Essays on Economic Laboratory and Field Experiments

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der

Wirtschafts- und Sozialwissenschaftlichen Fakultät

der

Universität zu Köln

2013

vorgelegt von

Diplom Volkswirt Felix Ebeling

aus Uelzen

Referent: Prof. Dr. Axel Ockenfels

Korreferent: Prof. Dr. Gerlinde Fellner

Tag der Promotion: 03.06.2013

Acknowledgements

The writing of this dissertation has been one of the greatest intellectual and personal challenges I have faced. It would not have been completed without the support, motivation and guidance of the following persons.

First, I thank my supervisor Axel Ockenfels for giving me the opportunity to work in an excellent academic environment and for guiding my research. His ideas, his knowledge and his commitment to the highest methodological standards inspired and challenged me and deepened my understanding of science.

I also thank Gerlinde Fellner for her valuable comments.

I thank my co-authors Gary Bolton, Gerlinde Fellner, Christoph Feldhaus, Johannes Fendrich, Axel Ockenfels and Johannes Wahlig for fruitful discussions and for close collaboration in our joint research projects.

I thank my present and former colleagues Julian Conrads, Christoph Feldhaus, Lyuba Ilieva, Jos Jansen, Sebastian Lotz, Johannes Mans, Julia Stauf and Peter Werner for the great working atmosphere, for countless discussions of academic, general and personal matters, and for making research fun.

I thank our student helpers, especially those supporting me conducting my field experiments. Furthermore, I would like to thank Krümel.

Finally, I am deeply grateful to my parents Lutke and Dorothee, my brother Florian, his wife Alexandra and my brother Philipp for their support, patience and confidence. This work is dedicated to you.

Table of contents

Part 1

Motivation and Research Questions

The core theory used in (neoclassical) economics builds on simple but powerful assumptions on individual behavior. Individuals maximize their utility (function) which only depends on their own payoffs in a time consistent, framing independent manner. Thereby, they accurately use all available information.[1] The core of behavioral economics is to refine these assumptions. It helps generate more realistic theories, make better predictions of field phenomena, and develop better policy suggestions.[2] This thesis is about behavioral economics. More precisely it is about experimental economics which represents a subarea of behavioral economics. In each part of my thesis, the main research method used to explore economic decision making is an experiment. Despite this methodological similarity, there is only a loose connection between different parts. Parts differ in characteristics as (a) systematic behavioral anomaly explored, (b) exact experimental method used to explore the anomaly, and (c) application. In the following, I will give a short overview about my thesis by illustrating the mentioned characteristics as well as the basic results for each part. Before starting, I will first elucidate characteristics in more detail.

Even if individuals do not always respond to incentives in a way predicted by standard (neoclassical) theory, their behavior is not foolish. Rather, their deviations from standard theory follow their own (bounded) rationality and have certain patterns – they are systematic (Bolton & Ockenfels 2012). Systematic deviations / anomalies can be explored scientifically and in turn be categorized. DellaVigna (2009) suggests three types of anomalies: (a) nonstandard preferences e.g. fairness, (b) nonstandard beliefs e.g. overconfidence, and (c) nonstandard decision making e.g. framing dependency. I will also use these types for anomaly classification.

---

[1] See DellaVigna (2009) p. 315.
[2] See Camerer & Loewnenstein (2004) p. 3.

Behavioral economists can analyze individual decision making theoretically, empirically, or even by simulations (for an introduction to simulation see e.g. Peichel 2009). The majority of my thesis is empirical research. As mentioned previously, experiments as a special form of data collection are at the core of each of my research projects. There are different forms of experiments. Following Harrison & List (2004) there are four types of experiments: (a) conventional laboratory experiment, (b) artifactual field experiments, (c) framed field experiments, and (d) natural field experiments. However, other researchers use a different taxonomy, e.g. Falk & ichino (2006).[3] I will not stick to any predefined taxonomy, but will simply explain our experimental design.

Experimental economics does not have to be applied. Experimental decision situations can be extremely stylized and far from realistic settings. Considering Roth's (1995) categorization of experiments by motivation, application strongly depends on the motivation. Roth differentiates between "Speaking to Theorists", "Searching for Facts", and "Whispering in the Ears of Princes". While the first is rather unapplied, the latter, which is a paraphrase for policy / firm advisory, has a stronger focus on the application.[4] My research is rather applied as practical applications (the "whispering") are not a byproduct, but the main motivation.[5] The starting point for each research project was an open question from an applied research field. As an example, consider my joint work on peer pressure with Gerlinde Fellner and Johannes Wahlig. On the one hand, previous economic literature on peer pressure revealed the opportunity to harness peer pressure as an incentive to increase productivity of workers. On the other hand, previous literature on monetary incentives revealed drawbacks on incentives at the workplace. In multi-tasking work environments, monetary incentives on single benchmarks lead to a

---

[3] Falk & Ichino (2006) use the term controlled field experiment in their paper, which is alike a framed field experiment in the taxonomy from Harrison & List (2004).

[4] However, even "Speaking to Theorists" can be applied, as several theories have important applications.

[5] The research article with Gary Bolton and Axel Ockenfels represents an exception. The basic purpose of the paper is to test game theoretical prediction of individual behavior in the laboratory.

crowding out of effort in other dimensions. Our work answers the naturally occurring question whether peer pressure as an alternative incentive is connected to similar problems and how employers should deal with it. I will present for each of my projects a corresponding application.

The presented parts are sorted in reverse chronological order. The first parts contain the latest research.

Part 2 is concerned with *nonstandard preferences*. More precisely, "default effects" arising due to the "status quo bias" (Samuelson & Zeckbauer 1988) are examined. Therefore, a randomized field experiment was conducted within a firm. On the company's website, households could enter into a typical consumer electricity contract. Website visitors were randomly allocated to one of the two treatments. Depending on the treatment, the 100% green energy check box was activated by default, or not. The results show a remarkable difference in consumer choice between treatments. When the check box for 100% green energy was activated by default, 69% of all consumers opted for the green contract. When the check box was not activated by default, only 7% opted for the green contract. Furthermore, the "conversion rate" (concluded contracts/ website visitor) does not differ between treatments. Hence, for the enterprise it makes no difference which treatment they implement. However, the important contribution of the paper is not the evidence that defaults influence decisions, but that the experimental environment allows examining why defaults work. Previous field evidence, especially from 401(k) saving plans (e.g. Madrian & Shea 2001, Caroll et al. 2009), suggests that defaults work due to time-inconsistent preferences. In contrast, my experiment suggests that defaults can also evoke strong effects due to other reasons. Even if the experiment does not allow identifying a single result driver, it indicates that defaults provide information or construct preferences. While such explanations were already known from non-incentivized laboratory experiments on defaults (e.g. McKensie et al. 2006, Dinner et al. 2011), to my best knowledge, this the first evidence indicating that defaults also trigger mentioned psychological processes in the field.

Part 3 presents a joint work with Gerlinde Fellner and Johannes Wahlig on peer pressure.[6] In a laboratory experiment we explore the role of peer pressure at the work place. DellaVigna (2009) defines individuals' behavior adjustment when observed by peers as *nonstandard decision making*. Following DellaVigna's argumentation, peer pressure or (more general) social pressure is considered as a pressure to conform, which, in turn, is considered as a result of the excessive impact of others' beliefs on individual behavior.[7] In our experiment participants had to describe pictures. For example, a picture of a garden with a lawn, some garden furniture and different kinds of flowers, trees etc. They could allocate up to ten labels to each picture. The number of labels / descriptions allocated to a picture represents the quality, while the number of pictures they described represents the quantity of work. In one treatment, we tried to harness peer pressure as an incentive to increase productivity.[8] We exposed participants to peer pressure by informing them about the work quantity (number of pictures) of a co-worker. In line with the prediction of a short theoretical model we developed for the project, the peer incentive in the quantity dimension increases the provided work quantity and productivity, but decreases quality. Beside this core result, data analysis revealed further insights with practical implications. Although peer pressure increases productivity, workers exposed to peer pressure show no higher

[6] Contributions: Idea & Theory: Ebeling; Experimental Design: Ebeling & Fellner; Experiment-implementation: Wahlig; Result-analysis & Writing: Ebeling & Fellner.

[7] Alternatively, it is also possible to consider individuals' effort adjustment due to peer pressure as a *nonstandard preference* for conformity (Bernheim 1994). Individuals do not want to be perceived as lazy and therefore increase work effort when exposed to peer pressure. Furthermore, in our experiment more egoistic individuals react stronger to peer pressure. Such systematic reaction patterns indicate the relevance of nonstandard (social) preferences for decision making under peer pressure. However, this example shows the close relation between behavioral anomalies.

[8] Productivity was measured by total number of labels.

stress levels in self-evaluations. This result might have implications for employers: Satisfaction of workforce does not change, but productivity rises when peer pressure is systematically exploited. Furthermore, our analysis sheds light on the optimal composition of workforce and thereby contradicts recommendations of previous literature from Falk & Ichino (2006) and Mas & Moretti (2009).

Part 4 is concerned with fundraising and altruism. It is a joint work with Christoph Feldhaus and Johannes Fendrich.[9] Most probably, fundraising is the economic research area most heavily drawing knowledge from field experiments. We also conducted a field experiment. In our experiment, we explored the role of the first giver's status in a fundraising campaign. There already exists theory and laboratory evidence on individuals concern for status. Ball et al. (2001) shows that individuals sacrifice consumption to associate with high status individuals in a laboratory market place. Kumru & Vesterlund (2010) transfer Ball et al.'s theory into a fundraising context. In a laboratory public good game they find similar results. When high status individuals donate first, low status individuals are more prone to donate to the public good. Despite the numerous field experiments on fundraising and the frequent anecdotal evidence about the role of celebrity donors in fundraising campaigns, there exists no scientific field evidence showing that potential donors indeed donate to associate with high status first givers. We provide field evidence for this *nonstandard preference*. For two weeks we trailed a homeless person asking for donations within Cologne's metro trains. Thereby we systematically varied the status of the first giver in the train. In the control treatment we did not intervene. In the low status treatment the first giver was always a (poor looking) low status person from our team and correspondingly in the high status treatment a (rich looking) high status person. The experiment provides two core results.

[9] Contributions: Idea & Experimental Design: Ebeling; Experiment-implementation: Fendrich; Result-analysis: Feldhaus; Writing: Ebeling.

First, in line with numerous previous research (e.g. Fischbacher et al. 2001, Kocher et al. 2008, Fischbacher & Gächter 2010) metro passenger are conditional cooperative and are more prone to give as soon as another metro passenger gives. Second, the probability of subsequent giving in the high status treatment is significantly higher than in the low status treatment. The practical application of our results is straight forward. Designers of fundraising campaigns should search for a high status first donor as it seems to be an advantageous method to increase subsequent giving.

In part 5, I further analyzed conditional cooperative behavior. As already shown in the previous part, conditional cooperation is one of the most persistent behavioral patterns in public good games and charitable giving. Nevertheless, there is rather little knowledge how heterogeneous incomes / endowments of potential donors affect conditional cooperation. The only existing results from laboratory public good games find that absolute amount of giving does not depend on endowment (see e.g. Buckley & Croson 2006), which is surprising, as it is intuitively appealing that those with higher endowments should give more (and those with lower endowments should give less). Furthermore, there is no research scrutinizing how ambiguity about other donors' endowment affects conditional cooperation. The experiment presented in part 5 tries to gain further insights on these two topics. Furthermore, the paper can be considered as groundwork for the paper in part 4, as income and status are closely related – those with higher status are often richer. The experimental results show that endowment heterogeneity affects conditional cooperation but ambiguity does not. Individuals donate less, when they think that other donors have a higher endowment. However, they do not deliberately overestimate others endowment in case of ambiguity to justify a lower donation. This result on endowment heterogeneity differs from existing literature which does not detect any effect of endowment on giving in laboratory public good games. A reason might be the different environment (classical public good game versus charitable giving environment) or the different methods used to detect the effect of endowment heterogeneity. Furthermore, it is interesting to compare this result with the result of part 4. While in part 4 a first giver with higher status (and also higher income) positively affects subsequent giving, in part 5 a higher endowment of previous givers lead to lower giving of subsequent donors. This comparison suggests that in the experimental environment in part 4 the status matters more than the income.

The last research article of my thesis in part 6 is a joint work with Gary Bolton and my supervisor Axel Ockenfels.[10] I am only a co-author of this article. In this research project we test predictions of the game theoretical model of Kreps & Wilson (1982) in a laboratory experiment. In their seminal paper, Kreps & Wilson (1982, p. 266) build a game theoretical model showing that the same reputation building equilibrium applies regardless of whether the incumbent is protecting a chain store monopoly from entry by one, repeat challenger or a series of one-shot challengers. This result only applies when reputation builder's record is freely available and agents comply with equilibrium strategy. But previous research showed actual behavior to be somewhat of the equilibrium path. In this case, repeated challengers might have additional incentives to test reputation. In our laboratory test, we find that individuals indeed show *nonstandard decision making* and systematic deviations from the equilibrium path. Most important, repeated challengers test reputation more often. This is intuitively appealing as they internalize potential gains from tests while this is not the case for one-shot challengers. While this result does not have direct applications, it might be helpful to better understand reputation / feedback systems as used by eBay.

[10] Contributions: Experiment-implementation: Ebeling; Result-analysis: Ebeling & Bolton; Remainder: Bolton & Ockenfels.

References

Ball, Sheryl, Catherine Eckel, Philip J. Grossman, and William Zame. 2001. “Status in Markets.” *Quarterly Journal of Economics, 116(1),* 161-188.

Bernheim, B. Douglas. 1994. “A Theory of Conformity.” *Journal of Political Economy*. 102(5), 841-877.

Bolton, Gary E. and Axel Ockenfels. 2012. “Behavioral economic engineering”. *Journal of Economic Psychology*. 33, 665–676.

Buckley, Edward, and Rachel Croson. 2006. “Income and wealth heterogeneity in the voluntary provision of linear public goods.”*Journal of Public Economics*, 90 (4-5), 935–55.

Camerer, Coln and GeorgLoewenstein. 2004. “Chapter 1: Behavioral Economics: Past, Present, Future”. *Advances in Behavioral Economics*. Eds. Camerer, Colin F., Georg Loewenstein and Matthew Rabin .

Caroll, Gabriel D., James J. Choi, David Laibson, Brigitte C. Madrian, Andrew Metrick. 2009. “Optimal Defaults and Active Decisions” *Quarterly Journal of Economics,*124(4), 1639-1674.

DellaVigna, Stefano. 2009. “Psychology and Economics: Evidence from the Field”. *Journal of Economic Literature*. 47 (2), 315-372.

Dinner, Isaac, Eric J. Johnson, Daniel G. Goldstein, Kaiya Liu. 2011. „Partitioning Default Effects: Why People Choose Not to Choose.“ Journal of Experimental Psychology: Applied, 17(4), 332-341.

Falk, Armin and Andrea Ichino. 2006. “Clean Evidence on Peer Effects.” *Journal of Labor Economics*, 24(1), 39-57.

Fischbacher, Urs and Simon Gaechter. 2010. "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Good Experiments."*American Economic Review, 100(1)*, 541–56.

Fischbacher, Urs, Simon Gachter, and Ernst Fehr. 2001. "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment."*Economics Letters*, *71(3),* 397–404.

Harrison, Glenn W. and John A. List. 2004. "Field Experiments". *Journal of Economic Literature*, 42 (4), 1009-1055.

Kocher, Martin G., Todd L. Cherry, Stephan Kroll, Robert J.Netzer, and Matthias Sutter. 2008. "Conditional Cooperation on Three Continents." *Economics Letters, 101(3),* 175–78.

Kreps, David M. and Robert Wilson. 1982. "Reputation and Imperfect Information." *Journal of Economic Theory*, 27, 253-279.

Kumru, Cagri S. and LiseVesterlund. 2010. "The Effect of Status on Charitable Giving."*Journal of Public Economic Theory, 12(4),* 709-735.

Madrian, Brigitte C., and Dennis Shea. 2001. "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior," *Quarterly Journal of Economics*, 116, 1149-1187.

Mas, Alexandre and Enrico Moretti. 2009. "Peers at Work." *The American Economic Review*, 99(1), 112-145.

McKenzie, C.R., M.J. Liersch and S.R. Finkelstein. 2006. „Recommendations Implicit in Policy Defaults." *Psychological Science*, 17(5), 414-420.

Peichel, Andreas. 2009. “The Benefits of Linking CGE and Microsimulation Models: Evidence from a Flat Tax Analysis”. *Journal of Applied Economics*,12 (2), 301-329.

Roth, Alvin E. 1995. “Introduction to Experimental Economics: The Use of Experiments.” *The Handbook of Experimental Economics*.Eds. John H. Kagel& Alvin E. Roth.

Samuelson, William and Zeckhauser, Richard. 1988. "Status Quo Bias in Decision Making." *Journal of Risk and Uncertainty*, 1(1), 7-59.

Part 2

Why do Defaults work?

-

Evidence from a Natural Field Experiment

Why do Defaults work?

-

Evidence from a Field Experiment

Felix Ebeling*

December 19, 2012

Abstract

In cooperation with an electricity provider we conducted a field experiment to scrutinize the impact of defaults on electricity customers' willingness to purchase a "green" electricity contract. We ran the experiment with more than 40,000 interested customers and tested how they react on different non-binding green energy defaults in the contract offer. The results show that (i) the fraction of consumers willing to purchase an electricity contract does not depend on the default, and (ii) the fraction of customers opting for a contract, delivering only electricity from renewable sources, strongly depends on the default. The fraction of "green" customers is less than 10% when the green energy check box is not set by default and around 70% when it is set by default- despite the higher price of the "green" contract. Importantly, in our experimental setting the costs of opting out of the default are negligible when purchasing the contract, but very high in later periods. Due to this cost structure, our results suggest that customers' reactions on defaults cannot be explained by switching costs or procrastination as done in previous literature on default effects.

JEL Classification: D03, D12, L94, Q41

Keywords: cooperation, optimal defaults, renewable energy demand, field experiment

* Felix Ebeling, University of Cologne, Department of Economics, ebeling@wiso.uni-koeln.de.

1. Introduction

Literature from economics and psychology documents a pervasive impact of non-binding default options on people's revealed preferences. Whatever the default in decisions on e.g. insurance contracts (Johnson et al. 1993), postmortem organ donation (Johnson & Goldstein 2003), retirement savings (Caroll et al. 2009), or even television program choice (Esteves-Sorenson & Perretti 2012), agents persistently stick to it. Explanations for the field evidence cited above either rely completely on switching costs and procrastination or suggest that a large portion of the effect size is due to this rationale. In contrast, the large default effects in our field experiment strongly suggest explanations other than switching costs. To our best knowledge, thus far, such rationales were unambiguously only detected in laboratory experiments.

This paper reports results from a field experiment that we conducted in cooperation with an electricity provider. On the electricity provider's website, prospective customers can purchase electricity household contracts. To receive a binding contract offer, interested customers only have to provide their zip code and their estimated yearly electricity consumption. Afterwards, customers can choose between several contract offers. Our focus is on contract offers' differences in the default of the "100% green" electricity option. 100% green means that electricity is purely generated by renewable sources as wind, solar or hydro. We conducted two treatments. In the control treatment, the 100% green check box is not set by default in the contract. In the green treatment, however, the check box is set by default.

We conducted the experiment for four weeks. During this period roughly 46,000 prospective customers inquired contract offers. 4720 purchased a contract. Our experimental analysis provides two main results. First, the conversion rate (fraction of website visitors buying a contract) does not differ between treatments. Second, there is a highly significant difference in fraction of customers buying a 100% green electricity contract. While in the control treatment less than 10% buy a 100% green energy contract, this percentage rises to around 70% in the green treatment.

Beside the contribution to the question why defaults work, this paper has some practical relevance. First, it contributes to behavioral economists efforts to nudge household to more sustainable energy consumption patterns. Schultz et al. (2007) and Alcott (2011) show that providing information on energy consumption norms induces individuals to conserve energy. This paper presents a complementary approach. Our intervention does not decrease energy consumption, but substitutes demand for carbon emitting electricity by demand for more costly, carbon neutral electricity. Second, it contributes to the publicly discussed question whether consumers are willing to pay high costs for the energy policy. [1] This paper shows that, at least in a certain range, there is no unique willingness to pay for green energy, but that consumers' preferences vary due to subtle changes in contract presentation.

The remainder of the paper is organized as follows. The next section presents an overview of the related literature. In section 3 we describe the experiment in more detail. Section 4 presents the results and section 5 concludes.

2. Relevant Literature

This section highlights the most important explanations for default effects and mentions the respective, relevant literature. Thereby, we categorize explanations in three categories (1. switching costs, 2. information provision and 3. constructing preferences).

[1] Several newspapers from different countries report consumer related problems with the green energy:
http://www.nytimes.com/2010/11/08/science/earth/08fossil.html
http://www.nytimes.com/2009/03/29/business/energy-environment/29renew.html
http://www.guardian.co.uk/money/2012/feb/08/higher-household-energy-bills-2020
http://www.spiegel.de/international/business/merkel-s-switch-to-renewables-rising-energy-prices-endanger-german-industry-a-816669.html
http://www.spiegel.de/international/germany/doubts-increasing-about-germany-s-switch-to-renewable-energy-a-844844.html

Furthermore, this section reviews past efforts of behavioral economists to nudge households to more sustainable consumption patterns.

First, default effects may occur because of *switching costs*. Assuming rational agents, defaults should have an effect on agents' decisions when costs of opting out of the defaults are higher than the benefits (see, e.g., Thaler & Sunstein 2003, Schwartz & Scott 2003). Such an explanation is especially feasible when agents do not really care about the default issue. In such cases opting-out costs quickly surpass benefits.[2] However, in the prominent studies on defaults in 401(k) investments, effects due to default-changes can hardly be explained by rational decisions due to switching costs. For clarification, consider the example provided by DellaVigna (2009). DellaVigna depicts that the mean employee from the Madrian & Shea (2001) study receives 1200$ yearly when opting out of the default contract. It is hard to imagine how employee's costs of the necessary phone call surpass this yearly benefit.[3] Therefore, studies on saving plans explain inertia generally with quasi-hyperbolic preferences of naïve individuals. Such agents stick with the default even when switching costs are lower.[4] Lately, Esteves-Sorenson & Peretti (2012) show that this reasoning even applies in situations where switching costs are seemingly negligible low. In their study, television viewers constantly postpone switching channels when the next program starts - despite the very low costs of switching.

Second, default effects may occur because they *provide information*. Assuming rational agents, defaults should have an effect on agents' decisions when a better informed decision designer potentially signals the best option through the default. For example,

---

[2] For example, it is possible - or at least imaginable – that the behavior of a certain fraction of agents in the prominent organ donor example (Johnson & Goldstein 2003) is rational. More precisely, there might be agents that simply do not care what happens to their body after death and therefore rationally do not engage in a costly change of the default option.
[3] Cost / benefit ratios in other studies about saving plans (Cronqvist & Thaler 2004, Choi et al. 2004, Caroll et al. 2009) are in the same range.
[4] See our model in Appendix B for a detailed explanation of quasi-hyperbolic preferences and naïveté.

rational workers in 401(k) saving studies might follow defaults because they think that their better informed employer provide the best solution via default. Similarly, rational car customers might follow defaults of a manufacturer when configuring their car, because they believe that the superiorly informed manufacturer provides recommendations about the optimal car customization via defaults (See Levav et al. 2010).[5] Several psychology papers show that agents indeed comprehend the information value of defaults (Brown & Krishna 2004, McKensey et al. 2006, Tannenbaum & Ditto 2011). Altmann et al. (2012) provide an economic approach with a formal model and laboratory evidence which extends findings from psychological research. They show that rational agents only unconditionally follow defaults when the decision designer is better informed and his interests are congruent with the agent's interests - an assumption which is arguable in the above mentioned car manufacturer example.[6] Furthermore, defaults might provide information on issues not related to the agent's monetary / rational interest, but following nevertheless maximizes the agent's utility. For example, defaults might signal what other agents do in the decision situation and are therefore appealing to conditional cooperators (see, e.g., McKensey et al. 2006, Altmann & Falk 2009). Alternatively, defaults might signal a social norm (e.g. about organ donations, see Johnson & Goldstein 2003), which is appealing for conformists (Bernheim 1994, Sliwka 2007). So, even in cases where defaults only provide information on non-rational issues, they might be valuable for the agent.

---

[5] In the field experiment of Levav et al. (2010), German car customers configure their vehicle previously to the purchase with the configuration software from the car manufacturer. Customers choose different options on motor, rims, sound system etc. They potentially follow defaults when they belief that the manufacturer has superior information about the optimal customization and signals best combinations via defaults (e.g. optimal gear transmission for a certain car).

[6] In fact, defaults as scrutinized by Levav et al. (2010) do not exist in real car customization processes because the German consumer advice center intervenes against defaults setting due to manufacturer's conflict of interest.

Third, defaults work because they *construct preferences*. The foremost explanation is that the default evokes an instant endowment-effect (e.g. Kahneman et al. 1991).[7] For an endowment effect, two assumptions on individual behavior have to be made. First, setting the default instantly alters the agent's perception and changes his reference point. Second, agents are loss averse. Under these assumptions, agents claim more compensation to drop a set default than they are willing to pay for an unset default. Hence, their willingness to accept (WTA) a loss is higher (more costly) than their willingness to pay (WTP) for a gain. In a contract offer such as in our experiment this implies that for a given price of an option a higher fraction of agents stick with it when it is the default option, than buy it when it is not the default option. Samuelson & Zeckhauer (1988) suggested such an explanation already for the status quo bias observed in their study. Kahneman et al. (1991), Johnson et al. (1993) and Camerer et al. (2003) take the same line in their reasoning for default effects. An alternative explanation how defaults construct preferences is that they eliminate uncertainty about individuals own preferences. Dhingra et al. (2012) provide a theoretical model and laboratory evidence for this approach. Furthermore, there is psychological research from Dinner et al. (2011) suggesting that defaults shape individuals' preferences for contract options. In a laboratory survey Dinner et al. (2011) show that participants evaluate more costly compact fluorescent light bulbs (CFL) much better and would buy them much more often, respectively, when CLFs are the default instead of incandescent light bulbs. We will discuss in the final chapter which reasoning is most likely for our results.

Furthermore, this paper is about sustainable consumption. Schultz et al. (2007) show in a small-scale experiment that providing information on social norms induces individuals

[7] Groundwork papers on endowment effects are Thaler (1980) and Kahneman & Tversky (1984).

to conserve energy. Inspired by these results, the company OPOWER launched a likewise experiment with nearly 600,000 households in the U.S.A. The company mails so-called Home Energy Report letters to households. Letters basically contain a comparison of own energy use and energy use of similar neighbors. As shown by Alcott (2011) the program reduces energy consumption on average by 2.0%. Our paper presents a complementary approach. It does not show how to decrease energy consumption, but how people are more prone to substitute carbon emitting electricity by more costly, carbon neutral electricity.

3. Experiment

3.1 Environment

Our natural field experiment (Harrison & List 2004) was conducted on an electricity provider's website in early summer 2012. The electricity provider is one of the ten biggest providers in Germany and delivers electricity to households throughout the whole country. On the firm's website, visitors can purchase typical consumer electricity contracts. Contracts are generally addressed to households with 1 to 6 persons.

We present data of a period of four and a half weeks (May 9th to June 11th, 2012). The actual experiment ran a week longer, but due to technical problems, the last week's data was not collected properly. There is no reason to assume that the last week's data differ from previous weeks' data or that overall results change if data was available.

3.2 Procedure

Independent of treatment, prospective customers have to scroll through three website screens to receive a binding contract offer from the electricity supplier by mail.

On the *first screen*, which can be found in Figure A1 in Appendix A in a translated and anonymized way, visitors have to type in their zip code and their yearly electricity consumption. The zip code is necessary, because power grid charges differ between German regions and consequently offered electricity prices also differ. The yearly electricity consumption is used to show customers on the second screen what they can save per year in comparison to the local electricity provider if they choose our provider.[8] After the prospective customer entered his zip code and his yearly consumption, they have to click the "show contract" button and the second screen with detailed contract offers for the region appears.

The *second screen*, which can be found in Figure A2 in Appendix A, offers two contracts. The structure of contracts is similar. In both contracts, households have to pay a consumption independent base price plus a certain price per unit. However, one contract provides less service (e.g. no telephone hotline) and is cheaper.[9] In the following, we will call these two contracts low-service and high-service contract. The base price is always 2€ lower in the low service contract. The price-per-unit is region dependent between 0.5€cent and 4€cent lower in the low service contract. Furthermore, each contract shows the yearly savings customers can achieve when they choose our provider instead of the local electricity provider. Customers can choose which of the two contracts they would like to order by clicking the "order now" button.

The *third screen* appears, after a customer ordered a contract. On the third screen, customers have to type in their name and address. They receive the chosen binding contract offer by mail a few days later. The consumer only concludes the contract, when she signs the contract that he received by mail and sends it back to the commodity supplier.

---

[8] In Germany, households receive electricity by default from the local public utility if they do not actively switch to another provider. The shown savings result from the comparison with prices of the local public utility. Unfortunately, our electricity provider does not calculate the savings, but receives the saving amounts from an external service provider. Therefore, we do not have access to the exact savings.
[9] Due to anonymity guaranty for cooperating firms we do not mention exact features.

3.3 Treatments

There are two treatments. The treatments differ in their default rule for the green energy option on the second screen. In the control treatment the check box for 100% green energy is not selected by default. In the green treatment the check box is selected by default. Independent of treatment, the price per kilowatt hour increases by 0.3 cent as soon as the box for 100% green energy is checked. Hence, the price increases by 0.3 € cent in the control treatment as soon as the check box for 100% green energy is set to on, and decreases by 0.3 € cent in the green treatment as soon as the checkbox is set to off. Depending on the zip code, 0.3 € cent represents additional costs of 1.15% to 1.46% per kilowatt hour for the consumer. This represents an additional payment of ≈ 9€/year for the average customers from our experiment. The yearly costs of green energy were visible to prospective customers when clicking the checkbox. As soon as they clicked, the yearly savings achievable in comparison to the local utility changed. When activating green energy, the savings decrease by the respective yearly costs, when deactivating green energy, the savings increase by the respective yearly costs. In case that customers order a non-green energy contract, they receive their electricity from the current standard electricity mix in Germany.[10] It is important to mention that the green energy check box was always similarly set for the low-service and the high-service contract. Either in both contracts the check box was set, or for none.

3.4 Technical Issues & Problems

While it is almost as easy to program an experiment for a website as for the laboratory, the environment provides less control. In the following, we will highlight two problems we had to deal with in our experiment.

---

[10] In this mix in 2011 43.5% of electricity came from coal, 19.9% from renewable sources, 17.6% from nuclear plants, 13.7% from gas and 5.3% from other sources.

Whether a website visitor saw the control treatment or the green treatment was decided randomly when he entered the website. To make sure that the visitor saw only one treatment in case he visited the website more than once, cookies were used. Nevertheless, due to the following two reasons, we analyze visits instead of visitors. First, it is not completely uncommon that visitors do not activate cookies or delete cookies. If we had analyzed visitors, we would have identified repeated visits of those visitors as new visitors. Second, Internet search engines, whose function and relevance we describe in the next paragraph, bypass cookies. As we don't know exactly how search engines interact with the electricity provider's website, allocation of visits to visitors is impossible. Fortunately, for our experiment, differences between visitors and visits are less important. First, since new visitors are randomly allocated to the treatments, we can deduce from an equal number of visits an equal number of visitors per treatment (even if we cannot exclude that some visitors saw both treatments). Second, and most important for our analysis, we mainly analyze the data of prospective customers that finally bought the contract. These visitors can be identified unambiguously, as each household can only buy one contract.[11]

Internet search engines, direct links or advertising links allow customers to buy a contract from our electricity provider without participating in the experiment. Indeed, a significant minority of customers who bought a contract from our electricity provider during the period of the experiment did not participate in the complete process of our experiment. As for most products offered in the Internet, there are search engines, which offer price comparisons for electricity contracts (e.g. www.stromsparer.de). Prospective customers who receive a contract from our electricity provider via search engines generate a visit on our page but most probably did not see the website, since the

[11] Even if a household order more than one binding contract offer to sign (what is rather unlikely), in the end with certainty only one contract is purchased.

visit is generated by the search engine instead of the customer directly.[12] Likewise, but to a minor extend, prospective customers may bypass the experimental procedure when they enter the website through a direct link or an advertising link. These links might bring prospective customers directly to the second screen. Fortunately, we are able to identify whether prospective customers ran through the whole experimental procedure or only saw part of it. We simply checked whether a visitor provided all data collected in the experimental process. Only when the visitor ran through the whole process, the firm's web-server was able to collect the complete data. However, for those who did not run through the complete experiment, we are not able to identify how much they exactly saw. Whether a prospective customer saw basically nothing of our experiment (because he used an internet search engine) or whether he saw the most important second screen but not the first (because a direct link led him directly on the second screen) remains unclear for us. Since we did not want to exclude "partial participants" completely from our analysis, we differentiate in the following sections between "direct" visitors who ran through the whole experimental process and "indirect" visitors who only saw part of it.

4. Results

Table 1 summarizes the collected data. The first row contains the visits of the second website screen including "indirect" visits. Overall, there were 63,768 visits on the second website screen, 33,341 in the control treatment and 30,427 in the green treatment. The firm estimates that each visitor makes on average 1.3 to 1.5 visits. This implies between 42,500 to 49,000. [13] The second row contains the number of "direct" visits on the second screen. The final row contains the means of the yearly consumption of customers that

[12] We checked for some of the most popular search engines how prospective customers receive a contract from them. Basically, customers who receive a contract from them do not see the original website of the actual electricity provider at all.
[13] From other tables we received from the company we could derive that the best guess is around 46.000 visitors.

purchased a contract through the direct process (based on consumers' own statement on the first website screen, which is only available if customer used the direct procedure).

TABLE 1. DATA SUMMARY

	Control	Green
Visits of second screen	33,341	30,427
Visits of second screen (direct only)	21,960	19,992
Purchased contracts	2,490	2,230
Purchased contracts (direct only)	1905	1607
100% green energy contracts	143	1,698
100% green energy contracts (direct only)	137	1115
Average yearly consumption (purchased & direct only)	3038kWh	3011kWh

First, we consider the "Conversion Rate". The conversion rate describes the fraction of prospective customers that finally purchase a contract and can be seen in Figure 1 (left). Unsurprisingly, the main focus of the firm was on the conversion rate. A two-tailed Chi-squared test cannot reject the hypothesis of similar ratios of visits to purchased contracts in both treatments when considering all visits ($p=0.51$). When differentiating between direct and indirect visits, there are significant differences. For direct visitors the conversion rate is higher in the control treatment (two-tailed Chi-squared test $p<0.05$), for "indirect" visits the conversion rate is higher in the green treatment (two-tailed Chi-squared test $p<0.01$). But as Figure 1 (left) shows, even if the conversion rate is generally higher for direct visits, for each visit-channel the differences between treatments are relatively small in size. Ultimately, for the firm the default doesn't matter, as the accumulated conversion rate does not differ between treatments.[14]

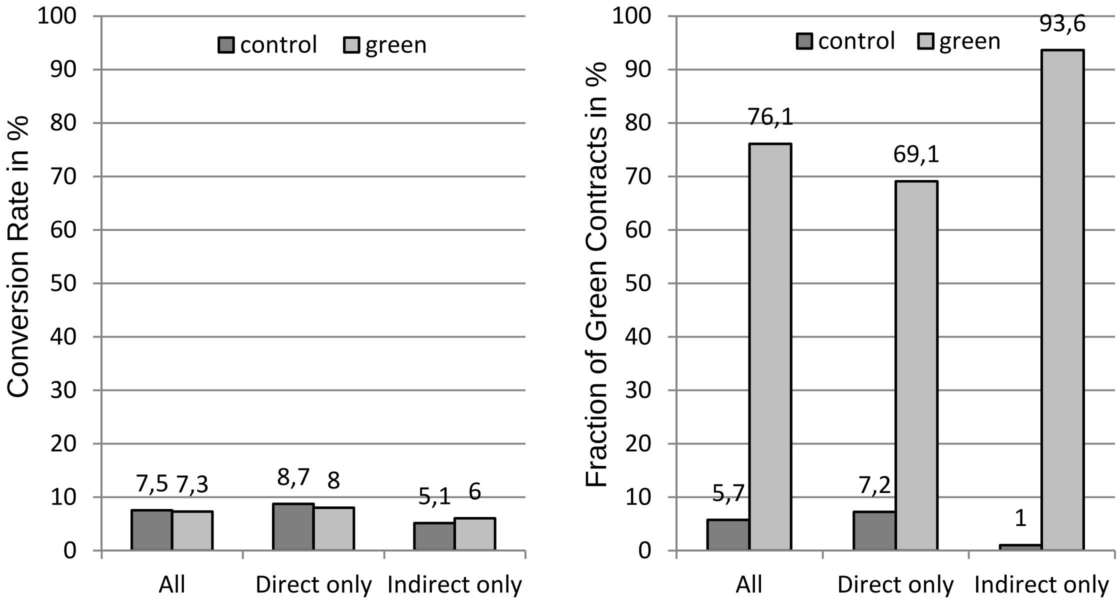


FIGURE 1. CONVERSION RATE AND FRACTION OF GREEN CONTRACTS

Second, we consider the fraction of green contracts. Figure 1 (right) displays that, the fraction of green contracts strongly differs between treatments. This difference is highly significant (two tailed Chi-squared test, p<0.01 for All, Direct only and Indirect only, respectively). However, we also see in Figure 1 (right) that there are differences between visit-channels. We will constrain our analysis to direct visitors only, because the more extreme result for indirect visits is most probably evoked by the simple fact that a large fraction of indirect customers never saw the second screen of our experiment and therefore could not change the default.

In the control treatment, of 1905 purchased contracts through the direct channel only 137 or 7.2% are green. In contrast, in the green treatment, of the 1607 purchased contracts through the direct channel 69.1 % or 1115 contracts are green. Even if the

[14] As the electricity market is highly competitive, the contribution margin for the standard energy mix and 100% green energy is similar. Hence, for the firm only the number of contracts matters, not the kind of contract.

difference is huge, more customers stick to the non-green default than to the green default. A two tailed Chi-squared test highly significantly rejects the hypothesis that the fraction of customers staying with the green contract in the green treatment is as high as the fraction of customers staying with the non-green contract in the control treatment ($p<0.01$). This implies that for a substantial fraction of customers the strength or stickiness of the default is not sufficiently high to make them willing to bear the additional costs of green energy. Hence, even if the result shows that defaults have a strong impact on customers, it also shows the limits of default effects.

Furthermore, we use probit analyses to examine customer's contract choices in more detail. Thereby, we benefit from differences of several contract variables between regions. Remember that customers have to type in their estimated yearly consumption as well as their zip code before receiving the contract offers. While the yearly consumption naturally differs between customers, there are also differences in the contract offers towards customers depending on the customer's zip code. The base-price, the price per unit as well as the difference in price per unit between the high-service and the low-service contract depend on the region. This allows us to test for the influence of these variables on green energy choice.

Table 2 presents the results of our probit analysis. The analysis supports that the treatment is by far the best predictor for the green energy choice. Besides, only the price difference between the low-service and high-service contract significantly affects the green energy choice (but to a far smaller degree than the treatment). The higher the difference in price-per-unit between the low-service and high-service contract, the higher the probability to choose a green contract. This might be due to a framing effect, a "Titchener Illusion" with numbers: The higher the price-per-unit difference between the low-service contract and the high-service contract, the lower customers perceive the mark-on for green energy and therefore are more prone to purchase the green energy contract. Furthermore, it is remarkable that yearly consumption is not significant. It would have been intuitively appealing, that those with a higher yearly consumption are more price-sensitive. Especially, as our probit analysis in Table 1.A in Appendix A shows that when it comes to the choice between the low-service and the high-service contract, the yearly consumption indeed has a highly significant positive influence on the propensity to choose the cheaper contract. A possible reason might be that bigger households have different socioeconomic values and are more open to green energy.

TABLE 2: GREEN CONTRACT CHOICE

	(1)	(2)	(3)
Treatment	1.968***	1.972***	2.009***
	(0.05)	(0.05)	(0.08)
Price Difference		0.121***	0.124***
		(0.04)	(0.04)
Contract Type			-0.025
			(0.09)
I-Contract-Treatment			-0.063
			(0.11)
Base Price			-0.028
			(0.02)
Price / Unit			0.013
			(0.04)
Yearly Consumption			-0.00
			(0.00)
Observations	3512	3512	3512
Pseudo R-squared	0.3520	0.3542	0.3555

Notes: probit regressions with green contract choice donations as dependent variable. Standard errors in parentheses. "Treatment" represents a binary variable for the green treatment, "Price Difference" represents the price-per-unit difference between the low-service and the high-service contract, "Contract Type" represents a binary variable indicating whether customer bought a low-service or a high-service contract, "I-Contract-Treatment" represents the interaction variable of "Treatment" and "Contract Type", "Base Price" represents the base price of the high-service contract, "Price / Unit" represents the price-per-unit of the high-service contract. Level of significance: *p<0.10, **p<0.05, ***p<0.01.

5. Discussion and Summary

In our experiment we did not implement a neutral frame.[15] The control treatment nudges the costumers towards no green energy, the green treatment nudges the costumers towards green energy. In the control treatment 7% opt for green energy, in the green treatment 69% opt for green energy. In consideration of all existing evidence about defaults, in a neutral frame the customer fraction opting for green energy should be in between these values. Hence, in our experiment at least $\approx 30\% = (69\% - 7\%)/2$ of the customers were nudged in one or the other direction. Which of the psychological processes reviewed in chapter two can explain changed decision of at least 30% of all customers?

An explanation based on switching costs and quasi-hyperbolic preferences is technically possible in our experimental setup.[16] Agents can order the default contract now and change the green contract feature later. If doing so, they would prefer to save very small switching costs now, but spend much higher switching costs later. Switching now is just a click, but switching later would represent an additional contract purchase, which involves calling or writing the electricity provider, signing a new contract offer and sending it back to the electricity provider.[17] In Appendix B we formalize this decision situation for individuals with quasi-hyperbolic preferences. Thereby, we use the standard theoretical framework, which was also used by Laibson (1997), O'Donoghue & Rabin (1999), Carroll et al. (2009) and others. Our formal approach shows that in our setting only individuals with very extreme time inconsistent preferences postpone

[15] A neutral frame would have been a contract offer (second screen) with two check boxes in each contract - one for green energy, one for non-green energy, none checked. Prospective customers would have been obliged to actively decide which green energy option they want if they want to purchase a contract.

[16] Explanations based on rationality and switching costs are not considered. Switching costs in our experiment are way too low for such reasoning.

[17] Furthermore, whenever customers switch, there are similar costs of product information, if customers do not know what 100% green energy means (a case rare in Germany). For these customers, there is an info link to detailed information in the "100% green" text next to the check box.

switching. This is due to the strong increase in switching costs after the initial period in our experiment: The stronger the increase in switching costs after the initial period, the higher the necessary degree of naïveté to explain the results with switching costs and quasi-hyperbolic preferences. Considering results of papers that measured actual degrees of time-inconsistency (e.g. Laibson et al. 2009, Paserman 2008, Skiba & Tobacman 2008), it seems to be very unlikely that a psychological process based on switching costs and quasi-hyperbolic preferences drives our results. If such a process indeed drove our experiment-results, the above-mentioned 30% of customers would have a degree of time-inconsistency that by far exceeds degrees measured in previous papers.

Another argument against the procrastination story is the similar fraction of customers sticking with the default in both contract versions. Due to the lower service, the switching costs in later periods are higher in the low service contract (e.g. because customers cannot call the electricity provider, but have to write an email). But higher switching costs in later periods should lead to less procrastination. We do not observe such results in our analyses in Table 2.

Does the default in our experimental setup provide information or construct preferences? With the existing evidence we are not able to disentangle these possible explanations. However, the laboratory survey from Dinner et al (2011) finds that defaults construct preference in an environment very similar to ours. In their survey they ask subjects whether they would prefer to buy either environment-friendly, but more expensive CFS light bulbs or environment-unfriendly, but cheap classical light bulbs. Depending on the default, choices change. If the default is on the CFS light bulbs, much more participants opt for these bulbs and vice versa. Furthermore, participants have to answer several questions on choice justification. The results suggest that in this particular case, the default does not provide information to the participant but construct their preferences. It would be easy to set-up a survey similar to the one of Dinner et al (2011), but focusing on green-energy choice as in our field experiment. Such survey could provide further insights on what mechanism drives our result.

Summarized, our experiment provides clear additional evidence for the strong impact of defaults on individuals' choice. Furthermore, to our best knowledge, this is the first field evidence showing that defaults can evoke strong effects on decisions, even if individuals' time-inconsistency can hardly explain the result. However, the setup does not allow differentiating between alternative explanations. Additional research as proposed in the previous paragraph is necessary to gain clearer insights on underlying psychological processes.

References

Alcott, Hunt. 2011. "Social norms and energy consumption." *Journal of Public Economics*, 95, 1082-1095.

Altmann, Steffen and Armin Falk. 2009. "The Impact of Cooperation Defaults on Voluntary Contributions to Public Goods." *Working Paper*.

Altmann, Steffen, Armin Falk and Andreas Grunewald. 2012. „Incentives and Information as Driving Forces of Default Effects." *Working Paper*.

Bernheim, B. Douglas. 1994. "A Theory of Conformity." *Journal of Political Economy*, 100, 841-847.

Brown, Christina L. and Aradhna Krishna. 2004. "The Skeptical Shopper: A Metacognitive Account for the Effects of Default Options on Choice." *Journal of Consumer Research*, 31(3), 529-539.

Camerer, Colin, Samuel Issacharoff, George Loewenstein, Ted O'Donoghue, and Matthew. Rabin. 2003. "Regulation for Conservatives: Behavioral Economics and the Case for "Asymmetric Paternalism" *University of Pennsylvania Law Review*, 151, 1211–1254.

Caroll, Gabriel D., James J. Choi, David Laibson, Brigitte C. Madrian and Andrew Metrick. 2009. "Optimal Defaults and Active Decisions" *Quarterly Journal of Economics,* 124(4), 1639-1674.

Choi, James J., David Laibson, Brigitte C. Madrian and Andrew Metrick. 2004. "For Better or for Worse: Default Effects and 401(k) Savings Behavior," in *Perspectives in the Economics of Aging*, David A. Wise, ed. (Chicago, IL: University of Chicago Press).

Cronqvist, Henrik, and Richard H. Thaler. 2004. "Design Choices in Privatized Social-Security Systems: Learning from the Swedish Experience," *American Economic Review*, 94, 424–428.

DellaVigne, Steffano. 2009. "Psychology and Economics: Evidence from the Field." Journal of Economic Literature, 47(2), 315-372.

Dhingra, Nikhil, Zach Gorn, Andrew Kener and Jason Dana. 2012. „The default pull: An experimental demonstration of subtle default effects on preferences." *Judgment and Decision Making*, 7(1), 69-76.

Dinner, Isaac, Eric J. Johnson, Daniel G. Goldstein and Kaiya Liu. 2011. „Partitioning Default Effects: Why People Choose Not to Choose." Journal of Experimental Psychology: Applied, 17(4), 332-341.

Esteves-Sorenson, Constanca, and Fabrizio Perretti. 2012. "Micro-costs: Inertia in Television Viewing." *The Economic Journal*, 563, 867-902.

Harrison, Glenn W. and John A. List. 2004. "Field Experiments." *Journal of Economic Literature*, 42(4), 1009-1055.

Johnson, Eric. J., and Daniel Goldstein. 2003. "Do Defaults Save Lives?" *Science*, 302, 1338–1339.

Johnson, Eric J., John Hershey, Jacqueline Meszaros and Howard Kunreuther. 1993. "Framing, Probability Distortions, and Insurance Decisions," *Journal of Risk and Uncertainty*, 7, 35–51.

Kahneman, Daniel and Amos Tversky. 1984. "Choices, Values and Frames." American Psychologist, 39, 341-350.

Kahneman, Daniel, Jack L. Knetsch and Richard H. Thaler. 1991. “Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias.” *The Journal of Economic Perspectives*, 5(1), 193-206.

Laibson, David. 1997. “Golden Eggs and Hyperbolic Discounting.” *Quarterly Journal of Economics*, 112, 443-477.

Laibson, David, Andrea Repetto and Jeremy Tobacman. 2009. “Estimating Discount Functions with Consumption Choices over the Lifecycle.” *Working Paper*.

Levav, Jonathan, Mark Heitmann, Andreas Herrman and Sheena S. Iyengar. 2010. “Order in Product Customization Decisions: Evidence from Field Experiments.” *Journal of Political Economy*, 118(2), 274-299.

Madrian, Brigitte C. and Dennis Shea. 2001. “The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior,” *Quarterly Journal of Economics*, 116, 1149-1187.

McKenzie, Craig R.M., Michael J. Liersch and Stacey .R. Finkelstein. 2006. „Recommendations Implicit in Policy Defaults.” *Psychological Science*, 17(5), 414-420.

O’ Donoghue, Ted and Matthew Rabin. 1999. “Doing it Now or Later.” *American Economic Review*, 89(1), 103-124.

O’ Donoghue, Ted and Matthew Rabin. 1999a. “Procrastination in Preparing for Retirement.” In Henry J. Aaron, eds., *Behavioral Dimensions of Retirement Economics,* Washington D.C.: Brooking Institution Press, 125-156.

Paserman, M. Daniele. 2008. “Job Search and Hyperbolic Discounting: Structural Estimation and Policy Evaluation.” *Economic Journal*, 118(531), 1418-52.

Samuelson, William and Richard Zeckhauser. 1988. "Status Quo Bias in Decision Making." *Journal of Risk and Uncertainty*, 1(1), 7-59.

Schultz, P. Wesley, Jessica M. Nolan, Robert B. Cialdini, Noah J. Goldstein and Vladas Griskevicius. 2007. "The Constructive, Destructive, and Reconstructive Power of Social Norms." *Psychological Science,* 18, 429–434.

Schwartz, Alan and Robert E. Scott. 2003. "Contract Theory and the Limits of Contract Law." *Yale Law Journal*, 113, 541-619.

Skiba, Paoge Marta and Jeremy Tobacman. 2008. "Payday Lonas, Uncertainty, and Discounting: Explaining Patterns of Borrowing, Repayment, and Default." *Working Paper.*

Sliwka, Dirk. 2007. "Trust as a Signal of a Social Norm and the Hidden Costs of Incentive Schemes." *American Economic Review*, 97, 999-1012.

Tannenbaum, David and Peter H. Ditto. 2011. „Information Asymmetries in Default Options." *Working Paper*.

Thaler, Richard H. 1980. "Toward a Positive Theory of Consumer Choice." *Journal of Economic Behavior and Organization*, 1, 39-60.

Thaler, Richard H. and Cass R. Sunstein. 2003. "Libertarian Paternalism." *American Economic Review*, 93, 175-179.

Appendix A. Figures and Tables

FIGURE A1. WEBSITE SCREEN FOR CONTRACT REQUEST

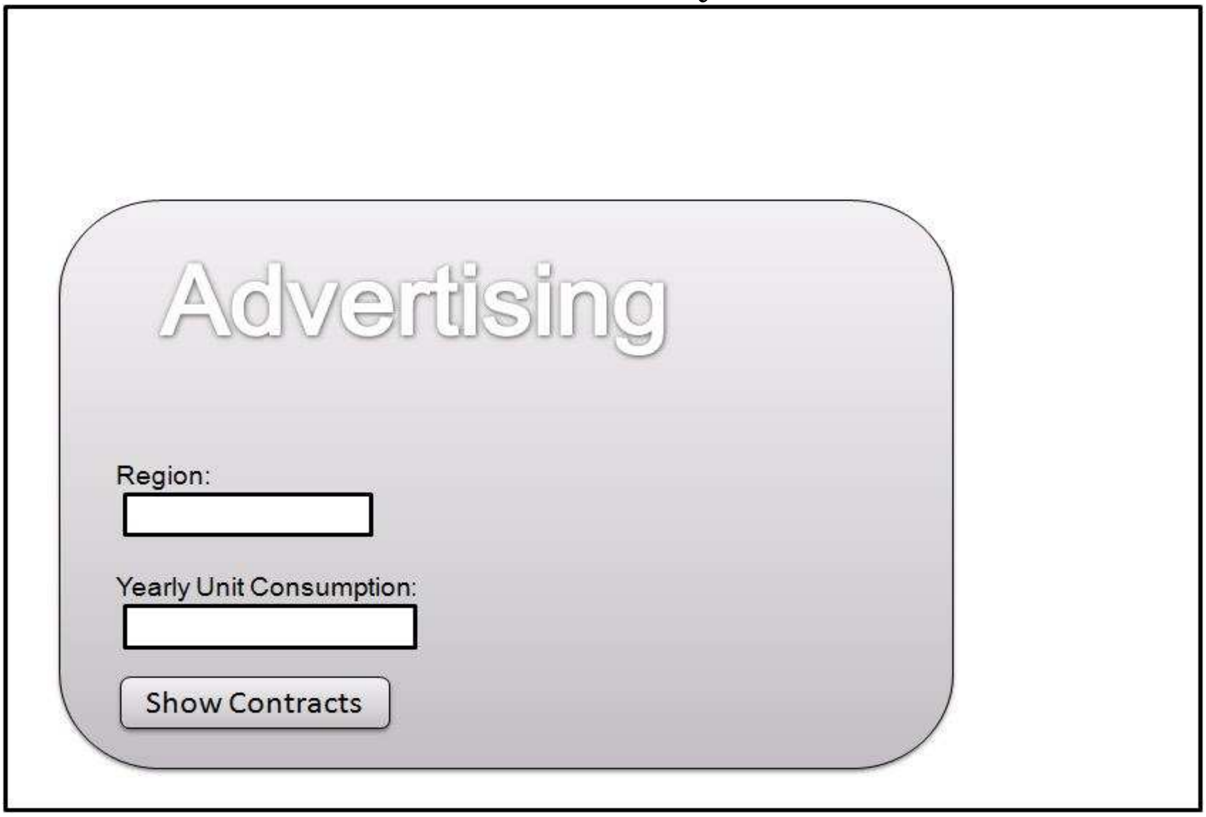
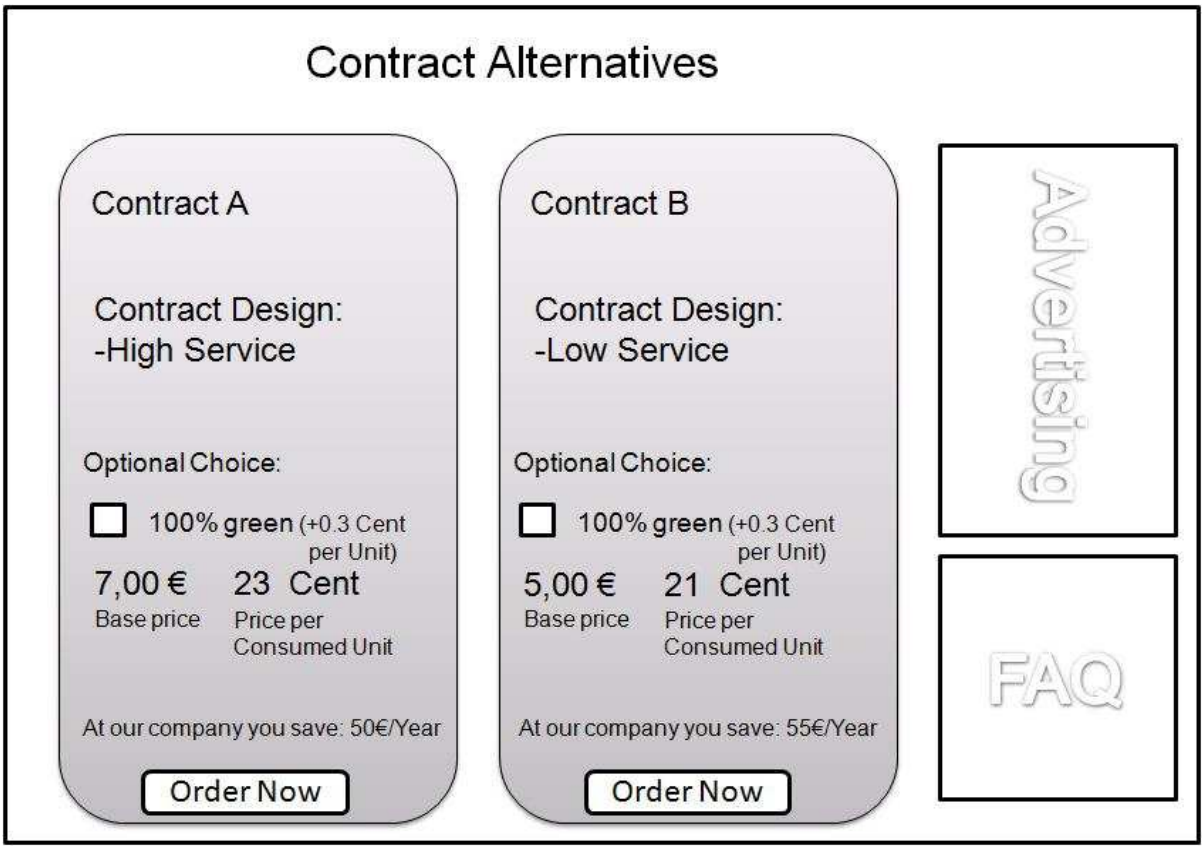
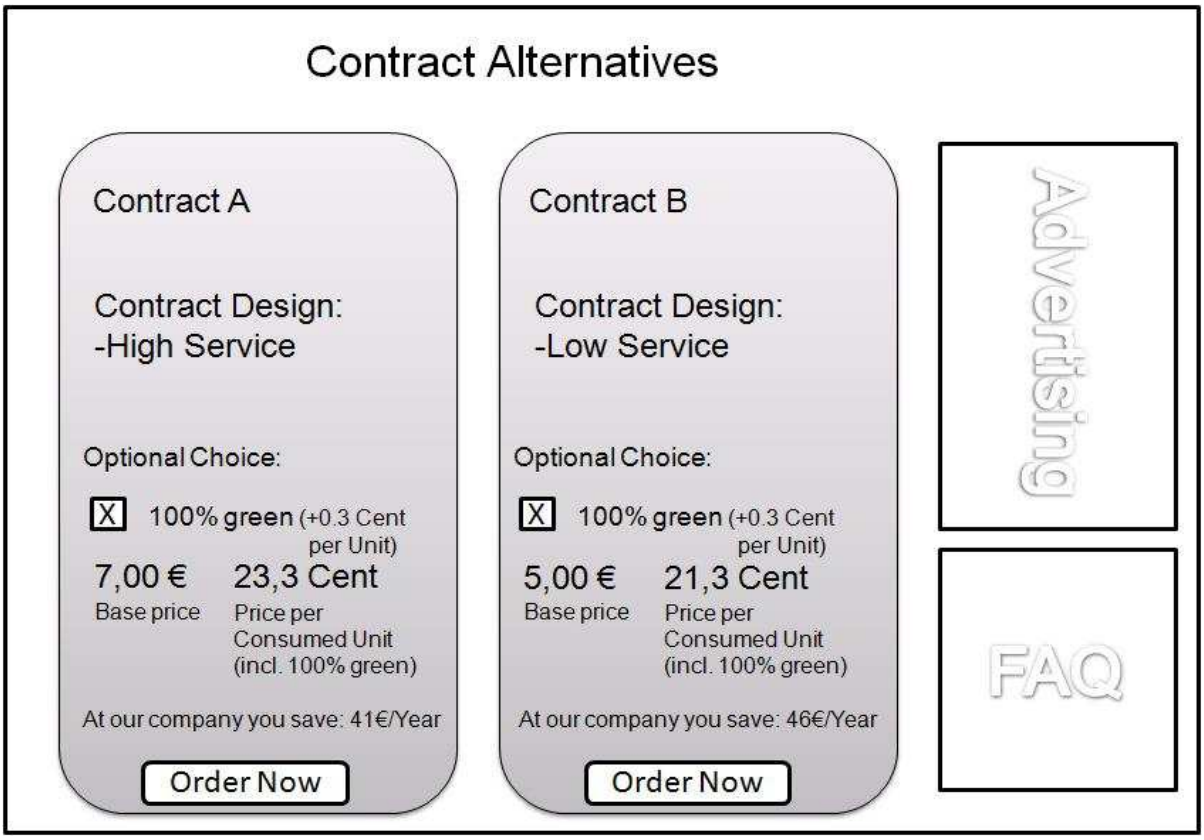


FIGURE A2. WEBSITE SCREEN FOR CONTRACT CHOICE IN CONTROL TREATMENT



Notes: Prices are exemplary

FIGURE A3. WEBSITE SCREEN FOR CONTRACT CHOICE IN GREEN TREATMENT



Notes: Prices are exemplary

TABLE 1.A: SERVICE CONTRACT CHOICE

	(1)	(2)	(3)
Treatment	-0.003	-0.016	-0.02
	(0.04)	(0.04)	(0.04)
Price Difference		0.69***	0.70***
		(0.03)	(0.03)
Base Price			-0.01
			(0.02)
Price / Unit			0.126***
			(0.03)
Yearly Consumption			0.0001***
			(0.00)
Observations	3512	3512	3512
Pseudo R-squared	0.0000	0.0979	0.1122

Notes: probit regressions with green contract choice donations as dependent variable. Standard errors in parentheses. "Treatment" represents a binary variable for the green treatment, "Price Difference" represents the price-per-unit difference between a low-service and a high-service contract, "Contract Type" represents a binary variable indicating whether customer bought the low-service or the high-service contract, "Base Price" represents the base price of the high-service contract, "Price / Unit" represents the price-per-unit of the high-service contract. Level of significance: *p<0.10, **p<0.05, ***p<0.01.

Appendix B. A Model of Procrastination

To formalize the described now or later choice for hyperbolic discounters, we use the simplifications for present-biased preferences developed by Phelps & Pollak (1968) and used by Laibson (1997), O'Donoghue & Rabin (1999) and Carroll et al. (2009) among others. Thereafter, an electricity customer has a long term discount factor δ and a bias for the present represented by β, where $0 < \beta, \delta \leq 1$. Customers utility in period 0,1,2,3…. is u_0, u_1, u_2, u_3,... leading to the simple two-parameter intertemporal-utility model $Ut \ = u_t + \beta(\delta u_{t+1} + \delta^2 u_{t+2} + \cdots)$. When $\beta < 1$ the customer has time inconsistent preferences (is a hyperbolic discounter). Furthermore, if time inconsistent customers are aware of their future self-control problems (= sophisticated), their belief $\hat{\beta}$about β is the actual value. If they are not completely aware about their time inconsistency (= naïve), they hold a belief $\hat{\beta} > \beta$ about their time inconsistency.

Now, we consider the at least ≈ 30% of customers who were nudged by the default. When the nudge is driven by procrastination, it implies that the customers stick with a sub-optimal choice. Formally, we assume that their utility from the default contract is b_d and their utility from the alternative is b_a; with $\Delta b = b_a - b_d > 0$. The costs for switching now are c_0 and the costs for switching later are c_τ, whereas τ is an undetermined period ≥ 1. We assume that an electricity contract runs infinitely or at least that a customer does not know when the contract will end. Hence, the discounted value of the alternative contract is $1/(1-\delta)\Delta b$.[18] The customer faces the following decision problem:

$$U_0 = \begin{cases} -c_0 + \Delta b + \beta \dfrac{\delta}{1-\delta} \Delta b & \textit{switch now} \\ -\beta \delta^\tau c_\tau + \beta \dfrac{\delta^\tau}{1-\delta} \Delta b & \textit{switch in } \tau \end{cases}$$

[18] For hyperbolic discounters the present value of $1/(1-\delta)\Delta b$ is $\Delta b + \beta\delta/(1-\delta)\Delta b$.

A calibration of the model showing when time inconsistent (& naïve) customers procrastinate would be almost exactly the same as the calibration shown in O'Donoghue & Rabin (1999a). We will not conduct an extensive calibration of our model at place. For our needs, it is sufficient when we focus on the impact of the peculiarity of our model, which is the switching cost structure.[19] In most previous papers analyzing hyperbolic discounting $c_\tau = c_0$, but in our model $c_\tau \gg c_0$. As we will see, this difference strongly tightens the possible values of the time inconsistency parameter β.

We consider the choice of completely naïve customers ($\hat{\beta} = 1$). Naïve customers are the customers who are most likely to (infinitely) postpone switching.[20] In each period naïves compare switching now to switching in the next period (=month). Thereby, they wrongly assume that they are time-consistent in the next period. But as they are not, in the next period they postpone switching again.[21] Naïves choose to postpone switching into the next period, if

$$c_0 - \beta\delta c_1 > \Delta b$$

Considering this equation and keeping in mind that $c_\tau \gg c_0$, it becomes clear that in our experiment even naïves only procrastinate for very small values of β and δ. We checked several papers concerned with a calibration of hyperbolic discounting models (e.g. Laibson et al. 2009, Paserman 2008, Skiba & Tobacman 2008). Estimated values strongly differ, but even a combination of the most extreme values for β and δ found in these papers cannot explain our results when $c_\tau \geq 3c_0$ - which is in regard of the actual difference in switching costs a small assumed ${}^{c_\tau}/_{c_0}$ratio.

[19] Furthermore, an exact calibration of a model would be extensive and we would have to estimate values for Δb, c_0 and c_τ, which would be a rather fuzzy task.
[20] Hence, for naïve customers the derived values for β are most relaxed. For time-inconsistent & non naïves β has to be even more extreme for not switching in the current period.
[21] Naïves "wrongly" compare switching in the current period with switching in the next period. It would be "right" if naïves compare switching now with switching never, since the latter is the outcome they finally realize.

It is arguable, whether all assumptions (e.g. infinite contract duration or that customers receive utility Δb already in the period of switching) exactly represent the switch now/later choice of our experiment. However, as long as the standard framework of Phelps & Pollak (1968) is used for formalization, the switching costs structure of our experiment will basically lead to the same extreme value predictions for β and δ.

Part 3

Peer Pressure in Multi Dimensional Work Tasks

Peer Pressure in Multi-Dimensional Work Tasks

Felix Ebeling, Gerlinde Fellner and Johannes Wahlig*

November 15, 2012

Abstract

We study the influence of peer pressure in multi-dimensional work tasks by means of a laboratory experiment. Work dimensions are (perfectly) substitutable and workers face peer pressure in only one work dimension. We find that effort provision increases in the dimension where peer pressure is introduced. However, not all of this increase translates into a productivity gain, since the effect is partly offset by a decrease of effort in the work dimension without peer pressure. This tradeoff is stronger for workers who run behind in the peer pressure dimension. Finally, we analyze the optimal group composition to harness peer pressure. Effort in the dimension of peer pressure and overall productivity are unaffected by group composition, but the effort reduction in the dimension not subject to peer pressure is stronger when workers' skills are highly diverse. While existing literature recommends maximizing worker-groups' skill diversity in one-dimensional work tasks, our results suggest mixing similar workers in multi-dimensional tasks.

JEL Classification: D03, D2, J21

Keywords: Peer Effects, Multi Tasking, Incentives, Laboratory Experiment

---

*Felix Ebeling, University of Cologne, Department of Economics, ebeling@wiso.uni-koeln.de, Gerlinde Fellner, University of Ulm, Institute of Economics, gerlinde.fellner@uni-ulm.de, Johannes Wahlig, University of Cologne, Department of Economics, wahlig@wiso.uni-koeln.de. We thank Bernd Irlenbusch, Sebastian Lotz and Axel Ockenfels for helpful comments. Furthermore, we thank the participants of the ESA European Conference 2012. We also gratefully acknowledge the financial support from the German Science Foundation (through Gottfried Wilhelm Leibniz Price of the DFG, awarded to Axel Ockenfels).

1. Introduction

Lately, several research papers discuss the opportunity to use psychological incentives, such as peer pressure, to increase work productivity. In particular, peer effects might serve as a substitute for explicit monetary incentives. The most parsimonious definition of peer effects at the work place is 'the shift of productivity of individual *i* when productivity of individual *j* changes and all else remains equal' (Falk & Ichino 2006, p.40). Recent empirical research also provides recommendations how to compose workforce to optimally exploit such peer effects (see, for instance, Mas & Moretti 2009, Bandiera et al. 2010).

So far, peer pressure has mainly received attention in one-dimensional work tasks. However, many work environments are characterized by a multiplicity of demands and facets, such as quantity and quality. Peer pressure can, however, by definition be only exerted in work dimensions that are subject to social comparisons, i.e. when monitoring by peers is possible. In multi-dimensional work tasks, however, social monitoring will be limited to work dimensions that are easily observable.

In this paper, we argue that peer pressure does not necessarily unfold strictly positive effects but might also have severe drawbacks. Our research centers on the presence of peer effects in more complex work tasks consisting of multiple dimensions. Especially in work tasks with several (substitutable) dimensions, peer pressure in a single dimension may crowd out effort in other dimensions, which is alike to the theoretical results of Holmstrom and Milgrom (1991) for monetary incentives in multidimensional work tasks.

Based on Holmstrom & Milgrom (1991), we provide a theoretical framework for this argument and put it to test in an experiment. Our results show that psychological incentives in the workplace seem to have similar drawbacks as pecuniary incentives: Incentives in an observable work dimension crowd out effort in a non-observable dimension. Consequently, we recommend a more cautious handling of psychological incentives when considering more complex work tasks. In a further step, we investigate how to organize a possibly heterogeneous workforce into work teams to optimally harness peer effects.

The paper is organized as follows. In section 2, we will present an overview of the related literature. By addressing the question of how peer pressure affects performance in multi-dimensional work tasks, we combine two, by now unlinked, sub-branches of research on incentives, that is, research on peer pressure and research on multidimensional incentive problems. To point out the contribution of our paper, we give a brief overview of the existing literature in these areas. In section 3, a theoretical framework is presented that derives the partially detrimental effects of peer pressure. Section 4 explains the setup of an experiment that tests the theory and presents the main hypotheses. Section 5 presents the results and section 6 concludes with a discussion.

2. Related Literature

2.1 Peer Effects

The analysis of peer effects has a long tradition in psychological and sociological research. The positive effect of the presence of peers and the observability of own behavior on effort and performance is denoted by social facilitation (Zajonc, 1965). In economic research, peer effects are distinguished according to their source. In particular, they might arise due to technological or psychological reasons. In former case, peer effects are purely based on rational considerations (Gould & Winter 2009). In contrast, psychological peer effects are based on phenomena as shame, social pressure or contagious enthusiasm that are represented in individuals' preferences. Due to our research focus, in the following we only consider literature about psychological peer effects.

Economists have only recently started to look at psychological peer effects in a variety of domains, like education (Sacerdote 2001, Zimmerman 2003, Graham 2008), sports (Guryan et al. 2009), crime (Glaeser et al. 1996), public good contribution (Falk et al. 2010) and classical work environments.[1] What this literature has in common is the focus on externalities of individual behavior or abilities on others. Researchers are interested in the question whether such externalities exist, and if so, whether efforts by different individuals are complements or substitutes. An illustrative example is provided by Sacerdote (2001). He uses the random assignment of college dorm roommates to measure peer effects in educational outcomes. Sacerdote finds a positive influence of roommates on each other: an increase in the roommate's first year average grade results in a significant increase of student's first year average grade.

We are concerned with psychological peer effects at the workplace. The existing research on peer effects at the workplace presents mixed results. More specifically, the prevalence of peer effects at the workplace seems to depend on the payment regime. While there is clear evidence for peer effects under fixed wages regime (Falk & Ichino 2006, Mas & Moretti 2009), the picture is less clear for other payment regimes, such as piece-rates. Guyran et al. (2009) suggests that this might be due to a crowding out of psychological incentives by monetary incentives: the more salient the monetary incentive, the less important are psychological incentives for performance. Their own findings from professional golf tournaments as well as laboratory findings by Eriksson et al. (2009) support this view. Both articles do not find evidence for peer effects in

[1] Furthermore we are aware of research on peer effects in welfare participation (Bertrand et al. 2000), unemployment insurance take-up (Kroft 2008) and retirement planning (Duflo & Saez 2003).

tournaments or piece-rate regimes.[2] However, this crowding-out rule might only be valid for individual incentives, since in case of revenue sharing between workers, there is a clear rationale for crowding-in (Kandel & Lazear 1992): the more own payoff depends on the performance of the co-worker, the stronger the peer pressure. Empirical evidence by Chan et al. (2010) supports this view.

We focus on a work environment of fixed wages, which applies to a large variety of jobs. Fixed wages are low powered individual monetary incentives. Intuitively appealing and corroborated by previous research results, such environments leave space for psychological incentives. To the best of our knowledge, there are two existing studies that are closely related to ours in that also analyze peer effects in fixed wage work environment. Falk & Ichino (2006) conducted a controlled field experiment where participants, consisting of high school students, received a fixed payment for stuffing letters into envelopes. The authors identify significant peer effects that increase productivity. Furthermore, in their study less productive workers were more affected by peer effects. They conclude that '"bad apples" gain quality from "good apples", but do not damage the latter one' (Falk & Ichino, 2006, p. 54). Another similar study was done by Mas & Moretti (2009) who investigated peer effects for supermarket cashiers who received fixed wages. They identified positive productivity spillovers of faster co-workers on slower ones and thus derived the recommendation to mix workers with a maximum of skill diversity to increase productivity. However, both studies focus on one-dimensional work tasks, or at least measure output only in the work dimension that is subject to peer pressure. In contrast, we focus explicitly on a work task that consists of multiple dimensions where peer monitoring and thus peer pressure can take effect only

[2] In contrast, Bandiera et al. (2010) do find evidence for peer effects in a piece-rate environment. However, they are less general and only observable for co-workers who are socially tied to each other.

in one dimension. This approach links research on peer pressure to multidimensional incentive problems.

2.2 Multidimensional Incentive Problems

Holmstrom & Milgrom (1991) were first to theoretically analyze the optimal payment scheme for workers in multidimensional work tasks.[3] They provide a rationale for a lack of incentives in contracts even if principals are able to monitor certain work dimensions. As long as there are some work dimensions that are excluded from monitoring, explicit incentives on the monitored dimensions might crowd out effort in non-monitored dimensions. Similarly, Pendergast (1999) emphasizes "dysfunctional behavioral responses" that arise when monetary incentives meet missing permanent holistic measures of the workers' contribution (p.8). This theory is supported by empirical evidence. For instance, Paarsch & Shearer (2000) compare quantity and quality data of workers planting trees under different payment regimes. When workers are paid according to the number of trees they have planted, quantity clearly increases in comparison to fixed wages. However, this productivity increase is partly offset by a reduction in the quality of work. Johnson et al. (2012) came to similar results when scrutinizing the effects of a payment system change for bus drivers in Chile. When wage system changes from fixed-wage system to per-passenger system, waiting time reduced by 13%, but accidents increased by 67%. The results of both papers are very similar to what we observe in our experiment, even though we do not change monetary incentives, but introduce peer pressure in one work dimension.

Let us now turn to the theoretical arguments on how peer pressure might affect effort choice in a multidimensional work task.

[3] Similarly, work might consist of several tasks.

3. A Simple Model of Peer Pressure in a Multidimensional Work Task

The primary goal of this section is to give some formal arguments on how workers' effort choice in a multidimensional work task reacts to peer pressure. In line with the literature, we assume that peer pressure arises when workers can compare themselves to their peers in terms of output in a specific work dimension. However, in a work task that consists of multiple dimensions, workers might be able to observe others' output or effort in some, but not in all dimensions. In the simple case of a work task that has a quantity and a quality dimension, it is plausible to assume that quantity output can be more easily observed and monitored than quality output. Thus, peer pressure is present in one work dimension, but not in others, but might still have spill-over effects to effort in work-dimensions that are not subject to peer monitoring.[4]

The baseline of our model is similar to Holmstrom & Milgrom (1991). We consider a work task that consists of *n* dimensions and is paid with a fixed wage *F*. A worker makes the choice of a vector of efforts $\vec{t} = (t_1, \dots, t_n)$ at strictly convex personal costs $C(t)$. We shall suppose that effort in the various work dimensions is perfectly substitutable in the agent's cost function, hence$C(t_1 + \cdots + t_n)$.Contrary to many other models, we shall *not* suppose that all work is unpleasant. A worker on the job may take pleasure in working up to some limit. Incentives are only required to encourage work beyond that limit. Formally, we assume that there is some $t > 0$ such that $C'(t) \leq 0$ for$t \leq \bar{t}$and $C(\bar{t}) = 0$.[5] From the F.O.C. of our utility function $U = F - C(t_1 + \cdots + t_n)$ follows an effort provision of $\vec{t}^* = t_1{}^* + \cdots + t_n{}^* = \bar{t}$.

[4] Note that in the model, we speak of 'effort' provided in the different work dimensions, where in fact, only 'output' is observable. If we assume that work output in a specific dimension is an increasing, strictly monotonic function of effort in that dimension, the argument made for effort changes directly translates to changes in observed output.
[5] The cost function is U-shaped with its minimum at $\bar{t}$.

However, contrary to Holmstrom & Milgrom (1991) we do not introduce a monetary incentive in our model, but a psychological incentive that arises from peer observability in dimension i. Formally, peer pressure is modeled as in Kandel & Lazear (1992) by introducing a negative term $P(\cdot)$in the utility function. Thereby, peer pressure decreases in provided effort in dimension i, i.e.,$\partial P/_{\partial t_i} < 0$. This changes workers' utility function to $U = F - C(t_1 + \cdots + t_n) - P(t_i)$.

Deriving the F.O.C. for this new utility function for t_i and t_{-i}, where $-i \neq i$, reveals a change of effort provision. Accumulated effort rises to $\vec{t}^{**} > \bar{t}$, with $t_i^{**} = \vec{t}^{**}$ and $t_{-i}^{**} = 0$. Hence, our model providesa corner solution. Effort in non-peer-observable dimensions is entirely crowded out, while effort in the peer dimension rises even above the previous level of accumulated effort.

4. Experimental Design and Hypotheses

The advantages of starting the investigation of peer pressure in multidimensional work tasks in a lab setting are twofold. First, it allows creating a work task that consists of not more and not less than two equally important dimensions, one that is observable and one that is not. As we ill show further on, the observability of only one dimension represents a realistic work situation. Second, the anonymous setting in the lab allows control over other factors that might serve as incentives to exert effort, like firm (group) affiliation or close personal relations between co-workers.

The work task that we implemented in the experiment required participants to describe pictures on the computer screen during a period of 25 minutes. For that, subjects earned a fixed wage of 5 Euros.[6] The pictures were selected from a pool of more than 200 million pictures under a creative common license from flikr.com which can be used

[6] Before the 25 minutes started, there was a learning phase of three pictures. In this way participants got to know the handling of describing pictures.

costless for non-commercial purposes (for a typical picture, see Appendix A). Describing digital pictures is a task that is easily executable by humans, but difficult for machines. Nowadays such “Human Intelligence Tasks” are typically done in online micro-labor markets as Amazon’s Mechanical Turk (Horton 2010) or Indian data entry firms (Kaur et al. 2010). Thus, this task represents a realistic work setting. Participants were asked to describe each picture with at least one label and at most ten labels. Within this range, participants could freely decide how many labels they wanted to assign to a pictures. We did not give any advice about the appropriate number of labels, but we instructed them that the descriptions are needed for research and teaching purposes. When subjects considered a picture as sufficiently described, they could press a button to get to the next picture on their screen. This work task is ideally suited for our purposes, since it clearly comprises two work dimensions that are clearly measurable: quantity, as the number of pictures, and quality as the (average) number of labels assigned to a picture.

To study the effect of peer pressure on work output, we let subjects experience the same work task in two subsequent phases of 25 minutes each but introduce two different between-subjects treatments, the peer treatment and the control treatment. For reasons explained below, we refer to the first work phase of each treatment as the *no-feedback work phase* and the second one as the *feedback work phase*.

In the first work phase, the control and the peer treatment did not differ. All participants worked individually and received no feedback about the work of other participants, hence *no-feedback work* phase. This ensures that peer pressure is absent in this phase, since subjects could neither observe (nor be observed by) others. At the end of the first work phase, subjects were asked to subjectively asses how hard they worked. More precisely, subjects had to give answers to the following three questions on a seven point scale: How much effort did you exert? How stressed do you feel now? How exhausted did you get? A value 1 corresponded to "not at all" and a value of 7 to "very strong".[7]

In the second work phase, subjects had to complete exactly the same task as before, i.e. describing (different) pictures for a period of 25 minutes for a fixed wage of 5 Euros, but now they worked under a *feedback*-condition. The peer and control treatment differed with respect to the kind of feedback they received.

In the *peer treatment* workers received feedback information about their own current quantity output (i.e., the number of pictures they have described so far in this work phase), as well as the quantity output by another participant, their co-worker. More precisely, the information about the number of pictures described so far in this phase was exchanged in real time between two matched workers.[8] During the whole work phase, subjects could see in the upper right corner of the computer screen the number of pictures that they described so far in this phase, as well as the number of pictures that a co-worker had described so far. As soon as one of the workers finished a picture, the picture counter increased by one. Thus, subjects could observe others and compare themselves only with respect to the quantity dimension. It was not revealed how many labels the co-worker has assigned to the pictures.

Rather particular for a laboratory experiment, the co-worker team, within which information was exchanged, got to know each other before starting the work. This was done after instructing participants for this phase by letting them stand up and allocating co-workers to laboratory seats next to each other. However, participants were not allowed to talk to each other. We decided to implement this unusual design feature basically for two reasons. First, workplaces are almost never completely anonymous.

[7] These questions are similar to the questions Falk & Dohmen (2011) used to find out how hard participants worked in their experiment.
[8] Hence, picture count starts at zero in the feedback-work phase.

Hence, our setting becomes more realistic and improves external validity of our results. Second, in previous studies about peer pressure (especially Falk & Ichino 2006, Mas & Moretti 2009, and Bandiera et al. 2010) workers also know each other. Due to non anonymity, our results are more comparable with theirs.

In the *control treatment,* subjects' feedback consisted only of their own quantity. During the whole phase subjects could see in the upper right corner of the computer screen how many pictures they have described so far. Similarly to the first work phase, subjects described pictures without any information exchange with other participants, but only got feedback, on their own quantity.[9] To keep the *control treatment* as similar as possible to the *peer treatment*, individuals were randomly reseated before the second work phase started.

In both treatments, subjects had to again submit a self-assessment on how hard they worked and on how stressed they felt at the end of the feedback work phase. Altogether, 181 subjects participated in the experiment that was conducted computerized with zTree (Fischbacher, 2007) in the Cologne Laboratory for Economic Research. 122 took part in the *peer treatment* and 59 in the *control treatment*. Subjects were students of the University of Cologne, 93 males and 88 females, from a wide variety of fields of study and were recruited with ORSEE (Greiner 2004). One session lasted between 75 and 85 minutes. The average payoff was 14.1 Euro.

Clearly, peer pressure works via a process of social comparison. To learn whether social preferences moderate the effect of peer pressure, we elicited an indicator for social preferences by using the social value orientation (SVO) slider measure (Murphy et al.

[9] In one session of the control treatment participants received no feedback in the second work phase. This was done to reveal possible experimenter demand effects: subjects might perceive feedback information about a specific work dimension as a hint about the experimenter's preferences for this particular dimension. However, there was no difference across control sessions with and without feedback about own quantity. Hence, we decided to pool the data for the control treatment and stick to the term "feedback work phase".

2011) at the beginning of the experiment. Subjects were asked to complete the six primary SVO Slider items (see Appendix B), that allows a classification of preferences from altruistic to egoistic / competitive. In each of the six items, subjects had to choose an allocation that assigns money to themselves and another, randomly chosen participant in the experiment. After completing this task, a dice was rolled to select one of the items to become payment relevant at the end of the experiment. The dice roll was done by one randomly chosen participant and visible to all participants. Participants within a session were randomly matched into pairs to allocate the payoffs according to the choice in the selected item.[10]

In light of the theory on peer effects in multidimensional tasks, presented in the previous section, we derive three main hypotheses.

First, peer pressure introduced by the exchange of information on quantity between two workers will lead to an increase of quantity output. However, as pure learning effects are expected to increase output as well from the first to the second work phase, we compare the difference in the quantity (i.e., the number of pictures) between the no-feedback to the feedback work phase across the peer and control treatment.

> ***Hypothesis 1.*** *In the peer treatment, output in the quantity dimension will increase more from first to the second work phase than in the control treatment.*

The effort increase in one work dimension due to peer pressure will crowd out effort, and thus output, in the other work dimension that is not observable. We therefore state

[10] Thus, every participant received a payoff from the SVO slider task as an allocator as well as a receiver.

__Hypothesis 2.__ In the peer treatment, output in the quality dimension will decrease more from the first to the second work phase than in the control treatment. In particular, under the assumptions made on individuals' effort cost function, output is reduced to the minimum, i.e. one label per picture.

Still, the effect of peer pressure on overall output that can be measured by the aggregate number of labels assigned is predicted to be positive, since the quantity increase due to peer pressure over-compensates the quality decrease.

__Hypothesis 3.__In the peer treatment, productivity will increasemore from the first to the second work phase than in the control treatment.

5. Results

We split up our main findings into three parts. In section 5.1, we investigate the general effects of peer pressure on workers' output and productivity. In section 5.2, we report how peer pressure individually affects different types of workers. In section 5.3, we analyze our data with regard to the optimal exploitation of peer effects.

5.1 Effort Provision and Peer Pressure

Our first result concerns the effect of peer pressure on effort provision in the two different work-dimensions, quantity and quality. In each work phase quantity is measured by the number of described pictures, whereas quality is measured by average number of labels per pictures. Furthermore we are interested in the impact of peer pressure on productivity, which is measured by the aggregate number of labels (or output in the quantity times output in the quality dimension).

In the no-feedback phase, subjects described on average 30.3 pictures (SD=1.77) with 5.5 labels (SD=0.26) in the control treatment and 33.1 pictures (SD=1.68) with 5.3 labels (SD=0.15) in the peer treatment. Hence, as to be expected, Mann Whitney U-test show no difference between quantity as well as quality output in the first work phase (p=.61 for pictures and p=.66 for labels). Note also, that although the fixed wage provides no monetary incentives to work at all, subjects obviously exhibit significant work effort which lends strong supports to the theoretical assumption that not all work is unpleasant.

Figure 1 summarizes the changes of quantity, average quality and productivity between the no-feedback and feedback work-phase for both the control and the peer treatment. In the control treatment, quantity increases by an average of 10.4 pictures, while in the peer treatment the increase was 18.2 pictures. At the same time, quality decreases, in the control treatment by an average of 0.32 labels and in the peer treatment by 0.75 labels per picture. Nevertheless, overall productivity increases on average by 31.3 labels in the control treatment and 46.4 labels in the peer treatment[11]. Note that also in the control treatment all dimensions significantly change, which can be considered the result of a learning effect. They represent a base level to evaluate the impact of peer pressure. Results from Mann-Whitney U test reveal that the differences in the peer treatment always exceed the differences in the control treatment, for the quantity and quality dimension (p<.01 and p=.04) as well as for the overall productivity (p<.01).[12] These first results confirm our expectations: peer pressure leads to an increase in quantity (supporting Hypothesis 1), and a decrease in quality, although we do not observe a complete crowding out of quality, as Hypothesis 2 suggests. On average, workers in the peer treatment assign an average of 4.5 labels (SD=0.16) to a picture in the feedback work phase, which is significantly more than the minimum of 1 (95% confidence interval is [4.22;4.87]). Hence, effort dimensions seem to be substitutes and

[11] All differences are significantly different from 0 (according to 95% confidence intervals around the mean observed differences).
[12] Cumulative distribution functions of the differences in quality, quantity and productivity between the no feedback and feedback work phase can be found in Appendix B.

quantity crowds out quality. Still overall productivity increases confirming the expectation formulated in Hypothesis 3.

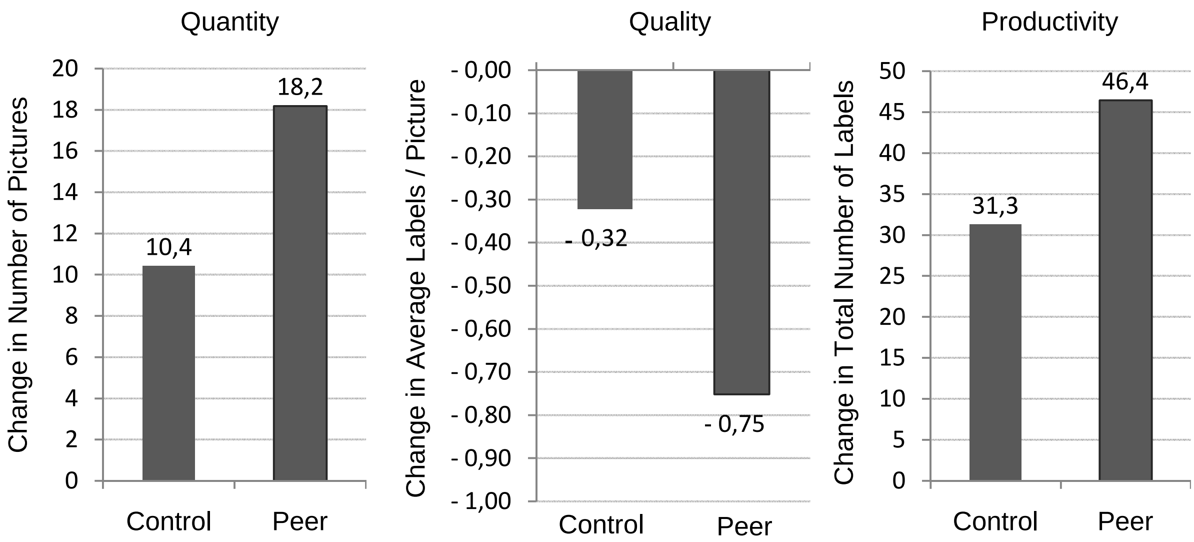


FIGURE 1. MEAN CHANGE FROM NO-FEEDBACK TO FEEDBACK PHASE

Notes: The figure shows, for both treatments, the mean change of quantity (=number of described pictures), quality (= average labels per picture) and productivity (=total number of labels) from the no-feedback to the feedback work phase. In case of quality it reports the unweighted mean of subjects' mean number of labels.

To substantiate our first results and control for (ex-ante) heterogeneity of workers we additionally conduct regression analyses, where the dependent variables are the differences in quantity, quality and productivity between the no-feedback and feedback-phase and the main independent variable is a dummy for the peer treatment. Moreover, we add the absolute amount of quality, quantity and productivity in the no-feedback phase as independent variables in specifications (2) to (4). Tables 1 to 3 present the results that confirm a stronger increase of quantity and productivity and a stronger decrease of quality in the peer treatment. Furthermore, regressions reveal the importance of ex-ante differences between workers. Independent of the treatment, ex ante more productive workers increase quantity more strongly, while workers who provide ex ante more quality stronger decrease quality, but also stronger increase productivity.

TABLE 1. QUANTITY DIFFERENCE BETWEEN TREATMENTS

	(1)	(2)	(3)	(4)
Peer treatment	7.76*** (2.63)	7.72*** (2.64)	7.47*** (2.57)	7.21*** (2.46)
Quality-NF		-0.29 (0.79)		
Quantity-NF			0.11 (0.09)	
Productivity-NF				0.08*** (0.02)
Observations	181	181	181	181
R-squared	0.0477	0.0487	0.0597	0.1155

Notes: OLS Regressions with "difference in number of pictures between feedback and no-feedback phase" as dependent variable. Robust standard errors in square brackets. Standard errors are clustered on groups. A group is a co-worker team in the peer treatment and a single person in the control treatment. Level of significance: *p<0.1, **p<0.05, ***p<0.01. "Peer" represents binary variable for peer treatment, "Quality-NF" represents variable for average number of labels per picture in the no-feedback work phase, "Quantity-NF" represents variable for number of picture in the no-feedback work phase, "Productivity-NF" represents variable for total number of labels in the no-feedback work phase

TABLE 2. QUALITY DIFFERENCE BETWEEN TREATMENTS

	(1)	(2)	(3)	(4)
Peer treatment	-0.43** (0.18)	-0.45*** (0.17)	-0.45** (0.17)	-0.42** (0.17)
Quality-NF		-0.14*** (0.05)		
Quantity-NF			0.01 (0.00)	
Productivity-NF				0.00 (0.00)
Observations	181	181	181	181
R-squared	0.0327	0.0826	0.0438	0.0386

Notes: OLS Regressions with "difference in average number of labels per picture between feedback and no-feedback phase" as dependent variable. Robust standard errors in square brackets. Standard errors are clustered on groups. A group is a co-worker team in the peer treatment and a single person in the control treatment. Level of significance: *p<0.1, **p<0.05, ***p<0.01.

TABLE 3. PRODUCTIVITY DIFFERENCE BETWEEN TREATMENTS

	(1)	(2)	(3)	(4)
Peer treatment	15.17**	16.27***	16.72**	14.55***
	(5.59)	(5.64)	(5.69)	(5.58)
Quality-NF		6.57***		
		(1.85)		
Quantity-NF			-0.56*	
			(0.26)	
Productivity-NF				0.09
				(0.07)
Observations	181	181	181	181
R-squared	0.0295	0.1105	0.0826	0.0437

Notes: OLS Regressions with "difference in total number of labels between feedback and no-feedback phase" as dependent variable. Robust standard errors in square brackets. Standard errors are clustered on groups. A group is a co-worker team in the peer treatment and a single person in the control treatment. Level of significance: *p<0.1, **p<0.05, ***p<0.01.

5.2 Individual Adjustment to Peer Pressure

As we have already established how peer pressure affects behavior on aggregate, we want to gain a more precise picture on how workers of different abilities adjust to peer pressure. Since we cannot observe ability directly, we proxy it by the quantity provided in the no-feedback work phase. Thereby, our experimental design allows measuring the influence of absolute and relative output provision in the no-feedback work phase on adjustment processes in the feedback phase. An *absolutely* high output implies that a worker's output (in the no-feedback work phase) is high in comparison to all other workers' output (in the no-feedback work phase), a *relatively* high output implies his output (in the no-feedback work phase) is high in comparison with the co-worker's output (in the no-feedback work phase) he is paired with later on. Thereby, we focus on relative differences in the quantity dimension. The idea is that a worker who provides less (more) quantity in the no-feedback work phase will soon run behind (ahead) her co-worker in the feedback phase. Peer pressure then might affect the behavior of the different workers differently. As the subsequent analysis shows, the ex ante relative differences in quantity are indeed the most important variable for adjustment processes.

Figure 2 visualizes the reaction of different workers to peer pressure. In the three graphs, the abscissa represents the absolute difference in quantity between the worker and her co-worker in the no feedback phase (before they were matched). For example, a worker with a value of 40 described 40 pictures more in the no-feedback phase than her co-worker. We label this difference as "Heterogeneity" since it describes how workers differ in their quantity provision before the exchange of peer information. The ordinate again represents the respective absolute differences in quantity, quality and productivity between the no-feedback and the feedback phase.

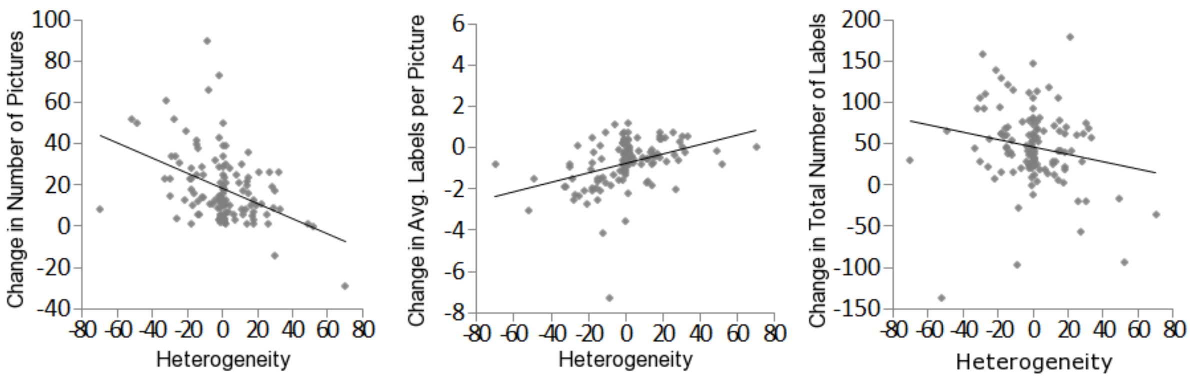


FIGURE 2. REACTION OF DIFFERENT WORKERS TO PEER PRESSURE

Notes: The figure shows the quantity, quality and productivity change for workers who provide more or less quantity in the no-feedback phase than their later co-worker. The solid line represents the least square regression slope.

The graphs in Figure 2 suggest a correlation of between heterogeneity and quantity, quality and productivity, which is confirmed by Spearman Rank Correlation tests. Quantity change is significantly negatively correlated to heterogeneity (spearman's rho=-0.37, p<.01) whereas quality is significantly positively correlated (spearman's rho=0.47, p<.01). This means that workers who are worse than their (later) co-workers in quantity provision increase their quantity under peer pressure more than workers who are better, but this is achieved at the expense of decreasing quality significantly more. Concerning productivity, the correlation with heterogeneity is significantly negative (spearman's rho=-0.18, p<.05). This means that workers who were ahead their co-worker increase their productivity less strongly under peer pressure.

We again substantiate our findings by regression analyses, while at the same time controlling for absolute levels of individual quantity, quality and productivity in the no-feedback phase.[13] Moreover, we control for heterogeneity in social preferences by adding the score in the social value orientation (SVO) slider measure as independent variable. Tables 4 to 6 present the results. While the impact of heterogeneity on quantity and quality is highly significant across specifications, the influence of heterogeneity on productivity becomes (partly) insignificant, in particular when controlling for the initial quantity and quality. The less significant correlation for productivity is intuitively appealing, as opposing correlations of quantity and quality with respect to heterogeneity neutralize the correlation for productivity.

Comparing this finding with existing evidence on adjustment processes (due to peer pressure), our results represent a refinement. For example, Mas & Moretti (2009) reveal a stronger influence of peer pressure on workers with below average productivity. However, in their paper productivity is the observable work dimension. Our results suggests that when work dimensions can be disentangled, productivity is no longer the most important variable for the adjustment processes, but rather the dimension in which co-workers compare themselves.

[13] Furthermore, controlling for these variables avoids conclusions that emerge due to "regressions to the mean".

TABLE 4. INFLUENCE OF HETEROGENEITY ON QUANTITY CHANGE IN THE PEER TREATMENT

	(1)	(2)	(3)	(4)	(5)
Heterogeneity	-0.36***	-0.38***	-0.43***	-0.53***	-0.41***
	(0.04)	(0.04)	(0.06)	(0.07)	(0.05)
SVO-slider-value		-0.33***	-0.30***	-0.32***	-0.32***
		(0.11)	(0.10)	(0.10)	(0.10)
Quality-NF			-1.57		
			(1.22)		
Quantity-NF				0.32***	
				(0.11)	
Productivity-NF					0.08***
					(0.24)
Observations	122	122	122	122	122
R-squared	0.1741	0.2241	0.2448	0.3113	0.3047

Notes: OLS Regressions with "difference in number of pictures between feedback and no-feedback phase" as dependent variable. Robust standard errors in brackets. Standard errors are clustered on groups. A group is a co-worker team in the peer treatment and a single person in the control treatment. Level of significance: *p<0.1, **p<0.05, ***p<0.01. "Heterogeneity" represents the difference between pictures described by the worker and her co-workers in the no feedback phase, "SVO-slider-value" represents the worker's SVO-slider score.

Another noteworthy observation in Table 4 is the significant relation between the SVO-slider-score and the quantity change due to peer pressure. Note that a smaller SVO-slider score indicates a more egoistic / competitive attitude. We find that the more competitive an individual is, the stronger she increases the quantity under peer pressure. Hence, competitive preferences seem to be a driver of peer effects. This fits nicely into previous findings about explanations for peer pressure, as Mas & Moretti (2009) already identified social pressure as a driver of peer effects. Competitiveness and the reaction to social pressure are both part of social comparison processes (Festinger 1954) and closely related.

TABLE 5. INFLUENCE OF HETEROGENEITY ON QUALITY CHANGE IN THE PEER TREATMENT

	(1)	(2)	(3)	(4)	(5)
Heterogeneity	0.02***	0.02***	0.02***	0.03***	0.02***
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
SVO-slider-value		0.01	0.01	0.01	0.01
		(0.01)	(0.01)	(0.01)	(0.01)
Quality-NF			0.09		
			(0.08)		
Quantity-NF				-0.01	
				(0.01)	
Productivity-NF					0.00
					(0.01)
Observations	122	122	122	122	122
R-squared	0.15151	0.1547	0.1688	0.1649	0.1691

Notes: OLS Regressions with "difference in average number of labels per picture between feedback and no-feedback phase" as dependent variable. Robust standard errors in brackets. Standard errors are clustered on groups. A group is a co-worker team in the peer treatment and a single person in the control treatment. Level of significance: *p<0.1, **p<0.05, ***p<0.01.

TABLE 6. INFLUENCE OF HETEROGENEITY ON PRODUCTIVITY CHANGE IN THE PEER TREATMENT

	(1)	(2)	(3)	(4)	(5)
Heterogeneity	-0.45**	-0.45**	-0.22	-0.11	-0.48**
	(0.17)	(0.18)	(0.22)	(0.39)	(0.21)
SVO-slider-value		0.06	-0.07	0.03	0.07
		(0.38)	(0.35)	(0.37)	(0.37)
Quality-NF			7.17**		
			(3.37)		
Quantity-NF				-0.69	
				(0.51)	
Productivity-NF					0.07
					(0.11)
Observations	122	122	122	122	122
R-squared	0.0362	0.0364	0.0946	0.0913	0.0432

Notes: OLS Regressions with "difference in total number of labels between feedback and no-feedback phase" as dependent variable. Robust standard errors in brackets. Standard errors are clustered on groups. A group is a co-worker team in the peer treatment and a single person in the control treatment. Level of significance: *p<0.1, **p<0.05, ***p<0.01.

When thinking about possible adaptation processes of workers to peer pressure it is interesting to consider the self-assessment and self-reported stress level. Does an increase in quantity and productivity due to peer pressure also lead to an increase in exhaustion and stress? We therefore turn to the self-evaluation of participants which they had to complete after each work phase. Table 7 summarizes how participants' self-evaluation changed from the no-feedback to the feedback work phase in both treatments.[14] Mann-Whitney-U tests detect no significant difference between treatments with respect to changes of subjective effort (p=0.83), stress (p=.63) and

[14] Recall that answers were given on a seven point scale.

exhaustion(p=.33).[15] This is remarkable, as one might expect higher values in the peer treatment due to the significantly stronger increase in productivity.

TABLE 7. MEAN CHANGE OF SUBJECTIVE EVALUATION

	Control	Peer
Effort	-0.20	-0.15
Stress	0.57	0.39
Exhaustion	0.61	0.90

5.3 Optimal Group Composition

In the previous section we only analyzed how individual behavior adjusts to peer pressure. For employers, it is of practical relevance how to compose co-workers to harness theses individual adjustment processes. As pointed out by Mas & Moretti (2009), an optimal group composition in terms of exploitation of peer pressure may decrease labor costs for a constant productivity level. In our setup, labor costs are constant, but for given labor costs, our experiment design allows to precisely investigate whether peer pressure has more positive effects in co-worker teams that are rather heterogeneous or homogeneous. To that end, we split co-worker teams in our peer treatment into two groups according to their absolute value of the "Heterogeneity" measure, i.e. the difference in quantity by the two co-workers in the no-feedback phase.[16] More specifically, the "homogenous" group (n=46) contains all workers in the 95% confidence interval around "Heterogeneity" mean of 0 which is [-3.5;+3.5]. The other, "heterogenous" group (n=76) thus contains individuals that rather dissimilar in their ex ante quantity provision.

[15] Additionally, ordered probit regressions to control for individual characteristics did not reveal any differences between treatments.

[16] We focus on the "Heterogeneity" measure as the previous section showed the prevalent importance of this variable on adjustments.

Figure 3 summarizes the changes of average quantity, quality and productivity between the no-feedback and feedback work-phase for heterogeneous as well as homogeneous group members.In the homogeneous groups, quantity increases by an average of 16.2 pictures, while in the heterogeneous groups the increase was 19.4 pictures. At the same time, quality decreases, in homogeneous groups by an average of 0.45 labels and in the heterogeneous groups by 0.93 labels per picture. Overall productivity increases on average by 48.9 labels in the homogeneous groupsand 45 labels in the heterogeneous groups. Results from Mann-Whitney U test reveal a significant difference between homogeneous and heterogeneous groups in terms of quality decrease ($p<.05$), while there is no significant difference for quantity and productivity ($p=0.2$ and $p=0.98$, respectively).

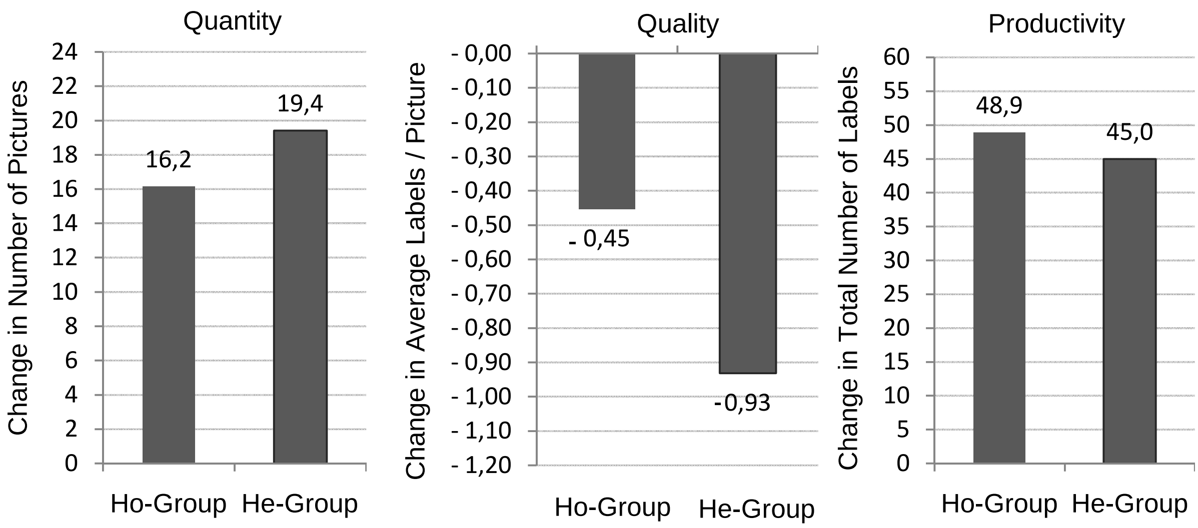


FIGURE 3. MEAN CHANGE FROM NO-FEEDBACK TO FEEDBACK PHASE FOR GROUPS

Notes: The figure shows, for both types of groups, the mean change of quantity (=number of described pictures), quality (= average labels per picture) and productivity (=total number of labels) from the no-feedback to the feedback work phase. In case of quality it reports the unweighted mean of subjects' mean number of labels.

We again substantiate our findings by regression analyses. The regressions in Tables 8 to 10 present the analysis of how team composition affects the change in quantity, quality and productivity in the peer treatment.[17] The independent variable "Hetero-Group" is a dummy variable that indicates that a worker is part of a heterogeneous team. Results confirm the significant effect of group composition on quality. Again, quantity and productivity seems to be unaffected by group composition. To check the robustness of these results, we extended the interval for defining the homogeneous groups up to an interval of [-10;+10] for the Heterogeneity variable. Results do not change. Hence, our data suggest that groups should be composed of rather homogenous workers to optimally harness peer effects. Although overall productivity does not seem to be affected much by the co-worker composition, the crowding out of quality by peer pressure is higher in heterogeneous teams. The reason for these results can easily be seen in Figures 4 to 6 in Appendix C. While Figure 2 in the previous chapter already indicates a stronger adjustment of individuals running behind, the Figure5 in Appendix C shows that this tendency is disproportionately high for the quality dimension, which is responsible for the significant lower quality output in heterogeneous groups.

[17] In the previous regressions in Tables 1 to 6 we did not include "difference in average number of labels per picture" as independent variable when measuring effects on "difference in number of pictures" and vice versa. We did so, because these regressions include participants from the control treatment. Since the independent variables "control", "difference in average number of labels per picture" and "difference in number of pictures" are highly correlated we had to exclude the latter ones to avoid multicollinearity. Since the regressions in Table 8 to 10 do not contain participants from the control treatment, we include these variables.

TABLE 8. INFLUENCE OF GROUP COMPOSITION ON QUANTITY

	(1)	(2)	(3)	(4)	(5)
Hetero-Group	-1.81	-0.93	-1.92	-2.03	-1.69
	(2.69)	(2.55)	(2.37)	(2.32)	(2.33)
Change in Quality	-10.62***	-10.46***	-11.20***	-10.77***	10.37***
	(0.93)	(0.85)	(0.86)	(0.81)	(0.81)
SVO-slider-value		-0.25***	-0.20**	-0.23***	-0.24***
		(0.09)	(0.08)	(0.08)	(0.09)
Quality-NF			-1.76*		
			(0.89)		
Quantity-NF				0.11	
				(0.10)	
Productivity-NF					0.05***
					(0.02)
Observations	122	122	122	122	122
R-squared	0.5067	0.5353	0.5630	0.55	0.5667

Notes: OLS Regressions with "difference in number of pictures between feedback and no-feedback phase" as dependent variable. Robust standard errors in square brackets. Standard errors are clustered on groups. A group is a co-worker team in the peer treatment and a single person in the control treatment. Level of significance: *p<0.1, **p<0.05, ***p<0.01. "Hetero-Group" represents whether worker exchanged peer information with a worker providing differing quantity in the no feedback phase, "Change in Quality" represents the difference in average number of labels per pictures between the no feedback and the feedback phase.

TABLE 9. INFLUENCE OF GROUP COMPOSITION ON QUALITY

	(1)	(2)	(3)	(4)	(5)
Hetero-Group	-0.32**	-0.29*	-0.35***	-0.38***	-0.32**
	(0.15)	(0.15)	(0.13)	(0.14)	(0.15)
Change in Quantity	-0.05***	-0.05***	-0.05***	-0.05***	-0.05***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
SVO-slider-value		-0.01	-0.00	-0.01	-0.01
		(0.01)	(0.01)	(0.01)	(0.01)
Quality-NF			-0.17***		
			(0.05)		
Quantity-NF				0.01*	
				(0.01)	
Productivity-NF					0.00
					(0.00)
Observations	122	122	122	122	122
R-squared	0.5227	0.5307	0.5937	0.5588	0.5438

Notes: OLS Regressions with "difference in number of pictures between feedback and no-feedback phase" as dependent variable. Robust standard errors in square brackets. Standard errors are clustered on groups. A group is a co-worker team in the peer treatment and a single person in the control treatment. Level of significance: *p<0.1, **p<0.05, ***p<0.01. "Change in Quantity" represents the difference in number of pictures between the no feedback and the feedback phase.

TABLE 10. INFLUENCE OF GROUP COMPOSITION ON PRODUCTIVITY

	(1)	(2)	(3)	(4)	(5)
Hetero-Group	-3.93	-4.35	-1.52	1.82	-5.07
	(8.76)	(8.99)	(8.86)	(8.54)	(8.43)
SVO-slider-value		0.13	-0.06	0.03	0.14
		(0.39)	(0.35)	(0.37)	(0.38)
Quality-NF			8.07***		
			(2.94)		
Quantity-NF				-0.76**	
				(0.32)	
Productivity-NF					0.05
					(0.09)
Observations	122	122	122	122	122
R-squared	0.0017	0.0027	0.0873	0.0901	0.0062

Notes: OLS Regressions with "difference in number of pictures between feedback and no-feedback phase" as dependent variable. Robust standard errors in square brackets. Standard errors are clustered on groups. A group is a co-worker team in the peer treatment and a single person in the control treatment. Level of significance: *p<0.1, **p<0.05, ***p<0.01.

6. Summary and Discussion

In this paper, we study peer effects in multi-dimensional work tasks. So far, peer pressure has received attention in the economic literature as an alternative or additional work incentivation and researchers have been concerned about optimal group composition to best utilize peer effects for productivity increase. However, the research so far has exclusively focused on peer pressure in one-dimensional work tasks, which is a far abstraction from many work settings. We argue that similar to crowding out of intrinsic motivation by monetary incentives, peer pressure that affects effort in one dimension of the work tasks might at the same time crowd out effort in other dimensions that are not observable by peers.

First, we present a simple theoretical model to predict the influence of peer pressure in a multi-dimensional work task that rests on the combination of models on peer pressure with models of incentives in multidimensional work tasks. In the model we assume that co-workers are only capable of comparing each other's output in one dimension. We then compare the case to the case without co-worker monitoring and peer pressure, respectively. With plausible assumptions on worker's effort cost function our model predicts that the introduction of peer pressure entails a (complete) crowding out of effort in the non-observable work dimension by effort in the observable one.

Second, we test the predictions in a laboratory experiment. In particular, we implement a two-dimensional work task in fixed wages environment that consists of separate quantity and quality dimensions. Peer pressure is introduced in the quantity dimension by exchanging information on the quantity output between two co-workers, while no information is exchanged about co-workers output in the quality dimension.

In accordance to the theoretical expectations, we observe a highly significant increase in quantity due to the introduction of peer pressure while quality highly significantly decreases. Thus, our results demonstrate that there is no invisible border preventing psychological incentives, such as peer effects, to have similar drawbacks as pecuniary incentives. However, aggregate effort rises when peer pressure is introduced.

Beside these core findings, our experimental data identifies competitiveness as a driver of peer effects and reveals that stress levels are unaffected by peer pressure despite the significant increase in productivity. In contrast, Falk & Dohmen (2011) find in their study that payment schemes that induce higher performance are accompanied by higher levels of subjective effort, stress and exhaustion. A possible explanation for the difference between their study and ours might be the smaller increase of productivity in our study. Peer pressure does not increase productivity to the same extent as monetary incentives do. Therefore, stress levels might not increase to the same extent. Alternatively, there might be differences in the subjective perception of the interventions. Hence, even in case of similar productivity gains by pecuniary and psychological incentives subjective assessment would only change in the first case. However, independent of the explanation, peer pressure seems to be a mean to increase productivity without increasing stress, which might have interesting implications for employers: The satisfaction of workers does not change due to the introduction of peer pressure. This means that the possibly negative consequences that accrue from a change of the payment regime, as suggested by Falk & Dohmen (2010), less likely appear under peer pressure despite significant productivity gains.

Moreover, we analyze whether different types of workers are affected differently by peer pressure. We differentiate workers by the ex-ante distance to their later co-worker in the work dimension that is subject to peer pressure and find that weaker workers who lag behind show a stronger increase in quantity due to peer pressure, but at the same time a stronger decrease in quality than workers who are ahead. Previous literature of Falk & Ichino (2006), Mas & Moretti (2009) and Guryan et al. (2009) discusses whether workers' susceptibility to peer pressure depends on workers attributes such as ability. They conclude that low-skilled workers respond stronger to peer pressure. However, in our multi-dimension task, a slightly modified conclusion is indicated: While the absolute skill or output level of the worker seems to be of minor influence, the ex-ante difference to the co-worker in the later observable work dimension is a significant predictor for workers' susceptibility to peer pressure. Generally, the effect of peer pressure is stronger for those providing less output in the dimension later exposed to peer pressure. However, the relation is less clear for the increase of productivity: those who increase quantity stronger also decrease quality stronger, leading to a roughly zero net gain of productivity.

We also investigate how to optimally harness peer pressure by analyzing whether employers should create teams with workers who are (ex-ante) similar or different in the level of quantity they achieve. We find that neither quantity nor productivity is sensitive to the mix of workers within a team. However, the decrease in quality is significantly stronger when teams are composed of heterogeneous workers. Hence, our data recommends composing work teams of rather similar workers. Comparing this result with results of previous studies on optimal group composition (e.g. Falk & Ichino 2006, Mas & Moretti 2009 and partly Bandiera et al., 2010) suggests that optimal group composition depends on work environment. While the mentioned literature on one-dimensional work tasks recommends mixing highly diverse workers, we arrive at the exact opposite conclusion in multi-dimensional work tasks.[18]

Finally, even if we find several particularities of peer pressure, drawbacks are generally the same as for monetary incentives. Hence, our findings might explain why peer pressure (or psychological incentives in general) is rarely utilized in real work environments, despite the fact that workers are highly responsive to it.

[18] Furthermore, our results most probably represent a lower bound for quality decrease. When the effort in the dimension observable by peers indeed rises stronger in heterogeneous groups as found by previous research, decrease in the other dimension might be even stronger.

References

Akerlof, George A. 1982. "Labor Contracts as Partial Gift Exchange."*The Quarterly Journal of Economics*, 97, 543-569.

Baker, George, Robert Gibbons and Kevin J. Murphy.2002, "Relational Contracts and the Theory of the Firm."*The Quarterly Journal of Economics*, 117(1), 39-84.

Bandiera, Oriana, IwanBarankay and Imran Rasul.2010. *"Social Incentives in the Workplace." The Review of Economic Studies*, 77(2), 417-458.

Bertrand, Marianne, Erzo F. P. Luttmer and SendhilMullainathan. 2000. "Network Effects and Network Cultures." *The Quarterly Journal of Economics*, 115(3), 1019-1055.

Chan, Tat, Jia Li and Lamar Pierce. 2010. "Compensation and Peer Effects in Competing Sales Teams." *Working Paper.*

Chen, Yan, F. Maxwell Harper, Joseph Konstan and Sherry XinLi. 2010. "Social Comparisons and Contributions to Online Communities: A Field Experiment on MovieLens." *American Economic Review*, 100(4), 1358–98.

Duflo, Esther and Emmanuel Saez. 2003. "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment." *The Quarterly Journal of Economics*, 118(3), 815-842.

Eriksson, Tor, Anders Poulsen and Marie Claire Villeval. 2009. "Feedback and Incentives: Experimental Evidence." *Labour Economics*, 16, 679-688.

Falk, Armin and Andrea Ichino. 2006. "Clean Evidence on Peer Effects." *Journal of Labor Economics*, 24(1), 39-57.

Falk, Armin and Thomas Dohmen. 2010. "You Get What You Pay For: Incentives and Selection in the Education System." *The Economic Journal*, 120, 256-271.

Falk, Armin and Thomas Dohmen. 2011. “Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender.” *The American Economic Review*, 101, 556-590.

Falk, Armin, UrsFischbacher and Simon Gächter. 2010. “Living in two Neighborhoods – Social Interaction Effects in the Laboratory.” *Economic Inquiry*, 48, 1 – 16.

Farrel, Jospeh and Carl Shapiro. 1990. “Horizontal Mergers: An Equilibrium Analysis.” *The American Economic Review*, 80(1), 107-126.

Fehr, Ernst, Simon Gächter and Georg Kirchsteiger.1997. „Reciprocity as a Contract Enforcement Device: Experimental Evidence.“ *Econometrica*, 65(4), 833-860.

Festinger, Leon. 1954. “A theory of social comparison processes.” *Human Relations*, 7, 117-140.

Fischbacher, Urs. 2007. “z-Tree: Zurich Toolbox for Ready-made Economic Experiments.” *Experimental Economics*, 10(2), 171-178.

Glaeser, Edward L., Bruce Sacerdote and José A. Scheinkman. 1996. “Crime and Social Interactions.” *The Quarterly Journal of Economics*, 111(2), 507-548

Gneezy, Uri and Aldo Rustichini. 2000. “A Fine is a Price.” *Journal of Legal Studies*, 29, 1-17.

Gould, Eric D. and Eyal Winter. 2009. “Interactions between Workers and the Technology of Production: Evidence from Professional Baseball.” *The Review of Economics and Statistics*, 91(1), 188-200.

Graham, Bryan S. 2008. “Identifying Social Interactions through Conditional Variance Restrictions.”*Econometrica*, 76(3), 643-60.

Greiner, Ben. 2004. “An Online Recruitment System for Economic Experiments.” In: Kremer, Kurt, Macho, Volker (Eds.), Forschung und wissenschaftliches Rechnen 2003, GWDG Bericht 63. Gesellschaft für Wissenschaftliche Datenverarbeitung, Göttingen, 79-93.

Guryan, Jonathan, Kory Kroft and Matthew J. Notowidigdo. 2009. “Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments.” *American Economic Journal: Applied Economics*, 1(4), 34-68.

Holmstrom, Bengt. 1999. “Managerial Incentive Problems: A Dynamic Perspective.” *The Review of Economic Studies*, 66(1), 169-182.

Holmstrom, Bengt and Paul Milgrom. 1991. “Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design.” *Journal of Law, Economics, and Organization*, 7, 24-52.

Horton, John J. 2010. “Employer Expectations, Peer Effects and Productivity: Evidence from a Series of Field Experiments.” *Working Paper*.

Johnson, Ryan M., David H. Reiley and Juan Carlos Munoz. 2012. “The War of Fare: How Driver Compensation Affects Bus System Performance.”*Working Paper*.

Kandel, Eugene and Edward P. Lazear. 1992. “Peer Pressure and Partnerships” *Journal of Political Economy*, 100(4), 801-817.

Kaur, Supreet, Michael Kremer and SendhilMullainathan. 2010. “Self Control and the Development of Work Arrangements.” *The American Economic Review: Papers and Proceedings*, 100, 624-628.

Kreps, David M. 1997. “Intrinsic Motivation and Extrinsic Incentives.”*The American Economic Review*, 87(2), 359-364.

Kroft, Kory. 2008. “Takeup, Social Multipliers and Optimal Social Insurance.” *Journal of Public Economics*, 92(3-4), 722-737.

MacLeod, W. Bentley and James M. Malcomson. 1989. “Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment.” *Econometrica*, 57(2), 447-480.

MacLeod, W. Bentley and James M. Malcomson. 1998. “Motivation and Markets.” *The American Economic Review*, 88(3), 388-411.

Mas, Alexandre and Enrico Moretti. 2009. “Peers at Work.” *The American Economic Review*, 99(1), 112-145.

Murphy, Ryan O., Kurt A. Ackermann and Michel J. J. Handgraaf.2011. “Measuring Social Value Orientation.” *Judgment and Decision Making*, 6(8), 771-781.

Paarsch, Harry J. and Bruce Shearer. 2000. “Piece Rates, Fixed Wages, and Incentive Effects: Statistical Evidence from Payroll Records.” *International Economic Review*, 41(1), 59-92.

Pendergast, Canice. 1999. “The Provision of Incentives in Firms.” *Journal of Economic Literature*, 37(1), 7-63.

Sacerdote, Bruce. 2001. “Peer Effects with Random Assignment: Results for Dartmouth Roommates.” *The Quarterly Journal of Economics*, 116(2), 681-704.

Zajonc, Robert B. 1965. “Social facilitation.”*Science*, 149, 269-274.

Zimmerman, David J. 2003. “Peer Effects in Academic Outcomes: Evidence from a Natural Experiment.” *The Review of Economics and Statistics*, 85(1), 9-23.

Appendix A. Instructions (Translated from German)

General Part

Welcome to today's experiment! In this experiment you will be able to earn money. The amount you will earn depends on your decisions and the decisions of other participants.

This experiment consists of three parts. The instructions concerning the first part have been handed out to you with this sheet. The instructions concerning the second and third part will be handed out at the beginning of the respective part. Additionally to your payout from all three parts of the experiment you will receive a fixed payment of 2.50€ for your presence.

Please turn of your mobile phone and abstain from communicating with other participants. Please raise your hand if you have any questions concerning the experiment. We will come to you and answer your question.

All decisions made during the experiment as well as all payments will be kept anonymous.

First Part

In this part you will be presented six decision situations. In each of them you can choose, which amount of money you will receive and which amount of money another participant will receive.

Please choose **one** distribution for each decision situation. All amounts of money in this part are denoted in **€-Cent**.

At the end of this part, one of the six decision situations will be chosen by chance. The decision you made in this situation will be paid out at the end of the experiment.

Please note that answers are neither right nor wrong in this part. You are exclusively asked to state your individual preferences. **Once you have made a decision, please mark the appropriate position on the center line and write the respective distribution of money onto the marks on the right**. With your decision you influence your individual payoff, as well as the payoff of the other participant. This participant will be referred to as "Other" hereafter.

Which other participant will receive the money from you will remain open for now.

<u>Example:</u>

You	30	35	40	45	50	55	60	65	70
Other	80	70	60	50	40	30	20	10	0

You 50

Other 40

1)

You	85	85	85	85	85	85	85	85	85
Other	85	76	68	59	50	41	33	24	15

You ________

Other ________

2)

You	85	87	89	91	93	94	96	98	10
Other	15	19	24	28	33	37	41	46	50

You ________

Other ________

3)

You	50	54	59	63	68	72	76	81	85
Other	10	98	96	94	93	91	89	87	85

You ________

Other ________

4)

You	50	54	59	63	68	72	76	81	85
Other	10	89	79	68	58	47	36	26	15

You ________

Other ________

5)

You	10	94	88	81	75	69	63	56	50
Other	50	56	63	69	75	81	88	94	10

You ________

Other ________

6)

You	10	98	96	94	93	91	89	87	85
Other	50	54	59	63	68	72	76	81	85

You ________

Other ________

Second Part

In this part of the experiment you will be presented photos on your screen. Your task will be to state attributes, with which you describe, what is depicted in the respective photo. An attribute can be located anywhere in the photo. E.g. also in the background or at the edge.

The described photos will be used for chair- and research-internal purposes.

You are asked to state at least one attribute per photo. You can state a maximum of 10 attributes, you see depicted in the respective photo.

If there are multiple kinds of one attribute to be seen in one photo, please state the respective attribute in **one row** with the respective count.

e.g. (correct):

1. **oneblackcow**
2. **threebrowncows**

instead of (wrong):

3. **blackcow**
4. **browncow**
5. **browncow**
6. **browncow**

Please state a **maximum of one** adjective per attribute:

e.g. one red rose

Concerning the input of the attributes:

Attributes have to be confirmed by pressing the “Enter” key. By doing so the attributes will be saved and an input field for a new attribute appears. If you do not want to enter a new attribute, please press the key “next photo”.

If you request a new photo **before** pressing “Enter”, the attribute will **not** be saved.

Before the actual start of this part, there will be a stage of testing. 3 photos will be presented to you, whereby you can familiarize yourself with the handling.

This part of the experiment will take 25 minutes. You will be paid 5 € for this part.

Please remember that communication with another participant is forbidden during this part of the experiment as well as during all other parts.

Third Part – Peer Treatment

In this part you will again be presented photos, and your task will again be to describe them.

The difference between this part and the former is that you will now be matched with another participant. This participant will be continuously informed about the number of photos described by you. You will also be continuously informed about the number of photos the other participant has described. You will find this information in the right top corner of your screen. Additionally, you will also see the number of photos you have described.

Before this part starts, you will be relocated so that you are sitting next to the participant who you are matched to.

This part of the experiment will take 25 minutes. You will be paid 5€ for this part.

Please remember that communication with another participant is forbidden during this part of the experiment as well as during all other parts.

Third Part – Control Treatment

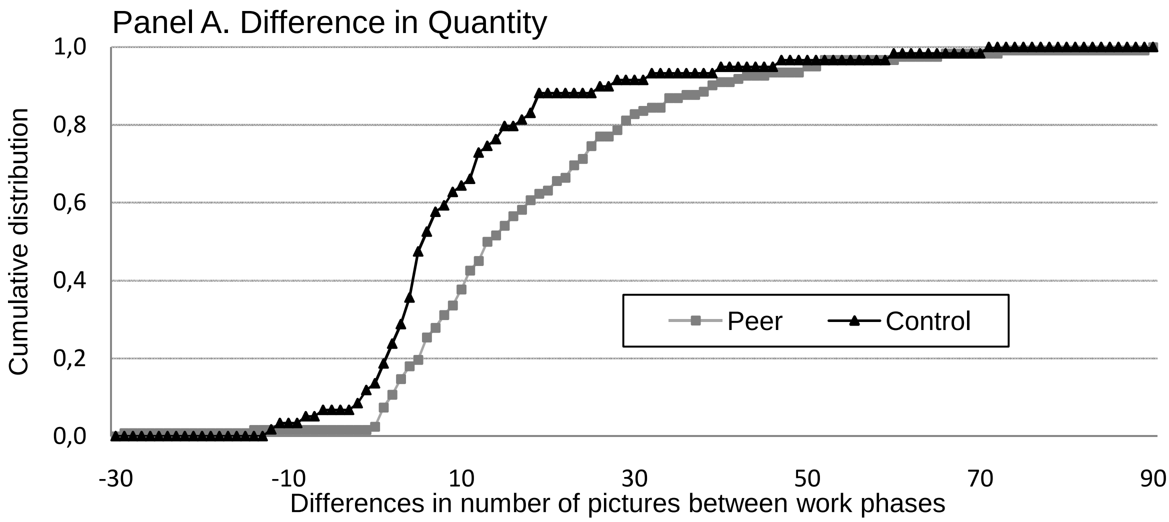
In this part you will again be presented photos, and your task will again be to describe them.

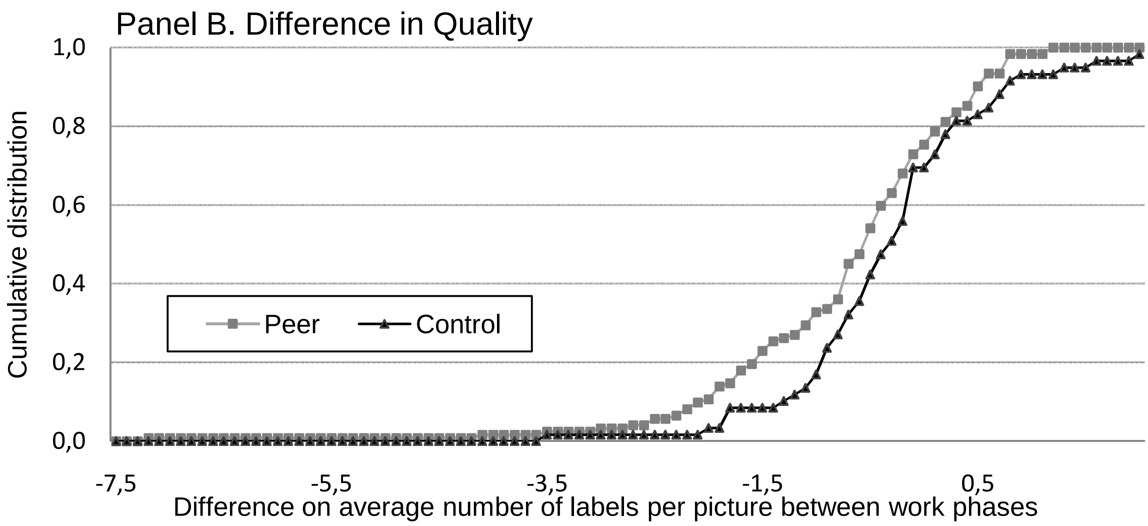
Before this part starts you will be relocated to a new cabin.

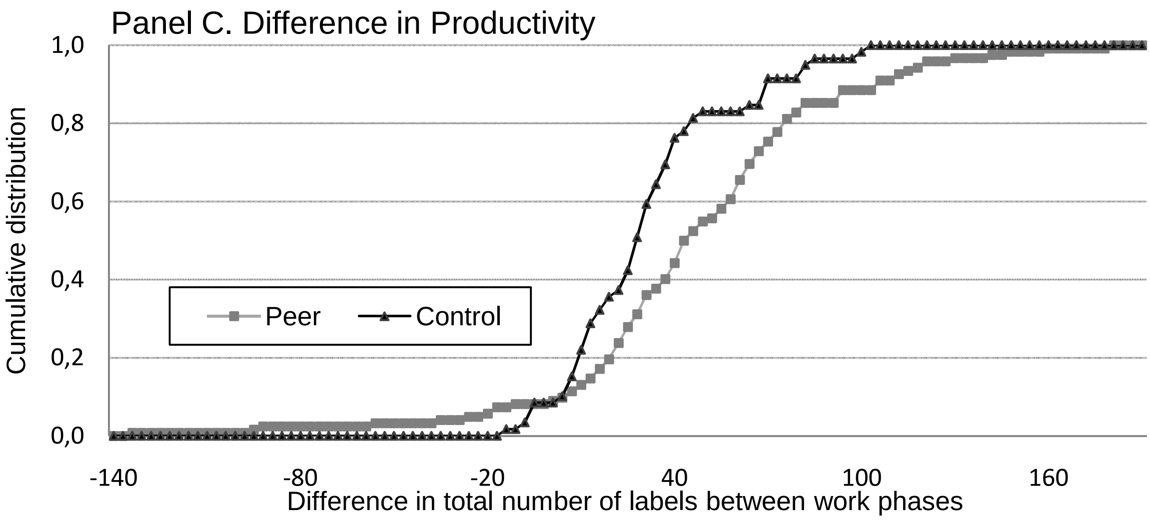
This part of the experiment will take 25 minutes. You will be paid 5€ for this part.

Please remember that communication with another participant is forbidden during this part of the experiment as well as during all other parts.

Appendix B. Cumulative distribution functions of changes in quality, quantity and productivity from the no-feedback to the feedback phase







Appendix C. Mean changes for different workers in different group types

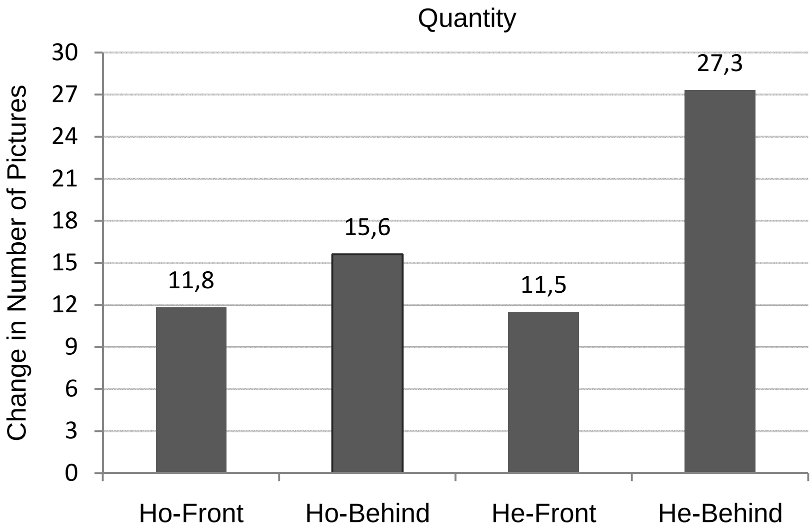


FIGURE 4. QUANTITY MEAN CHANGE FROM NO-FEEDBACK TO FEEDBACK PHASE

Notes: The figure shows for different worker types the mean change of quantity (=number of described pictures). “Ho-Front” represents workers with positive heterogeneity values in homogeneous groups, “Ho-Behind” represents workers with negative heteroneity values in homogeneous groups, “He-Front” represents workers with positive heterogeneity values in heterogeneous groups, “He-Behind” represents workers with negative heterogeneity values in heterogeneous groups.

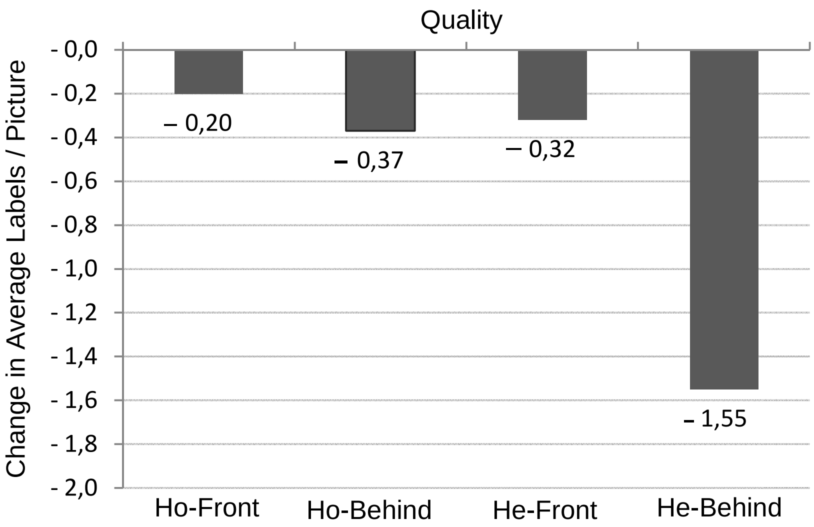


FIGURE 5. QUALITY MEAN CHANGE FROM NO-FEEDBACK TO FEEDBACK PHASE

Notes: The figure shows, for different worker types the mean change of quality (= average labels per picture).

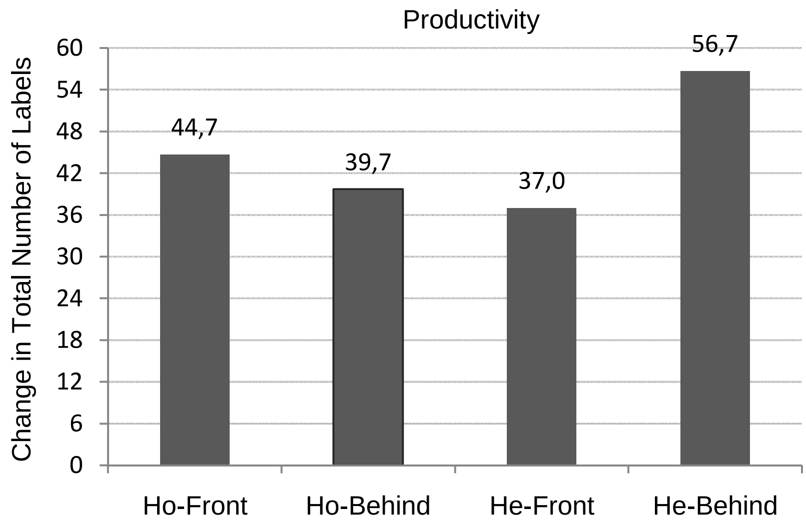


FIGURE 6.PRODUCTIVITY MEAN CHANGE FROM NO-FEEDBACK TO FEEDBACK PHASE

Notes: The figure shows for different worker types the mean change of productivity (= total numbers of labels).

Part 4

Follow the Leader or Follow Anyone

–

Evidence from a Natural Field Experiment

Follow the Leader or Follow Anyone – Evidence from a Natural Field Experiment

Felix Ebeling, Christoph Feldhaus, Johannes Fendrich*

June 12, 2012

Abstract

In a fundraising field experiment we show that individuals are not only conditionally cooperative, but that they are also more prone to donate to a homeless individual when the previous donor has a higher social status. We trailed a homeless person asking for donations within Cologne's metro trains for two weeks. Thereby we systematically varied the status of the first giver in the train. In the control treatment we did not intervene. In the low status treatment the first giver was always a (poor looking) low status person from our team and correspondingly in the high status treatment a (rich looking) high status person. In our experiment the average number of donations per train is 72% higher in the low status treatment than in the control treatment. Additionally, in comparison to the low status treatment, the number increases by 34% in the high status treatment. To our best knowledge this is the first study providing field evidence for the particular influence of high status individuals on others' donations.

JEL classification: C93, D64, H41

Keywords: Status, Fundraising, Field Experiment

---

*Felix Ebeling, University of Cologne, Department of Economics, ebeling@wiso.uni-koeln.de, Christoph Feldhaus, University of Cologne, Department of Economics, feldhaus@wiso.uni-koeln.de, Johannes Fendrich, University of Cologne, Department of Economics, fendrich@wiso.uni-koeln.de. We thank Constanca Esteves-Sorenson, Bernd Irlenbusch, Dean Karlan, Peter Matthews, Axel Ockenfels, Bettina Rockenbach, Dirk Sliwka, Abdolkarim Sadrieh and Matthias Sutter for helpful comments. Furthermore, we thank the participants of the GfeW Tagung 2011, the SMYE 2012, the M-BEES 2012 and the Workshop „Natural and Controlled Field Studies" in Holzhausen/Ammersee as well as seminar participants at the University of Cologne and the Yale University. We also gratefully acknowledge the financial support from the German Science Foundation (through Gottfried Wilhelm Leibniz Price of the DFG, awarded to Axel Ockenfels).

1. Introduction

Our research is concerned with the optimal solicitation order in fundraising campaigns when information on previous giving is available. Previous theoretical research on solicitation order suggests soliciting the most generous donors first (see e.g. Andreoni 1998, Versterlund 2003). Field-studies of List & Lucking-Reiley (2002) and Shang & Croson (2009) support these theoretical findings.[1] Our paper reveals the importance of first giver's status. Thus, fundraisers should not only be concerned with the amount of the first donation but also with the status of its donor. A higher status entails more subsequent donations.

In our study we designed a natural field experiment (Harrison & List 2004) to test the influence of the first donor's status in fundraising. Our experiment is concerned with donations to a homeless "street newspaper" seller. As in many larger western cities "street newspapers" are also sold in Cologne (Germany).[2] Sellers are mainly homeless people offering the newspaper at street corners or promoting it on the streets or on the metro. Despite the fact that they offer newspapers, sellers mostly receive donations while newspapers are sold rarely.[3] Frankly spoken, selling street newspapers is a polite way to ask for donations. For our experiment, we trailed a homeless newspaper seller for two weeks. Our focus was on metro wagon sales since every wagon provides an isolated environment that can be regarded as one independent observation. An observation started when the seller entered the metro wagon at a station. After the

[1] Andreoni (1998) and List & Lucking-Riley (2002) focus on the amount of seed money in fundraising campaigns for threshold public goods, Vesterlund (2003) is concerned with the optimal announcement strategy of previous contributions while Shang & Croson (2009) are interested in the optimal provision of social information. However, when it comes to the optimal solicitation order, all papers suggest soliciting large donations first.

[2] For general information on street newspapers see http://en.wikipedia.org/wiki/Street_newspaper. For information on the street newspaper in Cologne, see: http://www.querkopf-koeln.de/. A picture of the newspaper can be found in the appendix.

[3] This impression from everyday life is corroborated in our experiment. Our seller received roughly 25 times more through donations than by selling newspapers.

wagon doors closed he began to promote the newspaper. Basically the seller said: „Dear ladies and gentlemen, is anyone interested in a street newspaper or has a small donation for a homeless person?“. Afterwards, he walked through the train showing his collecting box and the newspaper to passengers. An observation ended, when the seller left the wagon at the next station. Our experiment involved three treatment variations. In the control treatment we did not intervene. We solely observed the seller making his tour through the train as described above. In the second and third treatment we manipulated the status of the first donor in the train. In the low-status treatment, the first donor was a poor looking person from our team. In the high-status treatment the first donor was a rich looking person from our team. The procedure in the low and high status treatment was identical to the control treatment with the exception that the donor from our team started giving directly after the seller's promotion.[4]

In our experiment about 10,500 individuals participated in 567 independent observations. In total we received 424 donations and raised 316.27€. The experiment provides four main results. First, donors in metro trains are conditionally cooperative. Compared to the control treatment, the number of donations per train-ride increases by 72 percent in the low-status treatment, and by 129 percent in the high-status treatment. These differences are highly significant. Second, donors are more prone to donate when the first donor has a higher status. Compared to the low-status treatment, the number of donations per train ride increases by 34 percent in the high-status treatment. This difference is (at least) significant. Third, there is some evidence for a crowding in of lower donations. Compared with the control treatment, values of single donations are lower in both other treatments. Compared with the low-status treatment, values of single donations are almost similar /slightly lower in the high-status treatment. Fourth,

[4] By clarifying in advance at which door the seller enters the train, we could easily ensure to be the first donor in the train.

we present data on donor characteristics. The characteristics show that our results are only partly in line with the standard status theory claiming that individuals in general like to associate with those of higher status.

The remainder of the paper is organized as follows. In the next section we review the related literature. Section three describes the experiment. Section four contains the results, and section five concludes.

2. Relevant Literature

This paper contributes to three fields of economic literature. It adds insight on fundraising research and on research on individuals' quest for status. Furthermore, it is partly related to research on leadership.

First, our research is concerned with fundraising. Within charitable giving/fundraising literature (apart from research relating status and fundraising) research on conditional cooperation is most important for us. Fischbacher & Gächter (2010) p. 541 describe conditional cooperation as "many people's propensity to cooperate provided others cooperate as well". Conditional cooperation is found robustly in laboratory studies (Fischbacher et al. 2001, Kocher et al. 2008, Fischbacher & Gächter 2010) as well as in charitable giving field studies (Frey & Meyer 2004, Shang & Croson 2009). In the field experiment of Frey & Meier (2004) students are more prone to contribute to charitable funds when knowing that many other students contribute. Shang & Croson (2009) provide additional evidence for conditional cooperation in a public radio fundraising experiment. In their experiment, participants receive information on previous donations whereby the amount of these donations vary; the higher the mentioned donation, the higher subsequent ones. Regarding our experiment, both, the high status and the low status treatment confirm the conditional cooperation hypothesis. Metro passengers' propensity to contribute is significantly higher in case of other contributors.

Second, our research is concerned with individual's quest for status. Different areas of economic research emphasize the role of status, e.g. research on consumer choice (Frank 1985, Hopkins & Kornienko 2004, Charles et al. 2009, Heffrtz 2011), organizations (Frank 1984, Moldovanu et al. 2007, Besley & Ghatak 2008) and fundraising (Harbaugh 1998a, Harbaugh 1998b, Kumru & Vesterlund 2010). Weiss & Fershtman (1998) p. 802 define social status as "a ranking of individuals […] in a given society, based on their traits, assets, and actions". Exact definitions in other research areas may deviate, but as Heffetz & Frank (2011) point out that it is hard to find a definition not related to "rank" or "position".[5] Most important for our research are the economic approaches towards status by Ball et al. (2001) and Kumru & Vesterlund (2010). Ball et al. (2001) examine prices in a competitive laboratory market. Participants of their experiment act as buyers or sellers and are attributed with a low or a high status. Independent of the market side assigned to high-status participants, they always capture a greater share of the surplus. Apparently, low status agents are willing to sacrifice consumption to trade with high-status agents or, more generally, they sacrifice consumption to associate with them. Kumru & Vesterlund (2010) transfer the idea of Ball et al. (2001) to a fundraising setup. In their sequential laboratory public good game individuals are also assigned with a high or a low status. When first movers are high status agents, public good contributions are significantly higher. Again, low status agents are willing to sacrifice consumption to associate with high status agents. The results from our field experiment corroborate Kumru & Vesterlunds' laboratory findings. Metro passengers contribute significantly more often to the homeless when the first mover has a higher status.

Third, our research is partly concerned with leadership. In economic literature, leaders are often defined as individuals with superior information (cf. Hermalin 1998, p. 1198).

[5] For research from other areas such as evolutionary sociology see for example Henrich & Gil-White (2001) and Boyd & Richerson (2002).

As Henrich & Gil-White (2001) argue, high status individuals often have superior information.[6] In so far, one might perceive the high status individual from our experiment to be a leader. Vesterlund (2003) and Andreoni (2006) investigate the influence of such superior informed leaders in fundraising campaigns. Following their theories, a donation of a high status individual or leader is a signal for the high quality of the charity and thus leads to higher subsequent contributions. Lately, Karlan & List (2012) find that mentioning the Bill and Melinda Gates Foundation as a matching donor (and quality indicator) significantly increase donations. However, considering our experimental set up, it is less likely that our leader (high status donor) is perceived as someone with superior information. The donation receiver is "well known" to all passengers and it is clear that he will use the money for his own consumption.

3. Experimental design

3.1 Environment

We conducted our experiment in summer 2011 in the metro trains of Cologne's municipal transport services "KölnerVerkehrs-Betriebe"(KVB). Cologne's metro train system consists of eleven lines and has a path length of 193.8 km. In 2010 Cologne's metro system had more than 200 million passengers. More than 300,000 customers are frequenters.[7] This represents more than one quarter of Cologne's inhabitants. Hence, passengers most likely represent a cross section of urban West-German society. To assure a subject pool representing these socio-demographic characteristics, observations in our experiment stem from different daytimes. We took trains from 9.15am to 12.15pm and from 5pm to 8pm. In the morning "shift" commuter traffic is

[6] Anthropologist literature as Henrich & Gil-White (2001), Chudek et al. (2012) and Panchanathan (2010) is about prestige biased learning. Accordingly, individuals have a tendency to adopt behavior from high status individuals, because their ex-post behavior seemed to be more successful (worthy to adopt).

[7] Source of figures: Website of "KölnerVerklehrs-Betriebe AG" (Cologne's Public Transport Enterprise) http://www.kvb-koeln.de/german/unternehmen/leistungsdaten/index.html

basically over and passengers are mainly non-working society (e.g. young mothers, students, pensioners). By contrast, at least in the early hours of the evening shift, commuters are a major fraction of passengers. Furthermore, to prevent effects arising from particular populations in different neighborhoods, we took nine of the eleven metro lines in different areas of Cologne.[8]

3.2 Homeless Newspaper Seller

In our experiment the receiver of the donations was a unique authentic homeless person. The most important facts about him are, first, that he had no permanent residence in the time period we conducted our experiment and, second, that selling street newspapers and receiving donations represented his main sources of income at that time.

Additionally to his earnings from newspaper sales and donations, we paid the homeless person 50€ each day. This is equivalent to approximately 1.5 to 2 times his daily income generated by newspaper sales and donations in the time span of our experiment.

Inevitably, the homeless person knew that we conduct an experiment to test for giving behavior. However, we never disclosed our main hypotheses to him.

3.3 Procedure

We trailed the homeless newspaper seller for two weeks in summer 2011. Both weeks were the first of their respective month.[9] Within a week we tried not to take trains with the same passengers more than once. For example, if we had taken a certain metro line every weekday at the same time, the probability to meet commuters more than once

[8] However, we want to emphasize that neighborhoods in German metropolitan cities aren't as heterogeneous in terms of social composition as e.g. cities in the United States.
[9] In interviews conducted previously to our experiment, homeless persons mentioned a decreasing tendency of donations over a month. To keep circumstances similar, we conducted our experiment in the first full week of July (4th - 8th) and the first full week of September (5th - 9th).

would have increased. Thus, in both weeks, we allocated only one morning and one evening shift to a particular metro line.[10] Since passenger compositions in metro trains in the morning and evening shifts substantially differ in terms of socio-demographic characteristics, this particular measure decreases the probability to encounter the same passengers more than once. Furthermore, these shifts on the same metro line were never consecutive. If we had used a particular line in the morning shift, we did not use the same line in the evening. And accordingly, if we had used a metro line in the evening, we did not use the same line the next morning. Thus, even if we encounter the same passenger twice our giving was inconspicuous, because giving in regular intervals (e.g. every second day) from a certain donor is rather the rule than the exception.[11]

Within a shift we shuttle on a certain track section of Cologne's metro network. The track section of a shift comprises three consecutive stations (e.g. the stations A, B, C). Treatments alternate between stations in a strict order. The following example clarifies the procedure: Recall that Cologne's metro consists of two wagons per train. At station A, the homeless newspaper seller and the high status person enter the train in the second wagon. In the meantime the low status person enters the first wagon. On the ride from station A to station B we conducted the high status treatment in the second wagon. At station B, the homeless newspaper seller leaves the second wagon and enters the first one. The high status person stays in the second wagon. On the ride from station B to C we conducted the low status treatment in the first wagon. At station C we left the train, changed the platform and went back to station A following the same procedure. Similarly, at station A, we left the train again and changed the platform. In the

[10] Actually, we used some metro lines twice in the evening and in the morning. However, the same line was only used a second time on a different track section rather remote from the first section. So we regard these "long lines" as different lines because passengers most probably change between different parts of the city.
[11] We interviewed several homeless newspaper sellers previous to our experiment about their income structure. All of them reported the importance of regular donors.

subsequent two rides from A to C we conducted two control treatments. The described six rides are one procedure cycle. At C we restart the procedure cycle, but with the low status treatment in the second wagon and correspondingly the high status person in the first wagon. Figure 1 shows the rides after two procedure cycles.

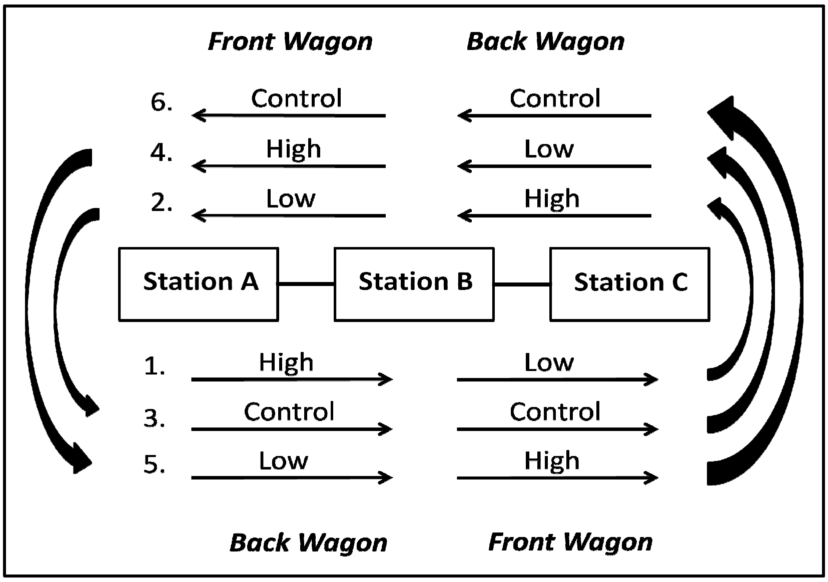


FIGURE 1. RUNNING ORDER, SHOWING TWO PROCEDURE CYCLES

After two procedure cycles we continue at station A with the high status treatment and so on. As can be seen in Figure 1, the described treatment alternation leads to an ongoing treatment change on every of the four station-connections (AB, BC, CB, BA). Most importantly, the procedure assures a minimal time difference between treatment observations from a certain station-connection.[12] Thus, the environment of observations is identical.

With one exception, we only shuttle between stations with parallel running lines because the frequency of trains is higher when two metro lines operate. Waiting time between taken trains shortens. It allows us to collect more data within a shift. The

[12] At a certain station, approximately 18 minutes pass from one observation to the next. In the first week we had some exceptions from the explained procedure. Basically, we extended our procedure to four instead of two stations (everything else, e.g. treatment order, remained similar). We tested whether results differ in these observations. We found no differences.

duration of a shift was approximately three hours. Any extension of a shift would have increased the probability to encounter the same passengers again (on their way back home), but the shorter the shift the lower the number of collected observations would have been. Three hours is the compromise we chose. We cannot rule out same passengers in some observations. However, the presented means reduce their number and thereby possible distortions of the results.

Within a train we implemented a standardized donation request of the homeless person. When conducting a treatment in the second wagon of a train, the homeless person always entered the train by the same door of the wagon. His request started after the doors closed and the train started. He announced his newspaper sale by: "Dear Ladies and Gentleman, is anyone interested in a street newspaper or has a small donation for a homeless person?" Afterwards, he walked through the train, showing his collecting box and the newspaper to passengers. His path through the train was predefined. He walked to the next door, turned around and left the train by the same door he entered. The distance to the next door was about 5 meters. When conducting a treatment in the first wagon, everything else remained equal, but the homeless newspaper seller started at the last door of the wagon. Due to the symmetrical structure of wagons, no further differences arose. In case of a low or high status treatment, the corresponding status person always entered the train by the same door as the seller or already waited at that door. Directly after the start of the donation request, the status person takes out some money from his pocket and puts it into the collecting box of the homeless person. Thereby, we could easily establish our status person as the first donor.

3.4 Treatments

We conducted three treatments in our experiment: A control treatment, a low status treatment and a high status treatment. We did not intervene in the donation request of the homeless person in the control treatment. In the low and high status treatment however, the first donor in the wagon was the respective status person.

In our experiment we had to attribute status visibly to a person. A successful assignment of status implies a substantial agreement among different members of a society on the hierarchical status of the person (Weiss & Fershtman 1998). We orientated our implementation to the prominent concept of socioeconomic status (SES). Kraus & Keltner (2009) p. 99 define SES "by material wealth, occupation, and participation in educational and social institutions".[13] We tried to choose persons and outfits clearly representing at a first glance different characteristics of these dimensions. In both weeks of our experiment the high status person was a 31 year old male. He was dressed in a suit, tie, shirt and leather shoes. He carried a laptop bag and a high profile national newspaper (Frankfurter Allgemeine Zeitung). Altogether, the price of his outfit was above 800€. The low status person changed between experiment weeks. In the first week, the low status person was an unemployed person in his mid-forties. He was dressed with raddled jeans, sneakers, an old hoodie and a cap. However, because of the higher age of the low status person we changed the low status person in the second week.[14] The second low status person was a 27 year old male, dressed in old tracksuit pants, camouflage sweater, tatty chucks and a cap. He carried a plastic bag from a discount supermarket and a tabloid paper. Photos of the persons can be found in the appendix.

[13] See Dutton & Levin (1989), Adler et al. (1994) and Oakes & Rossi (2003) among others for similar definitions.

[14] Individuals might associate a higher status to an older person. Thanks to Matthias Sutter for this helpful suggestion.

In both status treatments we donated exactly 50 Cent in five 10 Cent coins. Due to the short distance the seller walked through the train, all passengers should have been able to notice that the first givers donated some small coins. This is important from the economic perspective. Several theories about leadership in fundraising (e.g. Vesterlund 2003, Andreoni 2006, Hermalin 1998) are particularly concerned with the donation-amount of the first giver. Our study did not focus on this aspect. Since people in our experimental environment can roughly see and hear what others donate, beliefs about the donation-amount hardly differ between the treatments. Differences can be ascribed to status-modification.[15]

To control whether our visual implementation of high and low status individuals is in line with the general perception of these persons, we conducted a classroom survey. 319 bachelor students of the Faculty of Management, Economics and Social Science of the University of Cologne received pictures of the different characters. Each student received a picture of one person and had to estimate the level of education, occupational qualification, employment status and income level. In line with our expectations, the high status person exceeds by far both low type persons in all categories. Obviously, a change in appearance effectively change perceived socioeconomic status of a person. There are also significant differences between the two low type persons in some categories, but in comparison to the high type, these differences are small. A detailed description of the survey results can also be found in the appendix.

[15] This assumption receives support by the results reported in chapter 4.4, where we show that values of single donations are slightly lower in the high-status treatment than in the low status treatment. Almost similar single donation values in the low and the high status treatment are an indicator for similar perceived reference donation-values of the first donors. Especially, the values contradict the hypothesis that passenger give more in the high status treatment because they believe the high status person gave more.

4. Results

4.1 Overview

Table 1 presents the main descriptive results of the experiment. The first row shows the number of observations conducted per treatment. The second row reports the absolute number of donations received, the third row the average number of donations per train ride and the last row contains the average amount of money received per observation. While the absolute number of donations is 82 in the control treatment, it is 146 in the low-status treatment and 196 in the high-status treatment. This difference is also reflected in the average number of donations per observation across the treatments. These numbers are 0.45 in the control treatment, 0.76 the low-status and 1.02 in the high-status treatment. In comparison to the control treatment the average number of donations is 72% higher in the low-status treatment. Furthermore, in comparison to the low-status treatment, the average number of donations increases by 34% in the high-status treatment. The corresponding average amounts of money collected by the newspaper seller are 0.40€, 0.57€, and 0.69€, respectively.

Before analyzing the data in detail, we discuss some peculiarities of our rather unusual setup. In total, about 10,500 metro passengers participated in our experiment. These are all passengers in the trams we used in the course of our investigation. The average number of passengers per observation (per tram wagon) was 16.4 in July and 20.5 in September. The increase in passenger volume is most likely due to the rainy and cold weather in September.[16] In total, we conducted 567 observations for the experiment.[17]

[16] In July there was sunny weather with an average temperature of 20 degrees during our experiment week. In the September session there was rainy weather with an average temperature of 17 degrees. We assume that people prefer to take the metro in September.

[17] We excluded one observation from our analysis. The homeless seller received 16.22€ in this observation. This amount exceeded all other observation-amounts by far. The reasons for the exclusion are similarly to the reasons for exclusion of observations mentioned by Falk (2007). First, this observation

In 265 rides the homeless person received at least one donation. Altogether, he received 408 donations and sold 16 newspapers. The small fraction of newspaper sales corroborates our assumption regarding the sources of income of street-newspaper sellers. The lion's share of their earnings stems from donations.[18]

TABLE 1. DONATION PATTERNS IN DIFFERENT TREATMENTS

	Control	Low	High
Observations	184	191	192
Number of Donations	82	146	196
Average Number of Donation per Ride	0.45	0.76	1.02
Average Donation-amount per Ride	0.40€	0.57€	0.69€

4.2 Conditional Cooperation and Status Effects

To test whether passengers in Cologne's trams are conditionally cooperative and whether the probability of giving differs depending on the status of an initial giver, we analyze several dependent variables such as the number of donations per observation, the share of observations with at least one donation and the amount of money given to the seller per ride. Summing up, our results suggest (i) an effect of conditional cooperation as found in previous studies and (ii) a status effect suggesting that the characteristics of the initial giver are also of importance. People in the tram are more likely to donate to a homeless guy when there is an initial donor and even more so when he apparently has a high social status.

skews the analysis of the absolute donation level. Second, it is unlikely that such donations are due to our treatment variation.

[18] In the two weeks of the experiment our homeless person earned an accumulated amount of 316.27€. The recommended price for the newspaper is 1.50€ whereby 0.75€ are intended for the seller. Buyers often do not stick to the price and sometimes give a higher amount, e.g. 2€. The 316.27€ are earnings. We already deducted the 0.75€ wholesale price of the newspaper paid by the seller. One might subtract another 0.75€*16=12€ to receive a proper donation amount. Ultimately it does not matter. By far the biggest part comes from donations.

First, we analyze the number of donations per ride. Figure 2 shows the share of train rides with 0,1,2,… or 7 donations per treatment.[19] Observations with no donation occur more often in the control treatment than in the low-status and high-status treatment. Consequently, rides with 1,2,…,7 donations are more frequent in the low-status treatment and the high-status treatment. Results of Mann-Whitney U tests (MWU) confirm this impression. There are highly significant differences between the control treatment and the other two treatments (MWU, $p<0.01$).[20] Furthermore, there are fewer observations with no donation in the high-status treatment than in the low-status treatment, but more positive amounts when the initial giver has a higher status. This descriptive finding is supported by non-parametric statistics comparing the distribution between the two treatments (MWU, $p<0.05$).[21] These findings suggest both, an effect of conditional cooperation as well as an effect of the status of the initial giver.

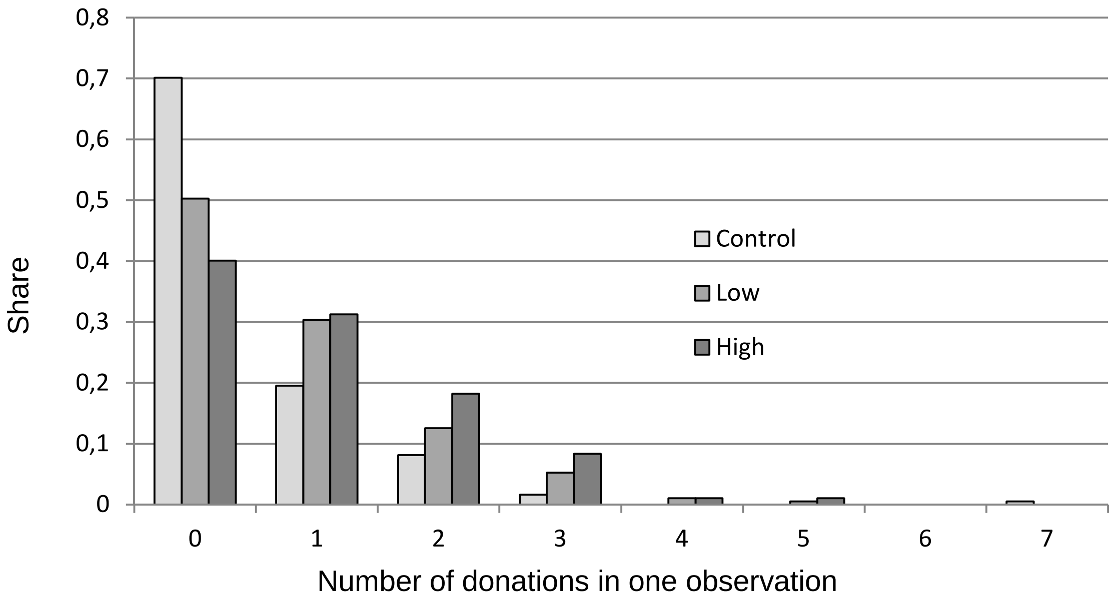


FIGURE 2. SHARE OF A CERTAIN NUMBER OF DONATIONS PER TREATMENT

[19] We never had more than seven donations in a train.
[20] The Mann-Whitney U test compares the distributions of number of donations per observation. In the distribution an observation / data-point takes the value "0" if no one in a train ride donated, "1" if one person in a ride donated, "2" if two persons donated and so on.
[21] We also compare the probability that any person donates per observation using a two-sample test of proportions. The results are identical. The probability is about 60% in the high-status, 50% in the low-status and 30% in the control treatment. All these proportions are different (two sample test of proportions, at least $p<0.05$).

To substantiate our main results and to control for possible confounding effects, we additionally conduct ordered probit analyses which can be found in Table 2. In our analyses the dependent variable is the number of donations per observation. The main independent variables are the treatment dummies. Thereby, the variable "low" is a dummy for the low-status treatment and the variable "high" is a dummy for the high-status treatment. Both are compared to the reference category "control". In Model (1) we regresses our dependent variable only on our treatment variations. In Model (2) and Model (3) we add variables that cannot be randomized in our field setting. Namely, Model (2) adds the variable "session" which is a dummy for the period (July or September) and Model (3) adds the urban district (area in the following). Finally, Model (4) adds some controls.

The results of our analysis in Table 2 show that in all models both treatment dummies are positive and highly significant (in both cases $p<0.01$). Furthermore, we use a wald-test to check for difference between the "low" treatment dummy and the "high" treatment dummy. For all models, the wald-test shows that the "high" dummy surpasses the "low" dummy significantly in size ($p<0.05$). Besides, while the models show no differences between our July and September session,[22] they reveal some significant area effects. However, this does not systematically change the effect of the treatments with respect to the dependent variable.[23] Summarized, the ordered probit analyses confirm the previous results from Mann-Whitney U tests. Passengers are conditional cooperative and a first giver with a higher status entails more donations.

In Table A.11 in the Appendix we present additional probit analyses where we constrain our focus on the first giver from the passenger crowd. Such constrain might make sense

[22] The interactions of treatments and sessions are also not significant. See Table A.10 in the Appendix.
[23] We also find this difference when controlling for interactions between the treatment and the area of the city and between the treatment and the session. See appendix Table A.10.

as giving in the train is a sequential process. We increasingly lose control over the giving process after the first donor out of the passenger crowd. For example, when there is more than one donor, subsequent donors may donate due to several reasons (e.g. status of first donor from the passenger crowd), but not due to our treatment intervention. However, the results in Table A.11 are similar to those presented here.

TABLE 2: EFFECTS ON THE NUMBER OF DONATIONS

	(1)	(2)	(3)	(4)
Low	0.470***	0.470***	0.468***	0.461***
	(0.123)	(0.123)	(0.123)	(0.124)
High	0.746***	0.745***	0.751***	0.799***
	(0.122)	(0.122)	(0.122)	(0.124)
Session	No	0.015	0.012	-0.011
		(0.097)	(0.099)	(0.103)
Area 2	No	No	0.110	0.255
			(0.160)	(0.165)
Area 3	No	No	0.283*	0.340**
			(0.160)	(0.161)
Area 4	No	No	0.255	0.269*
			(0.158)	(0.159)
Area 5	No	No	0.114	0.142
			(0.165)	(0.166)
Area 6	No	No	0.074	0.113
			(0.306)	(0.311)
Controls				
Daytime	No	No	No	Yes
Position	No	No	No	Yes
Passengers	No	No	No	Yes
Observations	567	567	567	567
Pseudo R-squared	0.029	0.029	0.032	0.048

Notes: Ordered probit regressions with donations per observation as dependent variable. Standard errors in parentheses. Level of significance: *p<0.10, **p<0.05, ***p<0.01, 6 cut-points were estimated (output excluded).

In the next step, we analyze the amount of money received per observation in order to investigate whether more donations do actually translate into more money. On a first glance, our figures support this view, as the average amount of money received per observation is 0.40€ in the control treatment, 0.57€ in the low status treatment and 0.69€ in the high status treatment. We use MWU tests to compare distributions of different treatments. In the following, our analysis is limited to the results of the second week as we did not gather these data in the first week. The MWU tests indicate a highly significant difference (MWU, $p<0.01$) between the control treatment and both status treatments. Furthermore, we find a weakly significant difference between the two status treatments (MWU, $p<0.10$). Even though we did not collect data on the amount received per observation in the first session, our data allow calculating the average amount received per observation across the whole week. In July, during the first session of the experiment, the average amount was 0.37€ in the control treatment, 0.50€ in the low-status treatment and 0.59€ in the high-status treatment. The corresponding values for the second session, in September, were 0.43€, 0.64€ and 0.78€, respectively. Although the amount received on average is generally lower in the first week, the pattern across treatments is similar:[24] Average amounts donated to the seller are higher in the low-status and in the high-status treatment in comparison to the control treatment. This indicates that more donations actually translate into more revenues. Furthermore, more money is raised when the initial giver has a high social status. We conclude that it is not only of importance that there is a first giver but also how he is perceived by other potential givers in terms of his social status.

4.3 Crowding In of relatively low Donations

As reported in the previous sections, starting with donations of low status or high status individuals crowds in additional donations. It is of interest to know whether motives of additional donors differ. Different values of single donations are an indication for different motives. On a first glance, the average values of single donations indicate differences. The average single donation is 0.90€ in the control treatment, 0.74€ in the

[24] Furthermore, the analyses of Table 2 and Table 3 show no difference in donation probabilities between weeks. There is no reason to assume systematically different distributions of donation amounts in July observations.

low status treatment and 0.68€ in the high status treatment. Unfortunately, our study is limited in the analysis of single donations. First, only in the second week of our experiment we gathered data on the donation amounts per observation. Second, due to technical reasons, we cannot properly disentangle the single values of donations when more than one passenger in a wagon donated.[25] The few remaining observations with only one donation do not indicate differences between single donation values when analyzed with non-parametric tests.[26] The second-best possibility to test for differences between single donation values is to calculate the average value of single donations for donations from observations with more than one donation and include them into the analysis. Now, a Mann-Whitney U test reveal highly significant differences between the single donation values of the control and the low-status treatment and between the control and the high-status treatment (in both cases $p<.01$), but no differences between the low-status and the high-status treatment ($p=.47$).

The results suggest some support for the hypothesis of a crowding in of low donations in the low-status and high-status treatment. However, we are aware of the fact, that our analysis has certain limits. First, including average values infringes statistical independence of observations, which is necessary for the application of the Mann-Whitney U test. Second, there are several economic explanations for reported differences in single donation amounts. For example, DellaVigna et al. (2012) reveal in their study social pressure as a motive for donations. Similarly, in our experiment a donation of a metro passenger might induce social pressure on other metro passengers. To circumvent social pressure, passengers reluctantly donate smaller amounts. However, pure altruism also explains smaller amounts of following donations, as the

[25] We were only able to count the accumulated donation amount in the beggar's collection box *after* a train ride and not within a train ride after each donation.

[26] There are only 18 observation in the control treatment, 30 in the low-status treatment and 32 in the high-status treatment left.

homeless person already received some money. Last but not least, follows might perceive the donation amount given by our team members as a reference value for appropriate giving. Since the low-status and high-status person give similar amounts, this would explain the almost similar donation amounts of other givers in this treatment.

4.4 Characteristics of Donors

We noted gender, perceived status and age of each donor. Even if age and status is often hard to estimate, it is worth to analyze the data, as it might provide insights on the psychological processes that drive our results.[27] Concerning status, we categorize donors into low, middle and high status. We used the outfits of our first donors as guideline for categorization. Neat persons in expensive clothes were categorized as high status person, unkempt persons in shabby clothes as low status person. All other persons were categorizes as middle status. In our analysis we transferred donor-status into a numeric variable. Thereby, the value 2 represents a high status donor, 1 represents a middle-status donor and 0 represents a low-status donor. Table 3 gives an overview about the collected data on donor characteristics.

TABLE 3. DONOR-CHARACTERISTICS

	All Donors			First Donors		
	Av. Status	Share Men	Av. Age	Av. Status	Share Men	Av. Age
Control	1.05	0.44	42.8	1.04	0.42	42.5
Low	1.01	0.45	42.7	1.02	0.45	42.8
High	1.11	0.34	44.2	1.15	0.35	44.6

When we consider the status variable in Table 3, we see that it is highest in the high status treatment and lowest in the low status treatment. We used an ordered-probit analysis to check whether the donor status significantly differs between treatments. Our

[27] Thanks to Dean Karlan for this helpful advice.

analysis shows that there is a weakly-significant difference ($p<.1$) when comparing the low status treatment with the high status treatment. This result holds when we consider first donors only.[28] Other treatment-comparisons do not show any significant status-differences.

When we consider the gender variable in Table 3, we see that the mean is lower in the high status treatment. Hence, the fraction of women is higher in the high status treatment. A probit-analysis shows that there is a significant gender difference between the low status treatment and the high status treatment when considering all donors ($p<.05$). For first donors only, this difference becomes insignificant ($p=.159$). Other probit-analyses do not reveal any significant gender-differences between treatments.

When we consider the age of donors, we see that there are only small differences between treatments. Only in the high status treatment the mean age is somewhat higher. However, a Mann-Whitney U test reveals no significant difference between treatments.

Furthermore, we analyzed interaction between different characteristics (e.g. whether the increase of female donors in the high status treatment is only driven by *old* ladies). We find that the increase in the donor status in the high status treatment is entirely driven by male donors. While the status for female donors is similar in all treatments, an ordered-probit analysis reveals a highly significant difference ($p<.01$) between the low status and the high status treatment for male donors. This result holds when we consider first donors only.[29] Furthermore, a consideration of the absolute numbers of

[28] Similarly, Mann Whitney U tests reveal weakly significant differences between the low status and the high status treatment for all donors as well as for first donors only.

[29] Similarly, Mann Whitney U tests reveal highly significant differences between the low status and the high status treatment for all donors as well as for first donors only.

male donors reveals that the status treatments not only crowd in donors of similar status, but also crowd out donors of the opposite status.[30]

What do these figures tell us about passengers' motives to donate? While the crowd in of female donors in the high status treatment is in line with the standard status theory from Ball et al. (2001) and Kumru & Vesterlund (2010), the characteristics of male donors cannot be explained by these theories. More precisely, the standard status theories assume that individuals in general like to associate with those of higher status and therefore cannot explain why our high status first donor crowd out low status male givers. A possible explanation for the male donor pattern might be psychological theories of social comparison (Festinger 1954, Mussweiler 2003). These theories suggest that people are more willing to compare with similar people and when similarities between people exist they are more willing to assimilate. In our environment, assimilation equates imitating the donation decision.

5. Conclusion

In our fundraising field experiment, we analyzed giving behavior of metro passengers towards a beggar. Thereby, we systematically varied the first donor's status to test for the particular influence of high status individuals on train passengers' propensity to donate. We find that the first giver's status matter. When we installed a high status person instead of a low status person as the first giver, the number of donations rises by 34%. Furthermore, in line with results of previous literature, we find that individuals are conditional cooperative: As soon as we installed a (low status) person as first giver, the propensity of another donation in the train increases by 72% compared to the treatment where we did not install any giver.

[30] Number of low status male donors is 14 in the low status treatment, but only 2 in the high status treatment. The number of high status male donors is 17 in the high status treatment, but only 9 in the low status treatment.

Beside this core result, our data provide two additional insights. First, there is some evidence for a crowding in of low donations when giving is fostered by donations of our team members. But as explained in the respective section, our study does not allow pinning down this observation to a unique reason. Second, and more fruitful in terms of clear interpretations, there is significant evidence that male train passengers are more prone to donate when the previous donor has a similar status. This is of particular interest, as it contradicts the (standard) hypothesis which claims that individuals in general like to associate with those of higher status. Rather, it supports theories of social comparisons which suggest that individuals only imitate behavior of their peer group. However, for female donors this is not true. Instead, we observe an increase of female donors of all status groups in the high status treatment, which is in line with the standard status hypothesis.

References

Adler, Nancy E., Thomas Boyce, Margaret A. Chesney, Sheldon Cohon, Susan Folkman, Robert L. Kahn and S. Leonard Syme. 1994. "Socioeconomic status and health: The challenge of the gradient" *American Psychologist, 49 (1)*, 15-24.

Andreoni, James. 1998. "Towards a Theory of Charitable Fundraising." *Journal of Political Economy, 106(6)*, 1186-1213.

Andreoni, James. 2006. "Leadership Giving in Charitable Fund-Raising." *Journal of Public Economic Theory*, *8(1),* 1-22.

Ball, Sheryl, CatherineEckel, Philip J. Grossman and William Zame.2001. "Status in Markets."*Quarterly Journal of Economics, 116(1),* 161-188.

Besley, Timothy and MaitreeshGhatak.2008. "Status Incentives." *American Economic Review, 98(2)*, 206-211.

Boyd, Robert and Peter J. Richerson.2002. "Group Beneficial Norms Can Spread Rapidly in aStructured Population."*Journal of Theoretical Biology, 215*, 287–296.

Charles, Kerwin Kofi, Erik Hurst and Nikolai Roussanov.2009. "Conspicuous Consumption and Race."*Quarterly Journal of Economics, 124(2),* 425-467.

Chudek, Maciej, Sarah Heller, Susan Birch and Joseph Henrich. 2012. "Prestige-biased Cultural Learning: Bystander's Differential Attention to Potential Models Influences Children's Learning."*Evolution and Human Behavior, 33(1)*, 46-56.

DellaVigna, Stefano, John A. List and Ulrike Malmendier. 2012."Testing for Altruism andSocial Pressure in Charitable Giving." *Quarterly Journal of Economics, 127(1)*, 1-56.

Dutton, Diana B. and Sol Levine. 1989. “Socioeconomic Status and Health: Overview, Methodological Critique and Reformulation”, In J.P. Bunker, D.S. Gomby and B. H. Kehrer (Eds.), *Pathways to Health: The Role of Social Factors*, 29-69, Menlo Park, CA: Henry Kaiser Family Foundation.

Falk, Armin. 2007. ”Gift Exchange in the Field.“ *Econometrica, 75(5)*, 1501-1511.

Festinger, Leon. 1954. “A Theory of Social Comparison Processes” *Human Relations*, 7, 117-140.

Fischbacher, Urs and Simon Gaechter. 2010. “Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Good Experiments.”*American Economic Review, 100(1)*, 541–56.

Fischbacher, Urs, Simon Gachter and Ernst Fehr. 2001. “Are People Conditionally Cooperative? Evidence from a Public Goods Experiment.”*Economics Letters*, *71(3),* 397–404.

Frank, Robert H. 1984. “Workers Paid their MarginalProducts?” *American Economic Review, 74(4)*, 549-71.

Frank, Robert H. 1985. “The Demand for Unobservable and Other Nonpositional Goods”, *American Economic Review, 75(1)*, 101-116.

Frey, Bruno S. and Stephan Meier. 2004. “Social Comparisons and Pro-social Behavior - Testing ‘Conditional Cooperation’ in a Field Experiment.” *American Economic Review, 94(5),* 1717–22.

Harbaugh, William T. 1998a. “What Do Donations Buy? A Model of Philanthrophy Based on Prestige and Warm Glow.” *Journal of Public Economics, 67(2),* 269-284.

Harbaugh, William T. 1998b. “The Prestige Motive for Making Charitable Transfers.”*American Economic Review, 88(2)*, 277-282.

Harrison, Glenn W. and John A. List. 2004. "Field Experiments." *Journal of EconomicLiterature, 42(4),* 1009-1055.

Heffetz, Ori and Robert H. Frank. 2011. "Preferences for Status: Evidence and Economic Implications." In Jess Benhabib, Alberto Bisin and Matthew Jackson, eds.,*Handbook of Social EconomicsVol.1A*, North-Holland: Elsevier, 69-91.

Heffetz, Ori. 2011. "A Test of Conspicuous Consumption: Visibility and Income Elasticities."*Review of Economic Studies, 93(4)*, 1101-1117.

Henrich, Joseph and Francisco F. Gil-White.2001."The Evolution of Prestige: Freely Conferred Deference as a Mechanism for Enhancing the Benefits of Cultural Transmission." *Evolution and Human Behavior, 22*, 165–196.

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

Hopkins, Ed and Tatiana Kornienko. 2004. "Running to Keep in the Same Place: Consumer as a Game of Status."*American Economic Review, 94(4)*, 1085-1107.

Karlan, Dean and John A. List. 2012. "How can Bill and Melinda Gates Increase Other People's Donations to Fund Public Goods?" *Working Paper*.

Kocher, Martin G., Todd L. Cherry, Stephan Kroll, Robert J.Netzer and Matthias Sutter. 2008. "Conditional Cooperation on Three Continents."*Economics Letters, 101(3),* 175–78.

Kraus, Michael W. and DacherKeltner.2009. "Signs of Socioeconomic Status: A Thin Slicing Approach."*Psychological Science, 20(1)*, 99-106.

Kumru, Cagri S. and LiseVesterlund. 2010. "The Effect of Status on Charitable Giving" *Journal of Public Economic Theory, 12(4),* 709-735.

List, John A. and David Lucking-Reiley. 2002. “The Effect of Seed Money and Refunds on Charitable Giving: Experimental Evidence from a University Capital Campaign.” *Journal of Political Economics, 110(8),* 215-233.

Moldovanu, Benny, AnerSela and Xianwen Shi. 2007. “Contests for Status.” *Journal of Political Economy, 115(2)*, 338-363.

Thomas Mussweiler. 2003. “Comparison processes in social judgment: Mechanisms and consequences.” *Psychological Review*, 110(3), 472-489.

Oakes, J. Michael and Peter H. Rossi. 2003. “The measurement of SES in health research: current practice and steps toward a new approach” *Social Science and Medicine, 56*, 769-784.

Panshanathan, Karthik. 2010. “The Evolution of Prestige-Biased Transmission” *Working Paper*, Department of Anthropology UCLA.

Reuben, Ernesto and Arno Riedl. 2011. “Enforcement of Contribution Norms in Public good Games with Heterogeneous Populations” *Working Paper*.

Shang, Jen and Rachel Croson.2009.“A Field Experiment in Charitable Contribution: The Impact of Social Information on the Voluntary Provision of Public Goods.”*Economic Journal, 119(540),* 1422–1439.

Vesterlund, Lise. 2003. “The Information Value of Sequential Fundraising.”*Journal of Public Economics*, *87(3-4),*627-657.

Weiss, Yoram and ChaimFershtman. 1998. “Social Status and Economic Behavior: A Survey.”*European Economic Review, 42(3-5)*,801-820.

Appendix A: Pictures of Characters and Newspaper

Low status individuals first week

Low status individual second week

High status individual in both weeks

Homeless street-newspaper seller

The street newspaper

Appendix B: Classroom Status Survey

We asked participants of a classroom-experiment (N = 319) which took place in November2011 in a bachelor course at the Faculty of Management, Economics and Social Science of the University of Cologne on their impression of the socio-economic status of the low – and high - status individuals of our experiment. Each participant received one photo and had to estimate the level of education, occupational qualification, employment status and net income of the characters. N=102 received the photo of the low status individual of the first week, N=122 the photo of the low status individual of the second week and N=95 the photo of the high status individual.

Specification of categories:

a) Level ofeducation
    a. Certificate of Secondary Education (Hauptschulabschluss)
    b. General Certificate of Secondary Education (Realschulabschluss)
    c. General qualification for university entrance (Abitur)
b) Occupationalqualification
    1. None
    2. Apprenticeship (Berufsausbildung)
    3. University degree (Hochschulabschluss)
c) Employmentstatus
    1. Unemployed
    2. Part time job
    3. Tenure
d) Net income
    1. To 1000
    2. 1000 – 1500
    3. 1500 – 2000
    4. 2000 – 2500
    5. 2500 – 3000
    6. 3000 – 3500
    7. 3500 – 4000
    8. 4000+

Results:

As can be seen in figure A.1 and A.2 the high status person exceeds by far both low type persons in all categories. This picture is corroborated by a T-test. It shows significant differences between the high status person and both low status persons in all categories. Apart from the category of education the T-test show significant differences between low type persons, too. The low status individual of the first week is on average perceived to have a higher occupational status, a higher employment status and more income. However, even though this difference is significant it does not seem to be substantial (with regards to the figures below). This is especially true in comparison to the huge difference of the low status persons to the high status person.

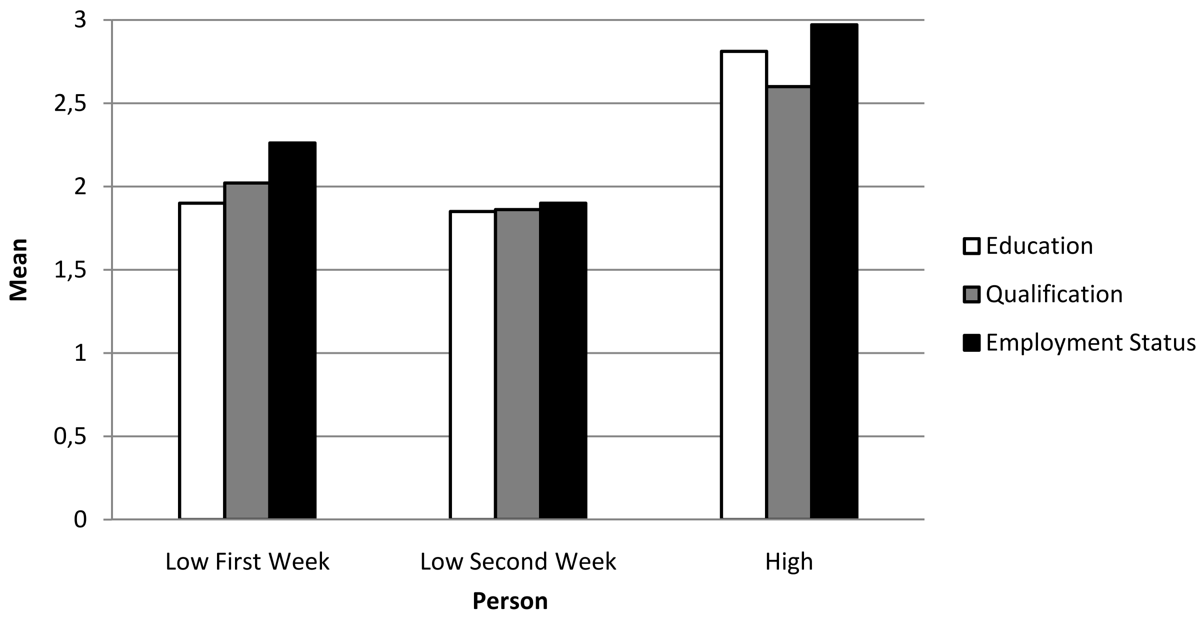


FIGURE A.1. MEAN OF ESTIMATED EDUCATION, OCCUPATIONAL QUALIFICATION AND EMPLOYMENT STATUS

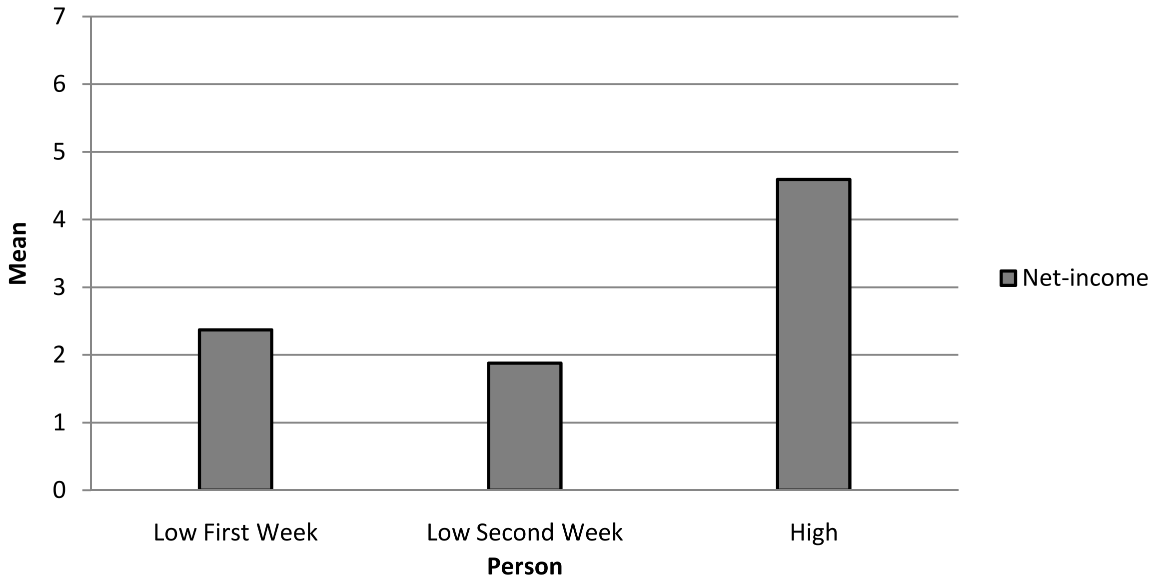


FIGURE A.2. MEAN OF ESTIMATED NET INCOME

Appendix C: Descriptive statistics

TABLE A.1. RIDES

	July	September	All
Control	87	97	184
Low Status	88	103	191
High Status	88	104	192
Total	263	304	567

TABLE A.2. RIDES WITH AT LEAST ONE DONATION

	July	September	All
Control	27	28	55
Low Status	44	49	93
High Status	52	63	115
Total	123	141	264

TABLE A.3. TOTAL NUMBER OF DONATIONS

	July	September	All
Control	43	39	82
Low Status	67	79	146
High Status	83	113	196
Total	193	231	424

TABLE A.4. TOTAL PROFITS

	July	September	All
Control	32.29€	41.33€	73.62€
Low Status	44.00€	65.78€	109.78€
High Status	51.65€	81.22€	132.87€
Total	127.94€	188.33€	316.27€

TABLE A.5. AVERAGE AMOUNT OF DONATION PER OBSERVATION

	July	September	All
Control	0.37€	0.43€	0.40€
Low Status	0.50€	0.64€	0.57€
High Status	0.59€	0.78€	0.69€
Total	0.49€	0.62€	0.56€

TABLE A.6. AVERAGE VALUE OF SINGLE DONATION

	July	**September**	**All**
Control	0.75€	1.06€	0.90€
Low Status	0.66€	0.82€	0.74€
High Status	0.62€	0.72€	0.68€
Total	0.66€	0.81€	0.74€

TABLE A.7. PASSENGERS

	July	**September**	**All**
Control	1539	1932	3471
Low Status	1436	2288	3724
High Status	1336	2021	3357
Total	4311	6241	10552

TABLE A.9. DAYTIME

	Rides		**Donations**		**Probability of Donation**	
	Morning	Evening	Morning	Evening	Morning	Evening
Control	96	88	36	19	37.50%	21.59%
Low Status	97	95	42	53	43.30%	55.79%
High Status	96	96	54	61	56.25%	63.54%
Total	289	279	132	133	45.67%	47.67%

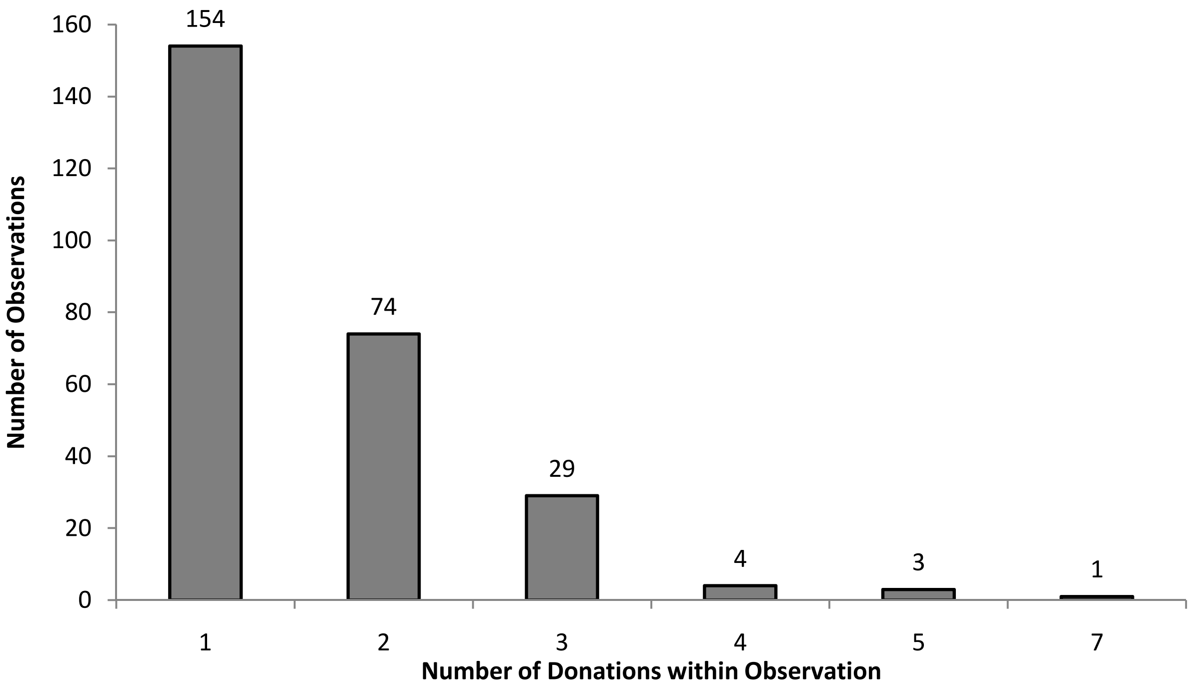


FIGURE A.3. NUMBER OF DONATIONS PER OBSERVATION WITH AT LEAST ONE DONATION

Appendix D: Additional Regression Analyses

TABLE A.10: EFFECTS ON THE NUMBER OF DONATIONS (interaction terms included)

	(1)	(2)	(3)
Low	0.461*** (0.124)	0.730** (0.315)	0.440** (0.181)
High	0.799*** (0.124)	1.013*** (0.310)	0.744*** (0.181)
Area*Treatment	No	Yes	No
Session*Treatment	No	No	Yes
Controls			
Session	Yes	Yes	Yes
Area	Yes	Yes	Yes
Daytime	Yes	Yes	Yes
Position	Yes	Yes	Yes
Passengers	Yes	Yes	Yes
Observations	567	567	567
Pseudo R-squared	0.048	0.048	0.055

Notes: Ordered probit regressions with donations per observation as dependent variable including treatment interactions with area dummies and session dummies. Standard errors in parentheses. Level of significance: *p<0.10, **p<0.05, ***p<0.01, 6 cut-points were estimated (output excluded).

TABLE A.11: EFFECTS ON THE PROBABILITY OF GIVING.

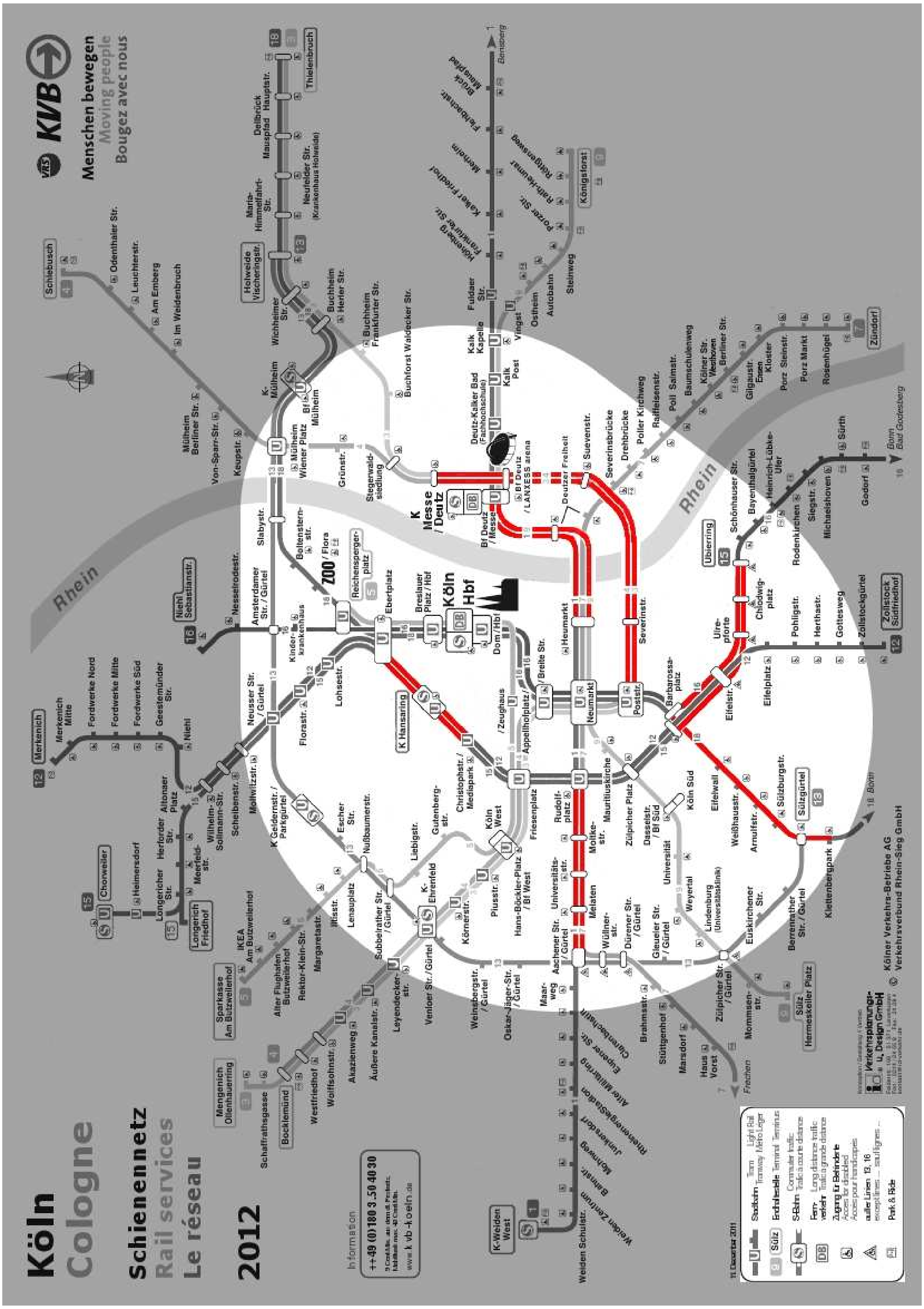
	(1)	(2)	(3)	(4)
Low	0.508***	0.508***	0.507***	0.501***
	(0.133)	(0.133)	(0.134)	(0.134)
High	0.778***	0.778***	0.786***	0.841***
	(0.134)	(0.134)	(0.134)	(0.137)
Session	No	-0.019	-0.020	-0.035
		(0.108)	(0.111)	(0.116)
Area 2	No	No	0.112	0.248
			(0.176)	(0.182)
Area 3	No	No	0.376**	0.420**
			(0.178)	(0.181)
Area 4	No	No	0.320*	0.333*
			(0.177)	(0.178)
Area 5	No	No	0.144	0.160
			(0.181)	(0.184)
Area 6	No	No	0.129	0.177
			(0.339)	(0.346)
Controls				
Daytime	No	No	No	Yes
Position	No	No	No	Yes
Passengers	No	No	No	Yes
Observations	567	567	567	567
Pseudo R-squared	0.045	0.045	0.054	0.075

Notes: Probit regressions with donations per observation as dependent variable. Standard errors in parentheses. Level of significance: *p<0.10, **p<0.05, ***p<0.01. Area 1: Aachener Str. /Gürtel – Rudolfplatz; Area 2: Poststr. – Koelnmesse; Area 3: Mediapark – Ebergplatz; Area 4: Barbarossaplatz – Chlodwigplatz; Area 5: Heumarkt – Bf Deutz; Area 6: Barbarossaplatz – Klettenbergpark

Appendix E: Variables used in the regression analysis

	Frequency	Percent
donation		
yes	264	46.56
no	303	53.44
treatment		
control	184	32.45
low	191	33.69
high	192	33.86
experiment		
July	263	46.38
September	304	53.62
daytime		
morning	289	50.97
evening	278	49.03
position		
back	293	51.68
front	274	48.32
passengers		
-10	98	17.28
11-20	266	46.91
21-30	152	26.81
31-	51	8.99
area		
1	99	17.46
2	120	21.16
3	110	19.40
4	117	20.63
5	108	18.17
6	18	3.17
Total	567	100.00

Appendix F: Map of Cologne Rail Services



Notes: We conducted the experiment on the red-marked lines.

Part 5

On the role of endowment heterogeneity and ambiguity on conditional cooperation

On the role of endowment heterogeneity and ambiguity on conditional cooperation

Felix Ebeling*

January 16, 2013

Abstract

Conditional cooperation (CC) is one of the most persistent behaviors in charitable giving. The laboratory experiment presented in this paper is designed to explore two questions: First, whether heterogeneous endowments of donors affect conditional cooperative giving. Second, whether potential donors exploit ambiguity about other donors' endowments in a self-serving manner to justify lower giving. We find that heterogeneous endowments affect giving in a way that suggests individuals concern for equality of donors' earnings after giving. Furthermore, the results do not confirm the exploitation of ambiguity about other donors' endowments. Individuals do not bias beliefs about other donors' endowments in a self-serving manner to justify lower giving.

JEL Classification: C91, D63, H41

Keywords: public good, donation, conditional cooperation, social norms, ambiguity

*Felix Ebeling, University of Cologne, Department of Economics, ebeling@wiso.uni-koeln.de. I also gratefully acknowledge the financial support from the German Science Foundation (through Gottfried Wilhelm Leibniz Price of the DFG, awarded to Axel Ockenfels).

1. Introduction

There is prevalent evidence that individuals' preferences for voluntary contributions in public good games or charitable giving strongly depends on others giving. Generally, individuals are more prone to give when others give. This phenomenon is called conditional cooperation (CC). Behavior in line with CC is observed in several public good laboratory experiments (Fischbacher et al. 2001, Kocher et al. 2008, Fischbacher & Gächter 2010) as well as in several field experiments on charitable giving (Frey & Meyer 2004, Shang & Croson 2009). Even if CC is not the only observed behavior peculiarity, it seems to be the most prevalent and robust behavior in these settings.

This paper explores how heterogeneous / asymmetric endowments of donors affect conditional cooperative giving. More precisely, the presented laboratory experiment is designed to test whether individuals show any concern for "equality of earnings" of donors when making their donation to a charity. It is intuitively appealing that individuals follow the idea that donors ought to have the same amount of money in their pocket after donating, or, to put it differently, that those with a higher endowment should donate more. However, existing economic research on laboratory public goods does not find such preferences.[1] The findings of Cherry et al. (2005), Buckley & Croson (2006) and Sadrieh & Verbon (2006) conflict with the idea of equality of earnings. Instead, in their laboratory public good games individuals with low endowments contribute the same absolute amount as individuals with high endowments. This is surprising, as fairness concerns are a robust phenomenon in many economic experiments (e.g. dictator games) and are also frequently debated in real-life negotiations on public good contributions. For example, the "fair contribution" of a

[1] As explained by Andreoni (2006), contribution to a charity out of altruism can be considered as contribution to a public good.

country has been among the most contentious topics in international negotiations on climate change. Even if laboratory public good games surely cannot reflect such complex real live negotiations, it is remarkable that preferences for equality of earnings are missed completely in abstract laboratory public good environments. Therefore, the laboratory experiment presented in this paper makes a new, methodically different, attempt to measure concerns for equality of earnings. With this method, which will be presented in detail in our experiment description in chapter three, we indeed find significant evidence for individuals concern for equality of earnings. However, also in our study the extent of this concern is rather small.

Furthermore, this paper explores whether potential donors self-servingly exploit ambiguity about other donors' endowment. We scrutinize whether participants that know the donation of others, but not their endowment, overestimate the latter to justify lower donations of their own. Despite the fact that we use the same method as Haisley & Weber (2010) to explore whether estimations / beliefs are self-servingly biased, we do not find any evidence that individuals bias estimations / beliefs in a self-serving manner in our environment. Considering existing literature, this is somewhat surprising as most studies find clear evidence for self-serving beliefs (e.g. Dunning et al. 1989 and Haisley & Weber 2010). However, the most closely related paper from Dahl & Ransom (1999) also finds onlymixed evidence for self-servingly biased beliefs.

The remainder of the paper is as follows: The next chapter explains the main research question in more detail and thereby reviews the relevant literature. Chapter three describes the experiment. Chapter four derives the hypotheses. Chapter five presents the results. Finally, chapter six discusses the methodological differences to previous studies and why we find different results than previous studies.

2. Relevant Literature

The first research question of this paper is how heterogeneous endowments of donors affect charitable giving. Thereby, we test whether individuals show any concern for "equality of earnings" of donors, when making their donation to a charity. There exists overwhelming literature about equity concerns in economic decision-making. From the methodological approach the papers by Reuben & Riedl (2011) and Nikiforakis et al. (forthcoming) are most closely related to our paper. Their studies scrutinize which normative rules individuals consider as *appropriate for others' behavior* in laboratory public good games and whether they enforce them by punishing deviators.[2] Our research considers whether individuals apply one of these rules, namely "equality of earnings", as *appropriate for their own behavior* when making their own decision with due regard to others' decisions. For example, individuals follow the rule of equality of earnings, when they donate less than donors with higher endowments, but more than donors with lower endowments. Such an internalized rule can be considers as a preference. For example, when donors follow the idea that all donors ought to have the same amount of money in their pocket after donating, they have a preference for equality of earnings. However, models concerned with inequity-averse preferences (e.g. Fehr & Schmidt 1999, Bolton & Ockenfels 2000) differ somewhat in their focus. These theories are concerned with decision situation where individuals can directly influence others' payment, e.g. in the dictator game the dictator can directly affect the receivers payment by giving him a certain amount of his endowment. As explained, for example by DellaVigna (2009), in the charitable giving context social preference theories focus on the interaction of donor and charity and not on the interaction between donors. Therefore, and due to the close relation to concepts scrutinized in the papers of Reuben

[2] For an excellent overview about possibly relevant norms in public good games, see Reuben & Ruben (2011).

& Riedl (2011) and Nikiforakis et al. (forthcoming), I stick with the term "concern for equality of earnings".[3] The meaning of this term is simple: Individuals' behavior follows the idea that individuals with a higher endowment should donate more, and vice versa, that individuals with a lower endowment belief they are "morally entitled" to donate less then individuals with a higher endowment.

Second, this paper explores whether potential donors self-servingly exploit ambiguity about other donors' endowment. The following example should clarify this thought: Assume a donor follows the above described 'equality of earnings' norm. Hence, she is interested in similar earnings for all donors after donations are done.[4] Furthermore, assume subjects only know the donations of others, but their endowment is not known. This is exactly the case in field experiments harnessing conditional cooperation to increase charitable giving (Frey & Meyer 2004, Alpizar et al. 2008, Shang & Croson 2009). If donors follow the 'equality of earnings' norm in this environment, they have to estimate other donors' endowment. Now, the basic question is whether donors self-servingly bias estimations about other donors' endowments. If so, in the given example donors overestimate others' endowment and manipulate their own "fair donation" downwardly in this way. This thought is in the spirit of Dana et al. (2007), who formulated on p.70: "Fair behavior is driven by comparisons against a standard, but that such a standard serves mainly as a constraint that individuals seek to circumvent rather than a goal that they seek to implement." Hence, individuals might 'outwardly' adhere to the concept of equality of earnings but 'secretly' try to circumvent the norm (via biased estimation). Several studies demonstrate self-serving interpretation of ambiguity. In the

[3] Previous literature of Buckley & Croson (2006) and Sadrieh & Verbon (2006) used social preference theories to analyze their results. In contrast to my experiment, they did not analyze a donation to a charity, but a classical public good game. Even if game structures are theoretically similar, in classical public good games interaction of contributors is more direct than in the charitable giving context. I will come back to this point in the final chapter.

[4] E.g. If other subjects' endowment is 10€ and others' contribution is 4€, her own contribution is 2€ in case her own endowment is 8€. So, in the end every subjects earning is equal to 6€.

psychological study of Dunning et al. (1989) individuals self-servingly assess their own abilities in case of ambiguity. Haisley & Weber (2010) demonstrate in an economic context that individuals interpret ambiguous risks in a self-serving manner. Similarly, Babcock & Loewenstein (1997) report various economic examples where people skew beliefs to line up with selfish interests. However, the study most closely related to ours by Dahl & Ransom (1999) only finds mixed results for self-serving beliefs. They test whether income situation and religious affiliation influence tithing. A tithe is a religiously motivated voluntary contribution equal to 10 percent of income. While individuals' income level does not affect their view on what represent income, religious affiliation does. Those with stronger affiliation to the church have a much more comprehensive view on what counts as income.

3. Experimental Design and Procedures

The experiment consists of three treatments. The design can be considered as a 2 + 1 treatment design. The focus will be on the 2-treatments. The +1-treatment was executed previously and its unique objective has been to collect data necessary to conduct the other 2-treatments. All treatments took place in autumn 2010. Participants were undergraduates from the faculty of Management, Economics and Social Science of the University of Cologne. Experiments were conducted in classrooms at the beginning of a course.

In the +1-treatment participants have to make one decision. At the beginning of a course 50 undergraduates received an endowment of 6€ and had the opportunity to donate to the German Red Cross. Decisions had to be made individually and talking was not allowed. On average students donated 2.98€.

The other two treatments were also conducted at the beginning of an undergraduate course. Treatments were conducted in the same classroom, but participants of different treatments were spatially separated within the classroom. All participants had to make two decisions. First, they had to estimate the endowment of the +1-treatment participants. Second, they received an endowment of 6€ and had the opportunity to donate to the German Red Cross. For the estimation, participants got to know the amount students donated on average in the +1treatment and had to deduce the endowment of the +1-treatment participants. Estimations were incentivized. If the endowment estimation differed less than 50 cent from the actual value, the subject earned 40 cent. If the estimation value differed less than 10 cent from the actual value, the subject earned 80 cent.

The 2-treatments differ in the *point of time* at which students receive information about the donation opportunity. Either they received the information at the beginning of the experiment or after they made their estimation about the behavior of the +1-treatment participants.[5] The different points of time at which students received information about their opportunity to donate were implemented by separation of instruction into different envelopes. In the "simultaneous" treatment (N=38), the complete instructions of the experiment were simultaneously given to the participants in one envelope. As soon as the experiment started and participants opened the envelope they knew that they had to estimate behavior of the +1-treatment participants and had the opportunity to donate. To make sure that participants indeed read the whole instruction, it was mentioned several times in the beginning of the experiment to read the whole instruction contained in an envelope before making any decisions. In the "sequential" treatment (N=40) instructions were separated into two envelopes to guarantee that

[5] As we will explain in the chapter four in detail, this difference changes incentives to bias estimations about +1-treatment participants' endowments.

participants received information sequentially.[6] The first envelope contained the request to estimate the behavior of the +1-treatment participants. The second envelope contained information on the opportunity to donate to the German Red Cross. After participants finished their estimation, they were instructed to put their estimation into the first envelope and close this envelope. Participants in this treatment were not allowed to open the second envelope until all participants had closed the first one.

4. Hypotheses

First, we consider the hypothesis concerning the effect of endowment heterogeneity on conditional cooperation. In our experiment, participants receive information on other donors' average donation and have to deduce their endowment. Furthermore, they have the opportunity to donate afterwards. Now, the basic assumption we made in our hypothesis H1 is that correlations in the experiment-results indicate concerns for equality of earnings. More precisely, we assume that, for the given own endowment and given average donation amount of other donors, a negative correlation between estimation of others' endowment and own donation implies concerns for equality of earnings.

> **H1.** *Individuals have a concern for equality of donor-earnings when making their donation - the higher their estimation about others' endowment, the lower their own donation.*

We are aware of possible false consensus effects (Ross et al. 1977) as found, for example, in the laboratory experiment of Selten & Ockenfels (1998). In our experiment individuals might look at their own behavioral inclinations in order to estimate others' behavior / endowment. However, this does not diminish the implications of the negative

[6] The procedure of "sequential" and "simultaneous" treatments is taken from Haisley & Weber (2010).

correlation between endowment estimation and donation for individuals concern for equality of earnings.

Second, we consider the hypothesis concerning the role of ambiguity about others' endowment. In our experiment, there are two different main treatments. In the simultaneous treatment, participants do know about the subsequent opportunity to donate while estimating other's endowment. In the sequential treatment, they do not know. Participants only have an incentive to upwardly bias their estimation in the simultaneous treatment, as it allows them to justify a lower giving in the subsequent donation decision. In the sequential treatment, there is no such incentive, simply because participants do not know about the subsequent donation opportunity (so there is no need to overestimate endowment for later justification of a low donation).

H2. *Individuals exploit ambiguity about other donors' endowments. Estimations about other's endowment will be higher in the simultaneous treatment than in the sequential treatment.*

5. Experimental Results

First, we find a significant negative correlation between estimation of other donors' endowments and own donation. Combining data of the 2-treatments, the coefficient of the Spearman nonparametric-test is -0.32 and highly significant ($p<0.01$). Even treatment-wise, we find a significant correlation coefficient of -0.32 for the sequential treatment ($p<0.05$), and a weakly significant correlation coefficient of -0.30 for the simultaneous treatment ($p<0.1$).

To substantiate our analysis, we additionally conduct regression analysis. We find significant negative linear correlation between estimation and donation. However, the regression coefficient of estimation is rather small. Furthermore, we find a significant treatment effect, that we cannot explain, but as the treatment-wise Spearman nonparametric-test shows significant negative regression, we do not think that this diminishes the basic message: The negative correlation between estimation and donation is a clear indication for individuals' concern for equality of earnings.

TABLE 1. LINEAR CORRELATION BETWEEN ESTIMATION AND DONATION

	Random effects regression		Tobit regression	
Model	(1.1)	(1.2)	(1.3)	(1.4)
Estimation	-0.09** (0.04)	-1.4** (0.64)	-0.12** (0.06)	-2.22** (0.97)
Treatment		-3.08** (1.26)		-4.81** (1.92)
Treatment × Estimation		0.33** (0.16)		0.54** (0.25)
Constant	-4.22*** (0.41)	15.87*** (4.87)	4.86*** (0.60)	23.17*** (7.47)
N	78	78	78	78
R^2 / *Pseudo* R^2	0.065	0.138	0.014	0.035

Notes: Random effects OLS Regressions with "donation" as dependent variable. */**/*** Significant at the 10/5/1 percent level. Standard errors in parentheses. Tobit models censor dependent variable on the left on 0 and on the right on 6.

Second, we do not find significant evidence for self-servingly exploitation of ambiguity about other donors' endowments. Comparing the distributions of estimations between "sequential" and "simultaneous" treatments reveals no significant difference. The p-value of two-sided Wilcoxon's rank-sum test is 0.1035.[7]

[7] One might argue that a one-sided p-value in case of an endowment estimation is weakly significant. However, a closer look at the data reveals that the results are driven by two outliers. Excluding these outliers leads to a p-value of 0.2014. Other results do not change significantly by exclusion of outliers.

6. Conclusion

First, our results suggest that individuals have a concern for equality of earnings. This result differs from previous studies by Cherry et al. (2005), Buckley & Croson (2006) and Sadrieh & Verbon (2006), but is intuitively appealing. A reason for our differing result might be our experimental method to measure equality concerns. For clarification, let us compare our setting with the experiment of Buckley & Croson (2006). In their laboratory public good game, in each (four player) group two players receive 25 tokens and the other two receive 50 tokens. But after each round individuals only receive information on average group giving. Hence, participants do not know what exactly the players with the same endowment or with the higher endowment earn. In contrast, in our two main treatments, participants hold a unique belief about average earnings of the +1-treatment participants. Individuals might be more inclined to include this (more unambiguous) information into their consideration for an appropriate donation. Alternatively, the framing might be the crucial difference. While cited literature considers classical laboratory public good games, our experiment uses a charitable giving context. Incentives might differ between contexts. For example, in laboratory public good games reciprocity concerns between players are a major incentive to contribute. Such concerns are most probably less relevant in our environment.

Second, we do not find evidence for self-servingly biased estimations. Considering existing literature, our results are somewhat surprising, but do not completely deviate. For example, Dahl & Ransom (1999) also find only weak evidence for the exploitation of ambiguity. It might simply be the case, that self-serving biased beliefs are a less persistent phenomenon. Alternatively, the norm we expected individuals would bias is not strong enough. We test whether individuals circumvent the equality of earnings norm. However, as our paper also shows, this norm, though identifiable, is not strongly pronounced. But if the norm is of minor importance, individuals do not have to exert effort to bend the norm; they can simply abandon it without high moral costs. Finally, estimation in our experiment might be over incentivized. Maybe for participants the chance to earn money with a precise estimation is more attractive than the opportunity to increase their earnings with a biased estimation.

References

Alpizar, Francisco, Fredrik Carlsson and Olof Johansson-Stenman. 2008. "Anonymity, reciprocity and conformity: Evidence from voluntary contributions to a national park in Costa Rica."*Journal of Public Economics*, 92(5-6), 1047–60.

Andreoni, James. 2006. "Philantrophy." In *Handbook of Giving, Altruism, and Reciprocity*, Vol. 2, ed. S.C.Kolm, and J.M. Ythier. Amsterdam, London & NY: Elsevier, 1201-65.

Babcock, Linda C. and George Loewenstein. 1997. "Explaining Bargaining Impasse: The Role of Self-Serving Biases." *Journal of Economic Perspectives*, 11(1), 109-26.

Bolton, Gary E. and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition."*American Economic Review*, 90(1), 166–93.

Buckley, Edward and Rachel Croson. 2006. "Income and wealth heterogeneity in the voluntary provision of linear public goods."*Journal of Public Economics*, 90(4-5), 935–55.

Cherry, Todd L., Stephan Kroll and Jason F. Shogren. 2005. "The impact of endowment heterogeneity and origin on public good contributions: evidence from the lab." *Journal of Economic Behavior & Organization*, 57(3), 357–65.

Dahl, Gordon B. and Michael R. Ransom. 1999. "Does where you stand depend on where you sit? Tithing donations and self-serving beliefs." *American Economic Review*, 89(4), 703–28.

Dana, Jason, Robert A. Weber and Jason Xi Kuang. 2007. "Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness." *Economic Theory*, 33(1), 67–80.

DellaVigna, Stefano. 2009. "Psychology and Economics: Evidence from the Field." *Journal of Economic Literature*, 47 (2), 315-372.

Dunning, David, Judith A. Meyerowitz and Amy D. Holzberg. 1989. "Ambiguity and Self-Evaluation: The Role of Idiosyncratic Trait Definitions in Self-Serving Assessment of Ability."*Journal of Personality and Social Psychology*, 57(6), 1082–90.

Fehr, Ernst and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 114(3), 817–68.

Fischbacher, Urs, Simon Gaechter and Ernst Fehr. 2001. "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment." *Economics Letters*, 71(3), 397–404.

Fischbacher, Urs and Simon Gaechter. 2010. "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Good Experiments." *American Economic Review*, 100(1), 541–56.

Frey, Bruno S. and Stephan Meier. 2004. "Social Comparisons and Pro-social Behavior - Testing 'Conditional Cooperation' in a Field Experiment." *American Economic Review*, 94(5), 1717–22.

Haisley, Emily C. and Robert A. Weber. 2010. "Self-serving interpretations of ambiguity in other-regarding behavior." *Games and Economic Behavior*, 68(2), 614–25.

Kocher, Martin G., Todd L. Cherry, Stephan Kroll, Robert J. Netzer and Matthias Sutter. 2008. "Conditional cooperation on three continents." *Economics Letters*, 101(3), 175–78.

Nikiforakis, Nikos, Charles Noussair and Tom Wilkening. 2012. "Normative Conflict and Feuds: The Limits of Self-Enforcement", *Journal of Public Economics*, 96 (9-10), 797-807

Reuben, Ernesto and Arno Riedl. 2011. "Enforcement of Contribution Norms in Public Good Games with Heterogeneous Populations." *Working Paper*.

Ross, Lee, David Greene and Pamela House. 1977. “The ‘False Consensus Effect’: An Egocentric Bias in Social Perception and Attribution Processes”, *Journal of Experimental Social Psychology*, 13, 279-301.

Sadrieh, Abdolkarim and Harrie A. A. Verbon. 2006. “Inequality, cooperation, and growth: An experimental study.” *European Economic Review*, 50(5), 1197–1222.

Selten, Reinhard, Axel Ockenfels. 1998. “An experimental solidarity game.” *Journal of Economic Behavior & Organization*, 34(4), 517-539.

Shang, Jen and Rachel Croson. 2009. “Field Experiments in Charitabel Contribution: The Impact of Social Influence on the Voluntary Provision of Public Goods.” *The Economic Journal*, 119(540), 1422–39.

Appendix: Instructions

+1treatment

You have been invited to participate in an experiment in the economics of decision making/ decision-making study/ experiment. At all times during the experiment the experimenter will answer any questions you may have. However, we ask you to refrain from talking to or in any other way communicating with other participants until all material has been collected at the end of the experiment.

You will be able to earn money during this experiment. Your respective amount will be cashed out to you after the next lecture. We will identify you only by your identification number, which you will find in the upper right hand corner of this sheet. Nobody, neither your fellow students nor the experimenter, will be able to match a decision to a particular person.

You are now allocated 6 Euros and may give any part of this money to the German Red Cross. Correspondingly, your payout from this decision will amount to the 6 Euros minus your specified donation.

This is not a hypothetical decision! Your specified donation will actually be donated to the German Red Cross. The accumulated amount of all donations made during this experiment will afterwards be transferred/ remitted to the German Red Cross by us.

Donation:

My donation to the German Red Cross: (correct to two decimal places please)

Other 2 treatments

You have been invited to participate in an experiment in the economics of decision making/ decision-making study/ experiment. At all times during the experiment the experimenter will answer any questions you may have. However, we ask you to refrain from talking to or in any other way communicating with other participants until all material has been collected at the end of the experiment.

You will be able to earn money during this experiment. Your respective amount will be cashed out to you after the next lecture. We will identify you only by your identification number, which you will find in the upper right hand corner of this sheet. Nobody, neither your fellow students nor the experimenter, will be able to match a decision to a particular person.

Please read the instructions to *both* decisions before making any choices.

Decision 1:

During a previous experiment 50 students of the Faculty of Management, Economics and Social Sciences were allocated an amount of X Euros. The participants had the possibility to donate any part of this money to the German Red Cross. Their respective payout thus amounted to X Euros minus their donation.

The participants donated an amount of 2,98 Euros on average. You are now asked to estimate the amount X, which the participants were allocated from the experimenter. The better your estimate, the higher your payout. You will be given 80 Cent, if your estimate diverges less than 10 Cent from the actual amount X and 40 Cent if your estimate diverges less than 50 Cent from the actual amount X. There will be no cashoutfor you from this part of the experiment, if your estimate diverges more than 50 Cent from the actual amount X.

Decision 2:

You are allocated an amount of 6 Euros and may give any part of this money to the German Red Cross. Correspondingly, your payout from this decision will amount to the 6 Euros minus your specified donation.

This is not a hypothetical decision! Your specified donation will actually be donated to the German Red Cross. The accumulated amount of all donations made during this experiment will afterwards be transferred/ remitted to the German Red Cross by us.

Estimate:

My estimate of the allocated amount X: (correct to two decimal places please)

Donation:

My donation to the German Red Cross: (correct to two decimal places please)

Part 6

Information Value and Externalities in Reputation Building

Information Value and Externalities in Reputation Building

Gary E. Bolton, Axel Ockenfels and Felix Ebeling*

18 March 2010

Abstract

In sequential equilibrium theory, reputation building is independent of whether the reputation builder is matched with one long-run partner or a series of short run "strangers". We observe, however, that reputation builders are significantly more challenged by long-run players in both laboratory chain store and buyer-seller games. Reputation builder behavior is more predictable than implied by equilibrium, and so reputation information has more economic value than implied by equilibrium. For short-run players, this reputation information value is an externality. For long-run players, the value of the information is internalized and so they have greater incentive to challenge the reputation builder.

JEL Classification: C71, C91, D83

Keywords: Cooperative Games, Reputation, Beliefs

* Correspondence to Gary Bolton, 334 Business Building, Smeal College of Business, Penn State University, University Park, PA 16802, USA; gbolton@psu.edu. Axel Ockenfelfs, University of Cologne, Department of Economics, ockenfels@uni-koeln.de. Felix Ebeling, University of Cologne, Department of Economics, ebeling@wiso.uni-koeln.de. We thank Ben Greiner, Torsten Weiland, and Brad Barnhorst for skillful research support, and Urs Fischbacher (1999) for z-Tree. We also thank two referees and seminar participants in Amsterdam, Bielefeld, Bonn, Boston, Cologne, Hagen, and Marseille for helpful comments. Bolton gratefully acknowledges the support of the National Science Foundation. Ockenfels gratefully acknowledges the support of the Deutsche Forschungsgemeinschaft.

1. Information Externalities in Reputation Building

Does the way that information about reputation disseminates in the market matter for the effectiveness of reputation building? Testing a reputation exposes the tester to risk, since the reputation builder may disappoint. So intuitively we might expect more testing of reputations in a market where long-run relationships are possible, because testers accrue future benefit from the production of this information. But in the sequential equilibrium theory of reputation building, this is not the case. Reputation is described as a matter of information, independent of the interaction pattern. In their seminal paper, Kreps & Wilson (1982, p. 266) observe that the same reputation building equilibrium applies regardless of whether the incumbent is protecting a chain store monopoly from entry by one, repeat challenger or a series of one-shot challengers. Or, in the context of a buyer-seller game with seller moral hazard, the same reputation building equilibrium applies regardless of whether the seller is building a reputation for trustworthiness with one, long-run buyer or a series of short-run buyers (Fudenberg & Tirole, 1991).

Part of the reason for why the pattern of reputation information does not matter in these models has to do with an equivalence in long and short-run challengers' decision problems: as long as the reputation builder's record is freely available and equally reliable, strangers and long-term associates are in the same position to reward the builder for defending his reputation or to punish him for deviating. But as we describe below, the argument also depends on conditions imposed along the equilibrium path, having to do with the reputation builder making himself unpredictable. Previous experimental studies (references below) find that human players often deviate from the equilibrium path in predictable ways. So we hypothesize that challenges to the builder's reputation will produce information with greater economic value than equilibrium implies for those who interact with the builder in the future. A long-run partner would then have a greater incentive to check the builder's reputation than would a short-run player, since the benefits of the information obtained are internalized in the former case while externalized – *information externalities* – in the latter. Reputation builders might then adapt their strategies in response. Interestingly, the pattern of adaptation should depend on the game because increased challenges make defending a reputation in the chain store game more costly, whereas they make defending in the buyer-seller game more profitable (less costly). By this hypothesis, information externalities produce a matching effect and so influence the way reputation information disseminates in markets and industrial organization.

In Section II, we briefly review the standard sequential equilibrium analysis of the chain store and the buyer-seller game, and describe how the two games differ with regard to the predicted economic value of reputation information. In Section III, we present a laboratory experiment on both games, to our knowledge the first controlled test for information externalities in reputation building.[1] We find evidence that information externalities and the pattern of agent interaction matter to the effectiveness of reputation building activities. Section IV concludes.

[1] Bolton, Katok and Ockenfels' (2004) study of Internet market feedback mechanisms maybe comes closest, even though they only looked at buyer-seller games and the experiment was exploratory, lacking various controls for the theory. Bolton et al. found evidence that partners are more likely than one-shot strangers to trust people who are new to buyer-seller transactions in the market.

2. Information Externalities in Theory: Irrelevance of Matching

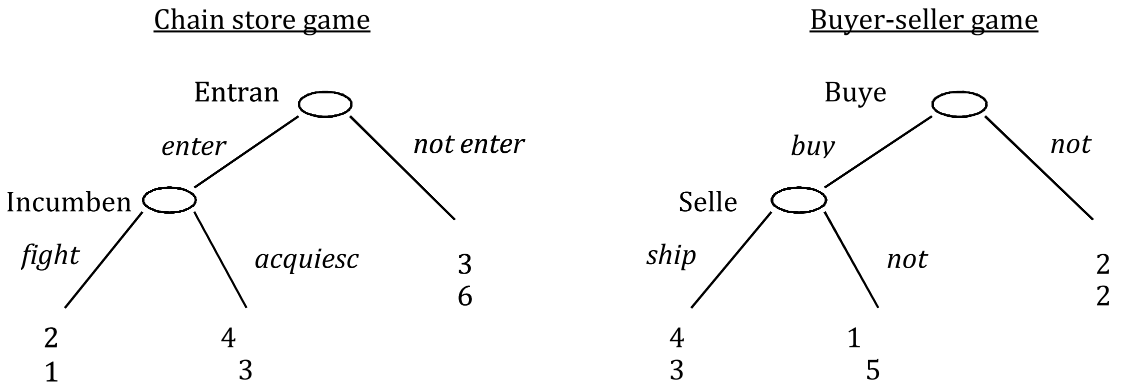


FIGURE 1. TWO REPUTATION BUILDING GAMES

We investigate information externalities with respect to the two reputation building stage games displayed in Figure 1 (the payoffs are consistent with the more general form of these games described in Appendix A). The chain store game is a specific case of the game studied by Kreps & Wilson (1982; see also Milgrom & Roberts, 1982), which in turn is a variant of a game introduced by Selten (1978). In each stage of this game, an entrant decides whether to enter the market of an incumbent monopolist. If the incumbent fights entry, it hurts the entrant but also hurts the incumbent relative to acquiescing. In each stage of the buyer-seller game, a buyer chooses to buy (by committing money) or not. The seller faces a moral hazard in that he is better off not shipping, keeping both money and good.

We suppose that the incumbent (seller) is the same player for all stages. Reputation is then introduced into these games by assuming there is a small probability δ, that, for the chain store game, the incumbent is "strong" in that he prefers fighting to acquiescing or, for the buyer-seller game, the seller is intrinsically trustworthy and prefers to ship than not to (Wilson, 1985). In the experiment, each game is played over 8 stages. Games are played in cohorts, such that in each cohort there are 8 entrants (buyers), 7 incumbents (sellers) and 1 'artificial' incumbent (seller) who is programmed to be always strong (trustworthy). It is commonly known that δ = 1/8. *Partner matching* refers to a game in which the entrant (buyer) is the same player for all stages. *Stranger matching* refers to a game in which the entrants (buyers) are randomly rematched from the pool of 8. Entrants (buyers) always receive information about the current opponent's play history before deciding. As is the convention, stages are numbered backwards: 8, 7, …, 1.

a. The Null hypothesis: Sequential Equilibrium Implies no Matching Effect

The aim of this subsection is to illustrate that the no-matching effect exhibited by the sequential equilibria in these games rests critically on the reputation builder disguising his type in the manner equilibrium prescribes.[2] Intuitively, the sequential equilibrium for these games unfolds in three phases, each phase characterized by how the reputation builder responds to challenges. For the chain store game: in Phase 1, with many stages still to go there is great advantage to the incumbent in having a reputation as being strong, and so the weak incumbent mimics the play of the strong incumbent, always playing fight if challenged. In Phase 2, reputation building is somewhat less advantageous and the incumbent plays a mixed strategy – he may play fight but also might reveal himself as weak (at which point he never plays fight again). Each of the first two phases typically lasts several stages. Phase 3 lasts exactly the final stage of the game: reputation no longer yielding any advantage, the incumbent reveals his true type if challenged. The three phases of the buyer-seller game are analogous.

Testing a reputation produces reputation information of economic value if possessing the information improves the possessor's continuation payoffs in the game; if so, we would expect different behaviors from short and long-run reputation testers (the payoffs of the latter are influenced by the continuation, the former are not). This is not

[2] We thank an anonymous referee for pointing out a mistake in an earlier draft.

what sequential equilibrium predicts, though. It is easy to see that Phase 1 of the equilibrium can produce no valuable reputation information, because all incumbents perfectly mimic the strong incumbent (and all sellers perfectly mimic the intrinsically trustworthy seller), so a challenge renders no information that can separate incumbent types. Phase 3 produces no valuable reputation information because it corresponds to the last stage of the game. The reason why no valuable reputation information is produced in Phase 2 is more involved, and can be grasped by considering the equilibrium conditions from the point of view of the entrant (buyer).

First consider the chain store game. Proceed by backwards induction: suppose the game has *proceeded to stage 1 (Phase 3*). For both partner and stranger matching, the entrant enters if

(2.1) $2p_1 + (1 - p_1)[2y_1 + 4(1 - y_1)] \geq 3$, where

p_n = prob(incumbent is strong in stage n);

y_n = prob(weak incumbent fights in stage n).

Because the weak incumbent has no incentive to fight in a single stage game $y_1 = 0$. It follows that it is optimal for the entrant to enter if $p_1 \leq 1/2$.

Now back up to stage 2 (and assume equilibrium play in stage 1). For the *stranger matching case*, there are no stage 1 continuation payoffs for the entrant to take into account. The entrant enters if

(2.2) $2p_2 + (1 - p_2)[2y_2 + 4(1 - y_2)] \geq 3$.

Reputation building goes on only when the entrant enters *and* the incumbent plays a mixed strategy. When the incumbent mixes, he chooses the probability weights such that (2.1) holds with equality and Bayes' rule, $p_1 = p_2/[p_2 + (1 - p_2)y_2]$, is satisfied. Combining conditions and substituting $p_1 = 1/2$, we find that $y_2 = 1/3$ and $p_2 = 1/4$.

Now consider stage 2 for the *partner matching case*, where stage continuation payoffs play a role in the entrant's decision. The entrant enters if

(2.3) $(2 + C)p_2 + (1 - p_2)[(2 + C)y_2 + (4 + E)(1 - y_2)] \geq 3 + D$, where

C = continuation payoff in case of no revelation of incumbent type

$= [2p_1 + 4(1 - p_1)]s_1 + 3(1 - s_1)$ with s_1 = entry probability in stage 1;

$D = 4(1 - p_2) + 2p_2$ as entrant always enters if no entry took place in stage 2;[3]

$E = 4$ continuation payoff in case of revelation of incumbent type.

As with the strangers case, we focus on the case where (2.3) holds with equality. On the equilibrium path $p_1 = 1/2$ and $C = 3$. Applying Bayes' rule and solving, we have $y_2 = 1/3$ and $p_2 = 1/4$. Hence the entrant enters under the same conditions as in the strangers matching case. The inclusion of future payoffs does not change entry incentives. The reader can verify that the conditions are unchanged on the equilibrium path even for the cases where the incumbent pursues a pure strategy (e.g., stages 8 to 4).

Continuing the backwards induction in the manner done for stage 2 shows that the conditions are unchanged for all 8 stages of the chain store game. In all cases, continuation payoffs from reputation information are neutralized along the equilibrium path. Figure 2, below, displays equilibrium frequencies of entry and fighting for each of the 8 stages.

For the *buyer-seller game* the equilibrium can also be derived using backwards induction. Again, suppose the game has proceeded to stage 1. The buyer buys, if

(2.1') $4q_1 + (1 - q_1)[4y_1 + (1 - g_1)] \geq 2$, where

q_n = prob(seller is intrinsically trustworthy in stage n);

g_n = prob(not intrinsically trustworthy seller ships in stage n).

[3] If no entry took place in stage 2, $p_1 = p_2 = \delta$. Because the entrant enters in stage 1 if $p_1 < 1/2$, he will enter in stage 1 if he chooses not to in stage 2 (see the general equilibrium conditions in Appendix A).

The non-intrinsically trustworthy seller will not ship in the last stage, hence $g_1 = 0$. The buyer buys if $q_1 \geq 1/3$.

Now back up to stage 2 (and assume equilibrium play in stage 1). For the *stranger matching* case, the buyer does not take continuation payoffs into account. He buys if

(2.2') $\quad 4q_2 + (1 - q_2)[4g_2 + (1 - g_2)] \geq 2.$

Reputation building goes on only when the buyer buys and the seller plays a mixed strategy.[4] In such a case, the seller mixes strategies such that (2.2') holds with equality. In the equilibrium $q_1 = 1/3$, $g_2 = 1/4$ and $q_2 = 1/9$.

For the *partner matching case* of stage 2, continuation payoffs are considered. The buyer now buys if

(2.3') $\quad (4 + C)q_2 + (1 - q_2)[(4 + C)g_2 + (1 + E)(1 - g_2)] \geq 2 + D$, where

C = continuation payoff in case of no revelation

$= [4q_1 + 1(1 - q_1)]s_1 + 2(1 - s_1)$ with s_1 = buy probability in stage 1;

$D = 2$ as buyer never buys if no buying took place in stage 2;

$E = 2$ continuation payoff in case of revelation of seller's type.

In equilibrium $q_1 = 1/3$, $C = 2$, $g_2 = 1/4$ and $q_2 = 1/9$. Again, buying conditions with and without inclusion of future payoffs lead to the same equilibrium results. Hence on the equilibrium path being matched with a long-run as opposed to a series of short-run strangers has no economic value. As with the chain store game, the reader can verify through induction that the result holds for all stages of the game. Figure 3, below, displays equilibrium frequencies of buying and shipping for each of the 8 stages.[5]

---

[4] Reasoning is similar to that for the chain store game.

[5] In other reputation environments, matching may play a more relevant role in equilibrium. Kreps & Wilson (1982) identify matching effects for the case of two-sided uncertainty. Fudenberg & Tirole (1991) survey studies on games with many long-run players with two-sided reputation building, where equilibria tend to be less robust (e.g., with respect to the exact nature of the incomplete information). Our observations are made in games with one-sided reputation building. Matching effects of a different sort are mentioned in work that examines equilibrium payoffs in repeated games with long horizons.

2.2 *Alternative hypothesis: Information Externalities produce a Matching Effect*

The above derivations illustrate that matching does not matter in equilibrium because reputation information is not produced in Phases 1 and 3, and the continuation payoffs that embody the economic value of reputation information in Phase 2 of the partner game are neutralized by the mixed strategies employed by the reputation builders. Given previous findings, however, we expect actual behavior to be somewhat off the sequential equilibrium path (e.g., Camerer & Weigelt, 1988, McKelvey, Neral & Ochs, 1992, Palfrey, 1992, Andreoni & Miller, 1993, Jung et al., 1994, and Brandts and Figueras, 2003).[6] Another line of experiments has shown that people have difficulties employing equilibria in mixed strategies, even in much simpler static environments. Importantly, there is typically more information leakage than predicted (e.g., Erev & Roth, 1998, O'Neill, 1987, and Shachat, 2002) although there is some evidence that professional athletes do better (e.g., Walker & Wooders, 2001).

The previous findings suggest the following alternative hypothesis: reputation builders will be more predictable than the sequential equilibrium implies. Anticipating this predictability, entrants (buyers) in the partner matching games will be more likely to test a reputation since they can capture the value of this information, unlike in stranger matching. Reputation builders should react, although they have differing incentives for doing so: with more intense entry, incumbents will find it more costly (relative to equilibrium) to defend a reputation and so should respond by doing *less* reputation building; that is, incumbents will be more likely to acquiesce. With a larger willingness to trust on the buyer side, a seller will find it more profitable to maintain a reputation,

Fudenberg & Levine (1989) analyze the interaction between short-run and long-run players, and Schmidt (1993) studies the case of two long-run players (see also Cripps & Thomas, 1995, and Cripps, Schmidt & Thomas, 1996, among others). In the context of his model, Schmidt mentions that if long-run players care about future payoffs, investing "in screening" the opponent might pay out in the long run. Such screening incentives are not present in our environment, though.

[6] Also see a recent paper by Grosskopf & Sarin (forthcoming) showing the sequential equilibrium predictions do better in reputation games when they agree with social preference implications.

thereby increasing sales, and so should respond by doing *more* reputation building; that is, they will be more likely to ship.

There is a second way in which chain store and buyer-seller games differ with respect to information externalities. From the previous subsection, observe that, in equilibrium, the continuation payoff the buyer receives is always equal to the payoff from not buying (from equation (2.3'), $C = E = 2$). That is, in equilibrium, the buyer faces the same expected payoff independent of whether the seller is revealed not intrinsically trustworthy. This is not true for the chain store game, where the equilibrium continuation payoff depends on whether the incumbent's type has been revealed or not; specifically the continuation payoff is higher if the incumbent is known to be weak than if not (from equation (2.3), C = 3 and E = 4). If reputation builder deviations from equilibrium are not too large, chain store entrants might then have relatively more incentive to challenge reputation builders than do buyers, and so the consequent matching effect would be larger in the chain store game than in the buyer-seller game.

3. Information Externalities in the Laboratory

In this section, we first describe the design of the experiment employed to test the null and alternative hypotheses developed in Section II. We then report the results of the experiment.

a. Design

Our experiment has a fully crossed 2×2 design (partners vs. strangers and chain store vs. buyer-seller game). In each of the four treatments, subjects play 20 sequences of 8 stage games. The chain store and buyer-seller games played are the same as those in Figure 1. In the partner treatments, there was no rematching within a sequence, while in the stranger treatments, players were rematched at random. In both partners and strangers, rematching *across* sequences was random. Also, regardless of matching scheme, first movers (either entrant or buyer) received full information about the current second mover's (incumbent or seller) play history within the current sequence before deciding. The history showed what move, if any, the second mover took in each of the preceding rounds (see Appendix C for the instructions used in the experiments).

Each of the four treatments of the experiment was run in two sessions. For each session there were 30 players, making for a total of 240 subjects. The 30 player groups were partitioned into two independent subgroups of 15 subjects, 8 first movers and 7 second movers; we then added an 'artificial' second mover to the pool of second movers, programmed to always fight or ship, respectively, which was public knowledge (similar to Neral and Ochs, 1992). Interaction was only within these subgroups, which was known to players, while which players were in the subgroup was not known.

The subjects were undergraduate students at the University of Jena and Cologne. At the beginning of the session, they read instructions and answered a questionnaire that checked their understanding of the rules. Actual matches were anonymous before, during and after the experiment. Subjects were paid a €2.50 show-up fee plus their earnings from all games. The average total payoff €17.9 (about $26.3 at the time of the experiment); the minimum earned was €14.1 and the maximum was €23.8.

b. Results

Figure 2 displays entry and fighting frequencies for chain store play during the reputation building phase of the game (that is, prior to any incumbent acquiescence). Observed frequencies are in black, expected equilibrium frequencies are in gray.[7] Not surprisingly, given what has been observed in earlier work, the quantitative fit with the

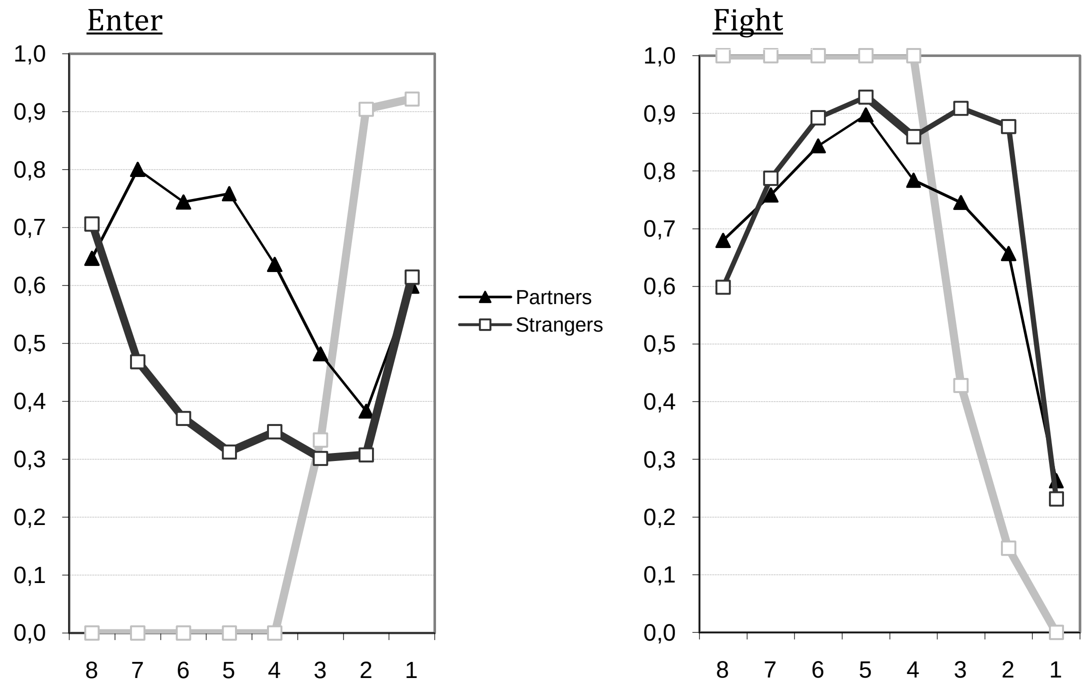
[7] Expected equilibrium frequencies are derived by Monte Carlo simulation (20,000 iterations) of the equilibrium equations in the Appendix.

theory is not good.[8] At the same time there is evidence for information externalities in that partners reputation builders are significantly more challenged than strangers reputation builders. For all stages, the frequency of entry weighted by the number of observations per stage averages 17.6% higher in partners than in strangers. At the same time, the frequency of fighting over the same stages is 2.3% lower, consistent with the increased tendency of entering.[9]

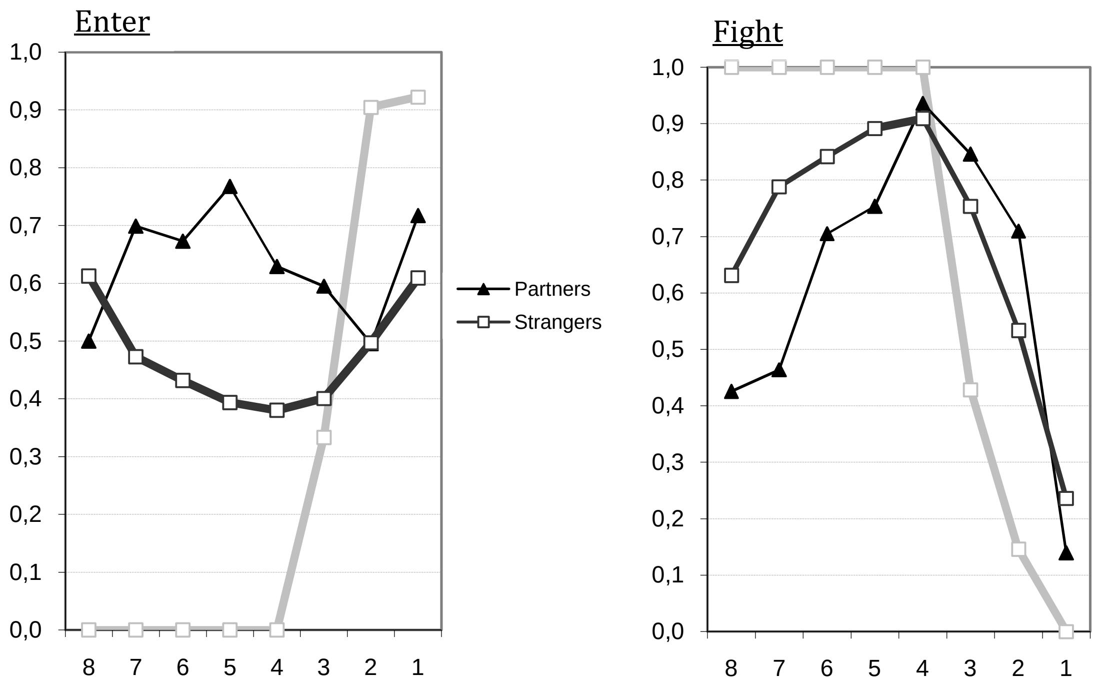
The data also shows some experience effects. For the first 10 sequences (the first half of the games played), the frequency of entry averaged 20.5% higher in partners than in strangers. For the second 10 sequences, the frequency of entry averages 14.5% higher in partners. With regard to fighting, the frequency is 4.4% *higher* for partner incumbents in the first 10 sequences, which then reverse to 11.7% lower for partners than strangers in the second 10 sequences. So, with experience, entrants enter more and incumbents fight less, consistent with our hypothesis concerning the influence of information externalities off the equilibrium path.

[8] Most strikingly, actual entry frequencies decline after stage 5 instead of increase as in theory. A closer examination of entrant behavior reveals a good deal of entrant heterogeneity. Nevertheless, for more than half the players there is a decreasing entry tendency after stage 5. One possibility is that entrants assume that, if the incumbent is not revealed as weak by a certain stage, they are facing either the strong incumbent or an extremely aggressive weak incumbent and do not enter any more. In fact, from the actual fight frequencies, this is not a bad assumption: 61% of the weak incumbents entered on by stage 5 reveal their type by stage 5.

[9] The 2.3 % may appear small given the values in the graph, and is largely due to the high entry frequency in the last stage of play.



A) First 10 game sequences



B) Last 10 game sequences

FIGURE 2. FREQUENCY OF ENTRY AND OF FIGHT BY STAGE, CONDITIONAL ON NO PREVIOUS ACQUIESCE

Notes: Path in black are frequencies from the data, averaged of individual movers. Data excludes robots. Path in gray are the expected equilibrium frequencies (excluding robots, too)

To check for statistical significance, we examine a probit model estimated from the reputation building phase data from both sequences of play. The model explains individual *i*'s action, y_{ijt}, in terms of the matching procedure (a dummy variable $Partners_i$), game stage (a dummy variable $Stage_j$ for stage *j*), sequence of play (t = 20, …,1 so that estimates of stage effects are for experienced players) and a random effect (u_i + v_{ijt}) to account for individual and session differences:

$$(3.1) \qquad z_{ijt} = \beta_0 + \beta_1 Partners_i + \sum_{j=2}^{8} \beta_j Stage_j + \beta_9 t + u_i + v_{ijt}$$

$$y_{ijt} = 1 \text{ if } z_{ijt} > 0\text{, and 0 otherwise,}$$

$$u_i + v_{ijt} \text{ error term.}$$

Estimates of the coefficient of the $Partner_i$ dummy, β_1, provide a baseline test for matching effects after experience, in the last sequence of play (recall that *t* is numbered in reverse). Estimating equation (3.1) with the entrant data, β_1 is significantly positive (two-tailed *p*-value = 0.015), while estimating with the incumbent data, β_1 is insignificant (two-tailed *p*-value = 0.377). The experience effects, as measured by the coefficients estimated for *t*, are much as implied by the descriptive statistics compiled in association with Figure 2, although the drop in entry is not significant. So, overall, there is some statistical evidence against the null hypothesis that matching is irrelevant. A full report of Model (3.1) estimates is given in Appendix C.

Model (3.1) is parsimonious but aggregates over potentially important information. A more detailed analysis is gotten by opening up (3.1) to test for information externalities in each stage of the game:

$$(3.2) \qquad z_{ijt} = \beta_0 + \sum_{j=1}^{8} \beta_j Partners_i Stage_j + \beta_9 Partners_i t + \sum_{j=2}^{8} \beta_{j+8} Stage_j + \beta_{17} t + u_i + v_{ijt}$$

Again, we estimate separate models for entrants and incumbents. The significance of information effects at each stage of the game again in the last sequence of play (that is, the significance of the $\beta_{8,}$ $_{\dots,}$ β_1 coefficients) are reported in Table 1. Most of the coefficients for entrants and incumbents are significant, and all those showing significance have the sign suggested by the alternative hypothesis.[10]

TABLE 1. RANDOM EFFECTS PROBIT ESTIMATES OF EQUATION (3.2)

Model	Chain store Enter		Chain store Fight		Trust game Buy		Trust Game Ship	
Dep var = *enter/fight/buy/ship*	Coeff	(StdErr)	Coeff	(StdErr)	Coeff	(StdErr)	Coeff	(StdErr)
Constant	0,639 ***	(.1861)	-1,997 ***	(.3163)	1,394 ***	(.2726)	-1,508	(.2732)
*partners_i*Stage_1"*	-0,001	(.2686)	-0,590	(.4552)	-0,895 **	(.3930)	0,082	(.3866)
*partners_i*Stage_2*	0,025	(.2655)	-0,580	(.4557)	0,049	(.3994)	-0,120	(.3100)
*partners_i*Stage_3*	0,560 **	(.2655)	-0,711	(.4588)	0,785 *	(.4313)	0,172	(.3128)
*partners_i*Stage_4*	0,825 ***	(.2650)	-0,967 **	(.4585)	5,651	(-370.1115)	0,749 **	(.3194)
*partners_i*Stage_5*	1,389 ***	(.2659)	-1,051 **	(.4581)	1,338 **	(.5258)	0,488	(.3094)
*partners_i*Stage_6*	1,014 ***	(.2623)	-0,936 **	(.4453)	1,589 ***	(.5470)	0,213	(.3114)
*partners_i*Stage_7*	0,903 ***	(.2575)	-1,057 **	(.4286)	5,559	(-347.8431)	0,473	(.3096)
*partners_i*Stage_8*	-0,396	(.2528)	-0,705 *	(.4197)	2,264 ***	(.5071)	1,266 ***	(.3238)
*partners_i*t*	0,007	(.0063)	0,044 ***	(.0115)	0,054 ***	(.0144)	0,043 ***	(.0085)
stage_2	-0,733 ***	(.1056)	1,454 ***	(.1778)	0,376 **	(.1736)	1,183 ***	(.2180)
stage_3	-0,976 ***	(.1069)	1,999 ***	(.2024)	0,804 ***	(.1857)	2,324 ***	(.2217)
stage_4	-0,927 ***	(.1055)	2,468 ***	(.2142)	1,302 ***	(.2066)	2,499 ***	(.2180)
stage_5	-0,974 ***	(.1047)	2,780 ***	(.2291)	1,443 ***	(.2110)	2,589 ***	(.2144)
stage_6	-0,789 ***	(.1026)	2,597 ***	(.2096)	1,380 ***	(.1968)	3,085 ***	(.2195)
stage_7	-0,530 ***	(.0986)	2,438 ***	(.1860)	1,383 ***	(.1919)	3,068 ***	(.2157)
stage_8	0,052	(.9456)	2,043 ***	(.1578)	0,567 ***	(.1526)	2,971 ***	(.2123)
t"	-0,006	(.0043)	0,021 **	(.0084)	-0,012	(.0075)	-0,048 ***	(.0060)
Rho	0,456	(.0495)	0,662	(.0543)	0,607	(.0669)	0,049	(.0553)
Number of obs	6578		2791		7116		5574	
Log-likelihood	-3387,5		-1157,9		-772,5		-1821,8	

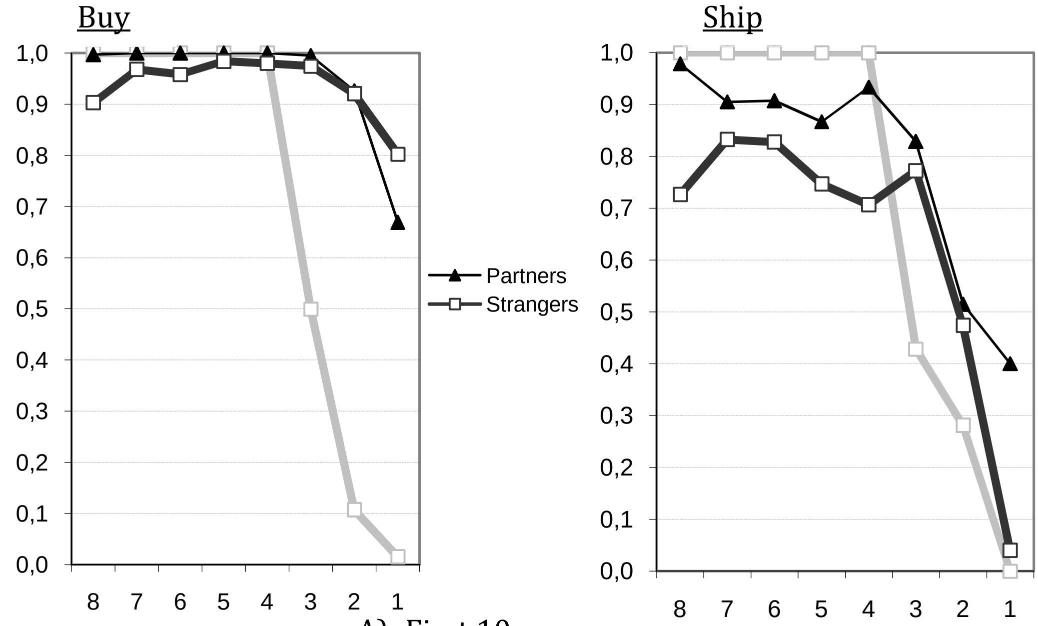
Notes: *, **, and *** indicate significance at the 10%, 5% and 1% level, respectively. t numbered in reverse of occurrence, so that the estimates of other variables are for experienced players.

Figure 3 suggests information externalities in the buyer-seller game as well, although smaller in magnitude than found in the chain store game. The buying frequencies are higher in all but the last two stages. For all sequences of play, the frequency of buying weighted by the number of observations per stage averages 1.8% higher in partners than in strangers. The shipping frequencies are higher in partners in all stages. Overall, the frequency of shipping weighted by the number of observations per stage is 8.8%

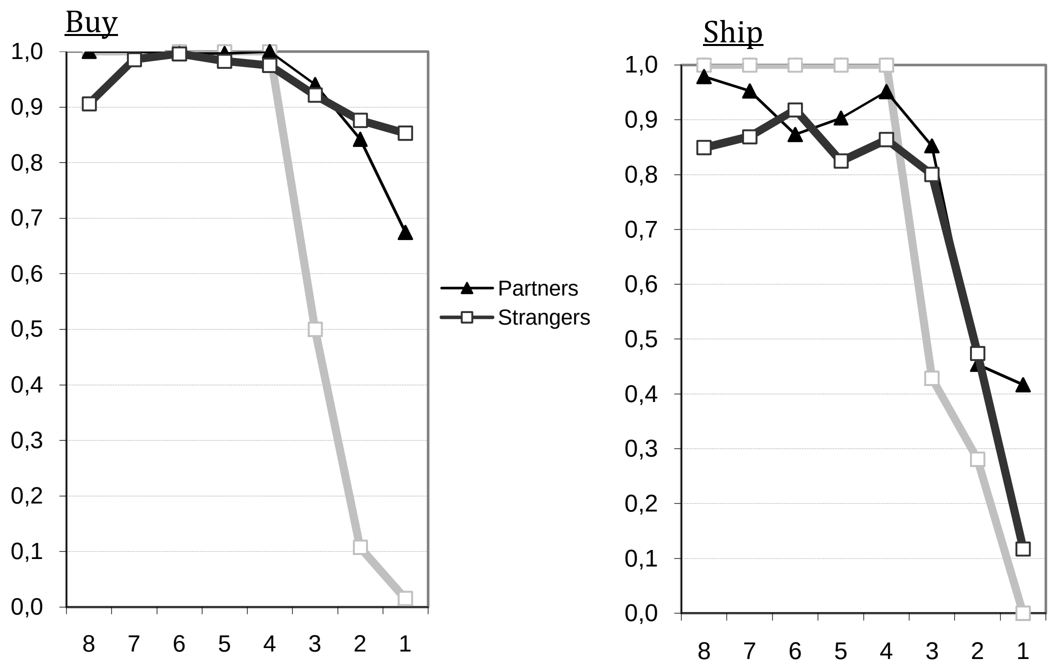
[10] Even if the interpretation of the model coefficients is not straightforward as is the interpretation for linear regression: The small coefficients of t indicate that behavior only change slightly for experienced players. Figures with the frequencies of the last five sequences are in the Appendix B and corroborate this result.

higher in partners than in strangers. For the last 10 sequenced (experienced play), the analogous values indicate 0.83% more buying in partners and 6.2% more shipping in partners. So there are no apparent differences with experience.

While the size of the effect for the buyer-seller game is modest, estimating Model (3.1) for this data finds they are significant (see Appendix D). For buyers, β_1 is positive with *p*-value = 0.024; for sellers, β_1 is positive with *p*-value = 0.001. Applying the more detailed Model (3.2) finds that almost all estimated stage effects (save for stage 1 of the shipping data) indicate both higher buying and shipping in the partners games, but most of the stage effects are insignificant, as shown in Table 1. It is only when aggregated across stages, as in Model (3.1), that information externality effects are evident for the buyer-seller game, after experience with play.



A) First 10 game sequences



B) Last 10 game sequences

FIGURE 2. FREQUENCY OF BUY AND OF SHIP BY STAGE, CONDITIONAL ON NO PREVIOUS FAILURE TO SHIP

Notes: Path in black are frequencies from the data, averaged of individual movers. Data excludes robots. Path in gray are the expected equilibrium frequencies (excluding robots, too)

Returning to Figures 2 and 3, observe that incumbent and seller behavior is more persistent than sequential equilibrium prescribes, in the sense that some reputation builders reveal themselves earlier than prescribed, while those who do not reveal themselves early are less likely to reveal themselves later as prescribed. Our alternative hypothesis suggests that this predictability should provide reputation testers (entrant or buyer) more incentive to challenge the builder's (incumbent's or seller's) reputation in the partner games than in the stranger games. Table 2 shows the actual expected payoffs of first movers (entrant or buyer) when challenging second movers (incumbent or seller) depending on the current reputation information (type revealed or not). The incentives implied by Table 2 are consistent with the alternative hypothesis and the observed behavior in the experiment. Specifically:

TABLE 2. EXPECTED PAYOFFS OF FIRST MOVERS (ENTRANT OR BUYER) WHEN CHALLENGING SECOND MOVERS (INCUMBENT OR SELLER) (ALL SEQUENCES)

		Chain store		Buyer-seller	
Stage	Type revealed	Partner	Stranger	Partner	Stranger
8	No	2.75	2.68	3.94	3.44
7	No	2.61	2.36	3.81	3.62
	Yes	3.77	3.96	3.00	2.05
6	No	2.32	2.22	3.71	3.70
	Yes	3.80	3.94	3.29	2.75
5	No	2.21	2.16	3.71	3.49
	Yes	3.73	3.94	3.71	2.48
4	No	2.23	2.18	3.86	3.54
	Yes	3.74	3.97	3.34	2.84
3	No	2.27	2.28	3.61	3.54
	Yes	3.81	3.94	2.44	2.69
2	No	2.34	2.53	2.78	2.93
	Yes	3.89	3.93	2.07	2.28
1	No	3.00	3.20	2.87	2.56
	Yes	3.83	3.98	1.12	1.55

Notes: Type revealed is *Yes* if incumbent (seller) acquiesced (did not ship) at least once before and *No* else. *Expected payoffs* are the expected first movers' payoffs in Euro computed on the basis of actual second movers' behavior including the artificial ones.

Regarding the chain store game, Table 2 shows a strong relationship between an incumbent's reputation and an entrant's expected payoff from entering. Since not entering yields a sure payoff of 3 to the entrant a stranger entrant is best off not entering in all but the last round whenever the incumbent's type is not yet revealed, but receives a higher expected payoff from entering whenever the incumbent has been revealed weak. The stage-by-stage numbers are similar for the chain store partner treatment but the situation is nevertheless quite different: since the incumbent is more likely to reveal himself weak in the early stages, the partnered entrant has a greater incentive to test the incumbent's reputation early, in order to reap the high expected payoffs from an incumbent who has revealed himself in later stages. For this reason a partner incumbent has less incentive to fight to protect his reputation in the early stages.

Regarding the buyer-seller game, Table 2 shows a similarly strong relationship between a seller's reputation and a buyer's expected payoff from buying. Because the payoff to not buying is 2, stranger buyers have an incentive to buy in all stages independent of whether the seller has revealed himself (with the exception of the last stage where buying from a revealed seller is a bad bet). Partnered buyers faced similar stage-by-stage expected payoffs from buying but nevertheless have a higher incentive to buy since early to capture the higher than equilibrium – and higher than not buying - continuation payoffs later in the game.

4. Summary

We show that the flow of information through the market can significantly influence the effectiveness of reputation building. The reputation building behavior we observe reveals more valuable reputation information than sequential equilibrium predicts. The value of this information is internalized in the partner matching case but becomes an externality in the strangers matching case. We observe partnered chain store entrants entering more often than corresponding stranger entrants; partnered incumbents fight less than corresponding stranger incumbents. For the buyer-seller game, partnered buyers buy more than stranger buyers, while partnered sellers ship more than stranger sellers.

While, so far as we are aware, no one has previously pointed out that information externalities play a role in the analysis of reputation building, information externalities have been recognized as a critical force in other areas. Porter (1995), for instance, describes how the strategic decision whether and when to explore oil fields are affected by drilling activities in neighboring areas. The results of such activities are publicly available and thus produce economically valuable information about oil field profitability to other firms, so that non-cooperative drilling may result in non-optimal exploration. Other examples include informational cascades and herding behavior, where a player's action may reveal economically valuable information about the state of nature to other players, resulting in too little information revelation (e.g., Chamley & Gale, 1994). Regarding reputation building, however, industrial economics and game theory textbooks often concentrate on stranger interaction only, mostly with the finite chain store market as an illustration. The partner case is not addressed, maybe because reputation theory has little to say about potential differences. However, our results suggest that the matching mechanism in these markets may play a significant role for the pattern of reputation building observed in these kinds of markets.[11]

Explaining why reputation information is more valuable than theoretically predicted seems to require new modeling approaches. In this regard, it is potentially instructive to point at where the current modeling approach falls down. One factor is that even a slight

[11] Of course, the results may have implications beyond our tested scenarios. For example, consider a repeated game in which in any stage a firm can choose to offer a high or low quality product and this quality is its private information. Our work suggests that the outcome depends on whether the firm sells to the same consumer or not.

deviation from the equilibrium strategies in a predictable way would lead to information externalities having value in a way that is consistent with the kind of matching effect we observe in our data. A second (not mutually exclusive) likely factor has to do with the fact that the current models are worked out only for quite specific configurations of payoffs, such that a single probability measure is a sufficient statistic by which to judge the reputation builder's history of play. An experiment can implement the required payoff structure only approximately since subject attitudes towards risk, inequality, etc., and other subjects' beliefs about these attitudes, cannot be entirely controlled for. Both factors suggest that actual reputation information need be characterized in a more subtle manner than current theoretical accounts.[12] The analysis here implies that a successful characterization will explain the economic value of reputation information and how it relates to the flow of this information through the market.

[12] See Bolton et al. (2009) for a similar conclusion reached from field data analysis involving the reciprocal nature of the process by which reputation information is produced.

References

Andreoni, James and John H. Miller. 1993. “Rational Cooperation in the Finitely Repeated Prisoner’s Dilemma: Experimental Evidence.” *Economic Journal*, 103, 570-85.

Bolton, Gary E., Ben Greiner and Axel Ockenfels. 2009. “Engineering Trust - Reciprocity in the Production of Reputation Information.” *Working paper*.

Bolton, Gary E., Elena Katok and Axel Ockenfels. 2004. “How Effective are Online Reputation Mechanisms? An Experimental Investigation.” *Management Science*, 50(11), 1587-1602.

Brandts, Jordi and Neus Figueras. 2003. “An Exploration of Reputation Formation in Experimental Games.” *Journal of Economic Behavior and Organization*, 50, 89-115.

Camerer, Colin and Keith Weigelt. 1988. “Experimental Test of a Sequential Equilibrium Reputation Model.” *Econometrica*, 56, 1-36.

Chamley, Christophe and Douglas Gale. 1994. “Information Revelation and Strategic Delay in a Model of Investment.” *Econometrica*, 62, 1065-1085.

Cripps, Martin W., Klaus M. Schmidt and Jonathan P. Thomas. 1996. “Reputation in Perturbed Repeated Games.” *Journal of Economic Theory*, 69, 387-410.

Cripps, Martin W. and Jonathan P. Thomas. 1995. “Reputation and Commitment in Two-Person Repeated Games without Discounting.” *Econometrica*, 63, 1401-19.

Erev, Ido and Alvin E. Roth. 1998. “Predicting how People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria.” *American Economic Review*, 88, 848-881.

Fischbacher, Urs. 1999. “z-Tree: A Toolbox for Readymade Economic Experiments.” *Working Paper No. 21*, University of Zurich.

Fudenberg, Drew and David K. Levine. 1989. “Reputation and Equilibrium Selection in Games with a Patient Player.” *Econometrica*, 57, 759-778.

Fudenberg, Drew and David K. Levine. 1992. “Maintaining a Reputation when Strategies are Imperfectly Observed.” *Review of Economic Studies*, 59, 561-579.

Fudenberg, Drew and Jean Tirole. 1991. Game Theory, The MIT Press: Cambridge, MA, Ch 9.

Grosskopf, Brit and Rajiv Sarin (forthcoming). ”Is Reputation Good or Bad? An Experiment.” *American Economic Review.*

Healy, Paul J. 2007. ”Group Reputations, Stereotypes and Cooperation in a Repeated Labor Market.” *American Economic Review*, 97, 1751-1773.

Jung, Yun J., John H. Kagel and Dan Levin. 1994. “On the Existence of Predatory Pricing: An Experimental Study of Reputation and Entry Deterrence in the Chain Store Game.” *The RAND Journal of Economics*, 25, 72-93.

Kreps, David M. and Robert Wilson. 1982. “Reputation and Imperfect Information.” *Journal of Economic Theory*, 27, 253-279.

McKelvey, Richard D. and Thomas R. Palfrey. 1992. “An Experimental Study of the Centipede Game.” *Econometrica*, 60, 803-836.

Milgrom, Paul and John Roberts. 1982. “Predation, Reputation, and Entry Deterrence.” *Journal of Economic Theory*, 27, 280-312.

Neral, John and Jack Ochs. 1992. “The Sequential Equilibrium Theory of Reputation Building: A Further Test.” *Econometrica*, 60, 1151-1169.

O'Neill, Barry 1987. “Nonmetric Test of the Minimax Theory of Two-Person Zero Sum Games.” *Proceedings of the National Academy of Sciences*, 84, 2106-2109.

Osborne, Martin J. 2004. An Introduction to Game Theory, Oxford University Press.

Porter, Robert H. 1995. “The Role of Information in U.S. Offshore Oil and Gas Lease Auctions.“ *Econometrica*, 63, 1-27.

Schmidt, Klaus M. 1993. “Reputation and Equilibrium Characterization in Repeated Games with Conflicting Interests.” *Econometrica*, 61, 325-351.

Selten, Reinhard. 1978. “The Chain Store Paradox.” *Theory and Decision*, 9, 127-159.

Shachat, Jason M. 2002. “Mixed Strategy Play and the minimax Hypothesis.” *Journal of Economic Theory*, 104, 189-226.

Waldman, Don E. and Elizabeth J. Jensen. 2001. Industrial Organization – Theory and Practice, Addison, Wesley, Longman.

Walker, Mark and John Wooders. 2001. “Minimax Play at Wimbledon.” *American Economic Review*, 91, 1521-1538.

Wilson, Robert 1985. “Reputations in Games and Markets.” Game-Theoretic Models of Bargaining, A. Roth (ed.), Cambridge University Press, 27-62.

Appendix A. Sequential equilibria for chain store and buyer seller games

We state the equilibria for the general game forms:

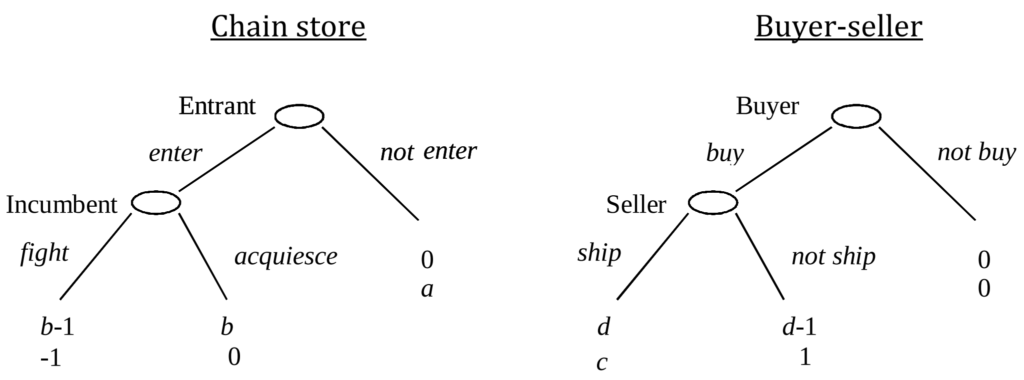


FIGURE A. BASE GAMES WITH PAYOFF STRUCTURE: a>1, 0<b, 0<c, d<1

The payoff structure of the games studied in the experiments are equivalent to those shown in Figure A up to affine transformation. Stages for all games are labeled in descending order: $n = N, \ldots, 1$.

Sequential equilibrium for the chain store game, both partners and strangers matching (Kreps and Wilson, 1982). For the chain store game in Figure 1, $b = 0.5$ and $a = 1.5$.

1) Let p_n= prob(incumbent is strong at the beginning of stage n). Set $p_N = \delta$.
2) Define $b_n = b^n$, where b is the payoff as stated in Figure A.

Updating p_n:

3) For $n < N$: If there is no entry in stage $n + 1$ then $p_n = p_{n+1}$. If there is entry in stage $n + 1$, and either this is met with acquiesces or $p_{n+1} = 0$, then $p_n = 0$.
4) For $n < N$: If there is entry in stage n+1 followed by fighting and $p_{n+1} > 0$, then $p_n = \max\{b_n, p_{n+1}\}$.

Incumbent's strategy (conditional on entry):

5) If $n = 1$, acquiesce.
6) If $n > 1$ and $p_n \geq b_{n-1}$, fight.

7) If $n > 1$ and $0 < p_n < b_{n-1}$, fight with probability. $\frac{(1-b_{n-1})p_n}{(1-p_n)b_{n-1}}$

8) If $n > 1$ and $p_n = 0$, acquiesce.

Stage n entrant strategy:

9) If $p_n > b_n$, stay out. If $p_n < b_n$, enter.

10) If $p_n = b_n$, stay out with probability $1/a$.

Sequential equilibrium for the buyer-seller game, both partners and strangers matching.

For the buyer-seller game in Figure 1, $d = 0.667$ and $c = 0.334$.

1) Let q_n = prob (seller is intrinsically trustworthy in stage n). Set $q_N = \delta$.

2) Define $d_n = d(1-d)^{n-1}$, where d is the payoff as stated in Figure A.

Updating q_n:

3) For $n < N$: If there is no buy in stage n+1 then $q_n = q_{n+1}$. If there is buying in stage n+1, and either this is met with no ship or $q_{n+1} = 0$, then $q_n = 0$.

4) For $n < N$: If there is buying in stage n+1 followed by shipping and $q_{n+1} > 0$, then $q_n = \max\{d_n, q_{n+1}\}$.

Shipper's strategy (conditional on buying):

5) If $n = 1$, do not ship.

6) If $n > 1$ and $q_n \geq d_{n-1}$, ship.

7) If $n > 1$ and $0 < q_n < d_{n-1}$, ship with probability. $\frac{1-d-q_n}{1-q_n}$

8) If $n > 1$ and $q_n = 0$, do not ship.

Stage n buyer strategy:

9) If $q_n < d_n$, do not buy. If $q_n > d_n$, buy.

10) If $q_n = d_n$, buy with probability $1 - c$.

Appendix B. Experimental Procedure [Translation of the Chain Store Game instructions from German; Buyer-seller Game instructions are analogous.]

Instructions This is an experiment in decision making. The German Science foundation has provided funds for this research.

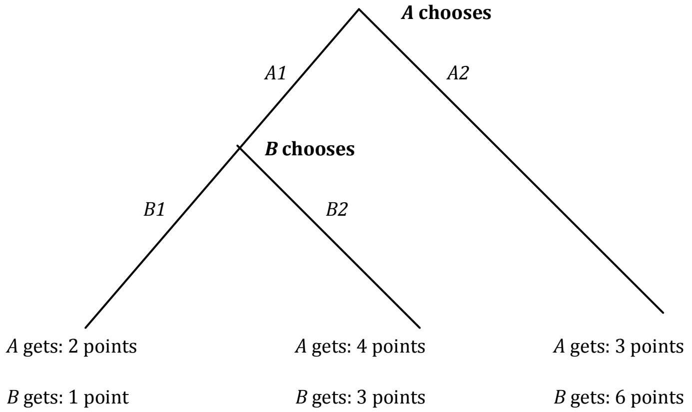
Each decision maker has been randomly assigned to be a member of one of two groups with 15 subjects each. Each group will play separately; there will be no interaction between them.

Each subject is assigned the role of an *A*-subject or a *B*-subject. The assignments will be the same for the whole session. Whether you are *A* or *B* will be determined randomly and shown on your computer screen once the experiment starts.

The decision situation

The experiment is divided into a series of 20 sequences. A sequence consists of 8 rounds. In each round, an *A* subject will be paired with a *B*-subject. Each round will proceed as follows. Each *A*-subject begins the round by choosing one of two alternatives. These alternatives are labeled *A1* and *A2*, respectively. If *A1* is chosen, *B* has to choose between alternatives *B1* and *B2*. If *A2* is chosen, *B* has no choice.

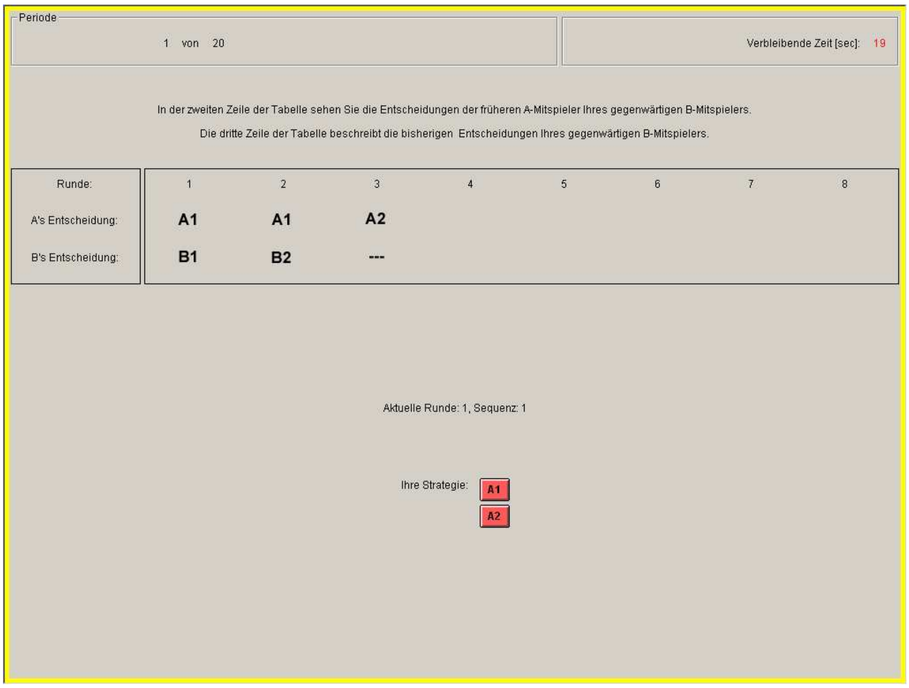
In each round, you can earn points according to the decisions made in this round. 33 points are worth 1 Euro, and all points are paid in cash along with your show-up fee at the end of the experiment. If *A* chooses *A1* and *B* chooses *B1*, then *A* gets 2 points and *B* gets 1 point. If *A1* and *B2* is chosen, *A* earns 4 points and *B* earns 3 points. If, finally, *A2* is chosen, then *A* gets 3 points and *B* gets 6 points. The following figure summarizes the payoff rules:



What is the matching procedure? [Partners; analogous for Strangers] Before each sequence (which consists of 8 rounds of the above described decision situation) you will be randomly paired with a new subject who is assigned the other role. Within a sequence, you are matched with the same opponent for all 8 rounds. The identity of your opponent, however, will not be revealed to you, neither during nor after the session.

Please notice that there are 8 *A*-subjects within your group, but only 7 *B*-subjects. The missing eight *B*-subject is a computer agent who is programmed to always choose *B1*. That is, if you are an *A*-subject, you might be randomly matched with an artificial *B*-subject (which happens with probability 1/8) who is programmed to choose *B1* whenever you choose *A1*.

Sequences Before making a choice, all *A*-subjects get a summary of the *B*-subject's decisions in the earlier rounds of the current sequence.



In this fictitious example, *A* is informed that *B* chose *B1* in round 1 and *B2* in round 2. In round 3, *B* had no choice, because *A* chose *A2*. (If *B* is our programmed computer agent, the history will, of course, never display *B2*.)

Summary

- This experiment consists of 20 sequences each consisting of 8 rounds. In each round, you will face the same decision situation as described above.
- Before each sequence, you will be matched with a new opponent. Within a sequence, however, you will be always matched with the same opponent. The identity of your opponent will not be revealed.
- One of the 8 *B*-subjects is a programmed computer agent. This agent will always respond to *A1* with *B1*.
- Before the *A*-subject is asked to make a decision, he will be informed about the behavior of the *B*-subject in the earlier rounds of the current sequence.

- All earned points will be summed up and paid in cash at a conversion rate of 33 points = 1 Euro at the end of the experiment.

If you have any question, now or during the experiment, please raise your hand and the monitor will be right with you.

Questionnaire. This questionnaire tests whether you fully understood the instructions. The experiment can only start when all subjects correctly answered all questions.

1. A sequence consists of how many rounds?
    a. 8
    b. 15
    c. 20
2. Within a sequence I'll be matched …
    a. always with the same opponent
    b. never more than once with the same opponent
    c. always with the programmed computer agent
3. The probability that an *A*-subject is matched with the programmed computer agent is …
    a. 1/10
    b. 1/4
    c. 1/8
4. If an *A*-subject observes that *B* chose *B2* he knows for sure that this *B*…
    a. is the programmed computer agent
    b. cannot be the programmed computer agent

c. neither a. nor b.

5. If *A* chooses *A1* and *B* chooses *B2*, than *A*'s payoff is:

   a. 1

   b. 3

   c. 4

6. Before making a choice, each *A*-subject receives information about the choices made by *B* in earlier rounds of the same sequence.

   a. true

   b. wrong

   c. not decidable

Appendix C. Estimates of Probit Models

Table C1. Random effects probit estimates (p-values) of Equation 3.1

	Chain store		*Buyer-seller*	
y_ijt = 1 if	*Enter*	*Fight*	*Buy*	*Ship*
Constant	0.38	-2.15	0.62	-1.84
	(0.027)	(0.000)	(0.005)	(0.000)
Partners_i	0.54	-0.34	0.67	-0.86
	(0.015)	(0.377)	(0.024)	(0.001)
*Stage_2**	-0.72	1.48	0.79	
	(0.000)	(0.000)	(0.000)	(0.000)
Stage_3	-0.73	2.00	1.48	2.31
	(0.000)	(0.000)	(0.000)	(0.000)
Stage_4	-0.57	2.33	2.22	2.77
	(0.000)	(0.000)	(0.000)	(0.000)
Stage_5	-0.39	2.55	2.25	2.73
	(0.000)	(0.000)	(0.000)	(0.000)
Stage_6	-0.35	2.44	2.20	3.07
	(0.000)	(0.000)	(0.000)	(0.000)
Stage_7	-0.13	2.19	2.25	3.20
	(0.078)	(0.000)	(0.000)	(0.000)
Stage_8	-0.18	2.01	1.46	3.32
	(0.008)	(0.000)	(0.000)	(0.000)
*T**	0.00	0.04	-0.00	-0.03
	(0.622)	(0.000)	(0.95)	(0.000)
Rho	0.43	0.66	0.52	0.48
	(0.000)	(0.000)	(0.000)	(0.000)
Number of observations	6578	2791	7116	5574
Number of individuals	64 entrants	56 incumbents	64 buyers	56 sellers
Log-likelihood	-3528.96	-1170.26	-851.74	-1867.59
Chi-squared p-value	0.000	0.000	0.000	0.000

*Numbered in reverse order of occurrence.

Erklärung

Ich erkläre hiermit, dass ich die vorgelegte Arbeit ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Aussagen, Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Im Folgenden werden für die nicht von mir allein verfassten Artikel die Beiträge der einzelnen Autoren kurz erörtert.

Peer Pressure and Multi Tasking

Grundidee & Theorien: Ebeling

Experimentdesign: Ebeling & Fellner

Experimentdurchführung: Wahlig

Auswertung & Niederschrift: Ebeling & Fellner

Follow the Leader of Follow Anyone

Grundidee & Experimentdesign: Ebeling

Experimentdurchführung: Fendrich

Auswertung: Feldhaus

Niederschrift: Ebeling

Information Value and Externalities in Reputation Building

Experimentdurchführung: Ebeling

Auswertung: Bolton & Ebeling

Rest: Bolton & Ockenfels

Weitere Personen neben den aufgeführten Koautorinnen und Koautoren waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt. Ich versichere, dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe.

Felix Ebeling