ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Englmaier, Florian; Gebhardt, Georg

Working Paper

Free Riding in the Lab and in the Field

SFB/TR 15 Discussion Paper, No. 344

Provided in Cooperation with:
Free University of Berlin, Humboldt University of Berlin, University of Bonn, University of Mannheim, University of Munich, Collaborative Research Center Transregio 15: Governance and the Efficiency of Economic Systems

Suggested Citation: Englmaier, Florian; Gebhardt, Georg (2010) : Free Riding in the Lab and in the Field, SFB/TR 15 Discussion Paper, No. 344, Sonderforschungsbereich/Transregio 15 - Governance and the Efficiency of Economic Systems (GESY), München, https://doi.org/10.5282/ubm/epub.13210

This Version is available at:
https://hdl.handle.net/10419/93907

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

GOVERNANCE AND THE EFFICIENCY OF ECONOMIC SYSTEMS
GESY

Discussion Paper No. 344

Free Riding in the Lab and in the Field

Florian Englmaier*
Georg Gebhardt**

*University of Munich
**Ulm University

September 2010

Financial support from the Deutsche Forschungsgemeinschaft through SFB/TR 15 is gratefully acknowledged.

Free Riding in the Lab and in the Field*

Florian Englmaier† Georg Gebhardt‡

September 24, 2010

Abstract

We run a public good experiment in the field and in the lab with (partly) the same subjects. The field experiment is a true natural field experiment as subjects do not know that they are exposed to an experimental variation. We can show that subjects' behavior in the classic lab public good experiment correlates with their behavior in the structurally comparable public good treatment in the field but not with behavior in any of two control treatments we ran in the field. This effect is also economically significant. We conclude that a) the classic lab public good experiment captures important aspects of structurally equivalent real life situations and b) that behavior in lab and field at least in our setting is driven by the same underlying forces.

JEL Codes: C91, C93, D01, D64

Keywords: Field and Lab Experiments, External Validity, Public Goods, Team Production

*We thank seminar participants in the MELESSA Brownbag Lunch at Munich, the University of Cologne, the Toulouse School of Economics, the Econometric Society World Congress 2010 in Shanghai, and the 2009 Workshop on "Natural Experiments and Controlled Field Studies" and in particular Robert Ulbricht for helpful comments and suggestions. The project was funded through SFB/TR-15.

†University of Munich, Department of Economics, Ludwigstr. 28 III VG, 80539 Munich, Germany, Phone: +49 89 2180 3916, englmaier@lmu.de

‡Ulm University

1 Introduction

Over the last decades a huge number of laboratory experiments have documented a large variety of behavioral patterns. Recently the discussion whether results obtained in laboratory settings can be readily extrapolated to field settings has intensified.[1] One particularly prominent strand of the laboratory literature has dealt with experimentally analyzing a large variety of public good situations where contributions are socially beneficial but unenforceable. We run a public good experiment in the field and in the lab with (partly) the same subjects to investigate the external validity of laboratory results and whether behavior in lab and field is driven by similar forces.

As the first part of our study we run a field experiment and hire students to register books in a university library. It is a natural field experiment according to John List's classification (see $http://www.fieldexperiments.com/$), as participants do not know that they participate in an experiment and work in a natural environment. In the end we invite them back for a lab experiment in which they play a public good game. We then compare behavior in the lab and in the field ("within subject"). The students register books in three institutional environments. In our main treatment they have the the opportunity to free-ride. Though the framing between lab and field is different (effort in, time out vs. tokens in, tokens out, labor market vs. neutral framing, ...), the underlying game theoretic structure is similar insofar as there are free riding incentives present. We document that in this treatment, productivity is highly correlated with contributions in the laboratory public good game. In two control treatments, without scope for free riding, it is not. This is consistent with a game theoretic model based on heterogeneous but stable (fairness) preferences.[2]

Furthermore, if we are willing to impose additional structure, we can quantify effects. In the field pubic good treatment, choosing a worker who contributes one additional standard deviation in the lab public good game, results in 14% more performance (measured as number of books registered per minute). To put this into perspective: For our task, registering books, ability (being used to use a computer, typing, etc.) is important and varies widely across subjects. Choosing a worker with a one standard deviation better test score results in 13% more performance (books per minute). I.e., selecting workers on *ability* and

[1] See e.g., the papers by Levitt (2007) or Levitt and List (2007).

[2] We will use the terminology "fairness preferences" rather loosely in most of the paper as we are not in the position to discern between different fairness theories like reciprocity (Rabin, 1993), inequity aversion(Fehr and Schmidt, 1999 and Bolton and Ockenfels, 2000), or (Charness and Rabin, 2002). We have in mind preferences that result in above selfish equilibrium contributions in public good settings. In Section 3.2 we will discuss the issue in some more detail.

fairness, as measured by lab contributions, have an impact in a similar order of magnitude. This seems to be broadly in line with the priorities of real world employers.[3]

Finally, we show that incentivizing workers has ambiguous effects depending on group heterogeneity with regards to ability: Incentives may even backfire in heterogeneous groups. We conclude that, at least in the case we study, selecting the right workers increases performance more reliably than incentivizing them.

Related Literature: Many variations of the classic public good experiment, Kim and Walker (1984) and Isaac et al. (1985), have been conducted to check the stability of the outcomes. For a survey of early variations, see the survey in the *Handbook of Experimental Economics* by Ledyard (1995) or the survey by Fehr and Gaechter (2000). One of the reasons why public good experiments have received a lot attention in the literature is that the simple public good game serves as a powerful metaphor for many economic situations which involve free riding, as classic public goods, intellectual property rights for informational goods, or team production in organizations.

Very recently, a few papers relate behavior in the lab and in the field. Fehr and Leibbrandt (2008) combine laboratory and field observations of the same individuals, for example, of fishermen who face a common pool resource dilemma and Benz and Meier (2008) match laboratory observations to donations data from the University of Zurich. These papers also find a positive correlation between field and lab behavior, but they do not have control and treatment groups in the field. Palacios-Huerta and Volij (2008) show that soccer players play mixing strategies both in penalty shootouts and experimental matching pennies games[4], and Palacios-Huerta and Volij (2009) show that chess grand masters play the Nash equilibrium in centipede games. These papers find similar strategic behavior in lab and field tasks, but there is selection into the field task (professional athletes and chess players).

The remainder of the paper is structured as follows. The next section presents the experimental set-up for the field and the lab study. Section 3 presents the results and Section 4 discusses the results and concludes.

[3]See for example Rosse et al. (1991) or Dunn (1995) for studies on the relative importance of different characteristics in hiring decisions.

[4]Though Levitt et al. (2010) are not able to replicate this finding in their study.

2 Experimental Set Up

2.1 General

As the first part of our study we run a field experiment and hire students to register books in a university library. It is a natural field experiment as participants do not know that they participate in an experiment and work in a natural work environment. At the end of the field experiment we pay the subjects their wage and invite them to come to our lab and participate in a lab experiment roughly two weeks later. We still do not reveal to the subjects that they participated in a field experiment. In the lab subjects play a public good game and answer a questionnaire. In the following we describe the two parts of our study separately in detail.

2.2 Field Experiment

The field experiment took place between October 1 and October 21 of 2008 at the Economics Department of the University of Munich. We recruited 103 subjects to register books in an institute library. The positions were advertised as one-time half-day jobs via posters all over the university campus, in student dorms, and on internet job portals. The job promised to pay a salary of €55 and the job duration was given with "up to 5 hours". Applicants had to apply via the recruiting website where they entered demographic information, contact details and, importantly, where they had to complete a 2 minute typing test. The requirement in the test mirrored their actual task as they were displayed pictures with book information which they had to type into an online form. The information we obtain from this test will be important, as we use it to control for heterogeneous ability across agents.

A research assistant (RA) invited the applicants to come to our library for a specific date and time. Every day we had subjects come in in two shifts, the morning shift starting at 8:30am, the afternoon shift starting at 1:30pm. The actual average working time was 3.5hrs. Each shift consisted of a group of four individuals. When inviting individuals to a certain shift, we used a first-come-first-serve rule and were careful not to sort along any observable dimension that might interact with our treatment variation. So allocation of subjects to shifts is mainly driven by subjects availability for a certain time subject to the constraint that we avoided putting subjects with the same last name, the same address or the same subject of study into the same shift to avoid any effects from prior relationships. We instructed our RAs to look for signs of prior relationship (eg. joint arrival, etc.), but did not find any evidence of that.

The subjects were welcomed and instructed together by an RA. After their task was carefully explained to them, the RA showed them the books they had to enter and they were led by the RA to four adjacent but separate offices that were equipped with laptop computers with an opened Excel-based entry mask. All entries subjects made in the Excel mask were saved on a central log file so we can track their performance over time. In principle, it was possible for the subjects to talk and interact, but we observed no signs of more than casual interaction. Also, to minimize interaction, our RA brought the books to the offices of the four subjects and gives them feedback on their performance every 30 minutes. See Figures 4, 5, and 6 in Appendix A for pictures of the physical set up.

Subjects (grouped per shift) were allocated into one of three treatments.

> Group: All subjects can leave once 680 books have been registered.[5] The feedback every 30 minutes is given with respect to group progress ("In the last 30 minutes the group entered x books. So in total the group has entered y out of a total of 680 books."). (private return from effort < social return)
>
> Single: Each subject can leave individually once 170 books have been registered. The feedback every 30 minutes is given with respect to individual progress ("In the last 30 minutes you entered x books. So in total you have entered y out of a total of 170 books."). (private return from effort = social return)
>
> No Incentives: All workers must work 3.5 hours, irrespective of the number of books entered. The feedback every 30 minutes is given with respect to individual progress ("In the last 30 minutes you entered x books. So in total you have entered y books."). (private return from effort = 0)

In Group, free riding of group members is an issue as the social benefit from effort is larger than the private benefit. In contrast, in Single and No Incentives free riding is no problem as there is no difference between private and social benefits from effort. We discuss this in more detail in Subsection 2.4. Note that we kept the 30 minute feedback structure constant across all treatments to get at the true treatment variation and not confound it with perceived variations in monitoring.

Only at the end of the experiment one of us entered the scene, introducing us (truthfully) as academic staff responsible for the library organization, to administer the final wage payments to the participants. After the payments were made we told them about the

[5] In the first sessions we had one session with 120 books and 2 sessions with 150 books per person. People were substantially faster than we expected so we increased the number of books to 170 per person. We control for this in the analysis.

opportunity to participate in paid laboratory experiments at *MELESSA (Munich Experimental Laboratory for Economic and Social Sciences)* and invited them to an experimental session roughly two weeks after the field experiment.[6] Subjects could either directly sign up for a session or we gave them information how to sign up for those sessions at a later point.

2.3 Laboratory Study

About 2 weeks after the field experiment, we run 5 sessions at MELESSA in Munich between October 28 and October 30, 2008. Roughly half of our subjects from the field (49 subjects) come to the lab and we fill up the sessions with people from the MELESSA subject pool. In total we have 84 subjects (4*16 and 1*20) in the lab where each session lasts roughly 90 minutes and the average earnings are 23.85€.

We run six rounds of experiments. Rounds 1-3 are non fully anonymous, and rounds 4-6 are fully anonymous. In the non fully anonymous rounds, the groups were randomized for each of the rounds. Before the round started, subjects were led out of the room and then each group was led separately to their respective computers. So subjects knew who was in their group, but did not know who does what within the group in the experiment. This treatment was meant to resemble the field situation where workers were welcomed and introduced together. In the fully anonymous treatments, groups are randomly reshuffled in each round an subjects can neither see the members of their group nor who does what within the group.

In round 1 and 4 we ran two standard public good games.[7] In this typical public good experiment subjects play in groups of four, get an initial endowment of 20 tokens and must decide how many tokens they put into a common pool. Tokens in the pool are multiplied by 1.6 and distributed among the group members. Tokens outside the pool belong to the subject. This game is repeated ten times with the same groups (partner treatment).

At the end of the experiment, people answered a questionnaire on demographic information with standard question from the world value survey (WVS) and the German Socio Economic Panel (G-SOEP).

[6] We deliberately chose to become visible only at that stage, to avoid that subjects made the link between the laboratory study and the prior employment.

[7] In rounds 2, 3, 5, and 6 we ran an inverse Public Good Game. This new experimental design is described in Appendix B. We will not discuss these experiments in detail, but suitably adjusted, the results using those experiments remain unchanged.

2.4 Comparison between Lab and Field

Lab: In each period each of the 4 subjects in a group receives an endowment of 20 tokens. A subject i can either keep these tokens for him- or herself or invest g_i tokens ($0 \leq g_i \leq 20$) into a project. These decisions are made simultaneously. The monetary payoff for each subject i in the group is given by $\pi_i = 20 - g_i + 0.4 \sum_{j=1}^{n} g_j$ in each period, where 0.4 is the marginal per capita return from a contribution to the public good. The total payoff is the sum of the period-payoffs over all ten periods.

Note that the definition of π_i implies that full free-riding ($g_i = 0$) is an individually rational dominant strategy in the stage game as $\frac{\partial \pi_i}{\partial g_i} = -1 + 0.4 = -0.6 < 0$. However, the socially efficient choice, maximizing aggregate payoff $\sum_{i=1}^{n} g_i$, is given by $g_i = 20$, i.e. fully cooperating. Hence a social dilemma situation arises.

Field: Though the game theoretic structure is similar, we have no clearcut theoretical prediction of behavior in the field. It is clear that in the No Incentives treatment effort should be lowest as there is no link between performance and any payoff relevant outcome variable (remuneration, leaving time). In the Single treatment subjects are residual claimants to their effort whereas in the Group treatment only a share of the marginal return to effort accrues to the subject. As compared to the social optimum we should expect inefficiently low contributions in Group.

3 Results

3.1 Overview of Lab and Field Results

Before we go into the joint analysis of lab and field results, we present a short summary of the basic results of the lab and field experiments. In Figure 1 and Figure 2, we present the results of the public good game in the lab.

Results from our **lab public good games** are representative of general public good results.[8] Even initially the contribution is not at 20 and the contribution level goes down over periods. In addition, there is substantial heterogeneity in contribution across subjects. This is true for the non-anonymous (PG1) and the fully anonymous (PG2) public good game. Figure 1 shows the average contribution levels over the ten rounds for PG1 on the left and PG2 on the right. After the experiment started again in PG2, subjects initially contribute more (restart effect), but then they reduce their contributions again. This restart

[8]Treatment assignment in the field does neither affect the probability to participate in the lab study nor the average public good contributions in the lab. Neither does participation in the field experiment affect lab contributions.

effect looks relatively small, but it is confounded by a treatment variation between the first and the second round: In the first round subjects knew with which other subjects they were in one group; in the second round, they did not. As we know from Fehr and Gaechter (1999) subjects contribute more if they meet the other people in their group before the experiment. This may explain why subjects contributed less in the second round, a behavior that looks like a smaller restart effect. In Figure 2 we present the average contributions combining PG1 and PG2. We will use this as our main measure of lab contributions. As discussed below, our results are robust to using varying measures.

Table 1 shows the descriptive statistics for the **field experiment**. We see that there are more subjects in Group as compared to the other treatments. The odd number of subjects in No Incentives is driven by a session comprising only 3 subjects due to one worker not showing up. The slight deviation from 170 in the average number of books entered in Group comes from some the three first sessions with 120 and 150 books respectively. The slight deviations from 170 in Single come from one subject who was so slow that he had not finished his share of books after 5hrs (the advertised maximum working time) and we allowed him to leave. It is noteworthy that the highest number of books were actually entered in No Incentives where also the variance in performance was highest. Women are over-represented in our sample in all treatments. But their share corresponds to the regular response rate for such jobs at the University of Munich.

Table 1: Descriptive Statistics Field

	Total	Group	Single	No Incentives
Number of Subjects	103	44	32	27
Number of Females	73	32	20	21
Avg. Speed (books/minute)	0.901	0.904	0.913	0.882
Std.Dev. Speed	0.199	0.182	0.217	0.211
Avg. # of Books entered	171.05	165.18	169,00	183,04
Std.Dev. Books entered	32.89	34,85	7,29	44,28

In Figure 3, we plot our measure of adjusted speed over time. We calculate adjusted speed as a measure of productivity by dividing speed (books per minute) by the testscore (characters per two minutes).[9] We plot the typing speed for the first two hours because

[9]We can also employ a more sophisticated approach to correct our speed measures for the varying ability levels of agents by regressing performance on testscore and using the residuals for our analysis. All our

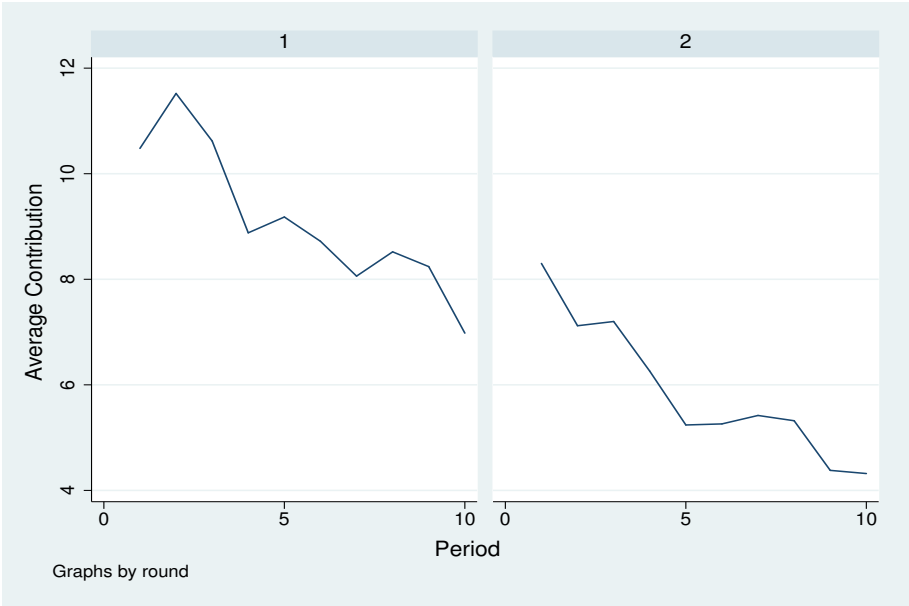


Figure 1: Avg. PG Contribution per Period; PG1 left panel, PG2 right panel

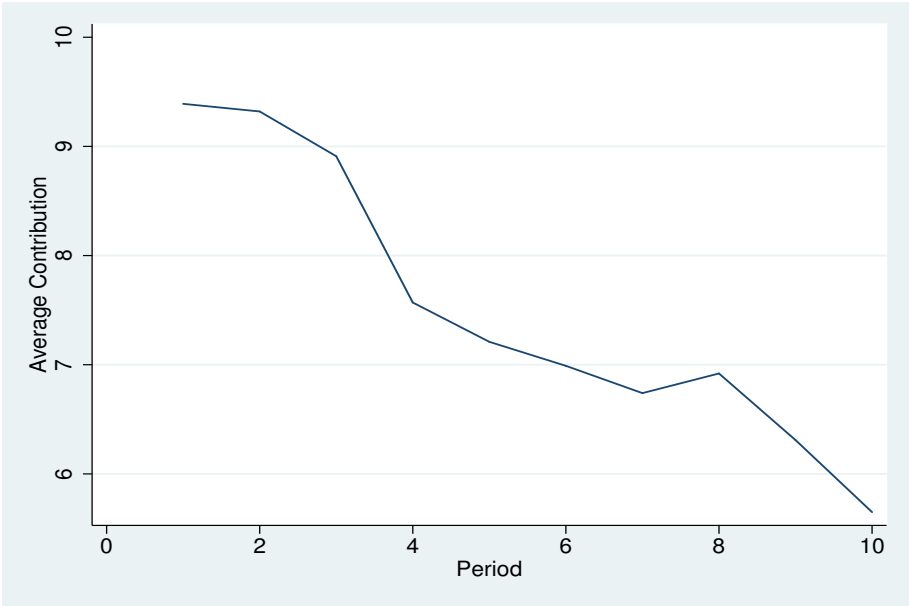


Figure 2: Avg. PG Contribution per Period; PG1 and PG2 merged

even the fastest participants worked for at least two hours. In later hours, the fastest participants in the SINGLE treatment were already finished so that we cannot compare average speeds. We calculate average speeds for four intervals of half an hour each because every half an hour the participants received feedback on their progress and the group's progress.

We find that adjusted speed changes little over time. In all three treatments subjects seem to learn a little initially so that they work faster in the second and third half an hour period. In the SINGLE and NO INCENTIVES treatment, the subjects continue to work as fast or even faster in the last period, while we find a slight drop off in the last period in the GROUP treatment. Overall, however, we find little dynamic over time apart from learning. Therefore, for the rest of this paper, we abstract from dynamic effects and present all results in terms of average (adjusted) speed over the whole working time of a subject.

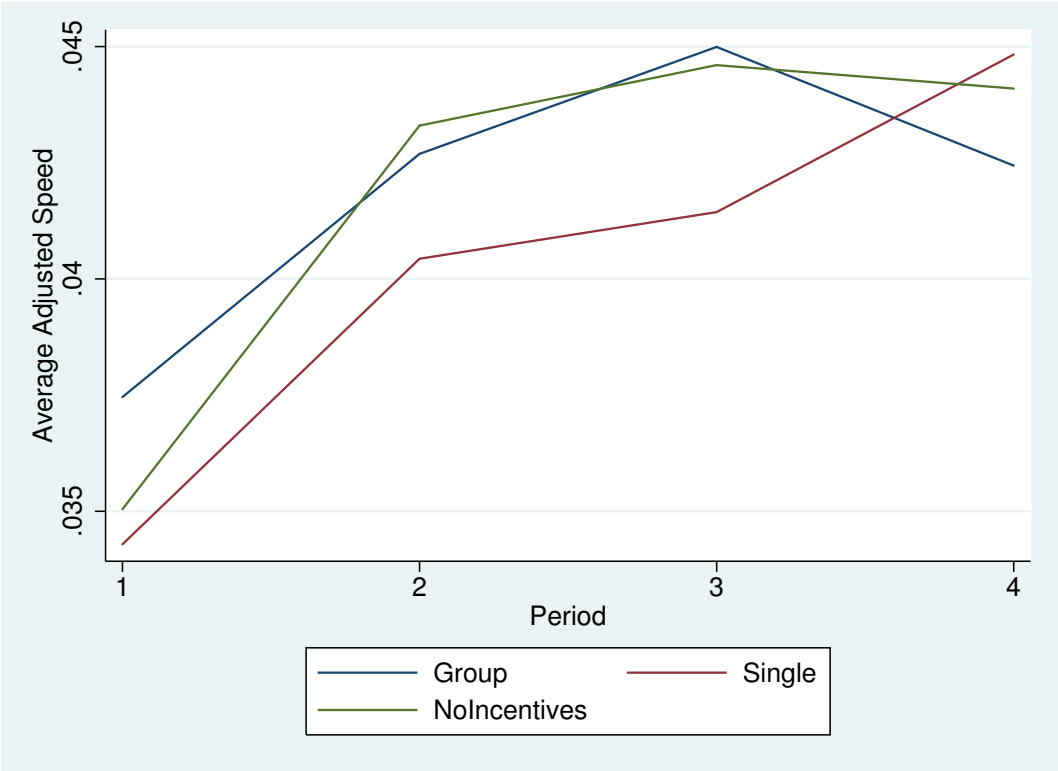


Figure 3: Adjusted Speed First Two Hours in the Field (books per minute/Testscore)

From a casual inspection of Figure 3 seems to suggest that there is little difference between the treatments in the field experiment. While this is true for the averages, we will demonstrate in the remainder of the paper that different countervailing effects lie behind the averages.

results remain robust when we use this alternative measure.

Table 2: Average Lab Contribution - Skill adjusted Speed $\left(\frac{Books/Minute}{Testscore}\right)$

	Full Sample	Group	No Group
Kendall	0.1559	0.3789	0.0230
	(0.1119)	(0.0212)*	(0.8724)
Spearman	0.2254	0.5654	0.0229
	(0.1156)	(0.0094)**	(0.9043)
N. obs.	50	20	30

In parenthesis, $Prob > |t|$ and $Prob > |z|$ respectively

* significant at 5%; ** significant at 1%; *** significant at 0.1%;

3.2 Correlation Lab and Field

In this section, we demonstrate qualitatively that subjects' behavior in the Group treatment, which allows free riding, in the field is correlated with their lab behavior. While the subject's behavior in the other two treatments, which do not allow free riding, is not correlated with their lab behavior.

A basic prediction of game theoretic models with stable (fairness) preferences is that subjects with a preference for fairness contribute more in a public good game in the lab than subjects who have purely egoistic preferences. The same prediction holds for the Group treatment, in which subjects can free ride, i.e. work slowly, so that the other group members have to enter into the computer system more of the books assigned to the whole group. No such prediction holds for the other two treatments (Single and No Incentives), as in both cases the work load of the other group members is independent of how fast another member works.

As the theoretic prediction for the lab field correlation is the same for the Single and the No Incentive treatment, we initially pool the results from the two treatments into the category No Group. Because we obtain a larger sample size for the No Group treatment compared to the two single treatments, this makes it more difficult for us to reject the null hypothesis that there is no correlation in the No Group treatments.[10]

In Table 2, we present the Kendall and Spearman correlation coefficients between the average contribution over all twenty periods in the two rounds in the lab experiment and

[10]In Table 3, we present the results for a three way split, i.e., we separately list the Single and the No Incentives treatment and we confirm that for neither of the two treatments the correlation is even close to being significant.

Table 3: Average Lab Contribution – Skill adjusted Speed $\left(\frac{Books/Minute}{Testscore}\right)$

	Group	Single	No Incentives
Kendall	0.3789	0.3250	−0.0769
	(0.0212)*	(0.0868)	(0.7418)
Spearman	0.5654	0.4356	−0.1189
	(0.0094)**	(0.0917)	(0.6855)
N. obs.	20	16	14

In parenthesis, $Prob > |t|$ and $Prob > |z|$ respectively
* significant at 5%; ** significant at 1%; *** significant at 0.1%;

the average adjusted speed in the field. For the full sample (column (1)) we find a slight correlation that is, however, not significant. When we split the treatments in Group and No Group, we find that there is almost no correlation in the No Group treatment. In the Group treatment however, we find a very strong correlation at least at the 5% level for both tests.

3.3 Regression Analysis of Lab and Field Relationship

In this section, we impose more structure and run a regression analysis to quantify the impact of the differences in fairness preferences as measured by the lab contributions; we find the impact is large. We estimate the two specifications with OLS, where each observation, indexed by i, is one subject that we have observed in the field and in the lab. In the main specification, we distinguish Group and No Group treatments and estimate

$$speed_i = \beta_0 + \beta_1 I_{Group} + \beta_2 score_i + \beta_3 contr_i + \beta_4 contr_i I_{Group} + \varepsilon_i,$$

where $speed_i$ is the log of the average speed (in books per minute), I_{Group} is an indicator variable that takes on the value 1 if the subject participated in the Group treatment in the field and 0 otherwise, $score_i$ is the log of the testscore (in characters in two minutes), and $contr_i$ is the log of the average contribution to the public good over the two rounds. Note, that we no longer use our skill adjusted speed measure but control for ability directly in the regression.

In addition, we use a specification with a three way split:

$$speed_i = \beta_0 + \beta_1 I_{Group} + \beta_2 I_{Single} + \beta_3 score_i + \beta_4 contr_i + \beta_5 contr_i I_{Group} + \beta_6 contr_i I_{Single} + \varepsilon_i,$$

Table 4: OLS Estimates of the Average Speed in the Field Experiment

	Books per Minute (Log)	
	Two Way Split	Three Way Split
Group (Dummy)	−0.245	−0.313
	(0.152)	(0.159)
Single (Dummy)		−0.282
		(0.133)*
Testscore (Log)	0.469	0.474
	(0.107)***	(0.113)***
Avg. Contribution (Log)	−0.004	−0.031
	(0.026)	(0.023)
Avg. Contribution (Log) * Group	0.166	0.195
	(0.074)*	(0.074)*
Avg. Contribution (Log) * Single		0.131
		(0.072)
N. obs.	50	50
R^2	0.452	0.489

In parenthesis robust standard errors clustered at session level
* significant at 5%; ** significant at 1%; *** significant at 0.1%;

where I_{Single} is an indicator variable that is 1 if the subject participated in the Single treatment in the field and 0 otherwise.

Qualitatively, we get the same results from the regression (Table 4) as from the correlation coefficients: Average speed is significantly correlated with lab contributions in the Group treatment, but neither in the No Group treatments taken together, nor in the Single or No Incentive treatments separately. In addition, we can demonstrate that our typing speed test captures a meaningful part of ability as the testscore is highly correlated with average speed.

We use the coefficient estimates of the main specification to obtain a measure of the economic significance of the effects: A one standard deviation increase in the log average contribution (71.7%) increases average speed by 13.9% in the Group treatment. We can compare this with the effect of the testscore: A one standard deviation increase in the

Table 5: Lab Contribution – Skill adjusted Speed $\left(\frac{Books/Minute}{Testscore}\right)$

	First Periods		Last Periods	
	Group	No Group	Group	No Group
Kendall	0.3579	0.0230	0.2316	0.0529
	(0.0292)*	(0.8722)	(0.1600)	(0.6930)
Spearman	0.5435	0.0147	0.3335	0.0536
	(0.0132)*	(0.9385)	(0.1508)	(0.7783)
N. obs.	20	30	20	30

In parenthesis, $Prob > |t|$ and $Prob > |z|$ respectively
* significant at 5%; ** significant at 1%; *** significant at 0.1%;

log testscore (28.0%) increases speed by 13.3%, i.e., if subjects have the possibility to free ride in the field, selecting them on the basis of their lab contributions yields the same productivity advantage as selecting them on the basis of ability, i.e., the testscore.

4 Discussion and Conclusion

4.1 Who drives our Effects

We can shed some more light on the motivations of the people who contribute in the lab and in the field. Results from various lab studies suggest that there is a (small) fraction of subjects who are genuinely willing to contribute, often called *altruists*. These subjects will typically contribute even in the last round. There are also other subjects who contribute as long as they believe that others contribute. These may be either purely egoistic subjects, *strategic egoists*, who contribute in the early rounds for strategic reasons to increase the contributions of others or genuine *conditional cooperators* who are willing to cooperate as long as others cooperate. In general, both of these latter types are observationally very similar and in equilibrium contribute initially but reduce their contributions towards the end, leading to the downward-sloping pattern of average contributions.

To analyze which of the two groups drives the correlation between the lab and the field, we correlate productivity in the field with the contribution levels in the first periods of the two rounds of the lab game and the last periods of the two rounds. If conditional cooperators and strategic egoists drive the correlation, we should find the correlation for the

first round but not for the last round. But if genuine altruists are behind the correlation, we should observe a significant correlation for both periods. In Table 5 we present the correlation coefficients for the the first and last periods. In neither specification, we find a significant correlation in the No Group treatments. In the Group treatment, however, we find a correlation between first period contributions and adjusted speed in the field but no significant correlation with the last period contributions. This is consistent with the hypothesis that strategic egoists and conditional cooperators drive the lab field correlations.

4.2 Regression Analysis Field: Group Heterogeneity

In this section, we return to the question why incentives seem to have so little impact on the speed in the field experiment. We show that incentivizing workers by allowing them to leave early if they work faster, has ambiguous effects depending on group heterogeneity with regards to ability: Incentives may even backfire in heterogeneous groups. We focus on the interaction of incentives and the heterogeneity of the groups with regard to ability, measured by our testscore. Because we did not select or group the participants in the field experiment on the basis of their testscores, some of the groups were very heterogeneous with regards to ability. This heterogeneity did not translate into differences in payment or working time in the Group and No Incentive treatments; everybody was paid the same and everybody came and left together. In the Single treatment, however, considerable differences in working time occurred: In some groups the fastest worker left after two hours, while his or her slower coworkers had to work for up to five hours. Even if the subjects could not directly see each other working, they would notice if one of their coworkers left. We conjecture that the presence of high performers lead to a collapse in motivation of the low performers in Single treatment. As this effect can only happen in the Single treatment, we suspect that this effect could drive the seeming ineffectiveness of incentives in our field experiment.

To investigate this hypothesis in a regression analysis, we need a measure of group heterogeneity. We cannot use the standard deviation of actual performances for obvious endogeneity problems. Instead, we use the standard deviation of testscores. We include for each subject the standard deviation of testscores in his or her group together with the (log)level of the testscore of the individual subject. We interact the standard deviation of the testscore with the treatment dummy, as we suspect that heterogeneity plays an especially important role in the Single treatment.

In Table 6 we present the results of this regression. We find that the point estimates of the coefficient on the standard deviation of testscores is negative in all three treatments:

Table 6: OLS Estimates of the Average Speed in the Field Experiment

	Books per Minute (Log)
Group (Dummy)	0.381
	(0.380)
Single (Dummy)	0.918
	(0.400)*
Testscore (Log)	0.455
	(0.084)***
Std. Testscore (Log)	−0.012
	(0.061)
Std. Testscore (Log) * Group	−0.075
	(0.083)
Std. Testscore (Log) * Single	−0.189
	(0.089)*
N. obs.	103
R^2	0.360

In parenthesis robust standard errors clustered at session level

* significant at 5%; ** significant at 1%; *** significant at 0.1%;

heterogeneous groups perform worse. But the size of the effect varies over treatments. It is very small for the No Incentive treatment, somewhat larger for the Group treatment, and by far largest for the Single treatment. It is statistically significant (at the 5% level) only in the Single treatment. If we control for group heterogeneity, we get different estimates of the treatment dummies. These dummies now capture the treatment effect for a perfectly homogeneous group. We find that the point estimates are now positive for the Group treatment (which provides some incentives) and even more positive for the Single treatment (which provides the strongest incentives). Only the coefficient for the Single treatment is significant (at the 5% level). The point estimates are large, in particular for the effect of Single treatment relative to the No Incentives treatment. The coefficient estimates implies a 250% increase in speed if we go from No Incentives to the Single treatment in a perfectly homogeneous group.

Our experimental set-up was not designed to explicitly answer the reasons behind the strong effect of heterogeneity. It could be that the slow members of a group feel unfairly treated because they must work much longer for the same amount of money than the fast members. Therefore, they may experience a collapse of motivation. Maybe, the Single treatment introduces a notion of competition that motivates group members that work at a similar speed, but discourages laggards. Whatever the reason for the interaction effects of treatment and heterogeneous ability, the effects are large. Whenever we introduce incentives, we reward not only effort but also ability, and we know little how this is perceived by workers, as most lab experiments abstract from heterogeneous ability. Our results suggest that this interaction should be added to the growing list of potential pitfalls associated with the provision of explicit incentives.

4.3 Summary

We run a public good experiment in the field and in the lab with (partly) the same subjects. The field experiment is a true natural field experiment as subjects do not know that they are exposed to an experimental variation. Our study offers various contributions. We can establish that lab behavior is externally valid, as behavior in the lab and in the field correlates, even within subjects. We can go beyond the existing literature, as we have treatment and control groups in the field. Using the placebo treatments, we can show that the correlation is only present in the public good treatment (Group) but not in the other treatments (Single, No Incentives). I.e., subjects behave similarly under "similar" incentive structures. We take this as indication for the external validity of public good experiments. This might also be indicative for the existence of stable preference

types. Furthermore, the simple game theoretic structure of the public good game seems to capture important aspects of real life situations. Moreover, we can show that the effect of "fairness" preferences on performance is economically relevant, in our setting comparable in size to the effect of ability. Finally we document the (detrimental) effect of heterogeneity w.r.t. ability on performance under explicit incentive structures.

References

Benz, M. and S. Meier (2008) "Do people behave in experiments as in the field-evidence from donations," *Experimental Economics*, Vol. 11(3), pp. 268-281

Bolton, G.E. and A. Ockenfels (2000) "ERC û A theory of equity, reciprocity and competition," *American Economic Review*, Vol. 90(1), pp. 166û193

Charness, G. and M. Rabin (2002) "Understanding social preferences with simple tests," *Quarterly Journal of Economics*, Vol. 117(3), pp. 817û869

Dunn, W., M.K. Mount, M.R. Barrick, and D.S. Ones (1995) "The Big Five personality dimensions, general mental ability and perceptions of employment suitability," *Journal of Applied Psychology*, Vol. 80, pp. 500-509

Fehr, E. and S. Gaechter (1999) "Collective action as a social exchange," *Journal of Economic Behavior & Organization*, Vol. 39(4), pp. 341-369

Fehr, E. and S. Gaechter (2000) "Fairness and Retaliation: The Economics of Reciprocity," *Journal of Economic Perspectives*, Vol. 14, pp. 159-181

Fehr, E. and Leibbrandt, A. (2008) "Cooperativeness and Impatience in the Tragedy of the Commons," IZA Discussion Papers 3625

Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition and cooperation. Quart. J. Econ. 114 (3), 817û868.

Fischbacher, U. (2007) "z-Tree: Zurich Toolbox for Ready-made Economic Experiments," *Experimental Economics*, Vol. 10(2), pp. 171-178.

Greiner, B. (2004) "An Online Recruitment System for Economic Experiments," GWDG Bericht 63, Göttingen : Ges. für Wiss. Datenverarbeitung, pp. 79-93

Isaac, R.M., K. McCue, and C.R. Plott (1985) "Public Goods Provision in an Experimental Environment," *Journal of Public Economics*, Vol. 26, pp. 51-74

Kim, O. and M. Walker (1984) "The Free Rider Problem: Experimental Evidence," *Public Choice*, Vol. 43, pp. 3-24

Ledyard, J. (1995) "Public Goods: A Survey of Experimental Research," in: *The Handbook of Experimental Economics*, Eds. A. Roth and J. Kagel, Princeton University Press, Princeton/NJ

Levitt, S.D. (2007) "Viewpoint: On the Generalizability of Lab Behavior in the Field," *Canadian Journal of Economics*, Vol. 40(2), pp. 347-370

Levitt, S.D. and J.A. List (2007) "What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World?," *Journal of Economic Perspectives*, Vol. 21(2), pp. 153-174

Levitt, S.D., J.A. List, and D.H. Reiley (2010) "What Happens in the Field Stays in the Field: Exploring Whether Professionals Play Minimax in Laboratory Experiments," *Econometrica*, Vol. 78(4), pp. 1413-1434

Palacios-Huerta, I. and O. Volij (2009) "Field Centipedes," *American Economic Review*, Vol. 99(4), pp. 1619-35

Palacios-Huerta, I. and O. Volij (2008) "Experientia Docet: Professionals Play Minimax in Laboratory Experiments," *Econometrica*, Vol. 76(1), pp. 71-115

Rabin, M. (1993) "Incorporating fairness into game theory and economics,' *American Economic Review*, Vol. 83(5), pp. 1281û1302

Rosse, J.G., H.E. Miller and L.K. Barnes (1991) "Combining personality and cognitive ability predictors for hiring service-oriented employees," *Journal of Business and Psychology*, Vol. 5(4), pp. 431-445

A Set-Up: Field Experiment

Figure 4: Workspace

Figure 5: Office location

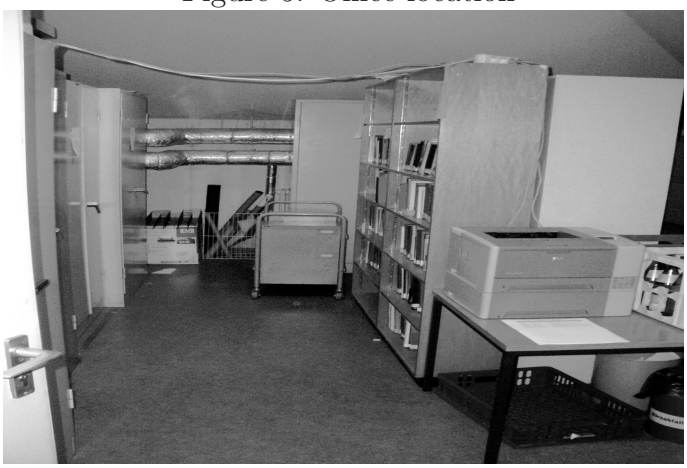


Figure 6: Books that had to be entered

B Instructions: Laboratory Experiment

Instructions (translated from German)

Welcome to the experiment and thanks a lot for your participation! From now on, please do not speak to other participants of this experiment any more!

General remarks on the procedure

This experiment serves to analyze economic decision making. You can earn money by taking part, which will be paid in cash after the experiment. During the experiment, you and other participants are asked to make decisions. Both your own decisions as well as those of other participants will determine your pay-off according to the following rules. The whole experiment lasts about 90 minutes. You will receive detailed instructions at the beginning. If you have questions regarding the instructions, please raise your hand. One of the conductors of the experiment will then come to you and answer your question personally. Each participant receives a fixed ID number by which he or she is identified during the experiment. During the experiment, we do not speak in terms of Euro, but Experiment-Points (EP). Your earnings will be calculated in EP in the course of the experiment. At the end of the experiment, all EPs you have earned will be converted into Euro using the following exchange rate:

1 Experiment-Point = 2 Euro-Cent

At the End: The experiment consists of 6 parts. It ends after the 6th part. The results from the individual rounds will be added up and converted into Euro. Your earnings will be paid out after the 6th part.

Anonymity: In each of the 6 parts, you will play in a new group. In each part your group will be re-matched randomly. Sometimes you know who is in your group, sometimes you do not. The other participants do not know during nor after the experiment which role you have taken or how much you have earned. We will evaluate the data from the experiment anonymously and delete all your personal and person-related information as soon as we have matched the data. At the end of the experiment you have to sign a receipt of your earnings which serves for accounting purposes only.

The Experiment

Part 1

Groups: At the beginning of part 1 you will be randomly divided into groups of four. In each later part your group will be re-matched randomly. You can see the other members of your group. But you cannot attribute decisions to other group members. Thus all group members stay anonymous. The 1st part of the experiment consists of 10 rounds in total. The composition of the group stays the same over the whole 1st part.

Budget and alternatives in each round: Each participant receives 20 Points in the beginning of each round. You can freely divide the points into two alternatives, X and Y:

1. You can put 0 to 20 points into pot X. The sum of the points in pot X will be multiplied by 1.6 and distributed equally among the group members. That is, for each point in pot X you will get 0.4 (=1.6/4) points. For example, if the sum of the points in pot X in your group is 60, each group member will get 60*0.4=24 points from pot X. If all group members together contribute 10 points to pot X, you and all other group members will get 10*0.4=4 EP from pot X.

2. You can put 0 to 20 points into pot Y. This amount goes then one-to-one into your earnings. So if you, for instance, contribute 6 points to pot Y, you get exactly 6 points. Your earnings per round is then the sum of your earnings from pot X and that from pot Y.

In mathematical terms, the result is:
Result (for member i) $= (20 - x_i) + (S * 1.6)/4$
x_i is the contribution of member i to pot X
S is the sum of the contributions of all group members to pot X
You will be asked on the screen how many points you want to contribute to pot X. The rest of the 20 points automatically goes to pot Y. Therefore, it is not possible to save points. You can enter only an integer number between 0 and 20 (i.e. 0; 1; 2; ...; 19; 20). After each round you will get to know the contributions of your group members to pot X and your total earning in this round in EPs. In the course of part 1 you will make 10 decisions according to the instructions above. Please remember: You will receive 20 points in each round, which you have to distribute to pot X and Y.
You will determine your deposit with the slider (See Figure 7). A click on the right-arrow increases your deposit by 1 point, whereas a click on the left-arrow reduces your deposit by 1 point.
You will get the instructions for the 2nd part of the experiment after end of the 1st part.

Part 2

Groups: At the beginning of part 2 you will be randomly assigned to a new group of four participants. In each later part your group will be re-matched randomly. You can see the other members of your group. But you cannot attribute decisions to other group members. Thus all group members stay anonymous.
Budget and alternatives in each round: In this 2nd part, each group has to deposit a sum of 240 points into a pot. The game ends as soon as the group has raised these 240 points. In the beginning, each member receives a budget of 220 points. From these points you must deposit at least 10 per round into the group pot. You can increase your contribution in steps of 2.5 points up to 60 points. If the necessary 240 points have not been deposited after a certain round, the experiment goes to the next round and you can again choose your contribution. The experiment continues to move to further rounds until the number of deposited points in the group pot reaches 240. For each round played, expenses of 20 points will be subtracted from every member.
Consider the following case:
In the first and second round the participants have deposited altogether 180 points into the group pot. This means that at the beginning of the 3rd round there are still 60 points to be raised. In the third round, two participants contribute 10 points each and the other two pay in 15 points each. Then, at the beginning of the 4th round, the number of points still to be raised decreases to 10 points (=60 points - 2*10 points - 2*15 points). In addition, the budget of the participants who have contributed 15 points decreases by 35 points (15 points contribution and 20 points expenses per round). The budget of the participants who have contributed 10 points decreases by 30 (10 points contribution and 20 points expenses per round). If in a round the participants raise more points than the total necessary amount, the points and the expenses will be subtracted proportionally from each participant.
Consider the following case:
At the beginning of the third round, the participants have deposited altogether 180 points into

the group pot. So they still need to raise 60 points. If now in the third round each participant deposits 30 points, then the total contribution would be 120 instead of the 60 points needed. In this case, the budget of each participant decreases on a pro-rata basis only by 15 points (=60/120 * 30 contributed points) and the 10 (=60/120 * 20) points of expenses for this round.

In the end you will get as payoff the points that remain from your budget after the last round. You will be asked on the screen how many points you want to contribute to the group pot. You can enter a number between 10 and 60 in steps of 2.5 (i.e. 10; 12.5; 15; . . . ; 57.5; 60). After each round you get to know the sum of the contributions of your group members to the group pot and your remaining budget in experiment points.

You determine your deposit with the slider (See Figure 8). A click on the right-arrow increases your deposit by 2.5 point, whereas a click on the left-arrow reduces your deposit by 2.5 point.

You will get the instructions for the 3rd part of the experiment after the end of the 2nd part.

Part 3

Groups: At the beginning of part 3 you will be randomly assigned to a new group of four participants. In each later part your group will be re-matched randomly. You can see the other members of your group. But you cannot attribute decisions to other group members. Thus, all group members stay anonymous.

Budget and alternatives in each round: Same as in part 2.

Part 4

Groups: At the beginning of part 4 you will be randomly assigned to a new group of four participants. In each later part your group will be re-matched randomly. You will not come to know the other members of your group, neither during nor after the experiment. So your decisions stay anonymous. The 4th part of the experiment consists of 10 rounds in total. The composition of the group stays the same for the whole 4th part.

Budget and alternatives in each round: Same as in part 1.

Part 5

Groups: At the beginning of part 5 you will be randomly assigned to a new group of four participants. In each later part your group will be re-matched randomly. You will not come to know the other members of your group, neither during nor after the experiment. So your decisions stay anonymous.

Budget and alternatives in each round: Same as in part 2.

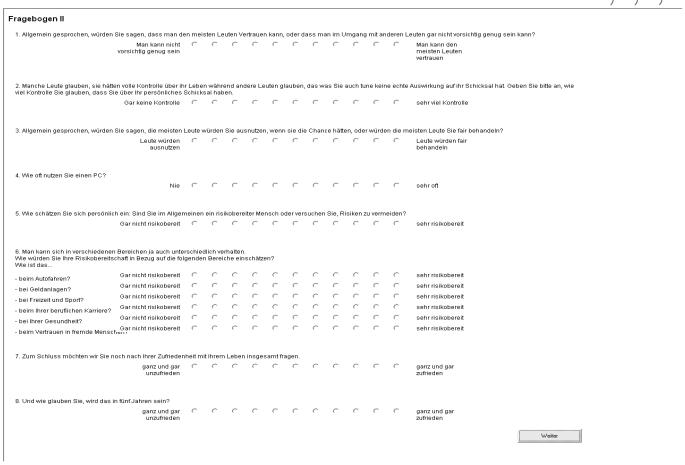
Part 6

Groups: At the beginning of part 6 you will be randomly assigned to a new group of four participants. You will not come to know the other members of your group, neither during nor after the experiment. So your decisions stay anonymous.

Budget and alternatives in each round: Same as in part 2.

Figure 7: Screenshot: Decision Screen for Part 1 and 4

Figure 8: Screenshot: Decision Screen for Part 2,3,5, and 6



Fragebogen II

1. Allgemein gesprochen, würden Sie sagen, dass man den meisten Leuten Vertrauen kann, oder dass man im Umgang mit anderen Leuten gar nicht vorsichtig genug sein kann?

Man kann nicht vorsichtig genug sein — Man kann den meisten Leuten vertrauen

2. Manche Leute glauben, sie hätten volle Kontrolle über ihr Leben während andere Leuten glauben, das was Sie auch tune keine echte Auswirkung auf ihr Schicksal hat. Geben Sie bitte an, wie viel Kontrolle Sie glauben, dass Sie über Ihr persönliches Schicksal haben.

Gar keine Kontrolle — sehr viel Kontrolle

3. Allgemein gesprochen, würden Sie sagen, die meisten Leute würden Sie ausnutzen, wenn sie die Chance hätten, oder würden die meisten Leute Sie fair behandeln?

Leute würden ausnutzen — Leute würden fair behandeln

4. Wie oft nutzen Sie einen PC?

Nie — sehr oft

5. Wie schätzen Sie sich persönlich ein: Sind Sie im Allgemeinen ein risikobereiter Mensch oder versuchen Sie, Risiken zu vermeiden?

Gar nicht risikobereit — sehr risikobereit

6. Man kann sich in verschiedenen Bereichen ja auch unterschiedlich verhalten.
Wie würden Sie Ihre Risikobereitschaft in Bezug auf die folgenden Bereiche einschätzen?
Wie ist das...

	Gar nicht risikobereit	sehr risikobereit
- beim Autofahren?		
- bei Geldanlagen?		
- bei Freizeit und Sport?		
- beim Ihrer beruflichen Karriere?		
- bei Ihrer Gesundheit?		
- beim Vertrauen in fremde Menschen?		

7. Zum Schluss möchten wir Sie noch nach Ihrer Zufriedenheit mit Ihrem Leben insgesamt fragen.

ganz und gar unzufrieden — ganz und gar zufrieden

8. Und wie glauben Sie, wird das in fünf Jahren sein?

ganz und gar unzufrieden — ganz und gar zufrieden

Weiter

Figure 9: Screenshot: Exit Questionnaire