's quality, an announcement strategy which leads to a sequential provision mechanism may be optimal.

[3]Rondeau, Poe and Schulze (2005) compare the relative performance of VCM and Provision Point Mechanisms (PPM). They analyze large groups of subjects with simultaneous moves and employ a refund role, and they find that the PPM increases contributions and is generally more efficient than the VCM.

An increasing experimental literature has been triggered by these theory contributions. Following Hermalin (1998) and Vesterlund (2003), Potters et al. (2005) study an experimental voluntary contribution mechanism (VCM) where some donors do not know the true value of the good. The authors conclude that sequential moves result in higher contributions of the public good. They also have a treatment where the sequencing of choices emerges endogenously. Potters et al. (2007) report that the leading-by-example approach depends on whether there is incomplete information in the experiment. This explains why some experiments have not found sequential moves to be superior (Andreoni, Brown and Vesterlund, 2002) while Potters et al. (2005) did. Our experiments differ to those of Potters et al. (2005, 2007) in that we do not include information asymmetries, and we do not employ the VCM.[4, 5, 6]

Gächter et al. (2010a) is also related to our study. They experimentally study the effects of a simultaneous vs. a sequential choice mode in a test of Varian's (1994) VCM model. Varian models a public-good setting where sequential contributions are predicted to be lower than simultaneous. In their experimental test of Varian's (1994) model, Gächter et al. find that sequential contributions are indeed lower than simultaneous-move contributions, although the difference in aggregate contributions across the two move

[4]Serra-Garcia et al. (2010) extend the analysis of Potters et al. (2007) by comparing an "action" treatment where a better informed first mover can give an example to a "word" treatment where the better informed player can verbally communicate the state of the world.

[5]Figuières et al. (2010) provide evidence that the "leadership effect" vanishes over time when subjects are randomly ordered in a sequence that differs from round to round. Gächter et al. (2010b) analyze the characteristics of effective leaders in a simple leader-follower public-good game. In a sequential VCM, they find that efficiency depends on the leaders' social preferences. Gächter and Renner (2005) demonstrate that there is a positive correlation between contributions of designated leaders (who act as first movers) and second movers' contributions. Teyssier (2009) analyzes a sequential VCM where the second mover's choice is binary (restricted to the first-mover's choice and zero). Güth et al. (2007) find evidence that leadership in the VCM setup is more efficient when the first mover has exclusion power.

[6]There also exist field experiments which demonstrate the efficiency of sequential designs. List and Lucking-Reily (2002) find that increasing seed money from 10 percent to 67 percent of the campaign goal lead to a sharp increase of subjects' contributions. Soeteven (2005) report evidence from field experiments in churches. He finds that people donate significantly more when total contributions are publicly observable. See also Huck and Rasul (2011) who analyze fund raising when donations are matched by a lead donor with field data. Their findings emphasize that linear matching schemes raise the total donations. In field setting without sequential moves, Glazer and Konrad (1996) report that donations to charities can also be interpreted as subjects' instrument to signal group membership.

structures is not as great as predicted, in part because second movers punish first movers who free ride in the sequential variant. While this is in contrast to our results, note that one of the major differences to our approach is that the authors test the Varian (1994) model, whereas we study a step-level setup. Even though we observe similar punishing behavior, the sequential-move variant is more efficient in our data.

Erev and Rapoport (1990) were the first to study simultaneous vs. sequential moves in a step-level public-good game with discrete choices. In their experiments, at least three of five players must contribute their endowments for the public good to be provided. Actions are minimal contribution sets, MCS, such that players either zero contribute or invest their whole endowment. They find that, with sequential-move choices, information about previous non-cooperative choices only is more effective in public good provision than information about previous cooperative choices. The main differences to our experiment are the discrete action space and the number of players (two in our case). Discrete contributions in a step-level game have also been studied in Schram et al. (2008) where at least three of five or seven players need to contribute their full endowment to meet the threshold. Coats and Neilson (2005) and Coats et al. (2009) add refund policies to this setting. Both studies use groups of four players and analyze sequential moves, and Coats et al. (2009) furthermore analyzes a simultaneous-move variant. The authors conclude that sequential moves are superior to the simultaneous-move case. The refund mechanism (which we do not study) stimulates efficiency.[7] Very few papers exist which analyze the efficiency of simultaneous versus continuous step-level public good designs (Cadsby and Maynes, 1998a; 1998b). In contrast to our setup the authors investigate group designs with refund mechanisms. Cadsby and Maynes (1998a) report that female participants make higher contributions to the threshold public good game than male ones. In Cadsby and Maynes (1998b) the authors find that students make contributions similar to the strong free-riding equilibrium compared to nurses who cycled around the efficient

[7] Less relevant to our study is Cadsby and Maynes (1999) which analyzes a simultaneous-move framework with a refund policy. The authors compare a continuous contribution mechanism with MCS. Their results show that continuous contributions increase efficiency. The refund guarantee also encourages provision.

threshold equilibrium.

The literature on public-good games with multiple step levels is much smaller.[8] Chewning et al. (2001) have a five-player experiment with one, two, three or five step levels. Compared to the baseline with one step level, treatments with multiple levels sometimes keep the social optimum constant and lower the Nash equilibrium contributions, sometimes—as in our case—they increase the group optima contributions but leave the Nash equilibria unchanged. In their treatment with two thresholds, a *lower* threshold is added and this threshold is feasible with the endowment of a single subject so that no cooperation at all is needed to reach this threshold. In this treatment, contributions drop compared to the baseline (where players need to corporate in order to reach the higher threshold), but, unsurprisingly, free riding is prevented throughout. When adding additional thresholds (which are higher compared to the baseline threshold), contributions decline even more sharply and may even result in averages below the lowest threshold. In our experiments, we find that the second threshold increases contributions but lowers welfare. Recently, Hashim et al. (2011) analyze a game with five levels and five players. Related to Croson and Marks' (1998) study of the effect of information feedback with one step level, they vary information feedback about other members' contributions to a subsample of group members. Results show improvements in coordination when information targeting is used. Providing information randomly does not improve coordination and eventually degrades towards free-riding over time. The authors argue that a random information provision approximates strategies used in practice for educating consumers about digital piracy, information targeting may be useful.

[8] Rauchdobler et al. (2010) study how different thresholds affect contributions in a VCM variant. Thresholds differ between $T = 0$ and $T = 57$, however, there is always only one threshold at a time. Moreover, higher thresholds do not improve efficiency per se here but merely serve as a minimum target for players which may be imposed exogenously or endogenously.

3.3 Experimental Design and Procedures

In our experiments, there are two players, player 1 and player 2, who each have a money endowment $e = 10$. They can make a voluntary contribution, c_i, to the public good, where $0 \leq c_i \leq e$.

In half of our treatments, there is *one threshold* for the provision of the public good. If the sum of contributions is at least 12, this yields an additional payoff of 10 to both players. Any contributions between 1 and 11 and beyond 12 are wasted. More formally, if x_i denotes player i's monetary payoff, then

$$x_i = \begin{cases} e - c_i + 10 & \text{if } c_1 + c_2 \geq 12 \\ e - c_i & \text{if } c_1 + c_2 < 12 \end{cases}$$

The other treatments involve an additional *second threshold* of 18. If $c_1 + c_2 \geq 18$, both players receive an additional payment of 5. That is, in these treatments, we have

$$x_i = \begin{cases} e - c_i + 15 & \text{if} \quad 18 \leq \; c_1 + c_2 \\ e - c_i + 10 & \text{iff} \quad 12 \leq \; c_1 + c_2 < 18 \\ e - c_i & \text{if} \quad c_1 + c_2 < 12 \end{cases}$$

Since $2e > 18$, both thresholds of the public good are feasible, but, due to $e < 12$, no player can meet the threshold on her own. Further, because $2 \cdot 10 > 12$ and $2 \cdot 15 > 18$, the provision of the public good at both provision points is collectively rational. Note that the return on contributing one Euro at each of the two levels is the same.

We have four treatments, labeled SIM_1, SIM_2, SEQ_1, and SEQ_2. The SIM labels refer to treatments where the two players make their decisions simultaneously whereas decisions are made sequentially (with player 1 moving first) in the SEQ treatments. The second treatment variable is the number of the thresholds (one or two). Table 3.1 summarizes our 2×2 treatments design.

		Order of moves	
		simultaneous	sequential
Step levels	one	SIM_1	SEQ_1
	two	SIM_2	SEQ_2

Table 3.1: *Treatments*

Subjects play this game over 10 periods. The payoffs of the above game were denoted in Euros in the experiments (so that the exchange rate was one to one). In each period, subjects were endowed with $e = 10$ Euros but the payoff of only one randomly chosen period was paid at the end of the experiment. (See also the instructions in the Appendix.)

We have three entirely independent matching groups per treatment. Each experimental session contained only one matching group. The size of the sessions or matching groups varied between 10 and 18 subjects. (We control for session size in our data analysis below). In each session and each period, subjects were randomly matched into groups of two players. In the SEQ treatments, also the roles of first and second movers were also random.

The subject pool consists of students from the University of Frankfurt from various fields. In total, we had 160 participants who earned on average 11.3 Euros. The experiment was programmed in z-Tree (Fischbacher, 2007). Sessions lasted about 60 minutes.

3.4 Predictions

Assumptions

We now derive the one-shot Nash equilibrium predictions for this public-good game. In addition to standard Nash predictions (selfish players who maximize their own monetary payoff), we will use Fehr and Schmidt's (1999) model, henceforth F&S. In their model, players are concerned not only about their own material payoff but also about the difference between their own payoff and other players' payoffs. Assumption 1 defines the two-player variant of their model.

Assumption 1. *Players' preferences can be represented by the utility function* $U_i(x_i, x_j) = x_i - \alpha_i \max[x_j - x_i, 0] - \beta_i \max[x_i - x_j, 0],\ x_i,\ x_j = 1, 2, i \neq j.$

Here, x_i and x_j denote the monetary payoffs to players i and j, and α_i and β_i denote i's aversion towards disadvantageous inequality (envy) and advantageous inequality (greed), respectively. Standard preferences occur for $\alpha = \beta = 0$. Following F&S, we assume $0 \leq \beta_i < 1$.

Using the specific functional forms of the step-level public good game for x_i above, we can write the F&S utilities as a function of contributions directly, so that we obtain $U_i(c_i, c_j)$. For the treatments with one step level, we obtain

$$U_i(c_i, c_j) = 10 - c_i + 10\chi_1 - \alpha_i \max[c_i - c_j, 0] - \beta_i \max[c_j - c_i, 0] \tag{3.1}$$

whereas, for the two-step-levels treatments, we get

$$U_i(c_i, c_j) = 10 - c_i + 10\chi_1 + 5\chi_2 - \alpha_i \max[c_i - c_j, 0] - \beta_i \max[c_j - c_i, 0], \tag{3.2}$$

$c_i,\ c_j = 1, 2; i \neq j$. The χ_k are indicator functions indicating whether a step level has been reached. We have $\chi_1 = 1$ iff $c_1 + c_2 \geq 12$ and $\chi_2 = 1$ iff $c_1 + c_2 \geq 18$.

Using this model, we will make *quantitative* predictions. We fully calibrate the F&S model using the joint distribution of the α and β parameters observed in Blanco, Engelmann and Normann (2011).[9] For each subject, they derive an α_i from rejection behavior

[9] We could have also calibrated the F&S model with decisions from our subjects. For our purposes, however, this is redundant as we are not interested in the individual-level consistency of decisions—this is the topic

in the ultimatum game and a β_i from a modified dictator game.[10] See Table 3.2. On average, $\alpha = 1.18$ and $\beta = 0.47$. There are no significant differences between the α distribution Blanco, Engelmann and Normann (2011) elicit and the one assumed in Fehr and Schmidt (1999). The β distribution differs significantly; however, one can argue that distributions still roughly compare and do not differ outlandishly.

Why rely on the specific α_i-β_i distribution of Blanco, Engelmann and Normann (2011)? While Fehr and Schmidt (1999) derive distributions for these parameters based on data from previous ultimatum-game experiments, here we need the *joint* distribution of the parameters. We are not aware of any joint distribution of inequality-aversion parameters for the Fehr and Schmidt model with the exception of Fehr, Kremhelmer and Schmidt (2008) who assume that there are 60 percent players with $\alpha = \beta = 0$ and 40 percent fair types with $\alpha = 2$ and $\beta = 0.6$—which seems too coarse for our purposes. Following Blanco, Engelmann and Normann (2011), several papers have used the same (or very similar) games to elicit the Fehr and Schmidt parameters. See for instance Dannenberg et al. (2007), Teyssier (2009), and Kölle and Sliwka (2011). The use of this joint distribution seems promising as it successfully predicts outcomes in several games (ultimatum game, sequential-move prisoner's dilemma, public-good game) in Blanco, Engelmann and Normann (2011) which have a similar complexity as the present game.

Assumption 2. *Players' inequality parameters are drawn from the joint α-β distribution in Table 3.2. This distribution is common knowledge. Players know their own type but not the type of the other player.*

of Blanco, Engelmann and Normann (2011). Moreover, asking subjects to play two more games, each with 21 strategy-method type of decisions, is not free (longer duration and grater complexity of the experiment, less salience of the main game). Hence, we decided to take a distribution from the literature rather than to derive our own.

[10] In Blanco, Engelmann and Normann's (2011) modified dictator game, dictators choose between 20-0 and equitable outcomes ranging from 0-0 to 20-20 (all denoted in £ (GBP)). A player i who is indifferent between payoff vectors (20, 0) and (x_i-x_i) has $\beta_i = 1 - x_i/20$.

Subject	α_i	β_i	Subject	α_i	β_i	Subject	α_i	β_i
1	0	0	22	0.409	0.175	42	0.929	0.8756
2	0	0.025	23	0.409	0.175	43	1.5	0.025
3	0	0.525	24	0.409	0.175	44	1.5	0.375
4	0	0.525	25	0.409	0.175	45	1.5	0.525
5	0	0.625	26	0.409	0.325	46	1.5	0.725
6	0	0.725	27	0.409	0.525	47	1.5	0.825
7	0	0.775	28	0.409	0.525	48	1.5	0.975
8	0	0.875	29	0.409	0.625	49	1.5	1
9	0	0.975	30	0.409	0.675	50	2.833	0.275
10	0.026	0	31	0.611	0.025	51	2.833	0.475
11	0.026	0	32	0.611	0.175	52	2.833	0.575
12	0.026	0.175	33	0.611	0.275	53	2.833	0.675
13	0.026	0.725	34	0.611	0.375	54	4.5	0
14	0.088	0.625	35	0.611	0.525	55	4.5	0
15	0.167	0.825	36	0.611	0.575	56	4.5	0.025
16	0.269	0.475	37	0.611	0.675	57	4.5	0.425
17	0.269	0.525	38	0.611	0.725	58	4.5	0.525
18	0.269	0.775	39	0.611	0.725	59	4.5	0.625
19	0.269	1	40	0.929	0.475	60	4.5	0.775
20	0.409	0	41	0.929	0.025	61	4.5	0.875
21	0.409	0.125						

Table 3.2: Blanco et al.'s (2011) joint α and β distribution

Sequential moves, one threshold

We start with the sequential-move variant with one threshold (SEQ_1). In the subgame perfect Nash equilibrium of this treatment, a second mover (S) with standard preferences will best respond to the first mover's (F) contribution, c_F, by choosing zero if $c_F < 2$ and by contributing $12 - c_F$ if $c_F \geq 2$. Anticipating this, the first mover will choose her payoff-maximizing contribution, which is $c_F = 2$.

Next, consider players whose preferences and beliefs are consistent with Assumptions 1 and 2. Even if $c_F \geq 2$, second movers with F&S preferences might choose $c_S = 0$ if the payoff inequality implied by c_F becomes too big. For $c_F \in [2, 6]$ and facing the decision between contributing $12 - c_F$ and zero, the second mover either obtains $U_S(12 - c_F, c_F) = 8 + c_F - \alpha_i(12 - 2c_F)$ or $U_S(0, c_F) = 10 - \beta_i c_F$. We find that $U_S(12 - c_F, c_F) > U_S(0, c_F)$

iff

$$c_F \geq \frac{2(1+6\alpha)}{1+2\alpha+\beta} \equiv \tilde{c}_F. \tag{3.3}$$

The $\tilde{c}_F$ in (3) is the *minimum acceptable first-mover contribution* for a given set of individual inequality parameters. Any contribution at least as high as $\tilde{c}_F$ will be met by $c_S = 12 - c_F$ and will result in the public good being provided. Any contribution lower than this threshold will face $c_S = 0$ as the second mover's best reply. Intuitively, $\tilde{c}_F$ is increasing in α and decreasing in β.

	First-mover contribution				
Second-mover contribution	$c_F = 2$	$c_F = 3$	$c_F = 4$	$c_F = 5$	$c_F = 6$
$c_S = 12 - c_F$ (PG level 1 provided)	21.3%	37.7%	67.2%	83.6%	100%
$c_S = 0$ (PG not provided)	78.7%	62.3%	32.8%	16.4%	0%
expected first-mover payoff	10.13	10.77	12.72	13.36	14.00

Table 3.3: Predicted second-mover responses conditional on first-mover choices and the resulting expected first-mover monetary payoff in the SEQ treatments

Based on our Assumption 2, we now predict the frequencies of public good provision as a function of c_F. For each player in that data set (see Table 3.3), we determine the $\tilde{c}_F$ as in (3). For subject #1 with $\alpha = \beta = 0$, for example, we obtain $\tilde{c}_F = 2$ as the minimum acceptable first-mover contribution, whereas subject #58 with $\alpha = 4.5$ and $\beta = 0.525$ has $\tilde{c}_F = 5.32$ as the minimum acceptable first-mover contribution and will thus only accept $c_F = 6$. Doing this for all subjects in Blanco, Engelmann and Normann (2011) allows us to predict how many players in *our* experiment will (not) provide the public good as a function of c_F.

Table 3.3 shows the results of this calibration. In contrast to the game of players with standard preferences, the likelihood of public good provision is strictly below 100 percent as long as $c_F < 6$. Table 3.3 also reveals that the expected monetary payoff of a risk neutral first mover monotonically increases in c_F and is maximized for $c_F = 6$ (the expected payoff from choosing $c_F = 0$ is 10). As $c_F < 6$ results in a lower likelihood of public good provision, lower payoffs, and greater payoff inequality, both selfish and inequality averse first movers will choose $c_F = 6$ in the perfect Bayesian equilibrium of this game. Thus we have

Proposition 1. *For treatment* SEQ_1, *the standard model predicts* $c_S = 0$ *if* $c_F < 2$, $c_S = 12 - c_F$ *if* $c_F \geq 2$ *and* $c_F = 2$ *for the first movers. The calibrated F&S model predicts the frequencies of second-mover responses as in Table 3, and* $c_F = 6$ *for the first movers.*

Sequential moves, two thresholds

Next, consider the *sequential-move variant with two thresholds* (SEQ_2). If the first mover contributes $c_F \leq 6$, the analysis is as above. But in the two-level game, the first mover may also choose her contribution in the range $c_F \in [8, 10]$ in order to make the second level feasible.

Players with standard preferences will not provide the public good at the second level in the subgame perfect equilibrium. Given $c_F \in [8, 10]$, second movers will respond with $c_S = 12 - c_F$ (yielding a monetary payoff of $8 + c_F$) but not with $c_S = 18 - c_F$ (which would yield $7 + c_F$). By backward induction, first movers will not choose $c_F \in [8, 10]$ but $c_F = 2$, as in the game with one step level. The second threshold is irrelevant in the subgame perfect equilibrium with standard preferences.

Now assume F&S players and begin with the second movers. With $c_F \in [8, 10]$, the second mover may choose $c_S = 18 - c_F$, $c_S = 12 - c_F$, or $c_S = 0$. Since $U_S(12 - c_F, 0) > U_S(0, c_F)$ for $c_F \in [8, 10]$, we can restrict the second-mover choices to $c_S = 18 - c_F$ and

$c_S = 12 - c_F$. First suppose $c_F = 8$. If the second mover chooses $c_S = 18 - c_F = 10$, we have $U_S(10, 8) = 15 - 2\alpha_i$. If she chooses $c_S = 12 - c_F = 4$, we have $U_F(4, 8) = 16 - 4\beta_i$. We obtain $U_S(10, 8) < U_S(4, 8)$ iff $1 - 4\beta + 2\alpha > 0$. This condition holds for 60.7 percent of the subjects in Blanco, Engelmann and Normann (2011). That is, if $c_F = 8$, the public good will be provided at level one with 60.7 percent probability and with 39.3 percent probability at level two. Then consider $c_F = 9$. If $c_S = 18 - c_F$, we obtain $U_F(9, 9) = 16$, whereas for $c_S = 12 - c_F$ we get $U_F(4, 8) = 17 - 6\beta_i$. We find that $16 < 17 - 6\beta_i$ iff $1 - 6\beta > 0$. In the data of Blanco, Engelmann and Normann (2011), 19.7 percent of the subjects meet this condition. That is, if $c_F = 9$, the public good will be provided at level one (two) with 19.7 (80.3) percent probability. Finally, the case $c_F = 10$ turns out to be identical regarding the second-movers' incentives. That is, $c_F = 9$ and $c_F = 10$ are equally likely to be "exploited" by the second mover, and the predicted frequencies of public good provision are hence the same. Table 3.4 summarizes the additional predictions in SEQ_2.

	First-mover contribution		
Second-mover contribution	$c_F = 8$	$c_F = 9$	$c_F = 10$
$c_S = 18 - c_F$ (PG level 2 provided)	39.3%	80.3%	80.3%
$c_S = 12 - c_F$ (PG level 1 provided)	60.7%	19.7%	19.7%
$c_S = 0$ (PG not provided)	0.0%	0.0%	0.0%
expected first-mover payoff	13.97	15.02	14.02

Table 3.4: Predicted second-mover responses conditional on first-mover choices between 8 and 10 and expected first-mover monetary payoff in SEQ_2

Consider next the first movers. $c_F = 10$ will never be chosen in a perfect Bayesian equilibrium by first movers because $c_F = 9$ triggers the same second-mover response as $c_F = 10$ (in terms of public good provision) but $c_F = 9$ yields a higher expected payoff and higher F&S utility than $c_F = 10$. As for the choice between $c_F = 8$ or $c_F = 9$, we find that $c_F = 8$ yields a lower expected monetary payoff than $c_F = 6$ (see Table 3.4) and accordingly an even lower F&S utility. Hence, a risk neutral first mover will never choose $c_F = 8$ in a perfect Bayesian equilibrium. The remaining possibilities are that first movers will either choose $c_F = 6$ or $c_F = 9$. Contributing $c_F = 6$ yields an expected utility of 14 and $c_F = 9$ gives an expected utility of $15.015 - 1.182\alpha$. Now $15.015 - 1.182\alpha > 14$ iff $\alpha < 0.859$. This is predicted to hold for 36 percent of the Blanco, Engelmann and Normann (2011) subjects.

Proposition 2. *For treatment* SEQ_2, *the standard model makes the same predictions as for* SEQ_1. *The calibrated F&S model predicts the frequencies of second-mover responses as in Tables 1 and 2, and that 64 percent of all first movers choose $c_F = 6$ and 36 percent choose $c_F = 9$.*

Taking second- and first-mover predictions together, we finally derive the prediction for the frequencies of public good provision. We expect the public good to be provided at step level 1 with a frequency of $0.64 + 0.36 \cdot 0.197 = 0.711$ and at step level 2 in the rest of the cases.

Simultaneous moves, one threshold

With simultaneous moves, there are multiple equilibria both in the standard model and in the F&S model. With standard preferences, both players contributing nothing and all allocations where $c_1 + c_2 = 12$ are the pure-strategy equilibria.[11] Perhaps somewhat surprisingly, all of these equilibria are also Nash equilibria with calibrated F&S preferences except for those where $(c_1 = 2, c_2 = 10)$ and $(c_1 = 10, c_2 = 2)$. (Proof available upon request.)

[11] There are also numerous mixed-strategy equilibria.

We believe that it is unlikely that entirely symmetric players will coordinate on asymmetric equilibria and we therefore focus on symmetric equilibria. The two symmetric pure-strategy Nash equilibria are $c_i = c_j = 0$ and $c_i = c_j = 6$, and the symmetric mixed-strategy equilibrium has both players contribute $c_i = 0$ with 40 percent probability and $c_i = 6$ otherwise with standard preferences.

With the calibrated F&S model, the symmetric pure strategy (Bayesian-Nash) equilibria $c_i = c_j = 0$ and $c_i = c_j = 6$ are the same but the best response correspondence changes both quantitatively and qualitatively. First of all, note that we can "purify" the mixed-strategy equilibrium (Harsanyi, 1973) as we have a population of 58 different types of players in the Blanco, Engelmann and Normann (2011) data.[12] We will analyze the mixed equilibrium such that each of these players chooses a pure strategy. From Assumption 2, players know the distribution of types and thus they also know how many of the other players will play which strategy in equilibrium. In the (Bayesian-Nash) mixed-strategy equilibrium with calibrated F&S utilities, 64 percent of the players contribute $c_i = 6$ whereas 36 percent choose $c_i = 0$. Hence, more types contribute $c_i = 6$ with F&S preferences in the mixed-strategy equilibrium.

There is, however, also a qualitative difference to the standard case. With standard preferences, all players have the same best reply: if less than 60 percent of the players are expected to contribute, nobody will contribute (and vice versa if more than 60 percent contribute). With the calibrated F&S model, it is not the case that all players have the same best response. If less than 64 percent of players are expected to contribute $c_i = 6$, some players will still contribute. Learning will is slower and the shape of the best response correspondence differs from the standard case. We discuss this in detail below.

Proposition 3. *In treatment* SIM_1, *the symmetric equilibria are* $c_i = c_j = 0$ *and* $c_i = c_j = 6$. *In the symmetric mixed-strategy equilibrium 60 percent of the players choose* $c_j = 6$*; and 64 percent in the case of F&S preferences.*

[12] Among the 61 players reported in Table 2, three types occur twice so that there are 58 types in total.

Simultaneous moves, two thresholds

We turn to the variant with *simultaneous-move game with two thresholds* (SIM_2). As argued above for SEQ_2, meeting the second threshold is not a Nash equilibrium with standard preferences. As the equilibria derived above for SIM_1 are unaffected by the introduction of the second threshold; with standard preferences, SIM_2 has the same Nash equilibria as SIM_1.

We now look for a Bayesian Nash equilibrium of players with F&S utilities where the second level of the public good is provided. Suppose that some types choose $c = 9$. Above, we have seen that, given $c_i = 9$, 80.3 percent of all types will reply with $c_j = 9$ whereas the rest plays $c_j = 3$. Hence, there cannot be a Bayesian Nash equilibrium where all types choose $c_i = 9$. We will therefore look for a Bayesian Nash equilibrium where p percent of all F&S types choose $c_i = 9$ whereas $1 - p$ choose $c_i = 3$.

The expected utility from playing $c = 9$ is $pU(9,9) + (1-p)U(9,3) = 16p + (1-p)(11-6\alpha)$, and the expected utility from playing $c = 3$ is $pU(3,9) + (1-p)U(3,3) = p(17-6\beta)+(1-p)7$. Contributing 9 yields a higher expected F&S utility than contributing 3 iff

$$p > \frac{6\alpha - 4}{6\alpha + 6\beta - 5}.$$

For F&S players with $\alpha = \beta = 0$, this condition is never met (as seen above); that is, selfish own utility maximizers will always choose $c = 3$. If p is sufficiently large, however, inequality averse players prefer $c = 9$. In the Blanco, Engelmann and Normann (2011) data, we find that for $p = 0.72$ exactly 72 percent of the players (44 players) have $pU(9,9)+(1-p)U(9,3) > pU(3,9)+(1-p)U(3,3)$ whereas for 28 percent (17 players) the inequality is reversed. Thus these strategies constitute a Bayesian Nash equilibrium.

It remains to check, though, whether it pays to deviate to any contribution other than 9 or 3. The only possible deviation is to contribute $c = 0$ since any other contribution is dominated either by $c = 0$ or $c = 3$. Contributing $c = 0$ yields an expected F&S utility of $10 - 3\beta - 0.72 \cdot 6\beta$. But the equilibrium action $c = 3$ yields $0.72(17 - 6\beta) + (0.28)7$ which is strictly larger for all $\beta \in [0, 1]$. Thus we have established:

Proposition 4. *The Bayesian Nash equilibria of* SIM_1 *are also equilibria in treatment* SIM_2*. With standard preferences, there are no additional equilibria. With the calibrated F&S preferences, 72 percent of the F&S types choosing $c = 9$ and the rest $c = 3$ is a Bayesian Nash equilibrium.*

Hypotheses

Based on Propositions 1 to 4, we will now derive two hypotheses regarding the impact of our two treatment variables. We will return to the propositions and the performance of the F&S model below.

Comparing the predicted public good provision in SIM vs. SEQ, we note that there are multiple equilibria in the SIM treatments and that the public good is not provided in all equilibria. By contrast, in the SEQ treatments, the equilibrium is unique and the public good is provided (at least at level one) in the unique equilibrium. This holds for both the one and the two-threshold case.

Hypothesis 1. *The public good will be provided more frequently in the* SEQ *treatments compared to* SIM.

Note that this hypothesis does not depend on assuming that players have F&S preferences. The F&S model makes the point that the $c_F = 2$ prediction with standard preferences in the SEQ treatments will be punished regularly by the second movers, but it also predicts that this case will not arise because first movers anticipate this. Our second hypothesis, though, does depend on assuming F&S preferences.

Propositions 1 to 4 show that public good provision can be improved if there is the second threshold. The case for improved public good provision in the SIM treatments is as follows. There are multiple equilibria in the SIM treatments anyway but there exists an equilibrium in which the second level is met with positive probability. For both SEQ_2 and SIM_2, we note that even if one player attempts to reach the second level but the other player exploits this, this does not harm payoffs that much as the first level of the public good is still provided. In both the simultaneous-move treatment and the sequential treatment with two levels, players may yield a higher payoff by achieving the second threshold level. Therefore they have an incentive to make higher contributions and public good provision will be more likely in the presence of two thresholds. If first movers make higher contributions in SEQ_2, fewer punishment and more second movers who contribute should occur such that the public good is provided at least at level one.

Hypothesis 2. *The public good will be provided more frequently in the treatments with two thresholds compared to one-threshold treatments.*

3.5 Main Treatment Effects

We present our results in two parts. Section 3.5 presents tests of Hypotheses 1 and 2. In addition to public good provision, we will also analyze contributions, payoffs (or efficiency) and coordination rates. Section 6 presents a more detailed analysis of the predictive power of the calibrated F&S model.

When we apply regressions analysis, we use Generalized Linear Latent and Mixed Models (gllamm; see Rabe-Hesketh et al., 2005) regressions, taking possible dependence of observations at the level of a (randomly matched) group and at the individual level into account. As dependent variables we use *sequential* (a dummy which is equal to one if the move order is sequential), *twolevel* (a dummy which is equal to one if there are two levels), *seq2* (an interaction term for the sequential treatment with two levels), furthermore we control for *period* and the *sessionsize*. Our results are robust to alternative specifications, such as linear, tobit and probit regressions that are clustered at the group level.

We typically report three regressions. Regression (1) reports the impact of the treatment variables *sequential* and *twolevel* only. Regression (2) includes the interaction *seq2*, and (3) adds *period* and *sessionsize*.

Overview

We start with a summary statistics of our four treatments in Table 3.5.

	Treatment			
Variable	SIM_1	SIM_2	SEQ_1	SEQ_2
PG level 1 provided (χ_1=1) in %	64.29	59.00	75.24	85.56
PG level 2 provided (χ_2=1) in %	-	6.00	-	16.67
PG not provided (in %)	35.71	41.00	24.76	14.44
Contributions	5.22	5.99	4.96	6.07
	(2.23)	(2.88)	(2.36)	(2.57)
First-mover contributions	-	-	4.76	6.41
	-	-	(1.53)	(2.23)
Second movers contributing $c_s = 0$ (in %)	-	-	18.57	10.00
Successful coordination (in %)	49.05	17.00	77.62	81.11
Payoff	11.21	10.30	12.56	13.32
	(3.86)	(4.27)	(2.92)	(3.18)

Table 3.5: Summary statistics of our four treatments. Note that the public good is provided at level 2 ($\chi_2 = 1$) only if it is also provided at level 1 ($\chi_1 = 1$).

It shows public good provision, contributions, frequency of successful coordination, and the resulting payoffs. Note in our treatments with two threshold levels we also count the cases where the second level has been achieved as a successful provision of *PG level 1*.

As can be seen, public good provision at the first level is most effective in the treatments with sequential moves. *PG level 1* is provided most frequently (85.56%) with the sequential-move order and two thresholds and thus *PG level 1* provision is also more effective in SEQ_2 compared to SEQ_1 where only 75.24 percent subjects manage to provide the public good. Only in 6 percent of SIM_2'S cases is the public good provided at the second threshold level. However, the second threshold level does come out better with sequential-move order (16.67% of PG level 2 in SEQ_2). The second threshold level leads to higher contributions in the simultaneous as well as in the sequential treatment. We define successful coordination as cases without wasteful contributions (that is cases where $c_1 + c_2 \in \{0, 12\}$, or $c_1 + c_2 \in \{0, 12, 18\}$ with two step levels). Coordination is best in the environment of sequential moves. Furthermore the sequential-move order also leads to higher payoffs compared to the simultaneous treatments.

A first look at the data in Table 3.5 thus suggest that we do find tentative support for Hypothesis 1. Regarding Hypothesis 2, the effect is ambiguous since the second level improves public good provision (at level one) in the SEQ treatments but not in the SIM settings.

Public Good Provision

Table 3.6 presents gllamm probit regressions of the frequency of PG provision. The first probit regression (left panel) is about public good provision at level 1. The dependent variable is equal to one if and only if the first threshold is met (that is, if and only if $\chi_1 = 1$). The second probit regression (right panel) has that dependent variable is equal to one if and only if the second level ($\chi_2 = 1$) is met. Note that the public good is provided at level 2 ($\chi_2 = 1$) only if it is also provided at level 1 ($\chi_1 = 1$).

The regressions in the left panel shows that *sequential* is significant. That is, the sequential-move order improves the PG provision at the first threshold. This supports Hypothesis 1. The implementation of a second threshold does not lead to a higher frequency of public good provision. The same is true for the sequential treatment with two thresholds. That is, we do not find support for Hypothesis 2 which predicts that the second threshold leads to more public good provision. In regression (3), we find that the coefficient of *sessionsize* is negative and weakly significant. That is, sessions with a higher numbers of subjects exhibit lower public good provision. This is consistent with findings in Botelho et al. (2009). In their paper repeated settings with "random strangers" and "perfect strangers" matching protocols are compared. The authors find that the assumption that subjects treat Random Strangers designs as if they were one-shot experiments is false. Our results indicate that the session size and hence the likelihood of meeting a random stranger once more has an impact on cooperation. We note, however, that the coefficient of *sessionsize* is small.

Table 3.6 also presents a gllamm probit regression of the frequency of public good provision of level 2. (Here, *twolevel* cannot be part of the regression analysis, of course.) *sequential* is again significant, that is, sequential-move contributions also stimulate the provision of the second level which is additional support for Hypothesis 1. Regression (2) reveals that public good provision at level two moderately decreases over time. The dummy *sessionsize* is not significant here.

	Public good provision level 1			Public good provision level 2	
	(1)	(2)	(3)	(1)	(2)
sequential	0.657***	0.390*	0.395**	0.638**	0.550**
	(0.184)	(0.229)	(0.200)	(0.300)	(0.277)
twolevel	0.123	-0.137	-0.151		
	(0.183)	(0.227)	(0.199)		
seq2		0.541	0.429		
		(0.330)	(0.297)		
period			-0.006		-0.466*
			(0.012)		(0.024)
sessionsize			-0.064*		-0.067
			(0.035)		(0.051)
constant	0.281*	0.408**	1.342***	-1.773***	-0.630
	(0.154)	(0.160)	(0.509)	(0.230)	(0.716)
# obs.	1,600	1,600	1,600	760	760

*** p<0.01, ** p<0.05, * p<0.1

Table 3.6: Probit gllamm regressions of public good provision

Contributions

We now analyze subjects' contribution levels. The left panel of Table 3.7 reports a linear gllam regression of the players' contribution. Contributions are *not* significantly influenced by the order of moves. Consistent with our Hypothesis 2, adding the second threshold leads to significantly higher contributions. The interaction of a sequential move order and two levels does not lead to further increased contributions. Over time, contributions get weakly smaller. *sessionsize* is significant, that is, in sessions with more participants contributions are slightly lower.

	Contributions			Payoffs		
	(1)	(2)	(3)	(1)	(2)	(3)
sequential	-0.00397	-0.126	-0.0959	2.140***	1.390***	1.389***
	(0.254)	(0.304)	(0.315)	(0.379)	(0.446)	(0.386)
twolevel	1.202***	0.904*	0.684***	-0.0103	-0.867**	-0.898**
	(0.257)	(0.424)	(0.244)	(0.393)	(0.439)	(0.392)
seq2		0.446	0.390		1.595***	1.404**
		(0.528)	(0.436)		(0.618)	(0.565)
period			-0.0330*			-0.00250
			(0.0175)			(0.0310)
sessionsize			-0.135***			-0.115*
			(0.0434)			(0.0671)
constant	5.039***	5.076***	7.151***	10.84***	11.21***	12.83***
	(0.169)	(0.174)	(0.639)	(0.254)	(0.323)	(0.99)
# obs.	1,600	1,600	1,600	1,600	1,600	1,600

*** p<0.01, ** p<0.05, * p<0.1

Table 3.7: Linear gllamm regressions of contributions and payoffs

Payoffs

In Table 3.7 (right panel), we report the results of a linear regression on subjects' payoffs. First, Table 3.7 shows that the sequential contribution mechanism significantly improves subjects' payoffs. This is due to the fact that public good provision is improved by the sequential-move order.

The second step level significantly reduces the payoffs. This can be explained by the fact that, on the one hand, two thresholds increase contributions but, on the other hand, the second level is rarely actually achieved. When we add the interaction *seq2*, we find that it significantly boosts subjects' payoff by 1.4 compared to the baseline SIM_1. The difference between SEQ_1 and SEQ_2 is, however, not significant as follows from a Wald test ($p = 0.22$). This emphasizes the overall negative impact of the second threshold on payoffs. Indeed, payoffs are worst in SIM_2. The size of the sessions is weakly significant, but, again, the coefficient is small. The time trend is insignificant here.

The payoff variable is the variable a social planner would ultimately be interested in. Payoffs reflect the combined effect of contributions, public good provision and avoiding excess contributions (coordination). The above regression confirms that the payoff differences reported in our summary statistics are significant. Specifically, it follows that SEQ_2 has the highest payoffs, followed by SEQ_1 and SIM_1, and SIM_2 has the lowest payoffs.

Coordination Rates

We finally report coordination rates. We define $C = c_1 + c_2$ and, as mentioned above, successful coordination occurs if and only $C \in \{0, 12\}$ and $C \in \{0, 12, 18\}$, respectively. We report descriptive statistics here only; a regression analysis of successful coordination is qualitatively very similar to the one on payoffs reported above.

Figure 3.1 compares coordination in the simultaneous and the sequential treatment with one threshold.

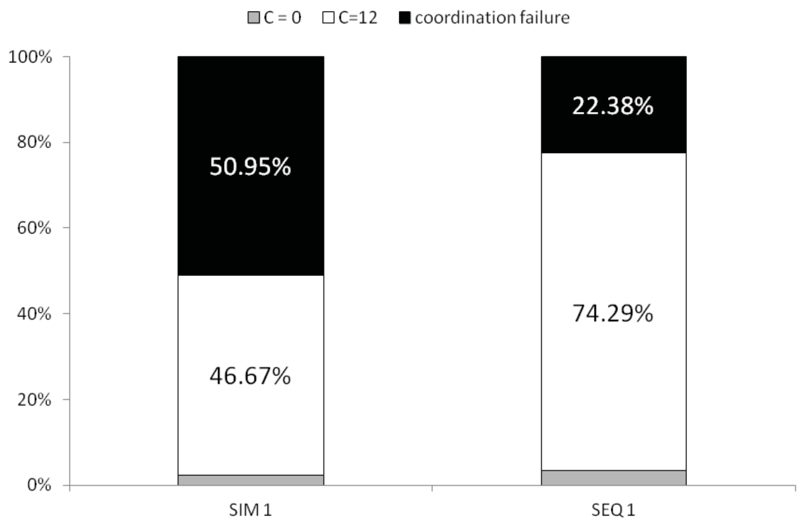

Figure 3.1: Frequency of the contribution sums (C) of both players in the simultaneous and sequential step-level public good game with one threshold.

In SEQ_1, $C = 12$ is the most frequent outcome. That is, subjects coordinate on $C = 12$ in 74.29 percent of all cases. By contrast, the simultaneous contribution mechanism only guarantees efficient contributions in 46.67 percent of the time. The difficulty of coordinating in SIM_1 is also documented by the aggregation of the cases (coordination failure) where the contribution sum is either too low ($0 < C < 12$) or too high ($C > 12$). As for the sum of these inefficient cases, we find that in SIM_1, 50 percent of the subjects do not manage to contribute efficiently. The remaining cases are those where $C = 0$, which is efficient in that no contributions wasted. In SEQ_1, there are only 22.38 percent inefficient cases. Mainly, these involve second movers punishing low first-mover contributions. Figure 1 shows that the increased payoff in SEQ_1 is mainly due to the fact that subjects' less contributions are wasted in the sequential environment.

Figure 3.2 compares coordination in the simultaneous and sequential treatment with two thresholds. This plot again documents the superiority of the sequential- over the

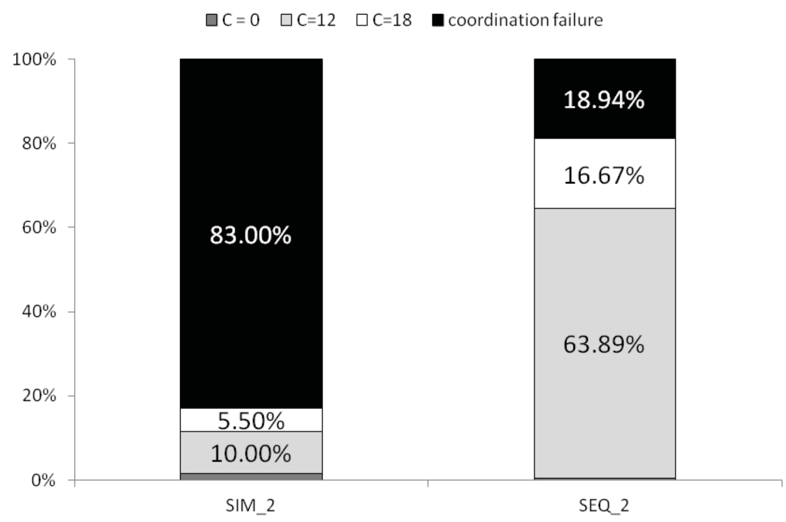

Figure 3.2: Frequency of the contribution sums (C) of both players in the simultaneous and sequential step-level public good game with two thresholds.

simultaneous-move variant. In SEQ_2, about 80 percent of all contribution sums are efficient. That is, subjects manage to exactly coordinate on the first threshold ($C = 12$) or on the second threshold ($C = 18$) without generating wasteful excess contributions.

This stands in strong contrast to the efficient cases in SIM_2 where only 15.5 percent of the contribution sums are efficient. In SIM_2, subjects seem to face a great deal of difficulty in terms of coordination. This leads to a high amount of wasteful contribution sums (the sum of all cases where either $0 < C < 12$, $12 < C < 18$ or $C > 18$) of 83 percent. Figure 2 therefore serves as an explanation of the fact that the second level leads to smaller payoffs. Especially in SIM_2 the second level leads to costly mis-coordination of the players.

However, two levels are efficient in the environment with sequential moves which explains the significance of our interaction term *seq2*. The result is driven by first movers contributing higher amounts in SEQ_2 compared to first movers in the one-level treatment. This is shown in Table 5 where average first-mover contributions of SEQ_1 and SEQ_2 are presented. It shows that first movers on average make higher contributions in the sequential treatment with two thresholds. In SEQ_2 first movers contribute more than half of the first threshold (6.41). Thus, second movers are not "exploited" that frequently and they only punish first-mover behavior in 10 percent of all cases. This is in contrast to the one-level treatment where first movers make average contributions below six (4.97) and second movers punish in 19 percent of all cases.

3.6 The predictive power of the calibrated F&S model

We now discuss the quantitative predictions of the F&S model in more detail. We begin with Proposition 1. Figure 3.3 contrasts the predictions made in Table 3.3 to the observations of the frequency of second movers who contribute $c_S = 12 - c_F$ in reply to first-mover contributions. The data underlying Figure 3.3 pools the c_F in both sequential treatments SEQ_1 and SEQ_2.[13] Using one-sample chi-square tests, we cannot reject that predicted and observed frequencies are the same (all $\chi^2_{(1)} < 2.38$ and $p > 0.123$). The F&S model predicts the second-mover responses amazingly well.

In SEQ_1, first movers should choose $c_F = 6$ in order to maximize payoffs (and F&S utilities). This is not the case as $c_F = 6$ is chosen only in 37.1 percent of the cases. In our SEQ_1 data, it turns out $c_F = 5$ is the (ex post) payoff maximizing strategy (yielding an expected payoff of 14.26, as opposed to 13.76 with $c_F = 6$) and it is chosen in 26 percent of the cases. While this rejects the F&S prediction, we note that similar observations have been made before (see below).

Figure 3.4 is a bubble plot of first and second movers in SEQ_1.[14] The modal outcome

[13] This is warranted because, firstly, the F&S model does not predict any differences and, secondly, we do not observe differences—with the exception of $c_F = 5$ where contribution rates differ significantly (two-sample chi-square test, $\chi_{(1)} = 8.579$, $p < 0.01$). Importantly, the minor differences we observe are not systematic. Contributions of $c_S = 12 - c_F$ are more frequent for $c_F = \{3, 4\}$ in SEQ_1 than in SEQ_2 but the other way round for $c_F = \{5, 6\}$. Note that, for $c_F = 6$, we cannot apply a test because the predicted frequency is 100 percent. Regarding $c_F = 2$, we only have two observations so we cannot test either (in one case the PG was provided).

[14] Here, we cannot include SEQ_2 data in the figure because behavior is predicted (and does) differ whenever

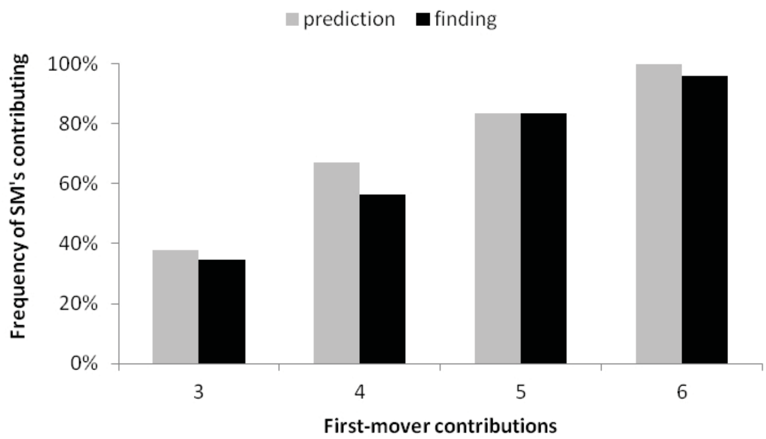

Figure 3.3: Predicted frequencies (based on the calibrated F&S model) and observed frequencies of second movers contributing such that the PG at level 1 is provided in SEQ treatments.

is (6, 6) as predicted, and many observations are on the Pareto frontier where $c_F + c_S = 12$. One can identify the punishing second movers on the vertical axis where $c_S = 0$. For the first movers in SEQ_2, Proposition 2 predicts that 36 percent contribute $c_F = 9$ and 64 percent should choose $c_F = 6$. In our data, 36.7 percent of the first movers choose 9—which seems a remarkable confirmation of the prediction. The remaining 63.3 percent choose $c_F \in [2, 6]$. While we do not find that 64 percent choose $c_F = 6$, this only restates the previous finding that first movers do not always choose the risk-neutral payoff maximizing action.

Intriguingly, the second mover prediction of Proposition 2 fails (whereas it was the first mover prediction of Proposition 1 that failed). The first mover in the two-level case is in a trust-game-like situation. If she chooses $c_F = 9$, she can be exploited by second movers. While the calibrated F&S model predicts that more than 80.3 percent of the second movers will be trustworthy, it turns out only 50.9 are. Predicted and observed share differ significantly (binomial test, $p < 0.05$). The failure of the theory seems surprising since the cost of being trustworthy are low here: second movers gain only one additional Euro by exploiting the first mover, but this costs the first mover five Euros. (See the

$c_F \geq 8$.

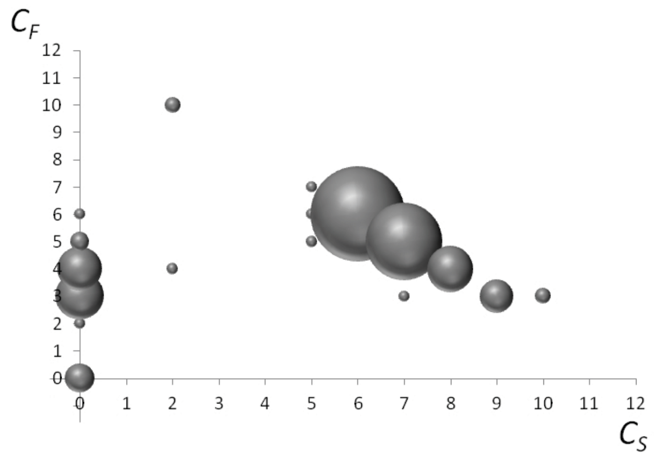

Figure 3.4: Frequencies of first- and second-mover choice combinations where the bubble size corresponds to frequency. The Pareto frontier can be found where $c_F + c_S = 12$ and $c_S = 0$ indicates punishing second movers.

discussion at the end of the section).

We finally turn to Proposition 3, the SIM_1 case. In SIM_1, we observe that in 81.4 percent of the cases subjects choose $c \geq 6$ and in 13.8 of the cases they choose $c = 0$.[15] Hence, both the standard model and the calibrated F&S model would predict that play converges to the pure-strategy equilibrium where both players choose $c = 6$. This is, however, not the case. There is no positive time trend, and some players persistently choose $c = 0$. Why do subjects not best respond? Figure 3.5 illustrates what might be going on. It shows the best-reply correspondences for standard selfish players, for F&S players and also for players with standard preferences but with a degree of risk aversion according to the findings in Holt and Laury (2002).

With selfish and rational players, the best-reply correspondence has a “bang-bang” property. If the belief is that player j chooses $c_i = 6$ less than 60 percent, *all* players will best respond with $c_i = 0$, and vice versa for a belief of more than 60 percent. With the calibrated F&S model, this is not the case. For beliefs between (roughly) 40 and 80

[15]These percentages are are based on data from periods 6 to 10 where we observe less heterogeneity in the data. Some subjects indeed choose $c_i > 6$, but—for our data—$0 < c_j < 6$ is never a best reply with standard or F&S preferences. Thus we focus on $c \geq 6$ and $c = 0$

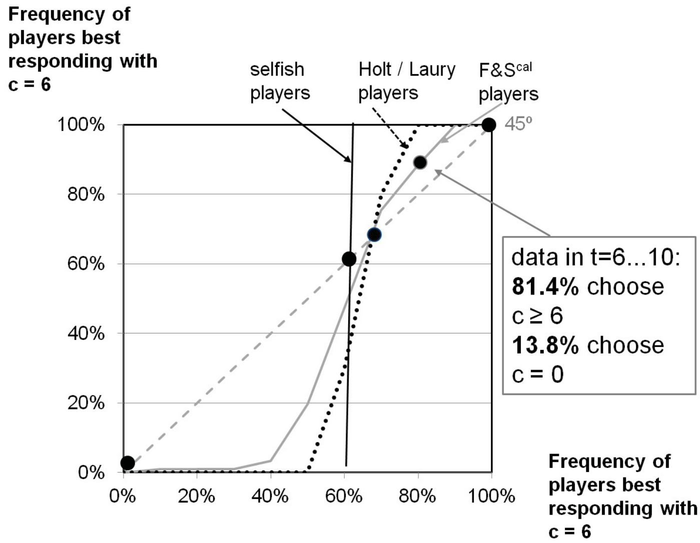

Figure 3.5: Best-reply correspondences for standard players, F&S players and Holt-Laury players in SIM_1.

percent, the best replies of the various F&S types differ. For example, given a belief that 70 percent of all players choose $c_i = 6$, only 75 percent of the players will best respond with $c_i = 6$ where 25 percent still choose $c_i = 0$.

As mentioned in Proposition 3, the share of players choosing $c_F = 6$ required such that $c_F = 6$ is a best reply that is slightly larger with F&S players. Inequality aversion has an effect similar, in fact a stronger effect, than risk aversion (on average, players in Holt and Laury, 2002, are slightly risk averse). We also see that the best replies differ from the case with standard preferences. Around the mixed-strategy equilibrium, the best replies are not vertical but somewhat "flat", implying that not all players will best reply once the fixed point of the mixed strategy is exceeded. This is what we see in the data.

What can we conclude from the analysis of the calibrated F&S model? First, we find remarkable confirmations of the predictions of the model. One may argue that, regarding SEQ_1, these are not so surprising because of the partial similarity of SEQ_1 to the ultimatum game (from which the alphas were elicited). However, the SEQ_1 prediction also depends on the joint distribution and not on the alpha only. Moreover, we also found confirmation of the calibrated F&S model for SEQ_2 and SIM_1. Hence, we conclude that the model is particularly powerful in our setup.

How about the two contradictions to the calibrated F&S model then? First, we found that first movers behaved too greedily to be consistent with Assumptions 1 and 2, providing $c_F < 6$ too often. This finding is not new. For standard ultimatum-game experiments, it can be argued that offering the equal split may be payoff maximizing (assuming risk neutrality), but about half the the proposers offer less than the equal split.[16] Huck, Müller and Normann (2001) show that, in quantity-setting duopoly, Stackelberg followers are inequality averse but the Stackelberg leaders still choose too high an output. The payoff maximizing (and inequality minimizing) output in that data set was the symmetric Cournot-Nash solution. In ultimatum games, the Stackelberg game and this study, risk-loving behavior can explain the first-mover behavior. However, it could also be that first movers feel entitled to more than 50 percent of the pie whereas second movers regard the equal split as fair. Social norms may be perceived differently by first and second movers.

We secondly saw that second movers exploit first-mover trust (that is, $c_F = 9$) too often in SEQ_2. We consider the following explanation plausible. In SEQ_2, first movers frequently choose $c_F < 6$ and, just as in SEQ_1, the second movers are in the weaker position. Whenever $c_F = 9$, second movers are suddenly in the stronger position. They can now ensure themselves the higher payoff and they often do so. It could be the low degree of trustworthiness is second movers scoring off greedy first movers, with a "now it is my turn" attitude (recall the game is repeated 10 times). In contrast to costly punishments of $c_F < 6$, responding with $c_S = 3$ to any $c_F = 9$ is free, in fact yields an even higher payoff. If so, second movers do not reflect that first movers contributing $c_F = 9$ are unlikely to be the same first movers who offered $c_F < 6$ in a previous round.

[16] In Blanco, Engelmann and Normann (2011), offering the equal split is actually (expected) payoff maximizing, but their ultimatum game was done with the strategy method which typically induces higher rejection rates.

3.7 Conclusion

How should a planner design, say, a fund-raising mechanism when the goal is that a certain threshold of contributions has to be met? Our experiments give two clear-cut answers to this question. First, we find that the sequential-move variant yields more frequent provision of the public good and higher payoffs. This confirms the literature on *leading by example* where, in our setup, it is the better coordination that renders the sequential mechanism superior in the threshold public-good game. Even though some low-contributing first movers (who actually give a *bad* example) are punished by second movers, higher provision rates and payoffs emerge. The finding is in contrast to Gächter et al.'s test of Varian's (1994) model. They find that sequential contributions are lower with sequential moves, but the difference is not as big as predicted. One reason for this is that, as in our setting, second movers sometimes punish first movers. In our setting, the sequential-move variant is more efficient.

Second, we find that an additional second step level does lead to higher contributions but the effect on public good provision is ambiguous and it even lowers payoffs. Hence, a fund-raising strategy like "we need to raise $10m dollars for a new hospital, but with $15m we could also build that additional cancer unit" is likely to backfire. The main reasons for the failure of the second step level to improve payoffs are that, with simultaneous moves, coordination failure becomes more frequent. With sequential moves, the two-level variant is at least not inferior but second movers exploit generous first movers too often to make the second threshold superior.

Based on the existing experimental data of Blanco, Engelmann and Normann (2011), we fully calibrate Fehr and Schmidt's (1999) model of inequality aversion to make ex ante predictions. We find that actual behavior fits quantitatively well with these predictions. Specifically, the F&S model predicts the second-mover responses amazingly well. While the predictive power on first-mover behavior is less impressive, similar findings have been observed before in other sequential games. The calibrated Fehr and Schmidt (1999) model also predicts behavior well in the sequential treatment with two step levels, and in the simultaneous-move case with one level.

References

[1] Andreoni, J. (1998). "Toward a Theory of Charitable Fund-Raising", The Journal of Political Economy 106, 1186-1213.

[2] Andreoni, J., Brown, P., and Vesterlund, L. (2002). "What Makes An Allocation Fair? Some Experimental Evidence", Games and Economic Behavior 40, 1-24.

[3] Andreoni, J. (2006). "Leadership Giving In Charitable Fund-Raising", Journal of Public Economic Theory, 8, 1-22.

[4] Blanco, M., Engelmann, D., and Normann, H., T. (2011). "A Within-Subject Analysis of Other-Regarding Preferences", Games and Economic Behavior 72, 321-338.

[5] Botelho A., Harrison, G., W., Costa Pinto, L., M., and Rutström, E., E. (2009). "Testing Static Game Theory with Dynamic Experiments: A Case Study of Public Goods", Games and Economic Behavior 67, 253-265.

[6] Cadsby, C., B., and Maynes, E. (1998a). "Gender and Free Riding in a threshold public goods game: Experimental Evidence", Journal of Economic Behavior and Organization 34, 603-620.

[7] Cadsby, C., B., and Maynes, E. (1998b). "Choosing Between a Socially Efficient and Free-Riding Equilibrium: Nurses versus Economics and Business Students", Journal of Economic Behavior and Organization 37, 183-192.

[8] Cadsby, C., B., and Maynes, E. (1999). "Voluntary Provision of Threshold Public Goods with Continuous Contributions: Experimental Evidence", Journal of

Public Economics 71, 53-73.

[9] Chewning, G., E., Coller, M., and Laury, S., K. (2001). "Voluntary Contributions to a Multiple Threshold Public Good", in: R.M. Isaac (ed.), Research in Experimental Economics 8, Emerald Group Publishing Limited, 47-83.

[10] Coats, J., C., and Neilson, W., S. (2005). "Beliefs about Other-Regarding Preferences in a Sequential Public Goods Game", Economic Inquiry 43(3), 614-622.

[11] Coats, J., C., Gronberg, T., J., and Grosskopf, B. (2009). "Simultaneous versus Sequential Public Good Provision and the Role of Refunds – An Experimental Study", Journal of Public Economics 93, 326-335.

[12] Croson, R., and Marks, M. (1998). "Identifiability of Individual Contributions in a Threshold Public Goods Experiment", Journal of Mathematical Psychology 42, 167-190.

[13] Dannenberg, A., Riechmann, T., Sturm, B., and Vogt, C. (2007). "Inequity Aversion and Individual Behavior in Public Good Games: An Experimental Investigation", ZEW Working Paper, No. 07-034.

[14] Erev, I., and Rapoport, A. (1990). "Provision of Step-Level Public Goods: The Sequential Contribution Mechanism", The Journal of Conflict Resolution 34, 401-425.

[15] Fehr, E., and Kremhelmer, S. and Schmidt, K. (2008). "Fairness and the Optimal Allocation of Ownership Rights". Economic Journal 118, 1262-84.

[16] Fehr, E., and Schmidt, K. (1999). "A Theory of Fairness, Competition and Cooperation", Quarterly Journal of Economics 114, 817-868.

[17] Figuières, C., Masclet, D., and Willinger, M. (2010). "Vanishing Leadership and Declining Reciprocity in a Sequential Contribution Experiment", Economic Inquiry forthcoming.

[18] Fischbacher, U. (2007). "z-Tree: Zurich Toolbox for Readymade Economic Experiments - Experimenter's Manual", Experimental Economics 10, 171-178.

[19] Gächter, S., and Renner, E. (2005). "Leading by Example in the Presence of Free-Riders", mimeo, Nottingham University.

[20] Gächter, S., Nosenzo, D., Renner, E., and Sefton, M. (2010a). "Sequential vs. Simultaneous Contributions to Public Goods: Experimental Evidence", Journal of Public Economics 94, 515-522.

[21] Gächter, S., Nosenzo, D., Renner, E., and Sefton, M. (2010b). "Who Makes a Good Leader? Cooperativeness, Optimism, and Leading-by-Example", Economic Inquiry.

[22] Güth, W., Levati, M., V., Sutter, M., and van der Heijden, E. (2007). "Leading by Example with and without Exclusion Power in Voluntary Contribution Experiments", Journal of Public Economics 91, 1023-1042.

[23] Glazer A., and Konrad, K., A. (1996). "A Signaling Explanation for Charity", American Economic Review 86, 1019-1028.

[24] Harsanyi, J.C. (1973). "Games with Randomly Disturbed Payoffs: A New Rationale for Mixed-Strategy Equilibrium Points", International Journal of Game Theory 2, 1-23.

[25] Hashim, M.J., Maximiano, S., and Kannan, K.N. (2011). "Information Targeting and Coordination: An Experimental Study", mimeo, Purdue University.

[26] Hermalin, B., E. (1998). "Toward an Economic Theory of Leadership: Leading by Example", The American Economic Review 88, 1188-1206.

[27] Holt, C., and Laury, S., (2002). "Risk Aversion and Incentive Effects", The American Economic Review 92, 1644-1655.

[28] Huck, S., Müller, and Normann, H., T. (2001). "Stackelberg Beats Cournot: On Collusion and Efficiency in Experimental Markets", The Economic Journal 111, 749-765.

[29] Huck, S., and Rasul, I. (2011). "Matched Fundraising: Evidence from a Natural Field Experiment", Journal of Public Economics 95, 351-362.

[30] Kölle, F., and Sliwka, D. (2011). "Inequality, Inequity Aversion and the Provision of Public Goods", Work in Progress.

[31] List, J., A., and Luckin-Reiley (2002). "The Effects of Seed Money and Refunds on Charitable Giving: Experimental Evidence from a University Capital Campaign", Journal of Political Economy 110, 215-33.

[32] Potters, J., Sefton, M., and Vesterlund, L. (2005). "After You - Endogenous Sequencing in Voluntary Contribution Games", Journal of Public Economics 89, 1399-1419.

[33] Potters, J., Sefton, M., and Vesterlund, L. (2007). "Leading-by-example and Signaling in Voluntary Contribution Games: An Experimental Study", Economic Theory 33, 169-182.

[34] Rabe-Hesketh, S., Skrondal A., Poe, G., L., and Schulze, W.D., (2005). "Multilevel and Longitudinal Modeling Using Stata", College Station, TX: Stata Press.

[35] Rauchdobler, J., Sausgruber, R., and Tyran, J.R., (2010). "Voting on Thresholds for Public Goods: Experimental Evidence", 34-64. FinanzArchiv / Public Finance Analysis 66(1).

[36] Rondeau, D., Poe, G., L., and Schulze, W., D., (2005). "VCM or PPM? A Comparison of the Performance of Two Voluntary Public Goods Mechanisms", Journal of Public Economics 89, 1581-1592.

[37] Schram, A., Offerman, T., and Sonnemans, J. (2008). "Explaining the Comparative Statistics in Step-Level Public Good Games", in: C.R. Plott and V.L. Smith (eds.), The Handbook of Experimental Economics Results, Volume 1, Amsterdam: North-Holland.

[38] Serra-Garcia, M., van Damme, and E., Potters, J. (2010). "Which Words Bond? An Experiment on Signaling in a Public Good Game", CentER Working Paper 2010 33, Tilburg University.

[39] Shang, J. and Croson, R. (2009). "Field Experiments in Charitable Contribution:

The Impact of Social Influence on the Voluntary Provision of Public Goods", The Economic Journal 119, 1422-1439.

[40] Soetevent, A., R., (2005) "Anonymity in Giving in a Natural Context? a Field Experiment in 30 Churches", Journal of Public Economics 89, 2301-2323.

[41] Teyssier, S. (2009). "Inequity and Risk Aversion in Sequential Public Good Games", Public Choice forthcoming.

[42] Varian, H., R. (1994). "Sequential Contributions to Public Goods", Journal of Public Economics 53, 165-186.

[43] Versterlund, L. (2003). "The Informational Value of Sequential Fundraising ". Journal of Public Economics 87, 627-657.

Appendix

Control questions

Before we continue with the experiment instructions, we want to make sure that everybody understands how payoffs can be earned. Please answer the questions below. Please raise your hand if you have a question. After some minutes we will check your answers.

(i) Assume you contribute 8 Euros to the joint project. The participant you are matched with contributes 5 Euros to the joint project.

  (a) What is the payoff from your private project?

  (b) What is the payoff from your joint project?

  (c) What is your entire income at the end of the round?

  (d) What is your matched participant's profit from her private project?

(e) What is your matched participant's bonus payment from the joint project?

(f) What is your matched participant's entire income at the end of the round?

(ii) Assume you contribute 9 Euros to the joint project. The participant you are matched with contributes 9 Euros to the joint project. (Six questions as above.)

(iii) Assume you contribute 5 Euros to the joint project. The participant you are matched with contributes 7 Euros to the joint project. (Six questions as above.)

(iv) Assume you contribute 1 Euro to the joint project. The participant you are matched with contributes 0 Euros to the joint project. (Six questions as above.)

How you will make your decisions

At the beginning of each round, you have to decide about the number of Euros you want to contribute to the joint project. You will do this by entering your chosen number. You have the possibility to type in any integer number between 0 and 10. Note that you and the participant you are matched with decide at the same time and independently of each other.

After the decisions have been made, both participants will be given an information screen at the end of the round. This information screen will show the participants the individually chosen contributions to the joint project in that round. Both participants get information about their individual returns from their private projects. Furthermore, the amount of the bonus payment will be displayed. Additionally, both participants are informed about their individual total payoff in that round.

Beginning the experiment

Please take a look at your computer screen and make your decision. If you have a question at any time, please raise your hand and we will come to your desk to answer it.

Part II

Group Decision Making

Chapter 4

Competition in the Workplace: An Experimental Investigation*

4.1 Introduction

Competition in the workplace is an astonishing and widespread phenomenon (e.g. Marino and Zabojnik (2004)). This kind of internal competition is characterized by several employees who are hired by the same employer and who try to outperform one another in order to receive higher wage or bonus payments. The phenomenon is of great empirical relevance as companies involving multiple employees are frequent in the field.

The standard workhorse for the analysis of experimental labor markets is the gift-exchange game [1]. The literature on gift-exchange games is huge (for a survey see Gächter and Fehr (2002) or Charness and Kuhn (2011)), but the standard version of the game is not suitable to control for competition in the workplace as there are two prerequisites for

*The research of this chapter is part of a joint project with Volker Benndorf.

[1] Fehr et al. (1993)

analyzing this phenomenon. The first one concerns the number of employees per firm. Most papers focus on setups where each employer hires exactly one employee,[2, 3] this is why these papers do not cover the full depth and breadth of these internal competition mechanisms. The second prerequisite relates to the timing of the game: in the standard workhorse model employees act as second movers, i.e., they have a strong incentive to exert low effort levels which, reduces their incentives to outperform each other.[4] We circumvent these problems by using the modified gift-exchange game introduced by Abeler, Altmann, Kube, and Wibral (2010) where two employees are matched to one employer and where the timing is reversed.[5] There is a growing literature on gift-exchange games with multiple employees[6], but none of them addresses competition in the workplace. Concerning employees' effort choices, most papers do not report significant changes compared to the standard gift-exchange game (e.g., Charness and Kuhn (2007)) or Maximiano et al. (2007)). Gächter and Thöni (2010) find that employees care greatly about disadvantageous wage inequality when workers receive a lower wage compared to their co-worker, they decrease their future effort levels. There are also studies reporting the workers' reactions to wage cuts (Gächter and Sefton (2008) and Cohn et al. (2011)). These papers show that workers' performance significantly decreases after the experience of a wage reduction when their co-workers' wage is held constant. Although all these studies do investigate multiple-employees setups they are sufficiently different to our paper. In contrast to these papers our main interest is not based on the consequences of unequal wage payments, but rather on the dynamics of the competition in the workplace phenomenon in multiple workers environments.

If employers pay higher wages to the employees who exert a higher effort level com-

[2]For example Fehr et al. (1998b), Charness (2000), Brandts and Charness (2004) and Pereira et. al (2006)

[3]Owens and Kagel (2010) even study a gift-exchange game with minimum wage restrictions.

[4]Engelmann and Ortmann (2009) find that responder behavior in gift-exchange games where employers first move is very sensible to parametrization, i.e., responders often tend to exploit proposers.

[5]This framework enables the *employers* to reciprocate high effort choices of the employees as in other papers such as Falk and Kosfeld (2006), Heinz et al. (2012) or Eriksson and Villeval (2011) who report gift-exchange results where employers can express non-monetary reciprocity to employees.

[6]In Altmann et al. (2009), the number of employees per firm is endogenous.

pared to their co-worker it can be considered as an implicit rank-order tournament where workers are paid according to their relative performance. Lazear and Rosen (1981) show theoretically that this kind of payment structure can result in optimal allocations if workers are risk neutral. Since tournament incentives are crucial for competition in the workplace we will control if employers in our experiment create these incentives by offering extra payments to workers who exert a higher effort level than their co-workers. These more productive workers would receive a payment similar to a bonus payment additional to the normal wage payment which only relates to the chosen effort level.

Our setup builds on Abeler at al.'s (2010) study which analyzes the effects of different payment regimes in a reversed gift-exchange game. The authors compare two different treatments called *Individual-Wage Treatment* and *Equal-Wage Treatment* to analyze the impact of possible *norm violations* on the average effort levels. They distinguish between disadvantageous and advantageous norm violations, where a disadvantageous (advantageous) norm violation is defined as a situation in which an agent exerts lower (higher) efforts but does not receive a higher (lower) payoff than the co-worker. The paper documents that norm violations lead to a substantial crowding out effect, that is, workers who face a disadvantageous norm violation lower their effort in the following period. This results in a significant treatment effect, as there are many more norm violations in the treatment where employers cannot discriminate in wages compared to the treatment where this possibility is given. Abeler et al. (2010) report that the average effort level of the Individual-Wage Treatment is roughly twice as high as the average effort level in the Equal-Wage-Treatment.

In this paper we replicate Abeler et al. (2010) Individual-Wage Treatment and compare it to a baseline treatment where only one employee is matched to one employer and where competition in the workplace cannot play a role. In both treatments the employees act as first movers and simultaneously decide about their effort choices, afterwards the employers choose the corresponding wages for each of the workers. The principals have the possibility to determine an individual wage payment for each employee and to therefore set tournament incentives. Put differently: as the employers can observe the employees' effort choices, they can reward employees who exerted higher efforts by paying them higher wages. Thus, in the treatment with two employees per firm, workers face competitive pressure while choosing their effort levels.

Competition in the workplace can also be regarded as a dynamic process which is connected to learning behavior. First, fast-learning employees will have significant advantages in the intra-company competition. Second, before new employees can possibly engage in the internal competition, they have to learn about the competitiveness of their new workplace. That is, they have to get to know if or to what extent their employer reciprocates competitive behavior in the workplace and they need to find out to what extent or on which occasions their co-workers compete with each other. This line of reasoning shows the importance of learning processes in the context of competition in the workplace.[7]

To the best of our knowledge our paper is the first which analyzes this kind of competition in an experiment. Other studies (e.g., Fehr et al. (1998a), Fehr and Falk (1999), Bartling et al. (2010) and Brandts et al. (2010)) address competition from another point of view: they analyze competition *for* the workplace where employees compete to be hired in contrast to the competition *in* the workplace phenomenon where employees who work for the same employer try to outperform one another in order they receive higher payoffs.[8]

Our results emphasize that the employers set the tournament incentives which allow competition in the workplace to occur. This is why individual effort choices are increased, especially in the early periods of the game, when a significant learning process can be

[7]Cabrales and Charness (2011) also report that teammates imitate their "colleagues" in a team production gift-exchange study. The authors report that low-skilled agents show enhanced social learning and are more likely to reject an offered contract menu after their teammate also rejected a contract menu in the previous period.

[8]For experimental evidence on the effects of competition see Niederle and Vesterlund (2007) or Flory et al. (2010) These settings are different from ours in that they do not analyze gift-exchange setups. They focus on gender effects due to competitional effects. Chen (2003) points out that too harsh internal competition may create incentives to sabotage the work of abler colleagues. Note that our framework instead focuses on the possibility to outperform colleagues in terms of effort choices.

found. Furthermore we show that imitation learning[9] serves as an explanation for this learning process. The data highlights that employers are also affected by competition between the workers. In the competitive treatment they significantly reduce their generosity towards the employees over time. In the non-competitive treatment this effect vanishes.

Our paper is organized as follows: in the next section we present our experimental design and our results are presented in Section 3, Section 4 concludes.

4.2 Experimental Design and Procedures

In this experiment we consider a two-stage game with two different types of players: employers and employees who are matched into firms. Compared to the standard gift-exchange game the timing of our game is reversed: in the first stage the employees choose an effort level and in the second stage the employers determine a wage payment for the employees. We choose this approach because in the standard gift-exchange game employees have the possibility to shirk, i.e., they can choose minimum effort levels despite of having received positive wage payments. This aspect complicates the formation of competitive pressure between the workers as they both have an incentive not to exert above minimum effort levels. This is why the reversed gift-exchange game is better suited to tackle our research questions.

Effort Level e	1	2	3	4	5	6	7	8	9	10
Costs $c(e)$	0	1	2	4	6	8	10	13	16	20

Table 4.1: Effort - Cost-of-effort relation

Effort is costly to employees and beneficial to the employer while wages are beneficial to employees and costly to the employer. The workers' and employer's actions and the corresponding payoffs are exactly the same as in Abeler et al. (2010). One unit of effort

[9] Imitation learning was theoretically introduced by Vega-Redondo (1997), for further evidence on imitation learning see, e.g., Huck et al. (1999), Offerman et al. (2010), Apesteguia et al. (2007)

increases the payoff of an employer by 10 units, whereas the employees' payoff is reduced by respective effort costs (see Table 4.1). Employees can choose effort levels between one and 10. The wages paid by the employers are added to the employees' payoff and subtracted from the employer's payoff. The wages chosen by the employer must not exceed 100.

We compare two different treatments: the *Single-Employee-Treatment (SET)*, and the *Multiple-Employees-Treatment (MET)*. Following Abeler et al. (2010), the base game of both treatments is repeated for 12 periods and a random matching routine is employed. In contrast to Abeler et al. (2010), the only difference between our treatments is the number of employees per firm: in SET there is only one employee per firm whereas firms consist of one employer and two employees in MET. We also applied a slight change in SET payoffs compared to the payoffs used in Abeler et al.: to avoid wealth effects and to ensure comparability, we doubled employers' payoffs' for SET as reported in Table 4.2 which summarizes the players' payoffs.

Treatment	Payoff Employer	Payoff Employee i
SET	$2 \cdot (10e - w)$	$w - c(e)$
MET	$10(e_1 + e_2) - w_1 - w_2$	$w_i - c(e_i)$

Table 4.2: Payoffs

In MET both employees are paid from the money generated by their effort choices. The employer has to decide on the wages for both employees. Evidence of prior gift-exchange games (e.g., Fehr et al. (1993)) suggests that due to reciprocal behavior employees exert higher levels of effort if employers pay higher wages. Due to the reversed timing, it should be the employers who show reciprocal behavior. However, as employers can observe the efforts of both workers it might be that "tournament incentives" are implicitly set - because of the reversed move order employers can easily reward higher effort choices. That is, they can pay higher wages to employees who exert a higher effort compared to their co-workers. In other words, the employees might compete for high wages by exerting high effort levels. The difference in the number of employees per firm therefore enables us to control for this competition in the workplace effect generated by the existence of a second employee.

The experiment was conducted in November 2010 at the *AWI Lab* of the University of Heidelberg using the z-Tree software package by Fischbacher (2007) and Greiner (2004) online recruitment system. At the beginning of the experiments participants were randomly placed into matching groups which remained constant for the whole session. Each matching group comprised three firms, i.e., three employers and three workers in SET and three employers and six workers in MET. At the beginning of each period the members of a matching group were randomly matched into firms. This procedure resulted in nine independent observations for the SET and four independent observations for the MET.

Additional data was provided by Abeler et al. who conducted prior sessions in April 2005 at the *BonnEconLab* of the University of Bonn. Table 4.3 compares the Abeler et al. (2010) data to our MET observations.

Dataset of	Avg. Effort	Avg. Wage
Abeler et al. (2010)	8.21	31.97
Our MET data	8.09	29.32

Table 4.3: Comparison of the datasets (Benndorf and Rau (2012); Abeler et al. (2010))

A Mann-Whitney test shows that there exists no difference between the average effort levels of Abeler et al. (2010) and our data (two-sided p-value = 0.999). The same is true when considering average wages (two-sided Mann-Whitney p-value = 0.865). Thus we successfully replicate Abeler et al.'s (2010) results. We pool the data elicited by Abeler et al. (2010) with our MET data in order to increase the informative value of the statistical analysis. The Abeler et al. sessions comprise eight independent matching groups, thus we analyze 12 independent observations for MET and nine independent observations for SET.

In total, 90 subjects participated in our experimental sessions. In SET 54 subjects took part and 36 subjects participated in our MET-sessions. In both treatments the base game was repeated 12 times and a session took about one hour. Each participant started with an endowment of 400 points which also served as show-up fee for the participants. The profits achieved by the participants were converted at an exchange rate of 0.01 Euro/point. This resulted in an average payoff of € 10.33 which corresponded to about $14.05 at that point in time.

4.3 Results

In this section we present the results of our experiment. First, we present the average effort levels and the corresponding statistical analyses. Afterwards we consider the development of the average effort levels and analyze the employers' behavior dependent on efforts exerted. Finally we give a short interpretation of our observations. We report two-sided p-values throughout.

4.3.1 Competition in the workplace

Table 4.4 summarizes the average effort levels in different periods of time. It shows that average effort is higher in MET compared to SET. That is, workers in the competitive treatment exert an average effort level of 8.17 in contrast to SET where average effort is only about 7.44. Nevertheless, this difference is not statistically significant (p-value = 0.255).

	First half			**Second half**			
	Periods 1-3	Periods 4-6	Increase	Periods 7-9	Periods 10-12	Increase	Avg.
SET	7.22	7.31	0.09	7.86	7.35	-0.51	**7.44**
MET	7.65	8.68	1.03	8.39	7.94	-0.55	**8.17**

Table 4.4: Average effort levels over time

However, considering the first half of the game (periods 1-6) , we observe a significant learning effect in MET: there is an intense increase in the effort levels (1.03 units of effort; Wilcoxon matched-pairs test p-value = 0.004), whereas effort levels are nearly unchanged in SET: here, the difference is not statistically significant, it is only about 0.09 effort units (Wilcoxon matched-pairs test p-value = 0.407). These findings can be interpreted as a consequence of internal competition between the MET workers. The result indicates that learning plays an important role for competition in the workplace.[10]

Result 1. *On average the introduction of a second employee does not result in increased effort levels. However, in periods 1-6 there is an intense increase of efforts in the competitive treatment whereas there is no such increase in the non-competitive treatment.*

The existence of a second employee could also induce the employers to set tournament incentives,[11] i.e., to pay a premium to those employees exerting higher efforts than their co-worker (henceforth "high types" and "low types"). We therefore analyze wage payments for high and low types in MET. We find evidence for tournament incentives set by the employers, that is, high types receive higher wages (36.69) whereas low types only receive a wage of 19.26. These differences in our results cannot be exclusively explained by the effort levels, it also seems that agents exerting a higher effort than their co-workers receive a premium simply for being the one exerting a higher effort. Table 4.5 reports the results of a regression with *wage* as a dependent variable[12] controlling for this phenomenon. Note that the regression is restricted to cases where the employees choose different effort levels, because tournament incentives do not cover cases where the employees choose identical effort levels.

The independent variables of the regression are *effort* (which represents the effort chosen by the employees), and *higher* (which is a dummy variable equal to one if a subject has chosen a higher effort than his co-worker). Furthermore, we control for the corre-

[10] We will return to learning behavior later on.

[11] We thank Matthias Wibral for pointing out this issue.

[12] The standard errors of this regression are adjusted for 36 clusters representing individual employers as all employers determine two wage payments per round.

	$wage$
$effort$	3.909***
	(0.508)
$higher$	4.427**
	(1.650)
$period$	-0.440*
	(0.235)
$constant$	-0.922
	(2.787)
# obs.	576
R-squared	0.344

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4.5: Regression results of tournament incentives

sponding *period* (1-12). The OLS regression reveals that *effort* and *higher* are significant. That is, higher effort levels increase employees' wages. It also shows that employers pay higher wages to workers who outperform their co-workers in contrast to workers who do not "win" the tournaments. We thus find support that tournament incentives are set by our employers. Finally, we find that later periods moderately lead to smaller wage payments for the workers (we will discuss this in more detail in section 4.3.4).

Result 2. *The employers try to trigger internal competition between the employees by making an extra payment to employees who exert a higher effort level compared to their co-worker.*

4.3.2 Behavior over time

The development of the average effort levels over time is significantly different across treatments which is illustrated by Figure 4.1. It comprises both treatments: the blue line which represents SET and the black line which depicts MET. In general, the MET effort levels are higher than the SET effort levels.[13] Considering the early periods (the first half of the game, periods 1-6) we find a steep increase of efforts in MET,[14] i.e., in MET there is a positive correlation between effort and period in the first half of the game (sign-test p-value < 0.01). There is no such correlation in SET. The sign-test p-value for periods 1 - 6 is 0.289, hence, in contrast to MET we cannot reject the null-hypothesis of no correlation between effort and period in SET.[15]

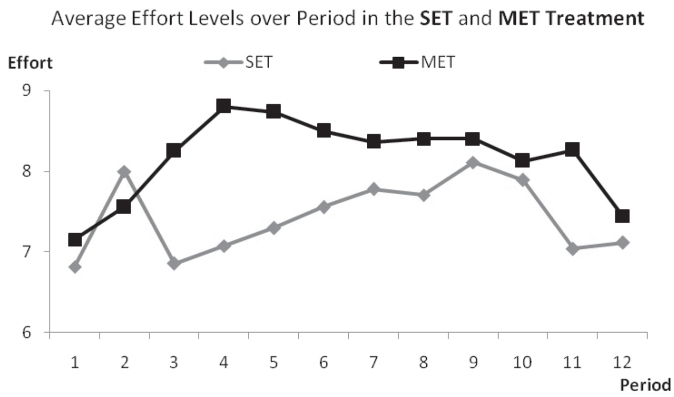

Figure 4.1: Average Effort Levels over Period

The different dynamics documented above support our intuition that the learning processes are more efficient in MET compared to SET. Focusing on the first period of the game we find no difference between the average effort levels of our treatments (Mann-Whitney p-value = 0.776). At this point in time neither MET nor SET employees have any information about the employers' reactions to their effort choices. The employees will first need to learn their employer's reaction and to gauge to what degree the employers reciprocate high effort levels. In MET this "early learning process" results in a significant increase of the average effort levels, but there is no indication for such a learning process in SET.

[13]Note that the black line is above the blue line in each single period except period 2.
[14]In the early periods, 11 out of 12 matching groups have Spearman's rank correlation coefficients between effort and period that are positive.
[15]The periods 3-9 seem to be a more promising choice to detect a correlation effort and period in SET (compare Figure 4.1). In this interval there is also no significant correlation, the corresponding p-value is 0.180.

Result 3. *In MET effort levels correlate positively and significantly with the period variable during the early periods of the game. This kind of correlation cannot be found in SET.*

4.3.3 Imitation learning

One possible driver for the different dynamics is imitation learning,[16] a simple learning process suggesting that players imitate the most successful action choice of the previous period. The concept of imitation learning cannot be applied in the SET treatment because there is only *one* employee who is employed by an employer. Therefore this worker would not get any information about the effort level a co-worker might chose. In contrast, MET employees have all the information necessary to make use of the imitation heuristic by monitoring the effort levels chosen by their co-workers. It is the aim of this subsection to figure out whether imitation learning is present in MET.

In this analysis we focus on a subsample of our dataset. We restrict our regression to observations where subjects earned less than their co-worker in the previous period. Otherwise imitation learning would suggest that a large fraction (about 67% of all cases in MET) of employees leave their effort choices unchanged and this may bias the results.

Using a random-effects model and adjusted standard errors for the 12 clusters of the MET match groups we estimate the following regression:

$$e_{i,t} - e_{i,t-1} = \alpha_0 + \alpha_1 \cdot imit + \alpha_2 \cdot imit\_nv + \alpha_3 \cdot nv + \epsilon$$

where $e_{i,t}$ and $e_{i,t-1}$ denote the effort choice of subject i in period t and period $t-1$, respectively. $imit$ denotes the difference between the most successful effort choice of the

[16] More detailed descriptions can be found in Vega-Redondo (1997), Huck et al. (1999), Offerman et al. (2002) or Apesteguia et al. (2007)

	(1)	(2)
	$e_{i,t} - e_{i,t-1}$	$e_{i,t} - e_{i,t-1}$
$imit$	0.464***	0.507***
	(0.076)	(0.0996)
$imit\_nv$		-0.743***
		(0.1156)
nv		-2.061***
		0.5661
$constant$	0.066	0.741
	(0.153)	(0.2496)
# obs.	261	
# subjects	65	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4.6: Regression results of imitation learning

previous period and the effort level chosen by the corresponding subject in the previous period, i.e., $imit$ always specifies the exact change of an imitation player's effort choice between period $t - 1$ and t. It contains integers between -10 and 10 if the employee was less successful than his co-worker. If both employees received the same payoffs in the previous period, we assume that imitation players consider only the lowest effort choice because it is less risky. $imit\_nv$ is an interaction term between $imit$ and the nv variable which is a dummy indicating whether the corresponding subject experienced a disadvantageous norm violation in the previous period (see Abeler et al. (2010)). Advantageous norm violations are not covered by this analysis as we consider only observations of subjects who earned less than their co-worker in the previous period.

The regression results are reported in Table 4.6. We find that the employees are prone to imitation. In our first regression we do not control for norm violations and find that the effort changes are largely in line with the imitation predictions. The $imit$ coefficient is about 0.5 and highly significant.

We also find that norm violations have a significant and substantial impact on the effort choices. The coefficient for the dummy is about -2 and the one for the interaction term about -0.74, both coefficients are highly significant. The sign of the sum of the $imit$ and $imit\_nv$ coefficients is particularly striking: it is negative, indicating that the employees still use the information about their co-worker to adjust their effort choices, but this adjustment based on the information about the other employee results in an increase of the effort levels and not in the decrease predicted by imitation learning.[17]

Our results do imply a decrease of an employee's efforts after a norm violation but this decrease is captured by the nv dummy which represents a general effort decrease that is not related to imitation learning. This is why we conclude that norm violations can be interpreted as a disturbance of imitation learning: the employees do not comply with the concept's predictions after experiencing a norm violation.

Result 4. *Imitation learning can explain the different developments of the average effort levels across treatments. Norm violations can be interpreted as disruptions of the imitation learning process.*

4.3.4 Employers' reactions

Our results in section 3.1 have shown that employers' behavior is also affected by the existence of a second employee. We showed in this section that due to the competitive environment the payment structure of MET employers enables incentives similar to those in tournaments.

However, the competitional environment between the employees might also reduce the employers' incentives to reciprocate high effort choices. The wage payment to the employee exerting less effort may be lower because imitation learning induces that em-

[17] Imitation will always predict effort decreases if there was a disadvantageous norm violation in the previous period. A negative coefficient of the interaction term therefore indicates a positive change of effort levels.

ployee to adapt the co-worker's effort choice anyway. The enhanced efficiency of the employee's learning process may also result in generally higher effort levels and therefore a reduction in the average wage payments. The following paragraphs analyze the employers' reactions to the existence of the second employee.

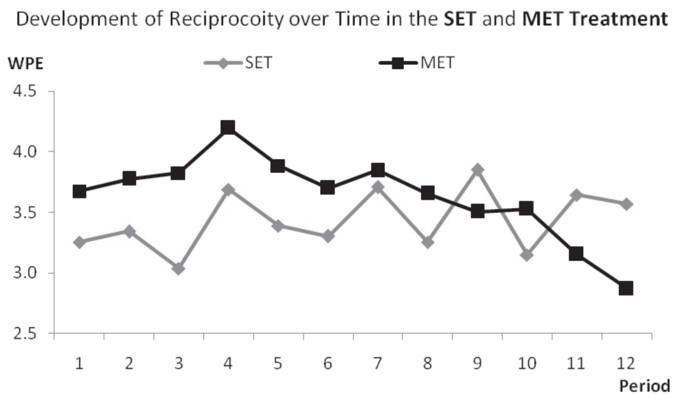

Figure 4.2: Development of Reciprocity over Time

Figure 4.2 shows the development of the average *wage payments per unit of effort exerted (WPE)*. Even though the WPE measure has comparable levels across treatments there is a substantial difference regarding the dynamics. In SET there is no distinct development of this measure: there is no apparent difference between the WPE of the early periods in SET compared to the last periods of the game. However, in MET a clear effect can be found, i.e., there is a decreasing trend of the average wage payments per unit of effort. In contrast to the early periods, employers significantly decrease their wage payments in the final periods. The time trend of the average wage payments per unit of effort exerted is also significantly different across treatments. In MET there is a significantly negative correlation between the period variable and the average wage payment per unit of effort exerted (sign-test p-value = 0.038), but there is no such correlation in SET (sign-test p-value = 0.508).[18] There is a decreasing trend in MET but there is no such development in SET.[19]

[18]In MET 10 out of 12 matching groups have negative Spearman's rank correlation coefficients. In SET only three of nine groups have negative coefficients.

[19]Note that in the first period, employers pay higher levels of WPE in MET compared to SET. However, this difference is not significant (Mann-Whitney p-value = 0.292 on individual level data).

Figure 4.3 supports this result: the left diagram presents MET-employers' wage payments per effort level (split into three effort intervals) in periods 1-6 compared to the wage payments per effort level in periods 7-12. The right diagram presents the same analysis for SET-employers' wage payments.

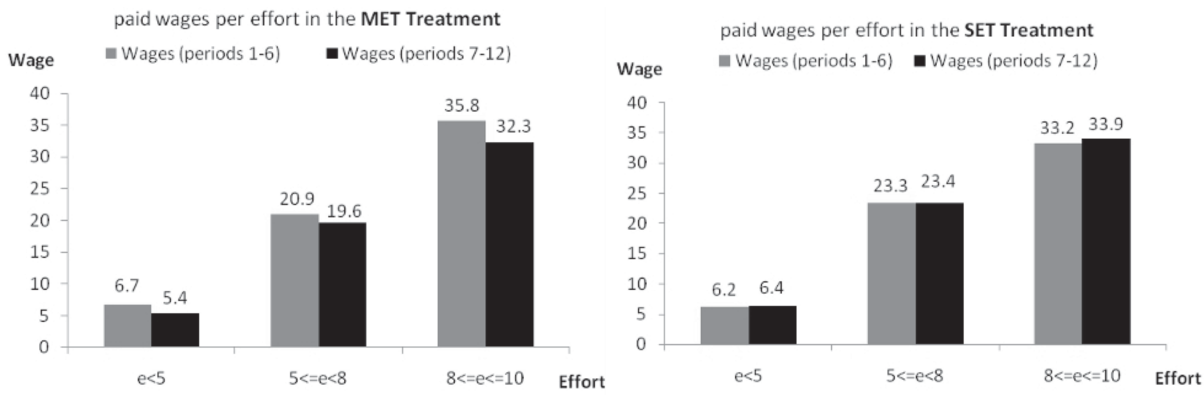

Figure 4.3: Paid wages per effort in the MET- and SET-Treatment (P:1-6 vs. P:7-12)

It can be clearly seen that in MET the wages paid per effort interval decreases in periods 7-12 compared to periods 1-6. This holds true for every effort interval. When efforts were below five, average wage payments decrease in periods 7-12 from 6.7 (periods 1-6) to 5.4 (periods 7-12). This difference is statistically significant (Wilcoxon matched-pairs test p-value = 0.028). For $5 \leq e < 8$ average wages also decrease from 20.9 (periods 1-6) down to 19.6 (periods 7-12). However, this difference is statistically not significant (Wilcoxon matched-pairs test p-value = 0.824). When employees' average effort level was $8 \leq e \leq 10$, average wages also decreased from 35.8 (periods 1-6) to 32.3 (periods 7-12). This difference is weakly significant (Wilcoxon singrank test p-value = 0.100). If we focus on the diagram for SET, we do not observe these findings at all: in each of the three intervals there is nearly no development over time. That is, employers always choose the same average wage for each of the three effort intervals. These results once more underline that the employers' reciprocity decreases over time in the presence of two workers. Because of competition in the workplace between workers in MET, employers seem to play the two workers off against each other by systematically decreasing individual wages.

Result 5. *In contrast to SET the wages paid per unit of effort decrease over time in MET. In this environment with two employees, employers decrease their reciprocal behavior in the second half of the game.*

4.4 Conclusion

How should an efficient workplace be designed? Our results may give answer to this question. We find that employing more than one worker results in a clear cut improvement of employees' effort levels over time because employers can induce tournament incentives by systematically rewarding workers who outperform their co-worker. Workers in the multiple employee treatment are therefore faced with competitive pressure and increase their effort levels compared to the single employee case. Even though competition in the workplace does not affect the aggregate effort levels, competition intensifies over time. In particular, there is a distinct, increasing development of the effort levels in the first half (periods 1-6) of the game if and only if there are multiple employees. This is a clear indication of the enhanced learning of employees who can compare their outcomes to those of a co-worker. The learning process in the environment of multiple employees is primarily driven by imitation. Considering only cases where imitation predicts a non-zero change of effort levels, we find that employees' behavior is in line with the imitation predictions. The results confirm those of the economic literature on peer effects at the workplace which report that workers increase the average effort levels over time in the presence of co-workers because of learning behavior and social pressure (compare, e.g., Falk and Ichino (2006); Mas and Moretti (2009)).

Analogously to Abeler et al. (2010) our findings document that norm violations play an important role in MET. The results reveal that norm violations are harmful for imitation learning. That is, employees systematically decrease their effort choices after prior norm violations. The fact is that employees are prone to imitation and competition adds valuable insights especially for organizational economics. Employers or managers should note that competition can substantially stimulate learning processes and that norm violations abate this enhanced learning.

The analysis also shows that a second employee working for the same employer influences employers' behavior. The increasing competitive pressure between the workers forces employees to raise their effort choices over time. The data documents that employers anticipate this. We find that employees systematically pay a smaller wage for each unit of exerted effort over time. This finding is striking because there has not been any other gift-exchange study which reports decreasing levels of reciprocity over time.[20] For future research it would be interesting to further analyze this finding, and studies testing this phenomenon with setups involving more periods would be particularly promising.

[20] The prevalance of reciprocity in gift-exchange games is well-documented in the literature (for surveys see Gächter and Fehr (2002), Charness and Kuhn (2011).

References

[1] Abeler, J., Altmann, S., Kube, S. and Wibral, M.(2010). “Gift exchange and workers’ fairness concerns: When equality is unfair”, Journal of the European Economic Association 8, 1299-1324.

[2] Altmann, S., Falk, A. and Huffman, D. (2009). “Contract enforcement, fairness, and involuntary unemployment”, Working Paper, IZA.

[3] Apesteguia, J., Huck, S. and Oechssler, J. (2007). “Imitation theory and experimental evidence”, Journal of Economic Theory 136, 217-235.

[4] Bartling, B., Fehr, E. and Schmidt, K. (2010). “Screening, competition, and job design: Economic origins of good jobs”, Institute for Empirical Research in Economics, Working Paper.

[5] Brandts, J. and Charness, G. (2004). “Do labour market conditions affect gift exchange? some experimental evidence”, The Economic Journal 114, 684-708.

[6] Brandts, J., Gërxhani, K., Schram, A. and Ygosse-Battisti, J. (2010). “Size doesn’t matter! gift exchange in experimental labor markets”, Journal of Economic Behavior & Organization 76, 544-548.

[7] Cabrales, A. and Charness, G. (2011). “Optimal contracts with team production and hidden information: An experiment”, Journal of Economic Behavior & Organization 77, 163-176.

[8] Charness, G. (2000). “Responsibility and effort in an experimental labor market”, Journal of Economic Behavior & Organization 42, 375-384.

[9] Charness, G. and Kuhn, P. (2007). "Does pay inequality affect worker effort? experimental evidence", Journal of Labor Economics 25, 693.

[10] Charness, G. and Kuhn, P. (2011). "What can labor economists learn from the lab?", in O. Ashenfelter and D. Card (eds), Handbooks in Economics Handbook of Labor Economics 4, 229-330.

[11] Chen, K. (2003). "Sabotage in promotion tournaments", Journal of Law, Economics, and Organization 19, 119-140.

[12] Cohn, A., Fehr, E., Herrmann, B. and Schneider, F. (2011). "Social comparison in the workplace: evidence from a field experiment", Working Paper, IZA.

[13] Engelmann, D. and Ortmann, A. (2009). "The robustness of laboratory gift exchange: A reconsideration", Working Paper.

[14] Eriksson, T. and Villeval, M. (2011), "Respect and relational contracts", Journal of Economic Behavior & Organization 81, 286-298.

[15] Falk, A. and Ichino, A. (2006). "Clean evidence on peer effects", Journal of Labor Economics 24, 39-57.

[16] Falk, A. and Kosfeld, M. (2006). "The hidden costs of control", American economic review 1611-1630.

[17] Fehr, E. and Falk, A. (1999). "Wage rigidity in a competitive incomplete contract market", Journal of Political Economy 107, 106-134.

[18] Fehr, E., Kirchler, E., Weichbold, A. and Gächter, S. (1998a). "When social norms overpower competition: Gift exchange in experimental labor markets", Journal of Labor Economics 16, 324-351.

[19] Fehr, E., Kirchsteiger, G. and Riedl, A. (1993). "Does fairness prevent market clearing? an experimental investigation", The Quarterly Journal of Economics 108, 437-459.

[20] Fehr, E., Kirchsteiger, G. and Riedl, A. (1998b). "Gift exchange and reciprocity in competitive experimental markets", European Economic Review 42, 1-34.

[21] Fischbacher, U. (2007). "z-tree: Zurich toolbox for ready-made economic experiments", Experimental Economics 10, 171-178.

[22] Flory, J., Leibbrandt, A. and List, J. (2010). "Do competitive work places deter female workers? a large-scale natural field experiment on gender differences in job-entry decisions", NBER Working Paper.

[23] Gächter, S. and Fehr, E. (2002). "Fairness in the labour market", Surveys in Experimental Economics 95-132.

[24] Gächter, S. and Sefton, M. (2008). "The impact of social comparisons on reciprocity, Working Paper, IZA.

[25] Gächter, S. and Thöni, C. (2010). "Social comparison and performance: Experimental evidence on the fair wage-effort hypothesis", Journal of Economic Behavior & Organization 76, 531-543.

[26] 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, 79-93.

[27] Heinz, M., Juranek, S. and Rau, H. A. (2012). "Do women behave more reciprocally than men? gender differences in real effort dictator games, Journal of Economic Behavior & Organization 83, 105-110.

[28] Huck, S., Normann, H. and Oechssler, J. (1999). "Learning in cournot oligopoly an experiment", The Economic Journal 109, 80-95.

[29] Lazear, E. and Rosen, S. (1981). "Rank-order tournaments as optimum labor contracts", The Journal of Political Economy 89, 841-864.

[30] Marino, A. and Zabojnik, J.. (2004). "Internal competition for corporate resources and incentives in teams", RAND Journal of Economics 710-727.

[31] Mas, A. and Moretti, E. (2009). "Peers at work", American economic review 99, 112-145.

[32] Maximiano, S., Sloof, R. and Sonnemans, J. (2007). "Gift exchange in a multi-worker firm", The Economic Journal 117, 1025-1050.

[33] Niederle, M. and Vesterlund, L. (2007). "Do women shy away from competition? do men compete too much?", Quarterly Journal of Economics 122, 1067-1101.

[34] Offerman, T., Potters, J. and Sonnemans, J. (2002). "Imitation and belief learning in an oligopoly experiment", Review of Economic Studies 69, 973-997.

[35] Owens, M. and Kagel, J. (2010). "Minimum wage restrictions and employee effort in incomplete labor markets: An experimental investigation", Journal of Economic Behavior & Organization 73, 317-326.

[36] Pereira, P., Silva, N., and Andrade e Silva, J. (2006). "Positive and negative reciprocity in the labor market", Journal of Economic Behavior & Organization 59, 406-422.

[37] Vega-Redondo, F. (1997). "The evolution of walrasian behavior", Econometrica 65, 375-384.

Appendix

Experimental Instructions: Multi-Employee Treatment (MET)

Please read these instructions carefully. At the end of these instructions you will be asked to answer several control questions. The experiment will begin as soon as each participant answered the control questions correctly. The experiment is anonymous, i.e., you will not get to know with which other participants you are interacting.

During the experiment you can earn "Experimental Currency Unit" (ECU). Your earnings depend on your decisions and on the other participants? decisions as well. After the experiment the ECUs will be converted into Euros at the following exchange-rate:

1 ECU = 1 Cent

Please wait at your desk until we ask you to come to receive your payment. After the experiment, please bring all the documents we handed out to the place where you will receive your payment.

You begin with a starting capital of **400 ECUs (€4)**. It increases if you make profits and it decreases if you experience losses during the experiment. Note, that you can always rule out the possibility of making losses by your own decisions.

Please also note that you must not talk to the other participants during the experiment. In this case we need to abort the experiment immediately. If you have any questions please raise your hand and we will answer them personally.

- In this experiment participants either act as an employer or as an employee. At the beginning of the experiment, you will be randomly assigned one of these roles.

- Your role does not change during the experiment.

- Each period comprises two stages. In the first stage employee 1 and employee 2 each choose an effort level.

  Their decision is independent of the other employee?s decision. There are ten different effort levels the employees may choose. The lowest possible effort level is 1 and the highest one is 10. Each unit of effort exerted by an employee produces 10 ECUs for the employer. For instance if the effort level is 1 the employer will receive 10 ECUs, if the effort level is 2 the employer will receive 20 ECUs, etc. If the effort level is 10 the employer receives 100 ECUs.

- Choosing an effort level is costly for the employees. The higher the effort level, the higher the corresponding costs.

However, the costs only depend on the effort level an employee chooses for himself. The effort level chosen by the other employee does not affect the costs. For an employee, the costs of choosing an effort are as follows:

effort level	1	2	3	4	5	6	7	8	9	10	
costs	0	1	2	4	6	8	10	13	16	20	ECUs

Table 4.7: Possible effort choices and corresponding costs.

Thus, choosing an effort level of 1 does not provoke any cost for the employee. Choosing a level of two costs 1 ECU, etc.; choosing a level of 10 costs 20 ECUs. All employees have the same cost table and it is the same for all periods.

In the second stage the employer is informed about the effort choices of employee 1 and employee 2. After that the employer chooses wage payments w_1 and w_2 for employee 1 and employee 2, respectively. The wage payments for the employees may either be equal or different. A wage payment for an employee must not be lower than 0 ECUs and it must not exceed 100 ECUs.

At the end of a period both employees and the employer are informed about the effort levels, about the wage payments and about the resulting profits.

Thus, in each period, a participant?s profit in ECUs is as follows:

Employer's profits =

10 x effort level chosen by employee 1 + 10 x effort level chosen by employee 2 - wage payment for employee 1 (w_1) - wage payment for employee 2 (w_2)

Employee 1's profits =

wage payment for employee 1 - cost of effort chosen by employee 1

Employee 2's profits =

wage payment for employee 2 - cost of effort chosen by employee 2

At the end of the experiment, you will receive your total profits. They consist of the starting capital and the sum of the profits earned in each period of the experiment. 1 ECU corresponds to €0.01.

At the end of each period, the employees are informed about their wage payment in the upper part of the feedback screen. In the middle of this screen a summary of choices and profits of the corresponding period is displayed. In the lower part, employees can track their total profits, i.e., their starting capital plus the sum of their earnings in previous periods. The screenshot below is an example screen for employees. The screen the employers face is similar but here, the upper part is empty.

Please raise your hand if you have any further questions.

Chapter 5

The Disposition Effect in Individual and Team Investments: Experimental Evidence

5.1 Introduction

Due to financial markets' high volatility and the fast speed of pricing processes, investment decisions became more and more complex. Investors have to decide under time pressure and therefore often make not the optimal portfolio decisions. These biases describe phenomenons where investors are rather driven by psychological forces than rational behavior (Tversky and Kahneman (1974), Kahneman and Tverksy (1979)).[1]

This paper focuses on one of theses biases: the disposition effect in individual and team investment decisions. The disposition effect is defined as a behavior where capital losses are still kept although stock prices are decreasing and at the same time it can be

[1] See Barberis and Thaler (2003) for a survey about behavioral finance literature tackling these biases.

observed that capital gains are sold (Shefrin and Statman (1985)). The disposition effect is empirically well-documented, e.g. in stock markets for private investors (Ferris et al. (1988); Odean (1998); Frazzini (2006)) and for professional traders (Garvey and Murphy (2004); Locke and Mann (2005)) and also for house owners (Genesove and Mayer (2001)). In contrast to these papers, only few papers analyze the disposition effect on an individual basis. Individual-level disposition effects are empirically and experimentally analyzed by Weber and Welfens (2007) who find them to be stable across investors.[2]

Weber and Camerer (1998) report in an experiment that investors in a laboratory experiment are also prone to the disposition effect. That is, 59% of their investors' sells are capital gains, whereas only 41% are sold as losses. Chui (2001) replicates these results for subjects in Macau. For single investment decisions, these papers highlight that the disposition effect is empirically and experimentally well-documented.

However, there is a significantly increasing fraction of daily investment decisions which is managed by portfolio management teams. Beaer et al. (2005) report that the percentage of US equity funds managed by teams, exploded between 1994 and 2003 from 5% to 46%. In the private sector there is also an increasing number of investors who tend to make their decisions jointly. The "National Association of Investors Corporation" (NAIC)[3] reports on its web page that in 2011 it encompassed around 13,000 investment clubs with 120,000 members. There only exist few empirical papers analyzing teams decision making in mutual funds Bliss et al. (2006); Baer et al. (2005)). Both papers do not focus on the disposition effect and report that team-managed funds take on less risk than individually-managed funds. Cici (2010) empirically investigates the disposition effect for funds managed by individuals and teams of portfolio managers. He finds that teams show a stronger tendency to realize more gains than losses compared to funds managed by a single portfolio manager. The paper argues that teams "gravitate" to *groupthink*, i.e.,

[2] Dhar and Zhu (2006) empirically investigate individual-level disposition effects and conclude that wealthier and professional investors are less prone to it.

[3] The NAIC is one of the most popular non-profit organizations assisting private investors. See http:\\www.betterinvesting.org

a tendency in groups that members reach their agreements without critically testing and evaluating their ideas. Hence, group members may not act objectively and adapt others' actions to reach unanimity.[4]

Although there is scarce empirical evidence about the disposition effect in team investments the individual motives (e.g. irrational behavior due to misjudgment of the stock developments[5]) remain largely unclear. Laboratory studies maybe useful to get more insights about team decisions. For instance, in economic experiments it can be controlled for investors' rationality or simplified environments can be applied to avoid possible biases.

The experimental literature about team decision making in general shows a tendency of increased rationality for teams (e.g Cooper and Kagel (2005); Kocher and Sutter (2005)). Focusing on behavioral finance, it appears that the findings are ambiguous. Rockenbach et al. (2007) find in a lottery choice task that team decisions do not differ to individual decisions in terms of expected utility theory. At the same time, teams' choices are better in terms of portfolio selection theory. Cheung and Palan (2009) show that teams of investors in an asset-market experiment generate weaker price bubbles compared to individuals. The experimental literature about risk preferences of individual and team decision-makers finds that teams' lottery choices are more risk-averse than individuals' (Masclet et al. (2009), Baker et al. (2008); Schupp and Williams (2008)). Focusing on loss aversion Sutter (2007) finds that single investors and teams both are prone to myopic loss aversion.[6]

Although there is experimental evidence for the disposition effect in individual decision making, there is no paper which addresses this issue for team decision making.

[4]Bénabou (2009) also higlights in a theoretical framework that groupthink might have been an impulsive factor for failures of companies like "Enron" and "Worldcom" or the financial crisis (see also Janis (1972).

[5]For instance if investors are reluctant to sell capital losses, it might be the case that they believe in mean reversion, i.e., they expect that they own good stocks which will soon increase in value.

[6]Gächter et al. (2007) report in their study that inindividual-level loss aversion is very stable. They find a high correlation between riskless and risky choices.

Because of the scarce and ambiguous evidence of team decision making in experimental asset markets, this paper focuses on the disposition effect in single and team investor decisions in the laboratory. It replicates Weber and Camerer's (1998) framework where investors can buy six risky assets to compare individual individual-level disposition effects of single investors and teams of *two* investors. The study compares this setup to a framework where *two* investors decide about a joint portfolio to analyze whether single and team investors show different degrees of the disposition effect.

The results in the baseline treatment adequately replicate Weber and Camerer's (1998) findings and document that single investors sell most of their stocks as capital gains (63%). Teams are also prone to a disposition effect and dispose 60% of their of their sells as capital gains. Furthermore they show a strong tendency to sell significantly more often after an increase of the last period's stock price in contrast to individual investors. The disposition effect in group decisions is also present when its calculation is based on the proportion of gains (losses) realized. Teams periodically realize a significantly higher proportion of their capital gains, whereas they are more reluctant to realize their paper losses in contrast to individuals. The results confirm Cici's (2010) empirical findings, showing in the laboratory that teams are prone to a higher disposition effect than individuals.

5.2 Experimental Design

The experiment builds on Weber and Camerer's (1998) study about the disposition effect where subjects individually make portfolio decisions in a framework with 14 periods. There were six different assets, called: "A", "B", "C", "D", "E", and "F" which were labeled with the neutral German word "Anteile" ("parts" or "shares").[7] The prices of the six risky assets followed a distinct random process which was carefully explained to the

[7]This was introduced by Weber and Camerer (1998) to avoid framing effects.

subjects within the written instructions. The stock price of each of the six risky assets was predetermined for all of the 14 periods before the experiment started. That is, subjects could not influence the stock prices with their trading actions. The subjects were told that the pricing process of every stock consists of two stages:

Pricing process: STAGE 1

In the beginning of each period it was first determined whether the price of each asset increased (decreased). For each asset there were different chances (which were fixed for each stock during the whole experiment) for a price increase. The subjects were told in the instructions that this relates to an asset's stock type. Subjects were not told which asset (A-F) followed these types. Each asset type had a different chance for a price increase. Subjects were told that exactly *one* asset was of *one* of these types: "$++$" (65% chance of an increase), "$+$" (55% chance of an increase), "$--$" (35% chance of an increase), "$-$" (45% chance of an increase). The remaining two assets were of the (neutral) "0"-type (50% chance of an increase). Since prices always changed, the probability of a price decrease was always one minus the chance of an increase.

Pricing process: STAGE 2

In the second stage of the pricing process the magnitudes of the price changes were randomly determined for each stock. There were three possible magnitudes: 1, 3 or 5 Talers. All three changes were equally likely to occur in period t, i.e., the likelihood that stock A-F changed by 1, 3, or 5 Taler was one third. The probability that a stock price rises was not correlated with the magnitude of the price change. The expected value of a price change for a randomly-chosen asset was zero.[8] Before the experiment started the participants received information about the stock prices (A-F) of periods $-3, -2, -1, 0$. This was done to give subjects a better understanding of the stock types. The Weber and Camerer (1998) framework easily allows subjects in each period to apply Bayesian Updating, based on observed price movements of past periods. Subjects can use a simple heuristic of count-

[8]Exactly the same stock data which was determined by Weber and Camerer (1998) was used. The authors predetermined these price sequences using a random number table.

ing the number of times a stock increased to determine its type. The asset whose price has increased most often is most likely to be of the ++ type. The share which had the second highest number of price increases must be of type + etc.[9] Figure 5.1 illustrates the stock movements over time.[10]

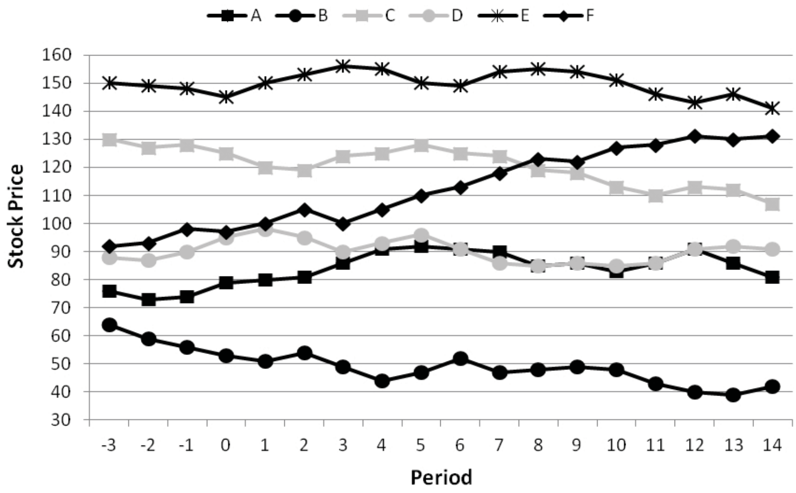

Figure 5.1: Price movements of stocks A-F over time

That is, if a subject knows which stock (A-F) is of type ++ it should be least eager to sell this stock. The opposite is true for the stock type −− which should be sold most frequently or should not be bought at all. Weber and Camerer ensure with this design that disposition effects should not occur in that setting, because they are clearly decision errors.

Two treatments were implemented: the *SINGLE*-investor treatment which is an exact replication of Weber and Camerer (1998) and the *TEAM*-investors treatment which is exactly the same as SINGLE. The only difference was that teams of *two* investors decide about a common portfolio. Before the experiment started, subjects in SINGLE and TEAM had to complete a loss-aversion elicitation task based on Gächter et al. (2007) which was

[9]If a Bayesian subject correctly applies this rule, it would have found out that after period 7, the stocks A-F were of the types: +, −−, −, 0, 0, ++. After period 14 they would judge the stocks A-F as: +, −, −−, 0, 0, ++.

[10]In the experiment the stocks were of the following types: A = +; B = −; C = −−; D = 0; E = 0; F = ++.

not incentivized.[11]

In SINGLE every subject received a show-up fee of 4 Euros and an endowment of 10,000 Taler. Subjects could only use their endowment for trading actions, but not the show-up fee. The exchange-rate of "Taler" was: 1000 Taler = 1 Euro. There were 14 periods and individual investors could buy (sell) stocks in periods 1-13.[12] Subjects did not necessarily have to invest their endowment and could not borrow money. When subjects buy (sell) an asset they pay (receive) the current stock price of the period for each asset which is traded. Subjects were not allowed to make short sales, i.e., they could only sell stocks which were in their portfolio. At the end of the experiment the subjects automatically received the value of their actual portfolio in Taler. Additionally they received the rest of their endowment in Taler. After periods 7 and 14 subjects were asked to guess the type (0, $+$, $++$, $-$, or $--$) of each of the six stocks. Subjects received 200 Taler (20 euro cent) for each correct guess. At the end of the experiment the subjects had to complete a short survey.

In TEAM exactly the same framework[13] was used. The crucial difference corresponds to the number of investors: in contrast to a single investor, rather a *team* of *two* investors decided about a joint portfolio. These teams were randomly composed, i.e. when entering the laboratory, each subject had to pick a ticket with a number indicating her matching partner. In every session of the TEAM treatment five teams of two investors were put in the laboratory.[14] In every period each team member was allowed to discuss quietly the strategies with their partner before trading took place.[15] In TEAM, each team was also endowed with 10,000 Taler. The only difference was that *each* team member received the

[11] In SINGLE this task was completed on z-Tree. In TEAM it was conducted as a "paper-and-pencil task". In TEAM I asked both team members separately.

[12] In period 14 the final stock prices were determined and all stocks were automatically liquidated.

[13] The team treatment also used exactly the same stock price movements.

[14] There also were three sessions where only four teams (instead of five teams) took part. This was due to the fact that only four teams showed up.

[15] To avoid that teams could understand what other teams were talking about, the teams were asked to sit down at predetermined desks. There were big gaps between these desks and great care was taken that subjects only talked quietly and could not understand the conversations of other teams.

total team payoff at the end of the experiment.

The experiment was programmed in z-Tree (Fischbacher (2007)). I conducted three sessions of the SINGLE treatment with a total of 55 participants (55 independent observations) and nine sessions of the TEAM treatment with a total of 84 subjects (42 independent observations). In total 139 participants took part in the experiments and were recruited with ORSEE (Greiner (2004)). The subject pool consisted of students from the University of Duesseldorf from various fields. The participants earned on average 15.89 Euros. The experiments were conducted in the *DICE Lab* in May 2011. Sessions lasted about 90 minutes.

5.3 Results

In this section the results are outlined. First, a short overview will be given, second, the statistical analyses will be conducted. I always report two-sided p-values throughout.

5.3.1 Overview

Table 5.1 depicts subjects' buying (selling) actions and the resulting payoffs.

	avg. stocks		sold (%)		kept (%)		profit	obs.
	bought	sold	as gain	as loss	as gain	as loss		
single	153	92	0.63	0.37	0.54	0.46	10333	55
	(98)	(78)	(0.25)	(0.25)	(0.34)	(0.34)	(486)	
team	163	100	0.60	0.40	0.60	0.40	10372	42
	(111)	(106)	(0.27)	(0.27)	(0.40)	(0.60)	(536)	
avg.	157	95	0.61	0.39	0.56	0.44	10350	97
	(103)	(91)	(0.26)	(0.26)	(0.37)	(0.37)	(506)	

Table 5.1: Summary statistics of the subjects' trading actions (standard deviations in parentheses)

A sold stock is referred as a “gain” stock, if its selling price is at least equal or higher than the purchase price, whereas stocks are called “loss” stocks when the selling prices are below their purchase prices. To account whether the stocks were above (below) its purchase prices the *Last-In-First-Out* (LIFO) principle is used.[16]

Single and team investors buy and sell nearly the same number of stocks and achieve the same payoffs. Single investors keep 54% of gain stocks and teams hold 60% gain stocks. Concentrating on the total number of sold gain and loss stocks, both investor types show a disposition effect and sell a significant higher fraction of gain stocks than loss stocks. Single investors sell 63% gain stocks and only 37% loss stocks (Wilcoxon matched-pairs test, $p - value <$ 0.001). Thus, the results confirm the findings of Weber and Camerer (1998) who find that subjects sell 64% gain and 36% loss stocks. Teams also dispose a significant higher fraction of gain stocks (60%) than loss stocks (40%) (Wilcoxon matched-pairs test, $p - value =$ 0.026).

Result 1. *The results in the SINGLE treatment adequately replicate Weber and Camerer's (1998) findings. Single and team investors sell a higher fraction of gain stocks compared to loss stocks.*

Subjects assessment of the stock types

Following Weber and Camerer (1998), subjects were asked *two* times for an assessment of the stock types $++, +, 0, -, --$ and afterwards individual guess scores for each stock were calculated. Table 5.2 presents the results of subjects' assessment of the stock types.

Each guess score (δ) is defined as the sum of the difference between the correct estimates and subjects' assessments. It nominally corresponds with the quality of subjects'

[16]When subjects sell stocks, the LIFO principle assumes that they first sell the stocks which were bought as a last resort. Weber and Camerer (1998) used both the LIFO and the *First-In-First-Out* principle as well, however they do not find any differences in the evaluation results.

guess score (δ)	after period 7	after period 14	avg.
single	2.15 (2.21)	2.15 (1.66)	2.15 (1.67)
team	1.17 (1.10)	1.67 (1.62)	1.42 (1.01)
avg.	1.72 (1.87)	1.94 (1.65)	1.83 (1.46)

Table 5.2: Subjects' assessment of the stock types (standard deviations in parentheses)

estimates where "0" ("12") is the best (worst) estimate. The table shows that subjects in general had a very good understanding of the stock types. Single investors have $\delta = 2.15$[17] and teams have $\delta = 1.42$ which is significantly smaller in contrast to single investors (Mann-Whitney test, $p - value = 0.030$). Thus, team membership increases the quality of the guess score. When comparing subjects' first and second guess, no difference can be found for single investors. The average of single investors' first and second guess score is 2.15 (Wilcoxon matched-pairs test, $p - value = 0.583$). Interestingly it changes for teams: in the course of the experiment teams' guess score increases from 1.17 to 1.67 (Wilcoxon matched-pairs test, $p - value = 0.101$).

5.3.2 Analysis of periodical Individual-Level Disposition Effects

Section 3.1 documents that investors are prone to a disposition effect, when controlling for the total amount of sold gains and losses. We will now focus on two methods to determine individual-level disposition effects periodically: (1) the reference-price based method (see Weber and Camerer (1998)) and (2) the percentage of gains (losses) realized (PGR/ PLR) (see Odean (1998)).

[17] The result confirms the finding of Weber and Camerer (1998): subjects in their session had an average δ of 2.27.

Disposition Effects determined by the reference-price method

The reference-price method assumes that investors use the last period's stock price as a reference point.[18] It controls, whether subjects realize more sells after an increase of last period's stock price. Weber and Camerer (1998) define a disposition effect coefficient (α) to measure disposition effects for every investor.

α_i is defined as:

$$\alpha_i = \frac{(S_+ - S_-)}{(S_+ + S_-)} \tag{5.1}$$

Where S_+ (S_-) is the number of sells after an increase (decrease) of last period's stock price. The coefficient is zero if the number of sells after price increases and decreases of the last period's stock price is the same. In this case the investor does not base her selling decisions on last period's stock prices. It is +1 (-1), if a subject only (never) sells after price increases of the last period.

The data reveals that in both treatments subjects on average have a positive α, i.e., subjects in all conditions are prone to a reference point driven disposition effect. Single investors have an $\alpha_1 = 0.22$ which is significantly greater than zero (t-test, $t(54) = 2.553$, $p-value = 0.014$). Thus, the data of the single treatment confirms Weber and Camerer's finding (1998).[19] Teams have an $\alpha_2 = 0.46$ which is also significantly greater than zero (t-test, $t(41) = 4.830$, $p-value < 0.001$). They even show a significantly stronger reference price disposition effect compared to individuals (Mann-Whitney test, $p-value = 0.046$). Figure 5.2 illustrates that the CDF's of the α's between single and team investors are significantly different (Kolmogorov-Smirnov test, $D = 0.309$; $p - value = 0.016$).

[18] In the Weber and Camerer (1998) framework, stocks either increase or decrease in each period.
[19] The mean of Weber and Camerer's individual-level disposition effect is $\alpha = 0.30$.

The diagram depicts that less single investors (64%) have a positive α in contrast to teams (79%).

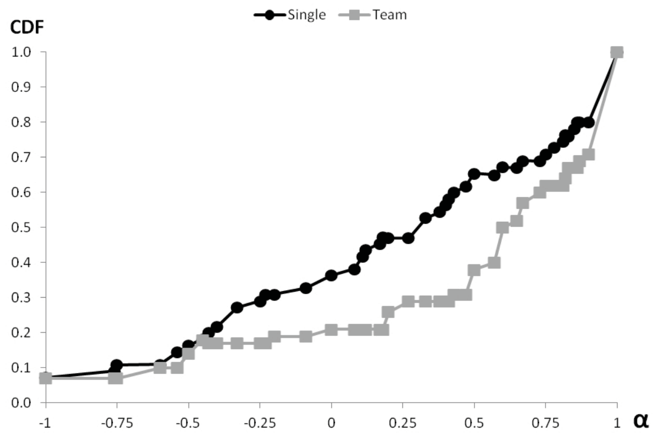

Figure 5.2: CDF of single and team investors' α

Result 2. *Single and team investors both are prone to a reference-price based disposition effect and sell more frequently after an increase of last period's stock price. This effect is significantly more pronounced for teams in contrast to single investors.*

Disposition Effects determined by the proportion of realized gains (PGR) and the proportion of realized losses (PLR)

The Weber and Camerer (1998) framework implies that in every period the stock price will either increase (decrease) in value. It follows, that if subjects own a stock with capital gains and if they do sell it, they are prone to a disposition effect in the gain domain. When subjects hold a stock with a capital loss and if they do not dispose it, subjects' behavior corresponds to a disposition effect in the loss domain. We introduce a measure based on stocks' purchase prices to control for the disposition to sell capital gains (losses) in every single period (see Odean (1998); Weber and Welfens (2007)). Following Odean (1998) it can be defined:

$$\textbf{Proportion of Gains Realized (PGR)} = \frac{Realized\ Gains}{Realized\ Gains + Paper\ Gains} \quad (5.2)$$

$$\textbf{Proportion of Losses Realized (PLR)} = \frac{Realized\ Losses}{Realized\ Losses + Paper Losses} \quad (5.3)$$

Where *Realized Gains* (*Realized Losses*) are defined as the aggregate number of stocks in the whole portfolio (stocks A-F) sold as gains (losses)[20] in every period. *Paper Gains* (*Paper Losses*) are the number of stocks of the *whole* portfolio (stocks A-F) whose price was above (below) their purchase price, but were not realized.[21] The individual-level disposition effect (DE) is the difference between PGR and PLR (Weber and Welfens (2007)):[22]

$$DE = PGR - PLR \quad (5.4)$$

The disposition effect is defined for values between -1 and 1. For instance if an investor has $DE = 1$ ($DE = -1$) she immediately realizes all capital gains (losses) but she is not able to realize any capital losses (gains) at the same time.[23] Table 5.3 illustrates subjects' PGR, PLR and the corresponding DE in both treatments.

	Mean PGR	Mean PLR	Mean DE	Freq. of DE > 0	Freq. of DE ≤ 0
single	0.16 (0.17)	0.18 (0.22)	-0.02 (0.29)	0.54	0.46
team	0.21 (0.22)	0.13 (0.13)	0.08 (0.28)	0.60	0.40

Table 5.3: Subjects' willingness to realize gains and losses

It can be seen that single investors realize a similar proportion of their paper gains

[20] Again, a gain stock is defined as a stock whose price is as least as high as its purchase price. Losses are defined as stocks whose prices are below their purchase price.

[21] When subjects do not own a stock which is currently in the gain (loss) domain, the PGR (PLR) is counted as a missing value. There was one team which never owned stocks with prices below the purchases prices.

[22] In this study the DE is calculated for every subject (see also Weber and Welfens, 2007; Dhar and Zhu, 2006), in contrast to Odean (1998) who derives the aggregate DE.

[23] An investor with $DE = 0$ either has the same PGR and PLR or she does not sell any stocks.

(0.16) and losses (0.18) which results in a balanced disposition effect (-0.02) which is not significantly different from zero (t-test, $t(53) = -0.496$; $p = 0.622$).

In contrast, team investors realize a considerable amount of their capital gains (0.21) and only 13 percent of capital losses, which leads to a disposition effect (0.08) which is significantly different from zero (t-test, $t(41) = 1.979$; $p = 0.055$). The lion's share of individual (54%) and team investors (60%) have positive disposition effects.

Regression Analysis of Subjects' Disposition Effects

We now estimate the following random effects model to analyze individual-level disposition effects in more detail:[24]

$$DE = \beta_0 + \beta_1\, team + \beta_2\, guess\_1 + \beta_3\, guess\_2 + \beta_4\, period + \beta_5\, team \cdot period + \epsilon \quad (5.5)$$

Where $team$ is a dummy variable indicating team membership (which is zero for single investors), $guess\_1$ and $guess\_2$ represent individuals' guess scores (between 0-12). $Period$ is an integer of the period number[25] and $team \cdot period$ is the interaction of $period$ and $team$. The cases where the subjects did not sell any stock (PGR = PLR = 0) are not covered by this analysis, as we only consider observations to study subjects' selling behavior.

Regression one and two reveal that teams show a significantly higher disposition effect than single investors. Hence, the experimental results confirm the empirical finding of **?**. The second regression shows that the coefficient of $team$ is 0.23. The positive sign indicates that team membership leads to a disposition effect which is higher by 23 percent compared to single investors.

[24] The model uses adjusted standard errors for the 97 subjects.
[25] Only periods where sells can occur (period 2-13) are considered.

	Disposition Effect (DE)			
	(1)	(2)	(3)	(4)
team	0.202**	0.225***	0.226***	0.395***
	(0.082)	(0.085)	(0.085)	(0.126)
guess_1		0.015	0.015	0.015
		(0.023)	(0.023)	(0.023)
guess_2		0.012	0.012	0.012
		(0.028)	(0.028)	(0.028)
period			-0.003	0.006
			(0.006)	(0.007)
team · period				-0.023*
				(0.012)
constant	-0.012	-0.074	-0.056	-0.115
	(0.052)	(0.077)	(0.089)	(0.095)
Observations	562	562	562	562
R-squared	0.038	0.044	0.044	0.050
Subjects	96	96	96	96

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5.4: OLS Regression of subjects' individual-level disposition effect

Astonishingly $guess\_1$ and $guess\_2$ are not significant, thus individual and team investors' guess scores do not impact the degree of investors' disposition effect. This result is in line with Weber and Camerer (1998) who also find that investors are prone to a disposition effect even though their investors have a good understanding of the stock types. Regression three and four document that the disposition effect does not change over time for single investors, whereas it moderately decreases for teams in later periods. Thus, teams show a weak learning effect, with smaller disposition effects in later periods.

Result 3. *The disposition effect dependent on subjects' willingness to realize capital gains (losses) is significantly higher for teams compared to single investors. The levels of the guess scores do not impact the level of DE.*

Frequency of Sells and Trade Size

The regression analysis emphasized that teams are prone to a significant higher individual-level disposition effect than single investors. To get more insights about the driver of this result, Table 5.5 gives an overview of single and team investors' average sell frequencies. It also compares the average trade size of individuals and teams when realizing gain (loss) stocks.

	frequency of sells		trade size	
	gain	loss	gain	loss
single	0.44 (0.27)	0.30 (0.20)	0.35 (0.24)	0.44 (0.28)
team	0.41 (0.28)	0.22 (0.12)	0.42 (0.25)	0.45 (0.25)

Table 5.5: Average sell frequency and average trade size

The table documents that when focusing on the frequency of sold gains, there is no significant difference between single investors and teams (Mann-Whitney test, $p - value$ = 0.465). However, single investors sell significantly more often their losses (30%) than teams (22%) (Mann-Whitney test, $p - value$ = 0.045). In contrast to single investors (35%), teams sell higher volumes (42%) when selling losses, whereas trade size is similar when selling gains. The analysis documents that the difference in the disposition effect between individuals and teams is mainly driven by the fact that teams behave more hesitantly when faced with capital losses. It seems that being member of a team complicates the selling process for capital losses, whereas there is no barrier when selling capital gains. The fact that teams use higher trade volumes when selling gain stocks, further drives the disposition effect.

5.3.3 Loss aversion and the role of regret

This section analyses the impact of subjects' loss aversion and the perceived regret. After the experiment subjects were asked in the survey to define on a scale between 1 (no regret) and 10 (strong regret) their feeling when observing capital losses. Table 5.6 gives an overview of subjects' loss-aversion coefficients,[26] and the perceived degree of regret. When analyzing teams I focus on the the mean of both team members' values. Furthermore the table depicts the minimum and the maximum of team members' values.

	loss aversion	regret
single	2.14 (1.30)	5.89 (2.42)
team (mean)	1.89 (0.66)	6.30 (1.98)
team (min)	1.49 (0.48)	5.17 (2.42)
team (max)	2.29 (1.03)	7.44 (1.80)

Table 5.6: Subjects' degree of loss aversion, and the perception of regret (standard deviations in parentheses)

Focusing on the maximum values of the team members, we find that they are higher for loss aversion and regret compared to single investors. On average single investors have a loss-aversion coefficient which is slightly higher than for team members. Interestingly perceived regret is higher in the presence of another team member. That is, teams have a significantly higher perception of regret (6.30) in contrast to single investors (5.89) ($\chi 2$ = 48.816, *d.f.* = 17, $p - value < 0.001$)

Table 5.7 depicts the impacts of investors' loss aversion and the perception of regret on the number of sales after price increases (α). For team members the table considers the average values. It reports spearman rank correlation tests.

	loss aversion		regret	
	spearman's ρ	p-value	spearman's ρ	p-value
impact on α				
single	0.172	0.227	0.162	0.239
team	0.032	0.843	0.406	0.008

Table 5.7: Impacts of loss aversion and the perception of regret on subjects' α

The table shows that single and team investors' degree of loss aversion does not influence the frequency of sells after price increases. When focusing on regret we find

[26] A higher loss-aversion coefficient denotes a stronger degree of loss aversion (see **?**).

that subjects' perception of regret also does not influence their α in the individual treatment. However, there is a highly significant positive correlation ($p = 0.008$) between team members' average perception of regret and the number of sells after price increases. It can also be found that the coefficient of team members (0.406) is higher compared to individuals (0.162).

Table 5.8 depicts the impacts of investors' loss aversion and the perception of regret on the percentage of losses realized (PLR). For team members the table considers the average values. It reports spearman rank correlation tests.

	loss aversion		regret	
	spearman's ρ	p-value	spearman's ρ	p-value
impact on PLR				
single	-0.046	0.747	-0.163	0.241
team	-0.213	0.180	-0.475	0.002

Table 5.8: Impacts of loss aversion and the perception of regret on PLR

It can be seen that individual investors' degree of loss aversion is not correlated with the percentage of losses realized. However, there is a weakly significant negative correlation between teams' average loss aversion coefficient and PLR (one-sided $p - value = 0.09$). The coefficient of teams is also much higher (-0.213) compared to single investors (-0.046). Thus, the degree of team investors' loss aversion also negatively impacts the percentage of realized losses. Focusing on regret we find no correlation between single investors' perceived regret and PLR. The opposite is true for teams: here we find a again a strong negative correlation between regret and PLR ($p - value = 0.002$). It is remarkably that teams coefficient (-0.475) is nearly three times higher compared to individual investors' (-0.163). Hence, Table 6 and 7 emphasize that regret has a stronger effect in teams.

5.4 Conclusion

This paper investigated individual-level disposition effects for single and team investors in a controlled laboratory experiment based on Weber and Camerer (1998). The data adequately replicates their findings for single investors, showing that individual investors sell more capital gains than losses. Interestingly, teams of two investors are also prone to a disposition effect. Although, they achieve lower guess scores than single investors and therefore should have a better idea about the stock types, they are not prone to a weaker degree of the disposition effect. Remarkably, the opposite is true: teams even show higher magnitudes of the disposition effect. On the one hand, teams are prone to a reference-price based disposition effect and sell more frequently after stock price increases of last period's stock price. On the other hand they are prone to a disposition effect based on the willingness to realize paper gains and losses in each period.

Hence, the laboratory results confirm the empirical findings of Cici (2010). The results indicate that increased rationality not necessarily leads to "better" results. It turned out that the drivers of this effect are teams' smaller trade frequency of loss stocks and the smaller trade volume of these assets. Teams also use higher trade sizes when trading gain stocks. The findings highlight that concepts as "groupthinking" may also influence group members to decide less objectively. That is, when selling stocks, an important prerequisite for team members is that they reach an agreement. The data shows that investors' degree of loss aversion has an impact for team members only. The same is true for investors' perception of regret when observing capital losses. When deciding alone it does not impact investors' selling decisions. However, when being member of a team, high degrees of regret complicate the realization of losses. This is driven by the fact that in the presence of a team member there is a stronger perception of regret.

The results maybe of importance for many real-life decisions. In times of the financial crisis, financial advice became a big topic. The huge amount of empirical data documenting that individual private investors are prone to significant disposition effects raises the question whether it is appropriate to decide jointly about portfolios. The laboratory results of this paper show that investing in teams even creates stronger disposition effects suggesting that investors should decide individually about their portfolio. For future research it will be interesting to investigate the circumstance how groups come to decisions (e.g. the role of "groupthinking") in more detail.

References

[1] Baer, M., Kempf, A. and Ruenzi, S. (2005). “Team management and mutual funds”, Unpublished working paper. University of Cologne.

[2] Baker, R., Laury, S. and Williams, A. (2008). “Comparing small-group and individual behavior in lottery-choice experiments”, Southern Economic Journal 367-382.

[3] Barberis, N. and Thaler, R. (2003). “A survey of behavioral finance”, Handbook of the Economics of Finance 1, 1053-1128.

[4] Bénabou, R. (2009). “Groupthink: collective delusions in organizations and markets, Technical report, NBER Working Paper.

[5] Bliss, R., Potter, M. and Schwarz, C. (2006). “Performance characteristics of individual vs. team managed mutual funds”, Unpublished Working Paper .

[6] Cheung, S. and Palan, S. (2009). “Two heads are less bubbly than one: team decision-making in an experimental asset market”, Experimental Economics 1-25.

[7] Chui, P. (2001). “An experimental study of the disposition effect: Evidence from macau”, The Journal of Psychology and Financial Markets 2, 216-222.

[8] Cici, G. (2010). “The relation of the disposition effect to mutual fund trades and performance”, Working paper Mason School of Business.

[9] Cooper, D. and Kagel, J. (2005). “Are two heads better than one? team versus indi-

vidual play in signaling games", American Economic Review, 477-509.

[10] Dhar, R. and Zhu, N. (2006). "Up close and personal: investor sophistication and the disposition effect", Management Science, 726-740.

[11] Ferris, S., Haugen, R. and Makhija, A. (1988). "Predicting contemporary volume with historic volume at differential price levels: Evidence supporting the disposition effect, Journal of Finance 677-697.

[12] Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments, Experimental Economics 10, 171-178.

[13] Frazzini, A. (2006). The disposition effect and underreaction to news, The Journal of Finance 61, 2017-2046.

[14] Garvey, R. and Murphy, A. (2004). "Are professional traders too slow to realize their losses?, Financial Analysts Journal 35-43.

[15] Gächter, S., Johnson, E. and Herrmann, A. (2007). "Individual-level loss aversion in riskless and risky choices", IZA Working Paper.

[16] Genesove, D. and Mayer, C. (2001). "Loss aversion and seller behavior: Evidence from the housing market, The Quarterly Journal of Economics 116, 1233-1260.

[17] 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, 79-93.

[18] Janis, I. (1972). "Victims of Group Thinking", Boston: Houghton Mifflin.

[19] Kahneman, D. and Tversky, A. (1979). "Prospect theory: An analysis of decision under risk", Econometrica 47, 263-291.

[20] Kocher, M. and Sutter, M. (2005), "The decision maker matters: Individual versus group behaviour in experimental beauty-contest games", The Economic Journal 115, 200-223.

[21] Locke, P. and Mann, S. (2005). "Professional trader discipline and trade disposition", Journal of Financial Economics 76, 401-444.

[22] Masclet, D., Colombier, N., Denant-Boemont, L. and Loheac, Y. (2009). “Group and individual risk preferences: A lottery-choice experiment with self-employed and salaried workers”, Journal of Economic Behavior & Organization 70, 470-484.

[23] Odean, T. (1998). “Are investors reluctant to realize their losses?”, The Journal of Finance 53, 1775-1798.

[24] Rockenbach, B., Sadrieh, A. and Mathauschek, B. (2007). “Teams take the better risks”, Journal of Economic Behavior & Organization 63, 412-422.

[25] Shefrin, H. and Statman, M. (1985). “The disposition to sell winners too early and ride losers too long: Theory and evidence”, Journal of Finance 777-790.

[26] Shupp, R. and Williams, A. (2008). “Risk preference differentials of small groups and individuals”, The Economic Journal 118, 258-283.

[27] Sutter, M. (2007). “Are teams prone to myopic loss aversion? an experimental study on individual versus team investment behavior, Economics Letters 97, 128-132.

[28] Tversky, A. and Kahneman, D. (1974). “Judgment under uncertainty: Heuristics and biases”, Science 185, 1124-1131.

[29] Weber, M. and Camerer, C. (1998). “The disposition effect in securities trading: An experimental analysis, Journal of Economic Behavior & Organization 33, 167-184.

[30] Weber, M. and Welfens, F. (2007). “An individual level analysis of the disposition effect: Empirical and experimental evidence”, Unpublished Working Paper.

Appendix

Experimental Instructions: Team treatment

Welcome to this experiment about decision theory. Please read these instructions carefully. At the end of these instructions there will be some control questions. Please answer these questions. After every participant answered the questions correctly, the experiment will start. In this experiment you will decide together with the other participant who was randomly matched to you. That is, all subsequent decisions will be decided together with your partner.

- During the experiment you are allowed to talk to the other partner
- In the experiment you and the matched participant have the to earn “Taler”

This depends on your joint decisions. At the end of the experiment you and the other participant earn these Talers. The exchange-rate is:

1000 Talers = 1 Euro

Here, **each** of the two participants earns this profit which was achieved commonly. For participating in this experiment each participant also receives 4€. After the experiment, please wait at your desk until we will ask you to come to get your payoff.

Please notice that you are only allowed to talk with your matched participant. If you will talk to the other persons the experiment will be finished. Please only talk quietly to the matched partner. If you have a question, please raise your hand. We will come to your desk to answer it individually.

The experiment consists of 14 periods. In every period you and your matched participant have the possibility to buy shares of the firms A, B, C, D, E, and F. Every share has a certain value in Talers in every period.

You start the experiment with an endowment of 10,000 Talers

Share price changes

The shares A-F will change in prices at the beginning of each of the 14 periods, i.e. in the subsequent period there will be no share which will have the same price as in the previous period. The share price changes have been predetermined before the experiment started. That is, all price changes of all shares are completely independent of all your buy and sell decisions. The same is true for all buy and sell decisions of the other participants of the experiment. Each of the shares A-F is of a certain type. The share types differ in their probability of increasing (decreasing) in value at the beginning of the period. The distributions of the types are given in the table below. In the experiment there will be exactly one share (of the shares A-F) which is of the type "++" and the same ist true for one share of the type "+", "-", and "–". There will be two types (of the shares A-F) of the type "0".

shares in the market	type	probability of increase	probability of decrease
1	++	65%	35%
1	+	55%	55%
2	0	50%	50%
1	-	45%	55%
1	–	35%	65%

Table 5.9: shares in the market

example:

- assume that share X is of type: "X"

- at the beginning of each period the probability of a price increase of X is: 65%
- at the beginning of each period the probability of a price decrease of X is: 35%

The final value of a share is determined as follows:

(i) At the beginning of each period a share either increases (decreases). The probability is dependent on the share's type (see table above).

(ii) Afterwards the magnitude of the price change (increase/ decrease) will be determined. The magnitude of the price change can either be of 1, 3 or 5 Talers. Every magnitude (1, 3 or 5 Talers) can happen with the same probability. That is, every magnitude (1, 3 or 5 Talers) can happen with the probability of one third. This is the same for every type, independently of its type.

Buying and selling actions of shares

In each of the 14 periods you and your matched participants have the possibility to buy and sell shares. In the following you are given an overview of the price changes of the shares A-F in the periods -3, -2, -1 and 0. This information is provided that you get an idea of the share types before the experiment starts. It is given in the table below:

	period -3	period -2	period -1	period 0
share A	76	73	74	79
share B	64	59	56	53
share C	130	127	128	125
share D	88	87	90	95
share E	150	149	148	145
share F	92	93	98	97

Table 5.10: stock price changes

For instance if you decide to buy shares of a firm then you have to pay for each share its current value. The sum of your expenditures cannot exceed your actual endowment.

Example:

- Share A's current price in period 1 is 110 Talers. Now you decide to buy five shares A.
- The expenditures for this transaction are given by: **5 * 110 Talers** = **550 Talers**
- This amount will be directly subtracted of your endowment

If you want to buy shares, you or your matched participant have to click the button labeled "Buy one share". If you want to buy more than one share, e.g., three shares, you or the matched participant have to click these button for three times.

If you already own some shares at the beginning of a period, then you have the possibility to **sell** these shares. You will receive the current value of each share which is sold. However, the numbers of sold shares cannot exceed the total number of shares owned.

Example:

- Share C's current price in period 5 is 90 Talers. Assume, you own a total of four shares C and decide to sell 3 shares C.
- This will lead to a payoff of: 3 * 90 Talers = 270 Talers.
- This amount will be directly credited to your endowment. Afterwards you will still own one share of C.

If you want to sell on share, you and your matched participant have to click the button labeled "Sell one share". If you want to sell more than one share, e.g., three shares you or your matched participants have to click these button for three times. The experiment ends after 14 periods. Then you and your matched participant do not have the possibility to buy or sell shares.

Afterwards all shares that you own at this point in time are automatically liquidated. The resulting money amount will automatically credited to your endowment.

After the end of period 7 and 14, you and your matched participant have to commonly guess the types ("++", "+", "0", "–', "-") of each stock A-F.

You will be credited 200 Talers to your endowment for every correct guess.

The total payoff you will earn in this experiment is given by:
Total payoff = the rest of your endowment (not invested) + value of the shares in your portfolio + earnings of your guesses

Note, that you and your matched participant both receive the total payoff earned in the experiment.

Questions

After you and your matched participant have correctly answered the control questions you will receive a questionnaire consisting of ten questions. Please answer the questionnaire. To answer each of the ten questions you will either have to chose "accept" or "reject". Note that you will not be paid for answering the questionnaire. Afterwards the experiment will start and you will have the possibility to buy (sell) shares in each of the 14 periods.

Now, please answer the control questions.

Questionnaire

What is your ID-number?

What is your gender?

Please answer the following ten questions

Assume that for each of the ten questions a coin is thrown. The coin can either land at “heads” or “tail”. To answer each of the ten questions you will either have to chose “accept” or “reject”. Note that all ten questions are hypothetical questions and thus will not influence your payoff.

1.) If the coin shows “heads” you will lose €2; if it shows “tail” you will win €10.
accept/ reject?

2.) If the coin shows “heads” you will lose €3; if it shows “tail” you will win €10.
accept/ reject?

3.) If the coin shows “heads” you will lose €4 if it shows “tail” you will win €10.
accept/ reject?

4.) If the coin shows “heads” you will lose €5; if it shows “tail” you will win €10.
accept/ reject?

5.) If the coin shows “heads” you will lose €6; if it shows “tail” you will win €10.
accept/ reject?

6.) If the coin shows “heads” you will lose €7; if it shows “tail” you will win €10.
accept/ reject?

7.) If the coin shows “heads” you will lose €8; if it shows “tail” you will win €10. accept/ reject?

8.) If the coin shows “heads” you will lose €9; if it shows “tail” you will win €10. accept/ reject?

9.) If the coin shows “heads” you will lose €10; if it shows “tail” you will win €10. accept/ reject?

10.) If the coin shows “heads” you will lose €11; if it shows “tail” you will win €10. accept/ reject?

Control questions

share X	period 1	period 2	period 3
price	80	83	82
bought (+) / sold (-)	5	-5	0

1.) You start with an endowment of 10,000 Talers. In periods 1-3 the transactions of the table (above) are processed.

(i) What is your endowment after period 1?

(ii) What is your endowment after period 2?

(iii) What is your endowment after period 3?

share X	period 1	period 2	period 3
price	100	95	90
bought (+) / sold (-)	10	0	-10

2.) You start with an endowment of 10,000 Talers. In periods 1-3 the transactions of the table above) are processed.

(i) What is your endowment after period 1?

(ii) What is your endowment after period 2?

(iii) What is your endowment after period 3?

Chapter 6

Conclusion

In the following chapter the main findings and further research directions will be discussed.

The first part of the thesis focuses on gender differences in reciprocal behavior and the impacts of social preferences and different move orders on coordination in step-level public good games. Chapter 2, entitled **"Do Women Behave More Reciprocally than Men - Gender Differences in Real Effort Dictator Games"** compares the impact of a real-effort task on reciprocal behavior between male and female decision makers. The results document that a real-effort task can increase dictators' reciprocity. Interestingly, this result does only hold for female dictators who significantly decrease their taking rates when recipients have solved a mathematical task. In contrast, male dictators always behave selfishly and take around 80 percent of the generated pot. The results emphasize the importance of the real-effort task because in the "Windfall" treatment there is no gender effect and dictators take on average the same amount. The chapter provides insights for Organizational Economics. The findings may be of importance in the presence of voluntary bonus payments. The data shows that workers' performance is only honored by female bosses. Hence, if workers should be motivated in the long run, it might be a good practice to install female bosses to decide about voluntary bonus payments.

Chapter 3, entitled **"Step-Level Public Goods: Experimental Evidence"**, discusses the efficient provision of step-level public goods. The experiment analyzes threshold public goods with one and two thresholds. The analysis focuses on different move orders (simultaneous vs. sequential). The findings illustrate that a sequential provision mechanism is appropriate to overcome the coordination problem. The sequential mechanism leads to a significant higher provision of the public good in contrast to the simultaneous environment. The data shows that applying a second threshold leads to ambiguous results: in the sequential environment it slightly increases public good provision, whereas in the simultaneous environment it harms welfare. The chapter also finds that subjects care about inequity aversion and second movers punish first movers for contributing too less. The results document that these social preferences can be explained by the Fehr and Schmidt (1999) model calibrated with Blanco et al.'s (2011) data. The findings of the chapter may be of interest when deciding about the provision to a public project like a research institute. Another application could be effort provision in companies. The results suggest that public good or effort provision should be done sequentially. Another finding is that setting higher "goals" only works in a sequential environment.

In the second part, the thesis focuses on the competitional effects in a gift-exchange game and on the disposition effect in individual and team investment decisions.

Chapter 4, entitled **"Competition in the Workplace: An Experimental Investigation"**, analyzes the role of competition between two workers employed to the same employer. The findings show that workers are affected by competition in a multi-employee environment. However, workers in a multi-employee environment need some time to monitor the co-worker's performance. That is, a significant learning effect between the workers can be observed. This learning process can be described by imitation learning. Interestingly, this behavior is initiated by employers who implicitly set tournament incentives. The data also shows that employers reduce their wage offers during the experiment.

Thus, the findings suggest for Organizational Economics that a multi-employee environment may lead to a higher performance of the employees due to pronounced learning effects.

Chapter 5, entitled **"The Disposition Effect in Individual and Team Investments: Experimental Evidence"** studies investment decisions for single and team investors. The chapter finds that individual and team decision makers both are prone to the disposition effect. Although teams show increased rational behavior and achieve better guess scores, they are prone to a higher degree of the disposition effect than individual investors. The results can be explained by regret aversion which significantly matters in teams, whereas it does not impact the selling decisions of individual investors. That is, in teams where both team members have to agree to reach a decision, regret aversion significantly hinders teams to sell their losses. The findings may be of importance for real life where financial advice became a big topic. It is often discussed whether "two heads do better than one head" and whether this might lead to better results. The finding in this chapter suggests the opposite: team membership leads to more pronounced disposition effects.

References

[1] Berg, J., Dickhaut, J., McCabe, K. (1995). "Trust, Reciprocity and Social History." Games and Economic Behavior 10, 122-142.

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

[3] Croson R., Gneezy, U., (2009). "Gender Differences in Preferences." Journal of Economic Literature 47, 448-474.

[4] Fehr, E., and Gächter, S. (2000). "Cooperation and punishment in public goods experiments." American Economic Review 90, 980-99.

[5] Fehr, E., Kirchsteiger, G. and Riedl, A. (1998). "Gift exchange and reciprocity in competitive experimental markets", European Economic Review 42, 1-34.

[6] Fehr, E., and Schmidt, K. (1999). "A Theory of Fairness, Competition and Cooperation", Quarterly Journal of Economics 114, 817-868.

[7] Isaac, R.M., and Walker, J.M. (1988). "Group Size Hypotheses of Public Goods Provision: An Experimental Examination", Quarterly Journal of Economics 103, 817-868.

[8] Kahneman, D., Knetsch, J., L., Thaler, R., (1986). "Fairness and the Assumptions of Economics." The Journal of Business 59, 285-300.

[9] Kahneman, D. and Tversky, A. (1979). "Prospect theory: An analysis of decision under risk", Econometrica 47, 263-291.

[10] Tversky, A. and Kahneman, D. (1974). “Judgment under uncertainty: Heuristics and biases”, Science 185, 1124-1131.