UNIVERSITÄT
BERN

Experimental evidence on human choices in organizations and markets

Inauguraldissertation zur Erlangung der Würde eines

Doctor rerum oeconomicarum

der Wirtschafts- und Sozialwissenschaftlichen Fakultät

der Universität Bern

vorgelegt von

Jonas Gehrlein

Bern, Juni 2019

In loving memory
of
my father & mother

Acknowledgements

First and foremost, I would like to thank my first supervisor, Frauke von Bieberstein, for her continuous support and guidance to finish this dissertation. Her professionalism and experience were reflected in many valuable discussions about research, and communication with companies. I have learned much from her, from which I benefited in my professional and personal lives. Personally, I felt a special connection to her, the institute, and the people working here right from the first day. Frauke is a great supervisor, and she generously provided me with all the necessary freedom and support to act on my broad research interest.

Second, I would like to thank Holger Herz, for dedicating his valuable time evaluating my thesis. Additionally, I would like to thank him for the numerous well-organized summer and winter schools, where he was a continuous source of inspiration. On these occasions, we had a great time hiking and skiing.

Third, I could not have enjoyed my time in Bern without great colleagues. For such a wonderful time, I especially thank Kathrin Friedrich, Stefanie Jaussi, Sebastian Berger, Andrea Essl, and Jessica Niedermair. There was always time to discuss research or personal matters, and many problems were solved this way. In addition, I would like to express gratitude to Andrea Probst, for all the fun we had joking around. Also, she solved every administrative hurdle I caused, and I always felt in good hands when dealing with regulatory matters. I would also like to give thanks to a good old friend, Ilija Braun, for all the great on-topic but especially off-topic, conversations.

Fourth, special thanks goes to Ann-Kathrin Crede, the best office neighbor imaginable, and a good friend. Aside from our great scientific work together, she was a source of motivation to do workouts during the lunch time or even before work. Our bilateral peer pressure made it possible to master the challenge to go (almost) weekly in the Aare during the 2018/19 winter. I enjoy being around a person who is up for any crazy idea, and therefore, would not have wished for anybody else to share office space with.

Above all, I would like to express my deepest gratitude to my partner Sandra, who always finds the time to support me in everything I do, or any funny idea I come up with. I do not know any other person who is more caring, has more energy, and is more reliable than her. I admire her strength, and I am happy to be at her side.

Bern, June 2019

Jonas Gehrlein

Contents

List of Figures

List of Tables

Executive Summary

Managerial and Behavioral Economics have received increasing attention in academia and in the private sector. For example, companies are incorporating behavioral findings into actively designing the environment for their employees, to account for social preferences, such as positive and negative reciprocity, or to enhance individual or team performance. Many of those approaches are successful, despite lacking a monetary incentive, and thus, conflict with standard economic theory. Therefore, fundamental behavioral research, as presented by three essays in this thesis, is important for uncovering basic human decision-making mechanisms. These essays utilize the experimental method, which offers control of various confounding factors, enabling the identification of causal relationships. The first three essays challenge rational economic models, and show results that can be explained only by incorporating behavioral theories. The last essay is a methodological approach to further develop experimental software. The common factor of all studies is an interdisciplinary perspective: Essay 1 investigates a longstanding question of Economics, pairing it with a mechanism prominent in Psychology, and measuring an outcome inspired by Philosophy. Essays 2 and 3 go beyond the model of the *Homo oeconomicus* to analyze teamwork and the role of intentions. Essay 4 concludes this dissertation, by including aspects of Computer Science to illustrate how current software can be extended to allow for interactive online experiments, which will be increasingly relevant in the future.

In essay 1, we investigate how the market mechanism influences moral decision making. This question is a longstanding one, which is receiving increasing attention due to our rapidly progressing society. We are driving towards a global economy, where markets are penetrating more and more aspects of daily life. For example, our social lives are dragged toward online platforms, which are launching efforts to monetize social interactions in the forms of likes and followers, and selling products in the process. Thus, it is important to understand the implications of an increasingly market-centric society for our moral standards. In a large online experiment, we expose a non-standard subject pool to either a market or a non-market condition, and elicit decisions about a subsequent moral dilemma. We hypothesize that markets foster a cost-benefit analysis mindset, which materializes in changing behavior in the dilemma. In comparison to the baseline, and in line with our hypothesis, we find a substantial effect after the participants are exposed to the market game. However, the non-market control setting shows a similar increase, and thus, excludes a treatment effect.

In essay 2, we analyze how social preferences translate into output in a modified teamwork setting. In most firms, the product of a teamwork situation is essential for successful operation. However, in a large number of cases, teamwork itself is a prerequisite for future individual performance. Examples involve helping on the job, internal knowledge sharing, or peer coaching. Those scenarios are still neglected in the literature. Therefore, we investigate how social preferences translate in such modified teamwork settings into future individual performance. In a lab-in-the-field setting, we observe students of Mathematics who work jointly on problem sets, to prepare them for future individual exams. Contrary to our hypothesis, we find that conditional cooperators are not more successful jointly or individually. Instead, people categorized as free riders excel in individual performance without causing negative externalities on their peers in the teamwork phase.

In essay 3, we investigate how intentions influence punishment behavior. Almost all decisions in daily life and within companies are made facing alternatives. In that sense, decisions can be evaluated regarding the underlying intentions. By actively framing the choice set, decision makers might be able to cast their actions in a better light than appropriate. In a laboratory experiment, we investigate the role of intentions in situations where it is possible to disguise the underlying motives. We find that assessing intentions plays a major role in punishing behavior. This seems to be common knowledge among humans, as the participants extensively make use of the possibility to disguise. Interestingly, the sanctions for such malign behavior are limited, even in cases of discovery, which makes this strategy quite efficient. The results shed some light on recent political developments, in which some politicians do not lose touch with voters, even when the politicians are obviously dishonest about their intentions.

In essay 4, we demonstrate a novel way to extend oTree to allow for real-time interactions in online settings, such as Amazon's Mechanical Turk. As a proof-of-concept, we run a series of double auction markets to show that the software works as intended, and that the results are comparable to the literature. We further share important insights into how to conduct interactive games online with large groups. The trend of online experiments paired with accessible software solutions provides various advantages for academic research: Experiments can be conducted with non-standard subject pools, and reduced costs make sufficiently powered studies affordable. It also provides researchers in various parts of the world the opportunity to participate in social science research. To facilitate this trend, we make the code accessible under an open-source license.

Essay 1: Do markets foster consequentialist decisions? Evidence from an online experiment

Nana Adrian, Ann-Kathrin Crede, Jonas Gehrlein*

Abstract

This paper investigates the influence of markets on morals. Whereas the current literature focuses on moral decisions *within* markets, little is known about how being exposed to markets shapes morals *outside* markets in unrelated environments. We adapt two concepts from philosophy to define morality: According to deontology, the morality of an action is evaluated by the action itself. According to consequentialism, the morality of an action is evaluated by its outcomes. In an online experiment, we expose participants to either a non-market or market environment, and elicit their subsequent decisions in a moral dilemma scenario. We hypothesize that the market environment induces cost-benefit analysis considerations, and thus, fosters consequentialist decisions. Compared to a baseline distribution of decisions in the moral dilemma, we find a substantial increase in consequentialist decisions in the market treatment. However, a similar increase can be observed in the non-market treatment, excluding a treatment effect of the market manipulation itself. We discuss potential explanations for these results, and suggest avenues for future research.

Keywords: morality, markets, deontology, consequentialism, oTree, online experiment

*Nana Adrian: adrian@ecoplan.ch, Ecoplan AG; Ann-Kathrin Crede: ann-kathrin.crede@iop.unibe.ch, Institute for Organization and Human Resource Management, University of Bern; Jonas Gehrlein: jonas.gehrlein@iop.unibe.ch, Institute for Organization and Human Resource Management, University of Bern. We thank Zita Spillmann, Christoph Zangger, Roden Safar, Annika Wyss, and Nicolas Hafner for their excellent research assistance. We are grateful to Sebastian Berger, Ilija Braun, Kathrin Friedrich, Stefanie Jaussi, Elisa Matthewes, and Frauke von Bieberstein for very valuable comments. We thank Jan Dietrich for technical assistance with the experiment. We also benefited from valuable comments at the CUSO Summer School (Kandersteg, 2016), a research workshop for young scientists (Erfurt, 2019) and the Innsbruck Winter School on Credence Goods (Kühtai, 2019). All remaining errors are ours.

2.1 Introduction

Today, markets are widely recognized as an efficient way to organize production and distribution in an economy (Satz 2010). At the same time, markets expand to more and more spheres of society, and it seems that you can find a market for almost everything. Many things that used to be considered with non-market values now have a price. For example, surrogate mothers in India offer to bear babies for US families, the European Union sells carbon emission certificates that enable companies to buy and sell the right to pollute, or lobbyists who want to attend congressional hearings pay line-standing companies that hire homeless people to line up (Sandel 2012). Therefore, Sandel skeptically notices that we have drifted from *having* a market economy to *being* a market society. If markets are ubiquitous, what does constant exposure to markets mean for decisions that we make in private life or other unrelated decision environments? That is, are there spillover effects?

The question how markets may affect morals is an old one, and has been controversially discussed since the beginning of the history of economic thought (e.g., Montesquieu 1748, Condorcet 1795, Marx 1872, Veblen 1899). Recent experimental literature discusses the question whether markets erode moral values or social responsibility *within* markets (e.g., Falk and Szech 2013, Bartling et al. 2015, Irlenbusch and Saxler 2015, Kirchler et al. 2016, Pigors and Rockenbach 2016, Bartling and Özdemir 2017, Sutter et al. 2019). For example, Falk and Szech (2013) show that participants in a laboratory experiment have a higher willingness to take money instead of preventing the death of a mouse when they are bargaining over the life of the mouse in double auction markets than when they are deciding individually. Therefore, the authors conclude that market interactions have a tendency to undermine moral values.

Whereas the experimental literature establishes a link between markets and morals *within* the institution of a market, less is known about the influence of markets on unrelated moral decisions *outside* markets. For example, imagine a passenger plane hijacked by terrorists is heading toward a packed soccer stadium. Should a fighter pilot shoot down the plane, killing 164 people to save 70,000 (von Schirach 2016)? This question arises beyond any market. We investigate whether people solve a moral dilemma differently due to increasing exposure to markets in many other areas of life. We run an experiment on Amazon Mechanical Turk ($n = 620$), exogenously vary whether participants are exposed to a non-market or market environment, and compare their subsequent moral decisions across treatments. Thus, we shed light on the question whether markets have consequences that go beyond the market sphere. For the current debate whether policy makers should limit the scope of markets, it is important to understand whether and how markets shape moral decisions in unrelated environments.

In our experiment, participants are randomly assigned to one of two treatments in a between-subject design: Participants in the non-market treatment play a repeated guessing game, whereas participants in the market treatment play a (payoff-equivalent) repeated double auction (DA) market game. Afterward, all participants make a decision in a moral dilemma scenario. In this moral dilemma, participants have to imagine a situation in which harm cannot be avoided. They can choose to stay passive, and thus, let three people die. Alternatively, they can choose to actively intervene, and thus, sacrifice one person to save the lives of the three other people. We define morality based on two concepts from philosophy: Following deontology, the morality of an action is evaluated by the action itself. Following consequentialism, the morality of an action is evaluated by its outcomes. Related to the moral dilemma scenario, we interpret staying passive as the deontological action and actively intervening as the consequentialist action, where neither is judged to be superior to the other.

Markets are based on cost-benefit analysis considerations, which might have spillover effects on unrelated moral decisions. If we are constantly weighing costs and benefits, and thus, focus on outcomes, are we looking through the same lens when we make decisions outside the scope of markets? Evidence from psychology on habitual behavior and routines suggests that people show similar patterns of behavior in similar patterns of circumstances (e.g. Weiss and Ilgen 1985, Gersick and Hackman 1990), supporting the idea that we may also focus on outcomes outside markets. We hypothesize that participants in the market treatment are more likely to choose the consequentialist action compared to participants in the non-market treatment. Compared to a baseline distribution of decisions in the moral dilemma scenario without a preceding economic game, we find a huge and statistically significant increase of 17 percentage points in consequentialist decisions in the market treatment. However, we observe a similar increase in consequentialist decisions in the non-market treatment (15 percentage points), ruling out that the market manipulation itself drives the result. It seems, instead, that the non-market and market manipulations share a common factor that drives consequentialist decisions. We discuss these potential factors, and suggest ideas for further research.

We proceed as follows: In section 2.2, we review the related literature. In section 2.3, we explain the experimental design and procedures. In section 2.4, we show the main results. In section 2.5, we discuss the results and suggest ideas for further research. In section 2.6, we conclude and give an outlook.

2.2 Related Literature

The early literature suggests two different views on how markets and moral values are related. Some scholars argue in favor of a market society, and stress the civilizing effect that markets, or more specifically, commerce, bring along (Hirschman 1982). For example, Montesquieu (1748) writes "commerce ... polishes and softens barbaric ways as we can see every day" (p. 81). Condorcet (1795) builds on this idea, and argues that "manners have become more gentle ... through the influence of the spirit of commerce and industry" (p. 238). Paine (1792) even makes a stronger statement, explaining that

> Commerce is a pacific system, operating to cordialise mankind, by rendering Nations, as well as individuals, useful to each other ... The invention of commerce ... is the greatest approach towards universal civilization that has yet been made by any means not immediately flowing from moral principles. (p. 215)

Another group of scholars takes the opposite view, and emphasizes that capitalist societies have a tendency to undermine the moral foundations on which they are based on (Marx 1872, Veblen 1899, Schumpeter 1942). Schumpeter (1942), for example, argues that "capitalism creates a critical frame of mind, which, after having destroyed the moral authority of so many institutions, in the end turns against its own" (p. 143). Taken together, the early literature clearly sees a connection between markets and morals, but remains unclear whether markets promote or undermine moral values.

The more recent economic literature sheds new light on this research topic, and yields several theoretical and empirical contributions. In a theoretical work, Bowles (1998) argues that preferences are endogenous, and that markets not only allocate goods and services but also influence the evolution of tastes and values. Similarly, Shleifer (2004) theoretically investigates the consequences of market competition, and finds that competitive pressure creates incentives for unethical practices (such as child labor) to reduce costs and guarantee survival in a competitive market. Opposing evidence comes from empirical, cross-sectional studies: Henrich et al. (2001) find that the higher the degree of market integration within a society, the more people cooperate in experimental games. In a more recent study, they find additional evidence that the spread of markets is also positively correlated with fairness (Henrich et al. 2010). Again, the more recent theoretical and empirical literature establishes a link between markets and moral or prosocial behavior, but yields opposite results.

The first experimental contribution on the interplay of morals and markets is the seminal paper by Falk and Szech (2013). They exogenously induce different institutions, and thus, establish a causal relationship between markets and moral decisions. In their experiment, participants are randomly assigned to one of three treatments: In the *individual* treatment, participants face the choice between taking 10 euros and killing a mouse, or not receiving the money and preventing the death of the mouse. In the *bilateral market* treatment, two participants are bargaining over the life of the mouse in a double auction market over 10 rounds. The seller is endowed with the mouse, and can offer prices for which he is willing to sell the mouse. The buyer is endowed with 20 euros, and can offer prices for which he is willing to buy the mouse. If the seller and the buyer agree on a price, that is, a split of the 20 euros, then the seller receives the price, the buyer receives 20 euros minus the price, and the mouse is killed. If participants do not agree on a price, or if one party refuses to bargain at all, then the mouse survives. The *multilateral market* treatment works the same, except that nine sellers and seven buyers bargain over prices (and the lives of nine mice). Results show that 45.9% of the participants are willing to kill the mouse in the individual treatment. This share increases to 72.2% in the bilateral and to 75.9% in the multilateral market treatment. Thus, the authors conclude that market interactions erode moral values.

The study by Falk and Szech (2013) received a lot of attention in the media (e.g. Spiegel 2013, Zeit 2013, SRF 2015) and in the academic world, starting a new wave of research on the interplay of markets and morals (e.g., Bartling et al. 2015, Irlenbusch and Saxler 2015, Kirchler et al. 2016, Pigors and Rockenbach 2016, Bartling and Özdemir 2017, Sutter et al. 2019). For example, Bartling et al. (2015) investigate a laboratory product market, in which producers and consumers can mitigate a negative externality affecting an uninvolved third party by incurring additional production costs. They find a substantial demand for, and supply of, socially responsible products across various conditions. However, comparing the level of socially responsible behavior in the market to an individual choice setting reveals that participants behave less socially responsible in the market compared to the non-market setting. Kirchler et al. (2016) build on the experimental design by Falk and Szech (2013) and test how different interventions affect moral behavior in an individual choice list versus a double auction market condition. In both conditions, participants can decide between taking money for themselves and forgoing a donation to UNICEF to finance measles vaccine, or not taking the money and thus, making the donation. In the individual choice setting, participants act as price takers, and decide for a list of 22 choices whether to take the money for themselves or to donate. In the double auction market, participants bargain over splitting an amount of money between themselves and making the donation. One market consists of six sellers and four buyers, where the sellers own the vaccine, and the buyers own money. If sellers and buyers agree on a price, the money is split accordingly, and no donation is made. If a seller does not trade, the vaccine is donated. The authors find that in both conditions, the potential threat of monetary punishment by an external observer promotes moral behavior, whereas removing anonymity by making participants identifiable promotes moral behavior only in the individual, but not in the market condition. The authors explain the latter result by the possibility to diffuse responsibility in the market condition, which cannot drive behavior in the individual choice list condition.

Some scholars are critical of the work by Falk and Szech (2013): Breyer and Weimann (2015) argue that Falk and Szech (2013) interpret their results incorrectly. Following Breyer and Weimann (2015), the individual treatment is what corresponds most closely to the kind of market we encounter in the real world; namely, that consumers act as price takers and do not bargain over prices. They further criticize that the treatment comparison between the individual and market treatments is not valid, as more than one variable was changed at the same time, for example the number of repetitions (one shot in the individual treatment versus 10 rounds in both market treatments). Bartling et al. (2019) test the results of Falk and Szech (2013) for robustness, and address the critical point that the number of repetitions varies across treatments. In their paradigm, participants face the choice between taking money for themselves and not having the experimenter finance a leprosy treatment in India, or not taking the money and thus, donating. Running the individual treatment over one round and the bilateral market treatment over 10 rounds as Falk and Szech (2013) did yields a comparable treatment effect: Participants in the bilateral market treatment choose to receive the money over donating statistically significantly more often than participants in the individual treatment. However, comparing the individual treatment over one round with the bilateral market treatment over one round does not yield a statistically significant difference, nor does the comparison of the individual treatment over 10 rounds with the bilateral market treatment over 10 rounds. Thus, Bartling et al. (2019) conclude that the adverse effect of markets on morals disappears if the number of rounds is held constant. Thus, overall, the explanatory power of the study by Falk and Szech (2013) remains open to debate.

One important feature of studying the interplay of markets and morals is the definition of what is considered *moral*. The experimental literature thus far has focused on moral behavior *within* the institution of the market, and mostly defined an immoral action as agreeing to trade at the expense of a third party, or put differently, as willingly causing a negative externality that harms an unrelated person or animal. Because we are interested in investigating moral decisions in decision environments *outside* markets, we need another approach to define morality: Following the principle of *deontology*, the morality of an action is evaluated by the action itself (Kant 1785). Following the principle of *consequentialism* (to which utilitarianism belongs), the morality of an action is evaluated by its consequences (Bentham 1789, Mill 1863). Whereas deontology prohibits any harmful action irrespective of its consequences, and emphasizes absolute and inviolable rights and duties, consequentialism aims at maximizing benefits and minimizing costs across affected individuals, and emphasizes the process of cost-benefit analysis (Greene et al. 2008, Cushman and Greene 2012, Barak-Corren et al. 2018). Importantly, we do not take a normative stance on the evaluation of the moral principles, and do not judge whether one is superior to the other.

The tension between the two principles can be captured in so-called moral trolley problems. These thought experiments stemming from philosophy represent a dilemma situation, as the only way to prevent harm to one group of people is to harm someone else or a smaller group of people (Bauman et al. 2014). In the original trolley problem (Foot 1967, Thomson 1985), a runaway trolley is heading toward five people, and about to kill them. In one version, one can save the five people by diverting the trolley onto a side track, where another person is standing, and will be killed instead. In the footbridge version, one can save the five people by pushing another person off a footbridge in front of the trolley, stopping the trolley, but killing the one person. A prototypical consequentialist would always become active, that is, killing the one person to save the other five people, to serve the greater good. A prototypical deontologist would never intervene, and consider killing the one person as an unacceptable violation of a right or duty (Greene et al. 2008). A robust result is that most people agree to hit the switch to divert the trolley to the other track, but disagree with pushing the person off the footbridge (Greene et al. 2001).

Thus far, economists have been reluctant to include the philosophical perspective when studying morality. One exception is the study by Chen and Schonger (2017), who present an economic approach to elicit consequentialist, deontological, and mixed consequentialist-deontological motivations. They suggest a revealed preferences approach to detect the different motivations, by varying the probability that a decision is implemented: A pure consequentialist always focuses on the outcomes, and does not react to varying probabilities with which decisions are implemented. For a pure deontologist, the decision is also independent of the probability, because the action per se determines what to do, independent of any consequences. Only mixed consequentialist-deontological motivations change a decision, as the probability that the decision is implemented varies. In another study, Chen (2016) examines the influence of the structure of employment on consequentialist versus deontological values. Participants in an online experiment are randomly assigned to a competitive or a piece-rate condition for a data-entry task in a between-subject design. Afterward, they make a decision in a moral trolley problem. Chen (2016) finds that experiences with a competitive work environment foster deontological decisions in the moral trolley problem. However, the impact of competition on deontological decisions depends on economic development: In rich countries, competition in the employment structure makes people more consequentialist. We take this finding as the very first hint that markets might generally foster consequentialist decisions, and design a new experimental paradigm to examine our research question.

2.3 Experimental Design

Our experiment consists of three stages: a manipulation, a moral dilemma, and a questionnaire. The experiment has two treatments: a non-market treatment and a market treatment. Whereas the moral dilemma and the questionnaire are identical for both treatments, the manipulation differs across treatments: Participants in the *non-market* treatment engage in a transcription task, and play a guessing game; participants in the *market* treatment play a DA market game.

2.3.1 Stage 1: Manipulation

Non-market treatment

In the first step, participants in the non-market treatment engage in a transcription task for 10 minutes. They see “lorem ipsum” sentences, and are asked to copy these sentences into an input field. If the participants commit more than two errors in one sentence, they are asked to correct the mistakes before they can proceed with the next sentence. We are not interested in the performance on the transcription task per se. However, the manipulation in the market treatment takes more time, and is cognitively more demanding than the guessing game. Therefore, we add the transcription task before the guessing game to keep the cognitive load similar across treatments. In the second step, participants in the non-market treatment play 10 rounds (plus 2 additional test rounds) of a guessing game, which works as follows: Participants are assigned to groups of nine. In each round, their task is to guess one number out of the set $G \in \{20, 30, 40, ..., 100\}$. Subsequently, a random device assigns each value of the set G once to one of the participants. If a participant's guess coincides with the randomly assigned number, this participant wins, and receives a payoff of $\pi_W = 50$.[1] Otherwise, the participant loses, and receives a payoff of $\pi_L = [0, 10, 20, 30, 40]$ with probabilities $p_L = [\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}]$. The expected payoff of one round of the guessing game is equal to $E[\pi_G] = \frac{1}{9} \cdot 50 + \frac{8}{9} \cdot \frac{1}{8}\ (10 + 20 + 30 + 40) = \frac{50}{3} \approx 16.67$. We will later show that we hold the expected payoff constant across treatments. For the treatment comparison, it is important that in the non-market treatment, the payoff of one participant does not depend on the interaction with another participant, but is determined only by luck. After each round, participants get feedback, and learn whether they won or not. At the end of the experiment, one round of the guessing game is randomly chosen, and accounts for payment.

[1] The currency used in the experiment is points. One point is worth \$0.15.

Market treatment

In the market treatment, participants play a continuous DA market consisting of 9 buyers and 9 sellers over 10 rounds (plus 2 additional test rounds). We assign participants randomly to the role of either a buyer or seller. Participants keep their role for the entire 10 rounds. In every round, they can trade a fictional good for 60 seconds. Every subject can trade at most once per round. At the beginning of each round, buyers privately learn their valuation of the good, and sellers privately learn their production costs of the good. Valuations and costs are randomly drawn from the sets $v \in \{30, 40, 50, ..., 110\}$ and $c \in \{10, 20, 30, ..., 90\}$. In each round, every value can appear only once among the buyers and sellers. While the distribution of demand and supply is common knowledge, the realization of v (for a buyer) or c (for a seller) is private knowledge to each market participant. In each round, sellers and buyers randomly receive a new display ID to avoid reputation effects.

Sellers can sell, and buyers can buy, one unit of the fictional good in each round. Once the market opens, sellers can submit asks, that is, the price at which they are willing to sell the product. Buyers can submit bids, that is, the price at which they are willing to buy the product. All asks and bids appear in the table "Current bids and asks," and are observable to all market participants (see Appendix A for a screenshot). A trade occurs if a seller makes an ask that is lower than a current bid or if a buyer makes a bid that is higher than a current ask. The trade is closed at the price of the bid, or the ask that was posted first. A trade is also possible by directly accepting a bid or ask that appears in the table. Sellers and buyers can modify their asks and bids until the market closes, as long as they have not traded yet. If a trade occurs, the payoffs are $\pi_s = price - costs$ for a seller and $\pi_B = valuation - price$ for a buyer. Production costs occur only when trading, which means that it is not possible that a seller produces the good at a personal cost but cannot sell it on the market.

Competitive equilibrium theory predicts an average trading price of 60 with a frequency of trades of between 5 and 6 per round. In equilibrium, only buyers with high valuations ($v \geq 60$) and sellers with low production costs ($c \leq 60$) end up trading. Before learning whether production costs are high or low, a seller expects to have production costs above the equilibrium price with probability $\frac{3}{9}$ (in which case, he would not trade and would receive zero payoff) and below or equal to the equilibrium price with probability $\frac{6}{9}$ (in which case, he can sell the product). A seller, therefore, has an expected payoff of $E[\pi_S] = \frac{6}{9} \cdot [p - c | c \leq 60] = \frac{50}{3} \approx 16.67$. The same logic holds true for the expected payoff of a buyer, that is, $E[\pi_B] = \frac{6}{9} \cdot [v - p | v \geq 60] = \frac{50}{3} \approx 16.67$. We keep the expected payoff of participants in the guessing game and in the market game constant, and thus, provide the same monetary incentives across treatments. After each round, sellers and buyers receive feedback, and see a table with all trades and prices for which goods were traded (see Appendix A for a screenshot). At the end of the whole experiment, one round of the DA market is randomly chosen, and accounts for payment (with one point worth $0.15).

Manipulation check

After the manipulation, we add a manipulation check to test whether being exposed to a subtle situational cue, such as a market environment, activates certain mental concepts (e.g., Cohn and Maréchal 2016). Therefore, we employ a word-completion task as used by Shu et al. (2012). We present participants 14 word fragments in a random order and ask them to complete the fragments as the first words that come to their mind. We chose the words such that nine of these words (e.g., _ O N E Y) can be completed as market-related words (MONEY) or neutral words (HONEY). Five additional words serve as control, and can (only) be completed with a neutral meaning (for the full list of words, see Appendix B).[2] We calculate the manipulation check score by counting the number of completed market-related words. We hypothesize that participants in the market treatment are more likely to complete the word fragments as market-related words than participants in the non-market treatment. Thus, we expect a higher manipulation check score in the market treatment compared to the non-market treatment.[3]

2.3.2 Stage 2: Moral dilemma

In the second stage, participants are presented with a moral dilemma scenario, and have to make a decision. We build on the classical moral trolley problem literature (Foot 1967, Thomson 1985), and present participants the footbridge (drop) version, as recently employed by Barak-Corren et al. (2018). In this scenario, participants have to imagine that they are working by the train tracks when they observe a boxcar breaking loose and speeding down the tracks. This boxcar is heading toward three workers who do not have enough time to get off the track. Participants further have to imagine that above the track there is a platform with another worker. This worker is not threatened by the boxcar, but he is standing over a trap door. Participants have to choose between two options: They can choose to *stay passive*, and let the boxcar head toward the three workers. The consequence is that the worker over the trap door stays unharmed, and

[2] We follow the framework by Koopman et al. (2013) to construct reliable and valid word fragments. Therefore, we pretested a list of 34 word fragments, and chose 14 words that participants completed with a neutral or a market-related meaning with sufficient variance. Importantly, we did not use words that appeared in either of the two instructions, to avoid participants completing the word fragments from their short-term memory.

[3] In a pretest, we elicited a baseline average score of 3.5. We hypothesize that this score increases if participants previously played the DA market game compared to the guessing game.

the three workers die. Alternatively, they can choose to *actively intervene* by using a switch that opens the trap door and drops the one worker in front of the boxcar. Thus, the worker's body gets caught in the wheels of the boxcar and slows it down. The consequence is that the one person dies, and the three workers stay unharmed. We present participants Figure 2.1 as an illustration next to the instructions (see Appendix C for the exact wording).

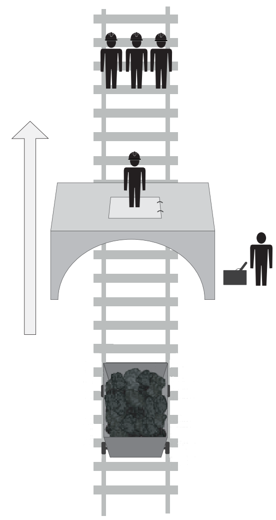


Figure 2.1: The boxcar dilemma (own illustration).

We ask participants if they would stay passive or actively intervene in the described scenario. We randomize the order of the answer choices to exclude any order effects. We interpret staying passive as deciding according to the deontological principle and actively intervening as following the consequentialist principle. We hypothesize that participants in the market treatment are more likely to actively intervene (consequentialist decision) than participants in the non-market treatment. The argument is that markets induce cost-benefit analysis considerations, which might have spillover effects on subsequent moral decisions.

2.3.3 Stage 3: Questionnaire

In the third stage of the experiment, and before participants get feedback on their payoff, they are asked to fill out a questionnaire. We first test whether participants understood the description of the moral dilemma correctly. Next, participants answer questions about their perceived performance in the game they played, the satisfaction with their decision, if they thought about their decision, and their mood. We further ask if participants have experience with negotiating. Additionally, we ask participants for their experience with moral trolley problems in general. Finally, we elicit information on risk and trust preferences and basic socio-demographic variables, such as gender and age.

2.3.4 Procedure

We preregistered the study in the American Economic Association's (AEA) registry for randomized controlled trials.[4] For this purpose, we ran a power analysis that suggested we should collect a total of $n = 700$ observations. For this power analysis, we elicited the baseline distribution of moral decisions. Therefore, we collected $n = 103$ observations including only the moral dilemma scenario.[5] We implemented the experiment with oTree (Chen et al. 2016), and used the DA market game of Crede et al. (2019). We ran the experiment online on Amazon Mechanical Turk between November 2018 and May 2019. We restricted participation to workers located in the US. Sessions were run between 11 a.m. (EST) and 6:30 p.m. (EST). Participants earned, on average, \$5.64 (\$3.00 participation fee plus the bonus from the guessing game/DA market game), and needed approximately 40 minutes to complete the experiment. Overall, we collected $n = 720$ observations in 26 sessions. In every session, we included the non-market and market treatments to minimize session effects. We had to drop 100 observations from participants who did not answer the control questions correctly,[6] resulting in a total of $n = 620$ observations for the data analysis (non-market: $n = 292$, market: $n = 328$).

2.4 Results

2.4.1 Manipulation check

We first look at the manipulation check, which we elicited only during the first 4 sessions, yielding $n = 106$ observations (non-market treatment: $n = 54$, market treatment: $n = 52$).[7] We did not include the manipulation check for all sessions, as we wanted to avoid the manipulation check itself manipulating participants' mindsets. To calculate the manipulation check score, we count the number of completed market-related words, and build the average within treatments. Figure 2.2 shows the results.

[4] https://www.socialscienceregistry.org/trials/2707/history/32548

[5] We elicited the baseline distribution of moral decisions on Amazon Mechanical Turk in December 2017. The baseline treatment yielded 35% of the decisions were consequentialist. Thus, we assumed 35% of the decisions were consequentialist for the non-market treatment and a 5 percentage point increase in consequentialist decisions for the market treatment. We further assumed a t-test, an alpha of 0.05, and a power of 0.8, which yielded the required number of observations of $n = 690$, which we rounded to $n = 700$.

[6] Results remain qualitatively the same if we include all observations.

[7] We ran a power analysis to determine the sample size for the manipulation check. Therefore, we assumed a t-test, a baseline score of 3.5 (as our pretest showed), an increase in the score of one word for the market treatment, an alpha of 0.05, and a power of 0.95, which yielded a total sample size of $n = 100$.

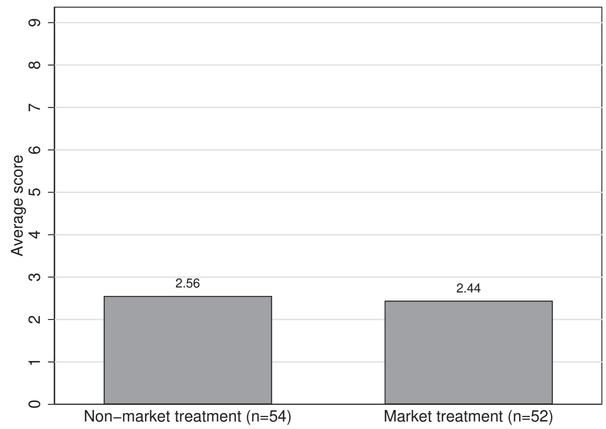


FIGURE 2.2: Results of the manipulation check.

The average score of market-related words is 2.56 in the non-market treatment and 2.44 in the market treatment. This difference is not statistically significant (Mann-Whitney U test, $p = 0.5577$). Thus, being in the market treatment compared to the non-market treatment does not seem to change participants' mindset such that they have different concepts in mind when they complete the presented word fragments.

2.4.2 Moral dilemma

In the next step, we look at the decisions participants made in the moral dilemma scenario. To get an idea of the baseline distribution of decisions for the power analysis, we presented participants only the moral dilemma scenario, without a previous manipulation stage. Figure 2.3 shows the distribution of decisions in the baseline.

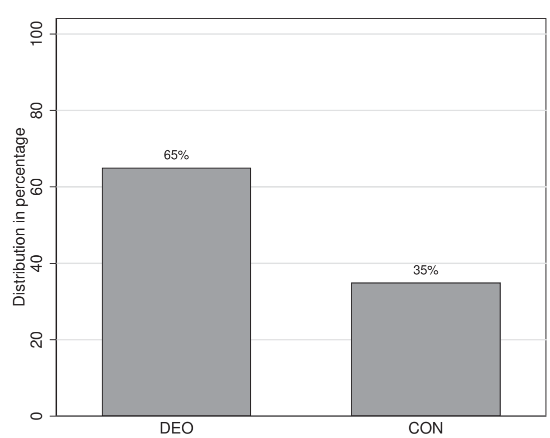


FIGURE 2.3: Distribution of decisions in the baseline
(DEO: deontological, CON: consequentialist).

As Figure 2.3 shows, 65% of participants decided to stay passive and act according to the deontological principle, whereas 35% of participants chose to actively intervene, and thus, followed the consequentialist principle. A recent study by Barak-Corren et al. (2018) yields similar results: In the corresponding treatment of their study, 59% of participants decided according to the deontological principle, whereas 41% of participants decided according to the consequentialist principle. Thus, we find a comparable baseline distribution for the footbridge (drop) dilemma.

In the market treatment, participants first engage in a DA market and trade over 10 rounds, before they make a decision in the moral dilemma scenario. We find an increase of 17 percentage points in consequentialist decisions between the baseline and the market treatment: Whereas 35% of participants chose according to the consequentialist principle in the baseline, this share goes up to 52% in the market treatment. This difference is highly statistically significant (t-test, $p = 0.0026$). This result supports our hypothesis that markets foster consequentialist decisions. However, taking into account the non-market treatment does not support this observation, as a similar increase in consequentialist decisions (15 percentage points) can be observed. Figure 2.4 compares the distribution of moral decisions in the baseline to the non-market and market treatments.

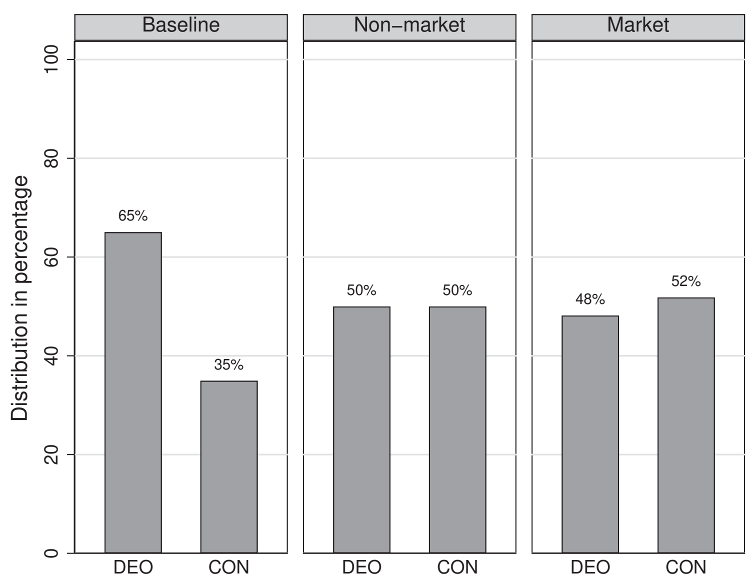


Figure 2.4: Distribution of decisions in the baseline versus the non-market and market treatments (DEO: deontological, CON: consequentialist).

As Figure 2.4 shows, 50% of participants in the non-market treatment chose to actively intervene, which yields a statistically significant increase in consequentialist decisions compared to the baseline (t-test, $p = 0.0083$). Table 2.1 summarizes the results.

TABLE 2.1: Pairwise comparisons between treatments

	Deontological	Consequentialist	Pairwise comparisons (t-test)
Baseline ($n = 103$)	65.05%	34.95%	Baseline vs. non-market: $p = 0.0083$
Non-market ($n = 292$)	50.00%	50.00%	Baseline vs. market: $p = 0.0026$
Market ($n = 328$)	48.17%	51.83%	Non-market vs. market: $p = 0.6499$

As can be seen in Table 2.1, the difference of 1.83 percentage points in consequentialist decisions between the non-market and market treatments is not statistically significant (t-test, $p = 0.6499$). Thus, we do not find support for our hypothesis that the market manipulation fosters consequentialist decisions. Instead, it seems that some characteristic (or a combination of several characteristics) that is common to the non-market and market manipulations drives the increase in consequentialist decisions. We will discuss these potential drivers in the next section. In the last step, we investigate whether additional factors influence the decision to act according to the consequentialist principle, and run probit regressions with the moral decision (0: deontological, 1: consequentialist) as the dependent variable. Table 2.2 shows the results.

As the regression confirms, the market treatment has no statistically significant impact on the decision to act according to the consequentialist principle. Experience with negotiating and a general willingness to take risks increase the likelihood of choosing the consequentialist action, whereas being male has a slightly negative impact on the likelihood of choosing the consequentialist action. The bonus points, perceived performance, satisfaction with the own decision, having thought about the own decision, mood, experience with trolley problems, and age do not have any influence on the moral decision.

2.5 Discussion

Summarizing the results, we do not find a statistically significant difference in the word completion task between the non-market and market treatments; that is, participants in the market treatment do not complete the word fragments as market-related words more often than participants in the non-market treatment. One reason could be that the manipulation did not work or was too subtle, meaning that the experience of the market did not activate certain mental concepts, compared to the experience of the guessing game. Another potential reason is linked to the current replication crisis, revealing that many effects uncovered in experiments cannot be replicated (e.g., Camerer et al. 2016, Verschuere et al. 2018). Especially the literature on priming has been criticized due to failed replications of some prominent studies (e.g., Yong 2012).

Table 2.2: Probit regression with the moral decision (0: Deontological, 1: Consequentialist) as dependent variable

	Model 1	Model 2	Model 3	Model 4
Market Treatment	0.046	0.090	0.090	0.042
	(0.101)	(0.104)	(0.105)	(0.108)
Bonus		0.000	0.000	-0.000
		(0.002)	(0.002)	(0.002)
Perceived Performance		0.119**	0.104*	0.055
		(0.052)	(0.057)	(0.058)
Satisfaction			-0.018	-0.037
			(0.033)	(0.035)
Thought			0.007	0.029
			(0.054)	(0.058)
Mood			0.043	0.006
			(0.061)	(0.063)
Experience Negotiation				0.193***
				(0.056)
Experience Trolley				-0.004
				(0.108)
Risk				0.058***
				(0.021)
Trust				-0.013
				(0.020)
Male				-0.180*
				(0.109)
Age				-0.008
				(0.005)
_cons	-0.000	-0.400**	-0.459	-0.362
	(0.073)	(0.197)	(0.416)	(0.519)
N	620	620	620	620
Pseudo-R^2	0.000	0.006	0.007	0.049

Notes: Robust standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Further, we do not find a statistically significant difference in the moral dilemma scenario between the non-market and market treatments; that is, participants in the market treatment do not choose the consequentialist action more often than participants in the non-market treatment. The small difference in consequentialist decisions of 1.83 percentage points between the two treatments goes in the direction of our hypothesis, but is far from statistically significant. Interestingly, however, we find a huge and statistically significant increase in consequentialist decisions between the baseline of the moral dilemma scenario and *both* the non-market treatment (15 percentage points) and the market treatment (17 percentage points).

Several reasons could drive these results. First, it could simply be that there is no effect of markets on subsequent moral decisions, which is why we do not find a difference between the non-market and market treatments. Another explanation could be that we cannot uncover a potential effect with our experimental design. One question is whether we chose an appropriate market manipulation to induce the experience of interacting in a market and to appeal to cost-benefit analysis considerations, or whether the effect does not persist until the moral dilemma stage is reached. Whereas some scholars argue that a double auction market is a very typical market institution, and use it to experimentally implement a market condition (e.g., Falk and Szech 2013), others argue that in real life, we act as price takers, and therefore, experience markets differently than represented by a double auction market (e.g., Breyer and Weimann 2015). Thus far, there is no unifying framework or definition determining what a market actually incorporates. It would be interesting for further research to disentangle the single components a market might include (like money, competition, diffusion of responsibility, etc.) to see if the market as a whole or single factors drive behavior. Another question is whether we chose the appropriate non-market manipulation. We designed the guessing game such that important characteristics of the manipulation are kept equal (e.g., the expected monetary payoff, being part of a group of nine, and playing over 10 rounds), while other aspects are in clear contrast to the market treatment (e.g., no interactions with other participants). Still, the challenge is to determine how the suitable control for a market should look.

The higher share of consequentialist decisions in both treatments suggests that one (or several) factor(s) that the non-market and market treatments have in common drive the change in moral decisions. One such factor could be cognitive fatigue: Both manipulations presumably fatigue participants cognitively, as they need to understand the rules of the game, answer control questions, and then play a game over 10 rounds. The cognitive load was lower in the baseline, as participants made only the decision in the moral dilemma scenario (which took, on average, eight minutes). Thus, we hypothesize that cognitive fatigue might increase consequentialist decisions. In a recent study, Timmons and Byrne (2019) examine whether moral fatigue affects people's deontological and consequentialist judgments. They find that participants who have completed a cognitively tiring task tend to judge that killing a person to save several others is *less* permissible compared to participants who have completed a less cognitively tiring task. Put differently, cognitive fatigue seems to reduce consequentialist actions. This result contradicts our hypothesis that cognitive fatigue could drive the increase in consequentialist decisions in both treatments. Other factors that might be common to both treatments could be a general focus on outcomes (as both treatments included a bonus), playing a game to earn money, a group feeling, the degree of perceived luck determining the payoff, or a general payoff uncertainty (as participants learned only at the very end how much they earned). For all these potential similarities across the two manipulations, we would need to run additional treatments. At this point, we cannot finally identify the driver of the increase in consequentialist decisions in the two treatments compared to the baseline.

2.6 Conclusion

The question whether markets influence morals is a longstanding one that is still important today. Given that markets capture more and more spheres of human life, a current debate raises the question whether policy should limit the scope of markets (Satz 2010, Sandel 2012). The far-reaching answer to this question requires robust empirical evidence. The current literature establishes a negative impact of markets on moral decisions, but the overall results are mixed, and policy implications are not clear. In addition, the existing literature focuses on what the influence of markets on moral decisions might be *within* the scope of markets. We go one step further by focusing on moral decisions *outside* markets, and by taking a non-judgmental philosophical perspective to define morality. Thus, we investigate how the constant exposure to markets influences moral decisions in unrelated decision environments.

To examine this research question, we exogenously assign participants to two different institutions in a between-subject design: In the non-market treatment, participants play a guessing game. In the market treatment, participants play a DA market game. We then compare the subsequent moral decisions made in a moral dilemma scenario. To the best of our knowledge, we are the first to use economic games to induce a market mindset. Our hypothesis was that interacting in a market environment triggers cost-benefit analysis considerations, and puts a focus on consequences, which might have spillover effects on unrelated moral decisions, and thus, foster consequentialist decisions. The results of this study do not support this hypothesis, as we do not find a difference between the non-market treatment and the market treatment. However, we discussed potential avenues for further research to get a more comprehensive answer to our research question.

Finding an answer to the question whether markets have an impact on the way we make moral decisions in environments outside the realm of markets is very important. Consider the example from the introduction: Imagine a passenger plane hijacked by terrorists is heading toward a packed soccer stadium. Should a fighter pilot shoot down the plane, killing 164 people to save 70,000? If we generally appreciate the fundamental value that one human life cannot be offset against another human life, we need to know if the exposure to markets changes how we react to such a moral dilemma. More specifically, it seems important to understand if markets shift our moral perspective such that we focus more on the outcome, and thus, disregard the action leading to this specific outcome.

References

Barak-Corren, N., Tsay, C.-J., Cushman, F., and Bazerman, M. H. (2018). If you're going to do wrong, at least do it right: Considering two moral dilemmas at the same time promotes moral consistency. *Management Science*, 64(4):1528–1540.

Bartling, B., Fehr, E., and Özdemir, Y. (2019). Do markets erode morals? *Presentation Innsbruck Winter Summit: (Un)ethical Behavior in Markets*.

Bartling, B. and Özdemir, Y. (2017). The limits to moral erosion in markets: Social norms and the replacement excuse. *CESifo Working Paper Series No. 6696*.

Bartling, B., Weber, R. A., and Yao, L. (2015). Do markets erode social responsibility? *The Quarterly Journal of Economics*, 130(1):219–266.

Bauman, C. W., McGraw, A. P., Bartels, D. M., and Warren, C. (2014). Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Social and Personality Psychology Compass*, 8(9):536–554.

Bentham, J. (1789). *An Introduction to the Principles of Morals and Legislation*. Clarendon Press.

Bowles, S. (1998). Endogenous preferences: The cultural consequences of markets and other economic institutions. *Journal of Economic Literature*, 36(1):75–111.

Breyer, F. and Weimann, J. (2015). Of morals, markets and mice: Be careful drawing policy conclusions from experimental findings! *European Journal of Political Economy*, 40(B):387–390.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., et al. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, 351(6280):1433–1436.

Chen, D. L. (2016). Markets, morality, and economic growth: Competition affects deontological judgment. *TSE Working Papers No. 16-692*.

Chen, D. L. and Schonger, M. (2017). Social preferences or sacred values? Theory and evidence of deontological motivations. *SSRN Working Paper*.

Chen, D. L., Schonger, M., and Wickens, C. (2016). oTree–An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

Cohn, A. and Maréchal, M. A. (2016). Priming in economics. *Current Opinion in Psychology*, 12:17–21.

Condorcet, M. d. (1795). *Esquisse d'un Tableau des Progrès de l'esprit Humain*. Bibliothèque nationale de France, Paris.

Crede, A.-K., Dietrich, J., Gehrlein, J., Neumann, O., Stürmer, M., and von Bieberstein, F. (2019). oTree: Implementing websockets to allow for real-time interactions–A continuous double auction as first application. *Working Paper*.

Cushman, F. and Greene, J. D. (2012). Finding faults: How moral dilemmas illuminate cognitive structure. *Social Neuroscience*, 7(3):269–279.

Falk, A. and Szech, N. (2013). Morals and markets. *Science*, 340(6133):707–711.

Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, 5:5–15.

Gersick, C. J. and Hackman, J. R. (1990). Habitual routines in task-performing groups. *Organizational Behavior and Human Decision Processes*, 47(1):65–97.

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., and Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, 107(3):1144–1154.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., and Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293(5537):2105–2108.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., and McElreath, R. (2001). In search of homo economicus: Behavioral experiments in 15 small-scale societies. *American Economic Review*, 91(2):73–78.

Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., and Ziker, J. (2010). Markets, religion, community size, and the evolution of fairness and punishment. *Science*, 327(5972):1480–1484.

Hirschman, A. O. (1982). Rival interpretations of market society: Civilizing, destructive, or feeble? *Journal of Economic Literature*, 20(4):1463–1484.

Irlenbusch, B. and Saxler, D. (2015). Social responsibility in market interaction. *IZA Discussion Paper No. 9240*.

Kant, I. (1785). *Grundlegung zur Metaphysik der Sitten*. J. F. Hartknoch Verlag.

Kirchler, M., Huber, J., Stefan, M., and Sutter, M. (2016). Market design and moral behavior. *Management Science*, 62(9):2615–2625.

Koopman, J., Howe, M., Johnson, R. E., Tan, J. A., and Chang, C.-H. (2013). A framework for developing word fragment completion tasks. *Human Resource Management Review*, 23(3):242–253.

Marx, K. (1872). *Das Kapital: Kritik der politischen Ökonomie*. O. Meissner Verlag.

Mill, J. S. (1863). Utilitarianism. In *Seven Masterpieces of Philosophy*, pages 337–383. Routledge.

Montesquieu, C. d. (1748). *De l'esprit des lois*. Barrillot & Fils.

Paine, T. (1792). *Rights of Man*. Broadview Press.

Pigors, M. and Rockenbach, B. (2016). Consumer social responsibility. *Management Science*, 62(11):3123–3137.

Sandel, M. J. (2012). *What money can't buy: The moral limits of markets*. Macmillan.

Satz, D. (2010). *Why some things should not be for sale: The moral limits of markets*. Oxford University Press.

Schumpeter, J. A. (1942). *Capitalism, Socialism and Democracy*. Routledge.

Shleifer, A. (2004). Does competition destroy ethical behavior? *American Economic Review*, 94(2):414–418.

Shu, L. L., Mazar, N., Gino, F., Ariely, D., and Bazerman, M. H. (2012). Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. *Proceedings of the National Academy of Sciences of the United States of America*, 109(38):15197–15200.

Spiegel (2013). Der Radikalforscher. *https://www.spiegel.de/lebenundlernen/job/professor-armin-falk-sucht-nach-der-moral-in-der-oekonomie-a-930093.html*.

SRF (2015). Fall Volkswagen: Warum die Moral so rasch mausetot ist. *https://www.srf.ch/kultur/gesellschaft-religion/fall-volkswagen-warum-die-moral-so-rasch-mausetot-ist*.

Sutter, M., Huber, J., Kirchler, M., Stefan, M., and Walzl, M. (2019). Where to look for the morals in markets? *Experimental Economics*, https://doi.org/10.1007/s10683-019-09608-z.

Thomson, J. J. (1985). The Trolley Problem. *Yale Law Journal*, 94:1395–1415.

Timmons, S. and Byrne, R. M. (2019). Moral fatigue: The effects of cognitive fatigue on moral reasoning. *Quarterly Journal of Experimental Psychology*, 72(4):943–954.

Veblen, T. (1899). *The Theory of the Leisure Class*. Routledge.

Verschuere, B., Meijer, E. H., Jim, A., Hoogesteyn, K., Orthey, R., McCarthy, R. J., Skowronski, J. J., Acar, O. A., Aczel, B., Bakos, B. E., et al. (2018). Registered replication report on Mazar, Amir, and Ariely (2008). *Advances in Methods and Practices in Psychological Science*, 1(3):299–317.

von Schirach, F. (2016). *Terror: Ein Theaterstück und eine Rede*. btb Verlag.

Weiss, H. M. and Ilgen, D. R. (1985). Routinized behavior in organizations. *Journal of Behavioral Economics*, 14:57–67.

Yong, E. (2012). Nobel Laureate challenges psychologists to clean up their act. *Nature*, http://dx.doi.org/10.1038/nature.2012.11535.

Zeit (2013). Was ist Ihnen das Leben dieser Maus wert? *https://www.zeit.de/2013/21/wirtschaft-markt-moral-experiment*.

Appendix A: Screenshots DA market game

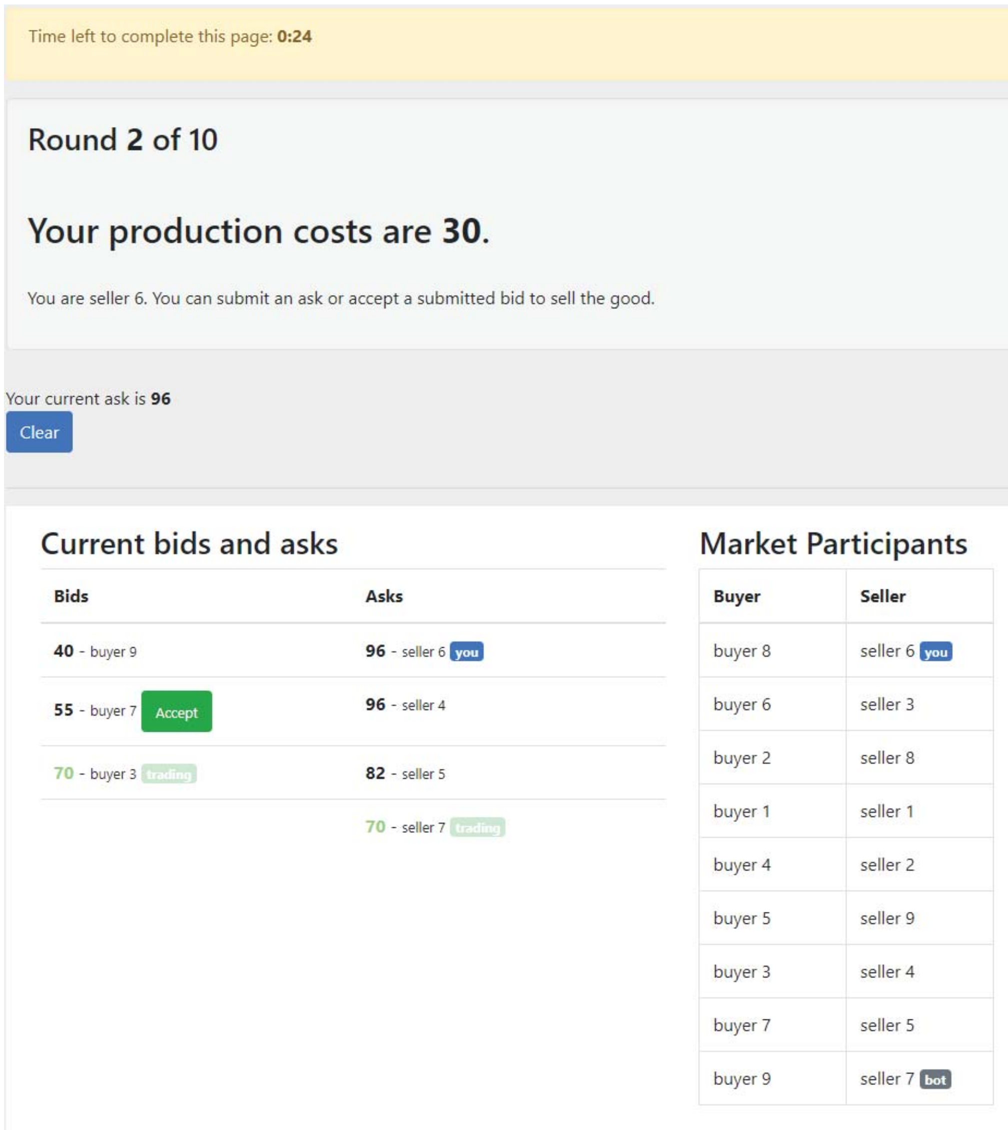


FIGURE 2.5: The Graphical User Interface of the DA market.

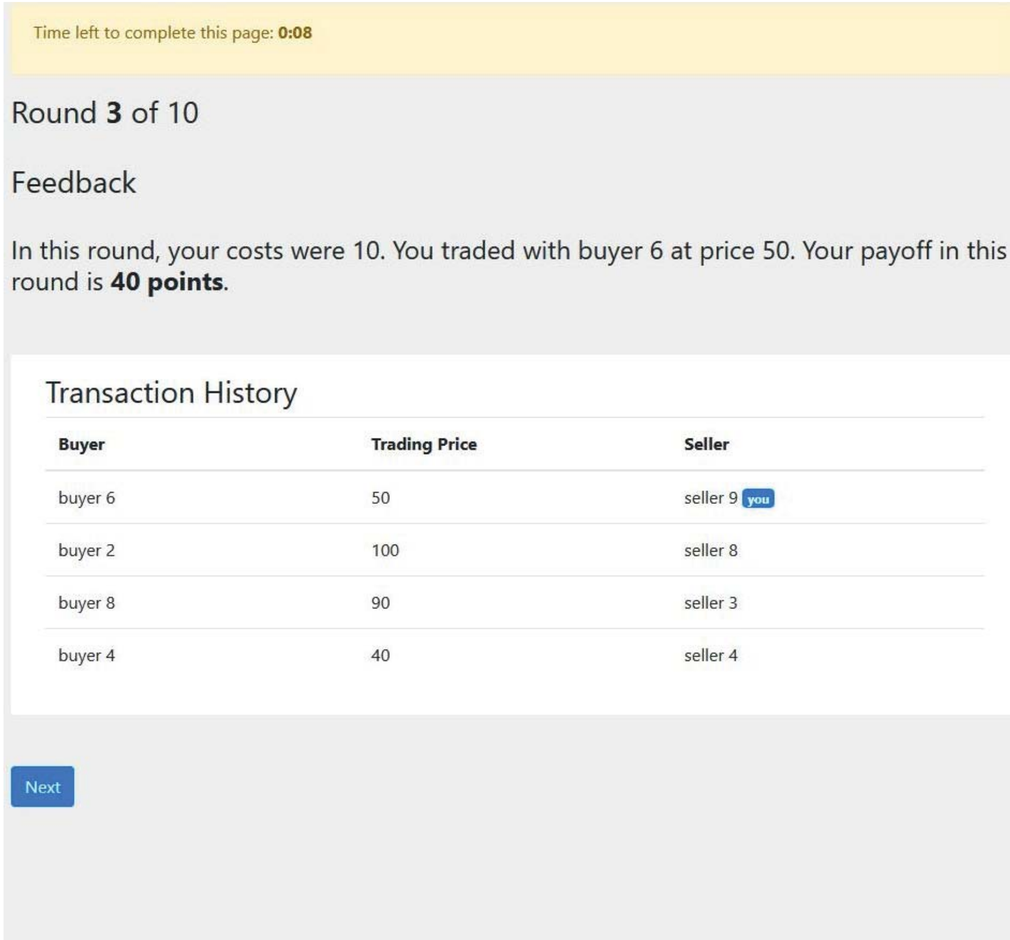


FIGURE 2.6: The feedback screen of the DA market.

Appendix B: Word completion task

Table 2.3: Full list of word fragments and the corresponding market and non-market solutions

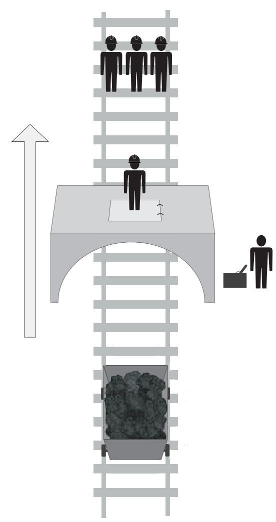
No.	Word fragment	Market-related	Non-market
1	M A _ L	MALL	MAIL
2	C A S _	CASH	CASE
3	_ O N E Y	MONEY	HONEY
4	_ A X	TAX	FAX
5	S U P P _ _	SUPPLY	SUPPER
6	S A L _	SALE	SALT
7	B R _ _ C H	BRANCH	BRUNCH
8	_ _ D G E T	BUDGET	WIDGET
9	S H _ P	SHOP	SHIP
10	F R _ _ T	-	FRUIT
11	T _ _ L E	-	TABLE
12	B E _ _	-	BEAR
13	B R _ _ C H	-	BREACH
14	C A B _ _	-	CABLE

Notes: Words 10–14 served as control and could only be completed with a neutral meaning. There might be additional solutions.

Appendix C: Moral dilemma scenario

In this part, please try to imagine the following situation:

You are working by the train tracks when you see an empty boxcar break loose and speed down the tracks. The boxcar is heading toward three workers who do not have enough time to get off the track. Above the track is a platform with another worker. This worker is not threatened by the boxcar. However, he is standing over a trap door.



You have two options:

Actively intervene
You use a switch that opens the trap door and drops the one worker in front of the boxcar. Thereby, the worker's body gets caught in the wheels of the boxcar and slows it down. That means the one worker dies and the three workers stay unharmed.

Stay passive
You stay passive and let the boxcar head toward the three workers. Thereby, the worker over the trap door stays unharmed and the three workers die.

Sidenote:
In any case, you are protected from the boxcar and stay unharmed. Furthermore, assume that you will not face any legal consequences for either action. Accept only the information given and try not to introduce additional assumptions that go beyond the problem as stated.

Essay 2: Teamwork revisited: Effect of social preferences on knowledge acquisition in the field

Frauke von Bieberstein, Jonas Gehrlein, Anna Güntner*

Abstract

Combining a lab-in-the-field experiment with field data, we study the effect of social preferences on performance in a modified teamwork setting, where public good production constitutes a prerequisite for individual performance, but is not a goal in itself. Examples of such modified team settings are knowledge sharing, peer coaching, and helping on the job – all highly relevant topics for organizations today. As opposed to a standard public good setting, we find that teams with conditional cooperators are not more successful jointly or individually. In contrast, selfish individuals tend to perform better individually, without generating negative externalities for their team partners.

Keywords: teamwork, social preferences, knowledge sharing, helping on the job, lab-in-the-field experiment

*Frauke von Bieberstein: frauke.vonbieberstein@iop.unibe.ch, Institute for Organization and Human Resource Management, University of Bern; Jonas Gehrlein: jonas.gehrlein@iop.unibe.ch, Institute for Organization and Human Resource Management, University of Bern; Anna Güntner: anna.güntner@mckinsey.com, McKinsey & Company. We are grateful to Leif Döring and his assistants, who supported us while we conducted this study and provided valuable feedback. We would like to thank Sebastian Berger, Andrea Essl, Lorenz Götte, and participants in the doctoral seminar at the University of Bern for their helpful comments. All remaining errors are ours.

3.1 Introduction

Teamwork is key for organizations today (Lazear and Shaw 2007, Delarue et al. 2008). From an economic point of view, teamwork is a public good, where individuals jointly manage an asset (the team production) from which all of them benefit, but that requires costly and unobservable individual resources to build or to sustain (Alchian and Demsetz 1972, Holmstrom 1982, Rob and Zemsky 2002). Group members' social preferences strongly influence the contributions of individual resources to the public good. Production is more successful when the group includes conditional cooperators – individuals who behave reciprocally – among its members (Fischbacher et al. 2001). This has been shown in the lab, where the common workhorse for these studies is a repeated public good game (Isaac et al. 1984), as well as in the field, where, for instance, groups with a larger share of conditional cooperators are more successful in forest commons management (Rustagi et al. 2010).

Given the findings above, it seems that firms and organizations relying on teamwork should try to hire as many conditional cooperators as possible. However, in practice, a widespread type of teamwork exists that is not fully captured by a public good setting: exchanges among workers with the objective of sharing knowledge and helping each other. This type of teamwork requires costly and unobserved individual effort, but is often unincentivized, because it does not directly produce tangible team output. Instead, teamwork in this modified sense creates a favorable environment for individual capability building, where *individual* performance *subsequent* to team production is measured, and incentivized. One situation of this kind is the exchange of best practice (internal knowledge) in firms (Tsai 2001, Szulanski et al. 2016), a crucial factor for generating, and continuously renewing, competitive advantage (Szulanski 1996). The modified team setting also covers helping, peer coaching, and cross-functional thinking, all important topics for organizations today.

In this article, we examine whether the positive effects of pro-social behaviors, such as conditional cooperation, carry over to the modified team setting, where individual output is measured and incentivized. On the one hand, prosociality could still be favorable, given that on-the-job helping and knowledge sharing require an investment of time and effort by the person who helps, without the certainty of receiving help back in the future. Thus, helping resembles a public good situation. On the other hand, compared to classical teamwork, it is not clear that maximum help is always optimal in the modified team setting in this study, because helping the team partner can take valuable resources from one's own production. Due to conditional cooperator's sense of reciprocity, they might be at risk of helping too much (even if the help is inefficient), if the team partner has been helpful in the past. In contrast, a free rider might choose more strategically when and how much to help.

We study the role of social preferences in such a modified setting, where team production constitutes a prerequisite for subsequent quantifiable individual performance. We are able to control for individual ability, and to obtain estimates for the intensity of the group exchanges. The subjects of this study are mathematics freshmen enrolled in a challenging mandatory course. Learner–learner exchanges constitute a core component of knowledge acquisition, according to the educational standards in mathematics (Ernest 2010), an idea communicated to the student by the faculty. In order to profit from positive effects of learner–learner exchanges (Eisenkopf 2010), the students in this study are required to hand in weekly homework assignments in pairs. We measure students' social preferences in a public good game (lab-in-the-field), using the standard procedure from Fischbacher et al. (2001).[8] In addition, we measure initial mathematical ability, and collect data on team performance (time spent on the task jointly and team members' self-reported levels of satisfaction with joint work). The main outcome variable is the student's individual performance on the final exam.

The public good game is especially suited to illustrate the situation the student participants encounter during their work on the assignments: Because the assignments are quite difficult, team production is important, for developing the necessary skills to solve the assignments, and ultimately, perform well individually on the exam. The public good is reflected in situations where one student has already obtained the necessary skill to handle a specific problem, but the other has not. As each team receives a collective grade for the assignment, there is no incentive to spend additional time to help a struggling team member, instead of just writing down and handing in the solution. However, over time, if helping is efficient, mutual help will maximize the skills of both team partners. Thus, taken together, the public good in this setting is the sum of the individual skills developed from teamwork.

We find that teams with conditional cooperators are not more successful jointly or individually in the modified setting. In contrast, free riders have a statistically significantly better individual performance on the final exam than other social types, controlling for individual ability. Furthermore, in this study free riders do not generate externalities, negative or positive, on their partners' individual performance. The most likely explanation for the results is that, contrary to classical teamwork, maximum help may not be socially optimal in the modified setting. A student might spend so much effort on helping her partner, that there might not be enough time left for her to proceed to the next level of understanding. This is especially true for conditional cooperators, who might

[8] The translated instructions are documented in Appendix A.

feel compelled to help their team partner a lot, if the team partner has been helpful in the past, even if helping is inefficient. In contrast, a free rider may be especially good at deciding when and how much to help, to secure the team partner's help in the future without losing too much time.

3.2 Related Literature

For the modified team setting, the economics literature has mostly examined the optimal task structure of specialization versus teamwork (Itoh 1991) and optimal incentives for helping on the job (Drago and Garvey 1998, Siemsen et al. 2007). Helping behavior in teams has been examined when agents have career concerns (Auriol et al. 2002). Helping and sabotaging have received special attention in literature on tournament incentives (Lazear 1989, Kräkel 2005). All of these models assume rational agents who are interested only in their own economic well-being. Helping in this context mostly emerges due to incentive schemes that are based on joint production.

The management literature, however, has paid substantial attention to the modified team setting. For instance, Flynn (2003) examines the effect of helping behavior on the job on individual productivity. He finds that employees who maintain an equitable balance between helping and being helped have the highest individual productivity. In addition, the vast literature on organizational citizenship focuses specifically on behaviors that are not part of one's narrow job description, but benefit the organization as a whole (Organ 1988, Podsakoff et al. 2000, DiPaola and Tschannen-Moran 2001).

The effect of social preferences on contributions in the public good setting has been studied extensively, mostly in the lab (Fischbacher et al. 2001, Fischbacher and Gächter 2010, Chaudhuri 2011). A common finding is that groups that count conditional cooperators among their members produce higher contribution levels. There are several reasons for this finding. First, conditional cooperators contribute to the common project, if they perceive that others are contributing as well. Second, conditional cooperators motivate selfish members, who rationally expect contributions from conditional cooperators in return, to participate more in the common project, at least in the first periods of a repeated game. Finally, the effects are amplified when punishment opportunities are given, as conditional cooperators negatively reciprocate free riding, punishing it even at a personal cost (Fehr and Fischbacher 2004).

There is ample evidence that social preferences are related to behavior in real life. Less selfish individuals are more likely to donate to charity (Benz and Meier 2008), and to participate in crowd-sourcing, such as Wikipedia (Algan et al. 2013). Gneezy et al. (2015) find that fishermen who must rely on teamwork due to environmental factors show more pro-social preferences across a range of experimental games than their neighbors who work individually. A detailed overview of literature on the generalizations of social behavior from the lab to the field, including its limitations, is given in Burks et al. (2016). This study shows that U.S. truck drivers' social preferences are related to their behavior toward their peers, but not toward the experimenters, with whom the truck drivers have fewer social connections.

Conditional cooperation (reciprocity) is particularly relevant for determining real-world outcomes. For example, cross-country skiers contribute more to the preparation of tracks, if they believe that others do so, too (Heldt 2005). Students' donations to a scholarship correlate with their beliefs about others' donations (Frey and Meier 2004). Reciprocity measured as second-mover behavior in a trust game (trustworthiness) is related to sales people's choice of selling strategy and success (Essl et al. 2018). Finally, according to a large-scale survey with employees in Germany, reciprocity influences tax morale (Frey and Torgler 2007) and effort exerted at work (Dohmen et al. 2009).

3.3 Study Design and Data Collection

The data was collected during the 2015/2016 winter semester in the course Analysis I at a major university in Germany. The 10 European Credit Transfer and Accumulation System (ECTS) course is mandatory for freshmen, and is the main course in mathematics in the first semester.[9] The course consists of two lectures per week, as well as a weekly tutorial in which take-home assignments are discussed. Class attendance is not compulsory. The typical exam failure rates in mathematics are about 30% to 50%, with 40% to 60% of students dropping out of university during the first few semesters. Thus, mathematics is generally a challenging university subject, and Analysis I is the first demanding course in mathematics.

To encourage learner–learner exchange, students are given weekly take-home assignments to solve in teams of two. The problem sets are designed to be difficult to the degree that most students cannot successfully solve them on their own. Accordingly, the instructor makes students aware of the importance of joint work as a prerequisite for a successful exam performance.

Each assignment is graded per team, meaning that the two team members receive the same number of points. At least 50% of the maximum attainable points must be obtained to gain exam admission. In the data set, no team failed to receive exam admission by at least the last assignment. In addition, because the solutions could always be copied from

[9] The curriculum for the first semester consists of 31 ECTS, with a total of 19 ECTS in mathematics and 12 ECTS in related subjects.

other teams, the team output is arguably not externally incentivized. However, the time required to reach the necessary points varies considerably. This opens up the possibility of judging the quality of the team's performance by the number of weeks needed to gain admission to the exam.

We conducted a lab-in-the-field experiment with the students, to assess their social preferences. In addition, we were given access to pseudo-anonymized data of individual exam grades, team composition, and scores for the homework assignments during the semester.[10] Figure 3.1 illustrates the timing of events and data collection.

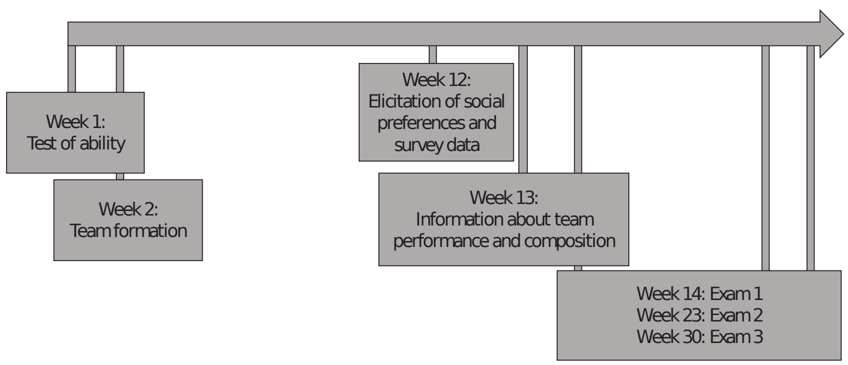


FIGURE 3.1: Timing of the study.

In the first lecture, students took a test based on high school math. The grade on this test was used as the measure of mathematical ability. Controlling for ability is important to avoid endogeneity in later analyses, as students with higher ability are more likely to earn better course grades.

In the second lecture, students were instructed by the lecturer to form teams of two. These teams had to turn in the homework assignments together. Due to fairness concerns, and as requested by the instructor, we had to keep the team formation procedure unchanged compared to previous semesters. Thus, we were not allowed to manipulate the formation of the teams. However, the teams formed at the beginning of the semester, and students had to sign up immediately after class with their partner. This reduces the probability that students selected their partners strategically. In the survey, 93% of students stated that they had not known the team partner before the course.

In week 12 of the semester, all students present in the lecture participated in a classroom paper-pen public good experiment that followed the procedure in Fischbacher et al.

[10]The students agreed that their student ID number would be matched with their player number, and stored physically at the university where the study was conducted. The list (or key) was used only once, at the end of the semester, to match the experimental data with the respective course data. Furthermore, students were assured that the key would be erased by the end of the study.

(2001). In total, 111 students out of the 158 enrolled students (roughly 70%) were present for this lecture and participated in the initial test of ability. All students present chose to participate in the study.[11] Participants decided on an unconditional contribution and a full (conditional) contribution table. Every participant was endowed with €5 (10 points). Integer contributions between 0 and 10 points were allowed. To capture the context of the take-home assignments, experimental groups consisted of two members. The efficiency factor of $\alpha = 0.7$ guaranteed a sufficient balance between individual and collective utility. To minimize social concerns toward their peers, students were informed that their decisions would be matched with those of another student from a different university.[12] After the experiment, students filled out a survey, stating how many hours per week they generally spent on the take-home assignment working individually and in the team, measured in 5-hour intervals, and how satisfied they were with the team collaboration and their partners' competence and engagement. In addition, students entered their demographic data, and indicated when they had first met their team partner. Finally, students signed a consent form allowing us to use their pseudo-anonymized data in the study.

There were three exam dates during weeks 14, 23, and 30, respectively. Students were free to register for one of these three dates (in case of failure, they could retake one of the subsequent exams), or they could even postpone the exam to the next semester. All tests were designed to be equally difficult. The variables used in the study are summarized in Table 3.1.

Table 3.1: The dependent and independent variables of the study

	Variable	**Measure**
Dependent	Individual performance	Points earned on the exam (min. 0; max. 60)
	Self-stated satisfaction with teamwork	Self-stated satisfaction with partner's engagement and competence, and team cooperation in general
	Team performance	Number of take-home assignments required to achieve admission to the exam (min. 6; max. 13)
Independent	Ability	Test of mathematical skills in the first lecture
	Social preferences	Experimental measures from public good game
	Individual effort, team effort	Self-reported hours spent on the assignments weekly individually and jointly with the team partner

A total of 158 students were enrolled in Analysis I. For 111 students, we have the measure of their ability from the first lecture and their behavior in the lab-in-the-field experiment.[13] For 85 of these students, we additionally have the experimental measures of

[11] One student refused to share his or her pseudo-anonymized exam data, and one failed to fill out the questionnaire. Both students' answers were excluded from the analysis.
[12] We collected the matched data two weeks later, and organized the payments shortly after.
[13] Note that lecture attendance is not mandatory.

their team partner. Both types of observations receive attention in the main analysis. 76 observations contain individual performance, ability, and experimental measures for *both* team members. This enables us to assess the quality of teamwork with the team members' perceived competence and engagement, and overall satisfaction, while controlling for ability.

3.4 Research Hypotheses

We assume a simple knowledge acquisition function where knowledge k_i for individual i is a function of *i's* ability a_i and the effort that i spends on homework assignments. This effort can be invested either in individual problem solving, ps_i, or in helping the team partner, h_i. Both types of effort increase an individual's knowledge; however, problem solving is more effective individually, while helping also benefits the team partner's knowledge. Finally, parts of the total available time T that each individual has at his or her disposal can also be spent completely unproductively on leisure, l_i. Thus, productive effort is given as

$$ps_i + h_i = T - l_i. \tag{3.1}$$

Taken together, the knowledge acquisition function looks as follows:

$$k_i = a_i + ps_i + \alpha \sum_{j=1}^{2} h_j, \tag{3.2}$$

where α is the marginal per-capita return (MPCR) of the public good, indicating the efficiency factor of helping each other. In common public goods with $n = 2$ individuals it is usually assumed that $0.5 < \alpha < 1$, indicating that contributions are individually detrimental (because the effort would be more productively spent on ps_i), but collectively efficient.

We start with standard predictions regarding the effect of ability and effort on performance ($\frac{\partial k_i}{\partial a_i} > 0$, $\frac{\partial k_i}{\partial ps_i} > 0$, $\frac{\partial k_i}{\partial h_i} > 0$).

Hypothesis 1 (The effect of ability and effort on individual performance)

i) Subjects with higher ability have better individual performance.

ii) Effort devoted to teamwork and individual work increases individual performance.

Now we turn to the hypothesis regarding the focus of the study: the effect of social preferences on performance in the modified team setting, where team production is not a goal in itself, but a prerequisite for individual performance. In this setting, helping becomes relevant when one team partner knows how to approach a problem, and the other does not. This situation resembles a public good game: Helping the team partner is individually costly, because it takes away valuable resources from one's own production ($\alpha < 1$), but if both partners help each other, the total knowledge of the team members increases ($n\alpha > 1$). It has been shown that production of a public good is more successful when the group includes conditional cooperators (Fehr and Fischbacher 2004, Rustagi et al. 2010), because conditional cooperators contribute if they think the other group members are contributing as well, and they are willing to punish deviators even at a personal cost. In this case of teamwork, punishment could mean withholding help for the team partner in the future, or it could mean fewer social interactions with the team partner.

Contrary to the classical public good setting, in this setting individual performance can be measured, and is the ultimate goal of the principal (the instructor). Given the argumentation above, conditional cooperators and their team partners invest more effort in helping each other (higher h_1 and h_2). Assuming that helping is efficient ($\alpha > 0.5$), conditional cooperators and their team partners, thus, should be more successful on the final exam compared to other social types.

Hypothesis 2 (The effect of social preferences on individual performance)

i) Conditional cooperators' individual performance is better than that of other social types, controlling for ability.

ii) Conditional cooperators' partners' individual performance is better than that of other social types, controlling for ability.

Finally, we predict that this positive effect on individual performance also translates into higher satisfaction with teamwork itself. Hypothesis 3 refers to literature about the intrinsic satisfaction of cooperation between workers. Several studies have shown theoretically and empirically that workers value team production in a cooperative environment (Kosfeld and von Siemens 2011, Rabin 1993, Hamilton et al. 2003).

Hypothesis 3 (The effect of social preferences on teamwork quality)

i) Individuals whose team partner is a conditional cooperator are more satisfied with their partners' engagement and competence, and the overall collaboration than individuals paired with a person of any other social type.

The following section contains analyses evaluating these hypotheses.

3.5 Results

Figure 3.2 shows the distribution of social preferences and ability among the 111 subjects: 48.7% ($N = 54$) are classified as conditional cooperators following the definition by Fischbacher et al. (2001). 17.1% ($N = 19$) contribute 0 in every conditional decision, and are classified as free riders. Around 17.1% ($N = 19$) show a hump-shaped contribution pattern. 2 subjects (1.8%) altruistically contribute all their endowment independently of the decision of the counterpart. For 15.3% ($N = 17$), social preferences cannot be classified according to standard definitions.

Ability is coded discretely between 1 and 5, with 1 the lowest grade and 5 the highest. A Kruskal-Wallis rank test[14] cannot reject the hypothesis that the distribution of ability does not differ across social types ($\chi^2_4 = 3.050$, $p = 0.55$); that is, social preferences do not explain differences in ability.

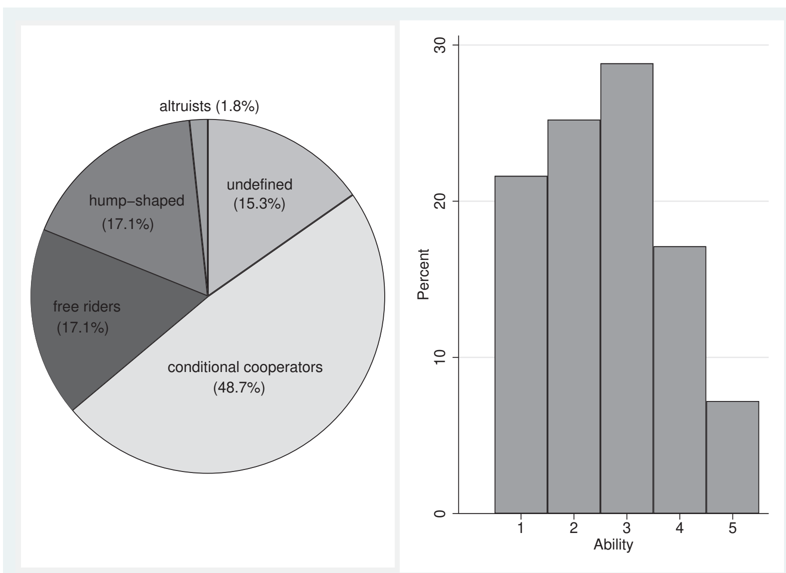


Figure 3.2: Distribution of social preferences and ability among subjects. Classification of social preferences according to Fischbacher et al. (2001).

Table 3.2 summarizes the number of students participating in each exam, the average number of points for those who passed (the main dependent variable), and the percentage of students who failed the exam.

[14] All tests are two-sided.

Table 3.2: Exam results

	Exam I	Exam II	Exam III	Total
Number of students	8	127	58	150
Average points for students who pass	38.8	36.1	31.0	34.9
Failure rate	37.5%	41.7%	53.4%	29.3%

Notes: Of the 158 totally enrolled students, 8 did not sign up for an exam, and 44 failed, of whom 43 students retook the exam. In the column *Total*, each student is counted once. Exams are designed to be of comparable difficulty. Differences in the failure rate can be explained by selection effects.

Table 3.3 contains the main regression analyses of this study related to hypotheses 1 and 2. The dependent variable is individual performance measured by the number of points obtained on the exam, distributed between 0 and 60, with 30 points as the minimum number necessary to pass.[15] In model (1), the independent variables are ability, and the self-reported number of weekly hours spent on the homework individually and working together with the team partner.[16] For both effort variables, the category "less than 5 hours per week" serves as reference. Model (2) considers the effect of the student's social type and model (3) the effect of the team partner's social type. In all cases the undefined social types serve as the reference category.[17] Changing the reference type to all non-freerider yield qualitatively similar results.

[15]We use Tobit regressions to fit the data to the lower and upper bounds as given by the exam. The coefficients and significance are qualitatively the same if we use an ordinary least squares (OLS) regression instead.

[16]Table 3.5 in Appendix B shows the same regression with combined effort levels. Results remain qualitatively the same, with more effort implying a better performance on the exam.

[17]Two students classified as altruists are also part of the reference category. Regressions in which these two observations are part of a separate category or in which they are dropped yield results similar to those presented here.

Table 3.3: Tobit regressions on the exam score

	(1) Effort&Ability	(2) (1)+Type	(3) (2)+Partner's type
Ability	3.703***	3.794***	4.039***
	(0.985)	(1.007)	(1.091)
Team effort: 5-10 hrs	-0.426	-0.679	3.948
	(2.586)	(2.603)	(2.724)
Team effort: > 10 hrs	6.386	7.265*	9.047**
	(3.910)	(3.666)	(3.905)
Indiv. effort: 5-10 hrs	9.596***	10.08***	9.876***
	(2.824)	(2.616)	(3.379)
Indiv. effort > 10 hrs	-1.815	-0.396	7.783
	(4.849)	(4.978)	(5.102)
Free rider		9.793***	6.925**
		(3.267)	(3.474)
Conditional cooperator		4.249	3.058
		(3.427)	(3.385)
Hump-shaped		3.449	4.832
		(3.874)	(4.361)
Partner, free rider			0.373
			(4.992)
Partner, conditional cooperator			-1.580
			(4.015)
Partner, hump-shaped			4.903
			(4.160)
Constant	17.33***	12.58***	10.95*
	(3.284)	(4.182)	(5.561)
Observations	111	111	85
Pseudo-R^2	0.027	0.034	0.037

Notes: Lower and upper bounds, 0 and 60, respectively. Robust standard errors, clustered on the team level, in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The first hypothesis predicts that the main output (individual performance) depends on one's ability and the effort invested in solving the assignments. Table 3.3 provides evidence in support of this hypothesis. The coefficient of *Ability* is statistically significant in all three models. For example, according to model (1), a student who obtains a one point higher grade on the ability test in the first week receives around an additional 3.7 points on the final exam. Regarding effort, it can be seen that individual effort is already useful at a lower level, while working together pays off in grades only after students spend more than 10 hours per week working together. This result is reasonable, because individual task-solving capacities are attained after a short phase in the beginning, whereas teams need to spend time with interacting socially and overcoming communication problems at first. Then, teamwork becomes extremely effective, with team members earning 7 to 9 additional points individually on the exam. Table 3.5 in Appendix B combines both types of effort for each student.[18] In all specifications, more effort is associated with a better individual performance. Evidence can be summarized as:

Result 1: Ability and effort, both individual and joint, increase individual performance.

The main hypothesis 2 examines the individual performance of the social types. According to models (2) and (3) in Table 3.3, conditional cooperators do not perform statistically significantly better than other social types. In addition, in model (3) the coefficient of *Partner, conditional cooperator* is negative and is not statistically significant, meaning that working together with a conditional cooperator does not lead to higher individual grades.

In contrast, free riders statistically significantly outperform other social types, earning, on average, around 7 to 10 more points on the exam.[19] Next, we examine whether the free rider's success comes at the expense of their team partners. In model (3), the coefficient of *Partner, free rider* is not significant, meaning that being on a team with a free rider does not lead to earning lower individual grades. Incorporating the partner's ability in models (1) to (3) does not change the results, while the respective coefficient is small and insignificant in every model.[20] The main result of the study can be summarized as follows:

Result 2: Conditional cooperators do not perform statistically significantly better on the exam compared to other social types. In addition, the conditional cooperator's team partners also do not perform significantly better. In contrast, free riders produce significantly higher individual output than other social types, and do not generate externalities for their partners. Controlling for ability, these results hold.

[18] Different specifications, such as controlling for the exam date, leave the results qualitatively the same.
[19] Results are similar if we pool all types who are not free riders in the reference group.
[20] Results available on request.

Hypothesis 3 concerns the quality of the teamwork, which is a precondition for individual success. Team partners' self-reported satisfaction with joint work serves as proxy a for this evaluation. Satisfaction with their partners' engagement and competence, and the overall collaboration is measured on a scale between 0 ("fully unsatisfied") to 3 ("fully satisfied"). Table 3.4 presents the regression results for each assessment.

Table 3.4: Tobit regressions on the team partner's perceived competence and engagement and overall satisfaction

	(1) Competence	(2) Engagement	(3) Satisfaction
Partner conditional cooperator	0.742	0.256	-0.326
	(0.494)	(1.038)	(0.208)
Partner free rider	1.280**	-0.058	-0.120
	(0.540)	(1.054)	(0.305)
Partner hump-shaped	0.066	-0.777	-0.054
	(0.598)	(1.089)	(0.327)
Partner's ability	0.475***	0.502	-0.083
	(0.175)	(0.324)	(0.093)
Ability of person who assesses	-0.131	-0.309	0.135
	(0.160)	(0.291)	(0.110)
Constant	1.258***	2.971***	1.930***
	(0.456)	(1.091)	(0.288)
Observations	76	76	76
Pseudo-R^2	0.061	0.027	0.017

Notes: Lower and upper bounds, from 0 ("fully unsatisfied") to 3 ("fully satisfied"). Robust standard errors, clustered on team level, in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The team partner's assessment of the conditional cooperators' competence is positive, but not statistically significant. In contrast, partners are significantly more satisfied with free riders' competence, compared to the reference category of unclassified individuals ($p < 0.05$). This result holds when controlling for ability; that is, ability is not the driving factor of the assessment. A student's ability itself is a significant factor of satisfaction with his or her competence, as assessed by his or her teammate ($p < 0.01$). Regarding the assessment of engagement and overall satisfaction, there are no significant differences across the partners' social types. In line with these results, Table 3.6 in Appendix B further examines team effectiveness as the number of assignments required to achieve exam admission. It indicates that group composition in terms of social preferences does not affect team production. This suggests that conditional cooperators, again, fail to thrive in this modified team setting. Furthermore, it strengthens the result that free riders do not exert negative externalities on their team members.

Summing up the evidence shown in Tables 3.4 and 3.6, the following is stated:

Result 3: Free riders are, independent of their ability, perceived as more competent by their team partners. There is no such effect for conditional cooperators or other social types. The assessment of the team partner's engagement and the overall collaboration does not depend on the partner's social type. Additionally, the team's composition with respect to social types does not influence the team's effectiveness.

3.6 Discussion and Conclusion

The present study tests the effect of social preferences on performance in a modified setting. In this setting, collective work is not externally incentivized, but is necessary to produce valuable individual output. This framework applies to a number of important situations in the organizational context, such as internal knowledge sharing, helping on the job, or peer coaching. The study context – a lab-in-the-field experiment with university students combined with field data – allows to reliably measure individual output (final grades), controlling for ability.

The initial hypotheses regarding the higher achievement of conditional cooperators with respect to individual performance and quality of teamwork must be rejected. We do not find better performances on the exam for conditional cooperators or their team partners. Given that teamwork is a prerequisite for individual performance in this setting, it is not surprising that conditional cooperators are also not perceived as more engaged or competent by their team partners, and their teams do not need less time to earn exam admission. In contrast, free riders perform significantly better on the final exam compared to other social types. As an additional result, in this study free riders do not generate externalities, negative or positive, on their partners' individual performance. Teams that include free riders do not need significantly more time to achieve joint production targets, and team partners are significantly more satisfied with free riders' competence than with the competence of other social types, even after controlling for ability.

Overall, based on the data, we conclude that when team exchanges are just part of a bigger picture, ultimately resulting in individual production, the presence of conditional cooperators does not have a positive effect, and the presence of free riders does not have a negative effect on team production, as in classical teamwork. On the contrary, free riders seem to be more successful in producing individual output without hurting others' performance, or the joint performance of the team. A possible reason for the success of free riders could be that they have a tendency toward rationality, which is an advantage for math studies. This explanation is, however, unlikely to be decisive in this case, as free riders do not perform better on the initial test, and we control for ability in the regression.

In our view, the most likely reason for these results is that, contrary to classical teamwork, maximum help may not be socially optimal in this modified setting. A student might spend so much effort on helping her partner, that there might not be enough time left for her to proceed to the next level of understanding. This is especially true for conditional cooperators, who might feel compelled to help their team partner a lot, if the team partner has been helpful in the past. In contrast, a free rider may be especially good at deciding when and how much to help, to secure the team partners' help in the future without losing too much time. This behavior could even result in free riders being perceived as particularly competent, if they sometimes explain their knowledge in detail, and sometimes just state the solution to a problem. In contrast, conditional cooperators might feel compelled to help, even if this is inefficient, if the team partner has been helpful in the past. In particular, in the case of helping, teamwork might sometimes be efficient, and sometimes be inefficient (formally, sometimes, $\alpha > 0.5$, and sometimes, $\alpha < 0.5$). For instance, if one team member absolutely does not understand a solution, it might take inefficiently long to explain just the basics of the solution to this person. A conditional cooperator might still feel compelled to spend this time on helping, but a free rider might choose more strategically when and how much to help.

By showing that free riders may work more efficiently in a modified team setting, the study provides clear managerial implications. First, hiring people with selfish social preferences for jobs that involve the modified team setting may prove useful based on the results of the study. In addition, for firms that like to capture the benefits of teamwork, but at the same time, aim to reduce the problem of free riding, it might be possible in some settings to transform teamwork in a modified team setting, by adding an individually measurable output. That way, organizations would not spend resources to screen out free riders at the recruitment stage, or give up on team output lost due to free riding, but strategically embed workers, based on their social preference, in suitable team environments.

In the given setting, several factors beyond social preferences and ability can influence performance. For instance, the degree of dynamic inconsistency (impatience) may vary among students, leading some of them to abandon coursework during the semester in favor of leisure and partying, despite the wish to earn a high grade (Augenblick et al.

2015). In addition, the level of trust toward the course instructor, and his claim that teamwork is important, may lead students to take the homework assignments more or less seriously. We chose not to control for these factors, for two reasons. First, because we are not aware of studies that show that these reasons are related to social preferences. Second, we aimed to keep the data collection simple to realize in a classroom environment during a lecture.

The study, of course, has several limitations. The analysis uses self-reported data on the team partner's assessment and on the effort invested, which might be biased systematically by different social types. In addition, we were not able to use random assignment to teams. Although 93% of students claimed that they had not known their team partner before the course, there might still be unobserved factors influencing team composition that we are not able to control for. Finally, because this study focused on only one course in the curriculum, we are not able to observe the effect of partnering with a free rider in a multitasking context, for example, taking multiple courses into account. It may well be that students in teams with free riders must invest more effort to study for the selected challenging mandatory course, and that their performance in other courses suffers. This, of course, is possible, however, unlikely, as the time reported spent working with the team was not lower for teams with free riders, and free riders' engagement was not reported to be lower than the engagement of subjects with other social preferences.

Future research on the modified team setting should use longitudinal data to study how the effects of social preference on performance develop over time, and take into account multitasking, as described in the previous paragraph. In addition, it would be interesting to study social preferences in the modified team setting with employees of a company where helping and knowledge sharing are important prerequisites for success. Finally, it would be interesting to conduct a laboratory experiment, where in the first part, subjects' social types are elicited using the standard procedure from Fischbacher et al. (2001), and in the second part, subjects play a repeated public good game, where contributions are sometimes inefficient, but still benefit the other player. Based on the present results, we would predict that conditional cooperators choose to contribute in the second part, even in inefficient situations, if the other player had contributed in these situations in the past.

References

Alchian, A. A. and Demsetz, H. (1972). Production, information costs, and economic organization. *American Economic Review*, 62(5):777–795.

Algan, Y., Benkler, Y., Morell, M. F., and Hergueux, J. (2013). Cooperation in a peer production economy experimental evidence from wikipedia. In *Workshop on Information Systems and Economics*, pages 1–31, Milan, Italy.

Augenblick, N., Niederle, M., and Sprenger, C. (2015). Working over time: Dynamic inconsistency in real effort tasks. *The Quarterly Journal of Economics*, 130(3):1067–1115.

Auriol, E., Friebel, G., and Pechlivanos, L. (2002). Career concerns in teams. *Journal of Labor Economics*, 20(2):289–307.

Benz, M. and Meier, S. (2008). Do people behave in experiments as in the field? Evidence from donations. *Experimental Economics*, 11(3):268–281.

Burks, S. V., Nosenzo, D., Anderson, J., Bombyk, M., Ganzhorn, D., Götte, L., and Rustichini, A. (2016). Lab measures of other-regarding preferences can predict some related on-the-job behavior: Evidence from a large scale field experiment. Working paper.

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Experimental Economics*, 14(1):47–83.

Delarue, A., Van Hootegem, G., Procter, S., and Burridge, M. (2008). Teamworking and organizational performance: A review of survey-based research. *International Journal of Management Reviews*, 10(2):127–148.

DiPaola, M. and Tschannen-Moran, M. (2001). Organizational citizenship behavior in schools and its relationship to school climate. *Journal of School Leadership*, 11(5):424–447.

Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2009). Homo reciprocans: Survey evidence on behavioural outcomes. *The Economic Journal*, 119(536):592–612.

Drago, R. and Garvey, G. T. (1998). Incentives for helping on the job: Theory and evidence. *Journal of labor Economics*, 16(1):1–25.

Eisenkopf, G. (2010). Peer effects, motivation, and learning. *Economics of Education Review*, 29(3):364–374.

Ernest, P. (2010). Reflections on theories of learning. In *Theories of Mathematics Education*, pages 39–47. Springer.

Essl, A. M., von Bieberstein, F., Kosfeld, M., and Kröll, M. (2018). Sales performance and social preferences. Working paper, University of Bern.

Fehr, E. and Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Sciences*, 8(4):185–190.

Fischbacher, U. and Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Economic Review*, 100(1):541–556.

Fischbacher, U., Gächter, S., and Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, 71(3):397–404.

Flynn, F. J. (2003). How much should I give and how often? The effects of generosity and frequency of favor exchange on social status and productivity. *Academy of Management Journal*, 46(5):539–553.

Frey, B. S. and Meier, S. (2004). Social comparisons and pro-social behavior: Testing "conditional cooperation" in a field experiment. *American Economic Review*, 94(5):1717–1722.

Frey, B. S. and Torgler, B. (2007). Tax morale and conditional cooperation. *Journal of Comparative Economics*, 35(1):136–159.

Gneezy, U., Leibbrandt, A., and List, J. A. (2015). Ode to the sea: Workplace organizations and norms of cooperation. *The Economic Journal*, 126(595):1856–1883.

Hamilton, B. H., Nickerson, J. A., and Owan, H. (2003). Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of Political Economy*, 111(3):465–497.

Heldt, T. (2005). Conditional cooperation in the field: Cross-country skiers' behavior in Sweden. In *Economic Science Association, European Meeting, Alessandria, Italien, September 15-18, 2005*.

Holmstrom, B. (1982). Moral hazard in teams. *The Bell Journal of Economics*, 13(2):324–340.

Isaac, R. M., Walker, J. M., and Thomas, S. H. (1984). Divergent evidence on free riding: An experimental examination of possible explanations. *Public Choice*, 43(2):113–149.

Itoh, H. (1991). Incentives to help in multi-agent situations. *Econometrica: Journal of the Econometric Society*, 59(3):611–636.

Kosfeld, M. and von Siemens, F. A. (2011). Competition, cooperation, and corporate culture. *The RAND Journal of Economics*, 42(1):23–43.

Kräkel, M. (2005). Helping and sabotaging in tournaments. *International Game Theory Review*, 7(2):211–228.

Lazear, E. P. (1989). Pay equality and industrial politics. *Journal of Political Economy*, 97(3):561–580.

Lazear, E. P. and Shaw, K. L. (2007). Personnel economics: The economist's view of human resources. *Journal of Economic Perspectives*, 21(4):91–114.

Organ, D. W. (1988). *Organizational citizenship behavior: The good soldier syndrome*. Lexington Books/DC Heath and Com.

Podsakoff, P. M., MacKenzie, S. B., Paine, J. B., and Bachrach, D. G. (2000). Organizational citizenship behaviors: A critical review of the theoretical and empirical literature and suggestions for future research. *Journal of Management*, 26(3):513–563.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review*, 83(5):1281–1302.

Rob, R. and Zemsky, P. (2002). Social capital, corporate culture, and incentive intensity. *RAND Journal of Economics*, 33(2):243–257.

Rustagi, D., Engel, S., and Kosfeld, M. (2010). Conditional cooperation and costly monitoring explain success in forest commons management. *Science*, 330(6006):961–965.

Siemsen, E., Balasubramanian, S., and Roth, A. V. (2007). Incentives that induce task-related effort, helping, and knowledge sharing in workgroups. *Management Science*, 53(10):1533–1550.

Szulanski, G. (1996). Exploring internal stickiness: Impediments to the transfer of best practice within the firm. *Strategic Management Journal*, 17(S2):27–43.

Szulanski, G., Ringov, D., and Jensen, R. J. (2016). Overcoming stickiness: How the timing of knowledge transfer methods affects transfer difficulty. *Organization Science*, 27(2):304–322.

Tsai, W. (2001). Knowledge transfer in intraorganizational networks: Effects of network position and absorptive capacity on business unit innovation and performance. *Academy of Management Journal*, 44(5):996–1004.

Appendix A: Instructions

Dear Student,[21]

By participating in today's game you take part in a scientific study about decisions. Please read this instruction carefully. The game is going to last approximately 45 minutes. You won't have any additional effort. Depending on your decisions you have the opportunity to earn a **payment** which you will receive in cash in Analysis 1 on December 9, 2015 .

You are not allowed to speak during the game. If you have a question, raise your hand and a person who conducts the game will come to your seat. A non-observance of these rules leads to an exclusion of the game and the payment.

Of course, you are free to choose whether to take part in the study or not. You have the possibility to withdraw your agreement at any time without giving reasons and without personal disadvantages.

During the game your payment is determined in points and will be converted to Euro at the end of the game. It counts

$$1 \text{ point} = 0.5 \text{ EUR}$$

[21] Translated from German.

General rules

You are in a **group of two**. Every group member needs to decide about the use of 10 points. The points can be put **wholly or partly** to a **private account** or can be invested in a **project**. Every point that is not invested in the project will be automatically put in the private account.

Earnings from the private account

For every point you put in the private account you earn exactly one point:

Earnings from private account = Own contribution to private account

Example: If you put 10 points in the private account, you earn 10 points out of it. If you put 6 points in the private account, you earn 6 points out of it. No person other than yourself earns out of your contribution to your private account.

Earnings from the project

Not just you, but also your other group member will earn from your investment to the project. Conversely, you earn from the other group member's investment to the project.

The earnings from the project for every group member are defined as followed:

Earnings from the project = Sum of the contributions to the project $\times$ **0.7**

Example: If every group member invests 10 points to the project, you and your other group member earn $20 \times 0.7 = 14$ points out of the project. If one group member invests 8 points and the other one invests 7 points, every group member earns $(8+7) \times 0.7 = 10.5$ points out of the project.

Total earnings

Your total earnings are the sum of your earnings from the private account and the project.

Earnings from the private account = 10 − Your contribution to the project

+ **Earnings from the project** = $0.7 \times$ Sum of the contributions to the project

---

= **Total earnings**

Comprehension questions

Please respond to the following comprehension questions. They are merely aimed at familiarizing you with the calculation of the earnings which occur with the differing decisions about the use of the 10 points.

1. Every group member has 10 points at his disposal. You invest 9 points to the project. The other group member contributes 6 points to the project.

   (a) How high are **your** total earnings in points?

   ☐ 10.0 ☐ 15.0 ☐ 11.5 ☐ 10.5 ☐ 6.5

   (b) How high are the total earnings of the **other** group member in points?

   ☐ 8.0 ☐ 11.5 ☐ 7.5 ☐ 14.5 ☐ 8.0

2. Every group member has 10 points at his disposal. The other group member contributes 5 points to the project.

   (a) How high are **your** total earnings if you invest - additionally to the 5 points invested by the other group member - 0 points in the project?

   ☐ 15.0 ☐ 13.0 ☐ 19.4 ☐ 13.5 ☐ 9.0

   (b) How high are **your** total earnings if you invest - additionally to the 5 points invested by the other group member - 10 points in the project?

   ☐ 12.5 ☐ 6.5 ☐ 10.5 ☐ 8.0 ☐ 6.0

The game

The game involves the decision situation as previously described. From now on you are part of a group of two. The other group member is a student from another university. **Neither you nor another participant knows who the other person of his group is.** The compositions of the other groups are also unknown.

Now you have 10 points at your disposal which you can put in your private account or invest to the project. Every group member has to make two types of contribution decisions. In the following they're called **unconditional** and **conditional contribution**.

1. In the first type of contribution decisions (**unconditional** contribution) you have to determine how many of the 10 points you want to invest.
2. Your second task is to complete a **contribution table** with your **conditional** contributions. For **every possible contribution** of the **other** group member you need to give an amount of points to the project in the contribution table as answer to this contribution.

Once all participants have made their unconditional and conditional contribution decisions one group member of each group will be randomly chosen. For the **randomly chosen group member** only the completed table is relevant. For the other not randomly chosen group member is solely the unconditional contribution relevant. Therefore the unconditional contribution of one group member and the corresponding conditional contribution of the other group member are included in the calculation of the payments

Your decisions

You and your other group member each have 10 points to spend. Please decide on the unconditional contribution and all possible conditional contributions.

Unconditional contribution:

1. How many points would you like to invest to the project? .... point(s)

Conditional contribution (Please take all eleven decisions):

Please fill **whole** numbers between 0 and 10.

1. If the other group member contributes **0** points: .... points
2. If the other group member contributes **1** points: .... points
3. If the other group member contributes **2** points: .... points
4. If the other group member contributes **3** points: .... points
5. If the other group member contributes **4** points: .... points
6. If the other group member contributes **5** points: .... points
7. If the other group member contributes **6** points: .... points
8. If the other group member contributes **7** points: .... points
9. If the other group member contributes **8** points: .... points
10. If the other group member contributes **9** points: .... points
11. If the other group member contributes **10** points: .... points

Please fill out the questionnaire on the next page. For this you receive additional **2 EUR** to your earnings from the game.

Appendix B: Supplementary analyses

Table 3.5: Tobit regressions on the final score (combined effort)

	(1) Effort&Ability	(2) (1)+Type	(3) (2)+Partner's type
Ability	4.160***	4.237***	4.397***
	(0.960)	(1.000)	(1.082)
Individual & team effort: 10-15 hrs	2.770	2.172	7.133**
	(3.054)	(3.010)	(3.484)
Individual & team effort: 15-20 hrs	6.750*	6.918*	8.271*
	(3.544)	(3.531)	(4.768)
Individual & team effort: > 20 hours	6.758	8.380	16.73***
	(6.123)	(5.924)	(4.367)
Free rider		8.929**	5.871
		(3.677)	(3.591)
Conditional cooperator		3.248	3.268
		(3.633)	(3.084)
Hump-shaped		0.934	3.664
		(3.844)	(4.280)
Partner, free rider			-0.594
			(4.586)
Partner, conditional cooperator			-3.437
			(3.255)
Partner, hump-shaped			3.110
			(3.464)
Constant	15.70***	12.42***	11.50**
	(3.279)	(4.239)	(5.233)
Observations	111	111	85
Pseudo-R^2	0.018	0.024	0.050

Notes: Lower and upper bounds, 0 and 60, respectively. Table contains new categories for combined effort. Robust standard errors, clustered on team level, in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.6: Tobit regression on number of assignments required to reach admission to exam

	Number of assignments
Ability: Both medium	-0.268
	(0.979)
Ability: Both high	-2.020**
	(0.806)
Ability: Low and high	0.05
	(0.668)
Ability: Low and medium	-0.207
	(0.722)
Ability: Medium and high	-0.591
	(1.160)
Individual & team effort: 10-15 hrs	-0.378
	(0.501)
Individual & team effort: 15-20 hrs	-1.644*
	(0.874)
Individual & team effort: > 20 hrs	-1.516
	(1.049)
At least one conditional cooperator	-0.072
	(1.242)
Both conditional cooperator	0.133
	(1.396)
At least one free rider and no conditional cooperator	-0.057
	(1.398)
Constant	10.590***
	(1.629)
Observations	38
Pseudo-R^2	0.096

Notes: Here, we investigate the impact of social preferences on team performance measured by the timing of admission to the exam. We note that this measure is not perfect, because there are no incentives attached to early completion. Nonetheless, on average, strong teams should earn the required amount of points earlier. The variance in the data (10 assignments on average needed to earn exam admission with a standard deviation of 1.5; minimum 6 assignments required), indicates that it can be used as a proxy for team performance. Lower and upper bounds, 6 and 13, respectively. Negative coefficients mean earlier admission. Robust standard errors, clustered on team level, in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Essay 3: The spin doctor in the lab: An experiment on perceived intentions

Gerald Eisenkopf, Jonas Gehrlein*

Abstract

In this paper, we study the role of intentions, and whether people act on the possibility to disguise them. We further investigate how recipients react to such behavior, by measuring punishment decisions. Using several variations of the mini-dictator game, the results show that intentions matter, and recipients tend to reduce punishment for unequal outcomes, if the alternative was even more unfavorable. Additionally, this result holds true for situations where participants know that the dictator has the possibility of casting their action in a better light than appropriate. Furthermore, even recipients who are informed about the attempted disguise do not substantially increase the punishment. Knowing this, most of the dictators make heavy use of spin doctoring, to disguise their malign intentions, and thus, benefit, on average, from higher payoff. The results shed some light, for example, on recent political developments, where politicians do not lose touch with voters, even when the politicians are obviously dishonest about their intentions.

Keywords: intention-based punishment, disguise, inequity aversion, dictator game

*Gerald Eisenkopf: gerald.eisenkopf@uni-vechta.de, University of Vechta, Faculty I; Jonas Gehrlein: jonas.gehrlein@iop.unibe.ch, Institute for Organization and Human Resource Management, University of Bern. We thank Zita Spillmann, Nicolas Hafner, and Fabienne Zwick for their excellent research assistance. We are grateful to Frauke von Bieberstein, Ann-Kathrin Crede, Stefani Jaussi, Kathrin Friedrich, and two anonymous referees for valuable feedback. All remaining errors are ours.

4.1 Introduction

Few people admit that they have harmed others to increase their own payoff. We often try to disguise our true intentions, and emphasize constraints that do not allow for more benevolent outcomes. For example, managers rarely admit that they reduce wages to increase their bonus payments. Instead, managers rather refer to the firm's vulnerability, and point out that lower wages help save jobs. Similarly, workers blame faulty equipment or other adversary conditions, rather than their own laziness, if they performed poorly. One motive behind such window dressing is the deflection of anger. People punish less intensely if they do not detect bad intentions behind a harmful decision (see, for example, Charness and Levine 2003, 2007, Falk et al. 2008, Rand et al. 2015). However, it is unclear under which circumstances such disguising is actually effective. Fired workers may not believe the managers' statements. Moreover, if obviously untrue statements add insult to injury, then the worker's anger might actually intensify. This reasoning implies that potential window dressers face a trade off, which we study in this paper. We look at whether people disguise their true intentions in situations where they act unkindly. We also investigate how their "victims," or the recipients respond to such behavior. More specifically, the experimental design allows for a clean distinction between recipients who face disguised intentions, and those who learn that window dressing occurred. In this within-subjects experiment, participants play three independent mini-dictator games, each involving one dictator and three receivers. In each game, all players know that there are four potential allocations of equal total size that differ regarding the payoff distribution between the dictator and the receivers. At first, the computer eliminates two randomly chosen allocations. Then, the dictator can choose between the two remaining alternatives. The first of the three games ends at this stage. It generates information about the distributional preferences of the dictator. The second game provides an extension, and enables the receivers to punish the dictator. The receivers see the remaining two allocations, and can withdraw points from the dictator at some cost. Thus, receivers can incorporate intentions into their punishment decisions, as the receivers see whether the dictator could have chosen a kinder alternative or not. In the third game the dictator can disguise her intentions. The receivers still see two out of four allocations and they know that the dictator has chosen one of them. However, the dictator can choose which alternative she will display alongside the chosen one. This choice can facilitate disguising true intentions, for example, by displaying a very unkind alternative along with the chosen unkind one. Ahead of the punishment decisions, one of the three receivers learns about the dictator's actual choice set, and whether she has disguised her intentions. The results show that dictators extensively disguise their intentions. The vast majority of dictators generate the choice set that minimizes perceived unkindness. This behavior reduces the expected punishment by deceived (uninformed) receivers. Interestingly, unsuccessful deception does not trigger substantially higher punishment by informed receivers, compared to the situation without the possibility to disguise.

Thus, this paper contributes to two strands of literature. One strand has shown that the perception of intentions matters when people decide about punishment (Charness and Levine 2003, 2007, Falk et al. 2008, Rand et al. 2015, Cushman et al. 2009), or in similar prosocial behavior, such as trust and gift-giving (Alempaki et al. 2016, Orhun 2018, Toussaert 2017). Prominent theories of reciprocal behavior incorporate this evidence (Dufwenberg and Kirchsteiger 2004, Falk and Fischbacher 2006, Cox et al. 2007). The perception of intentions depends strongly on the available alternatives. The implementation of a relatively unequal allocation may be kind, if the available alternatives imply a larger inequality. The same decision appears unkind, if an egalitarian outcome is feasible. We contribute to this literature by showing that people try to disguise in this decision-making context, and how other people respond to this window dressing. This approach is in contrast to Friehe and Utikal (2018), who investigate a variation of the dictator game with punishment in which the dictator can spend money to hide information about her unfair intentions. Their results show that many recipients respond to an attempted cover-up with increased punishment.

We also contribute to the literature on honesty in economic interactions.[22] Many people punish others who do not tell the truth. Disguising intentions is a borderline case of lying, and literature has shown that people punish others who do not tell the truth (Brandts and Charness 2003, Sánchez-Pagés and Vorsatz 2007, 2009, Lundquist et al. 2009). People are clearly not honest about their intentions, and individuals may make untrue statements in the process. Moreover, the deliberate act of disguising reveals that people are aware of the social consequences of their distribution decisions. However,

[22] Rosenbaum et al. (2014), Jacobsen et al. (2018), Gerlach et al. (2019) and Abeler et al. (2019) provide recent and extensive reviews of this literature.

disguising has no immediate material impact. It “only” violates the social norm of honesty, but causes no material harm.

4.2 Experimental Design

The within-subjects design of the experiment contains three variations of a one-shot mini-dictator game with one dictator (Player A) and three recipients (Players B1–3, the translated instructions are documented in Appendix A). In the first game, each subject decides as dictator; whereas, for the subsequent two games, all subjects are assigned randomly into groups of four, containing one dictator and three recipients. Once the roles are determined, they remain fixed in the subsequent two mini-dictator games. At the end of the experiment, subjects receive a show-up fee of €3 and an additional payment according to the outcome of one randomly chosen game. In each game, it is common knowledge that the computer selects two out of four distribution options. The dictator then chooses one of the remaining two options. As Table 4.1 shows, each option implies an aggregate payoff of 24 points (1 point = €2.5), but they differ with respect to inequality between the dictator and the receivers. Note that all receivers always get the same payoff.

TABLE 4.1: The four potential distributions in the mini-dictator games

Player/Option	Option 1	Option 2	Option 3	Option 4
Label	Altruistic	Egalitarian	Unfair	Very Unfair
Player A	0	6	12	18
Player B1	8	6	4	2
Player B2	8	6	4	2
Player B3	8	6	4	2
$\sum$	24	24	24	24

4.2.1 Game I: Dictator game

The first game is a simple mini-dictator game, as described above. We make extensive use of the strategy method, because we want to measure the distributional preferences of the participants. That is, every subject must decide as a dictator in different circumstances. The subjects face five different binary choices in random order. Each binary choice involves at least one of the two unfair options.[23] Because receivers cannot punish at this stage, and subjects do not learn about the actual assignment of the roles and the outcome of this game until the very end of the experiment, the choices provide an undistorted measure of the underlying distributional preferences. After the first game ends, subjects learn about their prospective roles in the next two games. Each subject gets the information whether she will decide as a dictator or as a receiver. Subjects know that this information does not indicate whether one of their decisions as dictator in Game I will become relevant for payoffs.

4.2.2 Game II: Dictator game with punishment

In the second game, we eliminate the strategy method for the chosen dictators. The computer randomly picks one of the five choice sets for each dictator. The dictator then chooses one of the available options. Simultaneously, the receivers learn about the eliminated options, and decide individually about a punishment for each of the two available options. We use the following punishment technology: Each receiver decides whether she wants to take away points from the dictator. Each eliminated point implies a cost of .5 points for the group of receivers. Note that all three receivers share the costs of punishment, which eliminates inequity and free-rider concerns among the receivers. Thus, the per-capita costs of a one-point punishment are 1/6. The budget for the punishment stems from the option which the receiver is punishing, and is constrained, to leave the subject with a non-negative payoff. At the end of the experiment, the computer selects one of the three receivers. It will implement her punishment decision for the option that has been chosen by the dictator. All other decisions become irrelevant. This second game measures how receivers assess the dictator's kindness. Consider the case in which the dictator faces the choice between the fair option and the unfair option (Options 2 and 3). In this context, the choice of Option 3 is very unkind. However, if the dictator faces the choice between the unfair option and the very unfair option (Options 3 and 4), the choice of Option 3 is a kind action.

[23] Thus, subjects face the choice between Options 1 and 3, or 1 and 4, or 2 and 3, or 2 and 4, or 3 and 4. They never face a choice between Options 1 and 2.

4.2.3 Game III: Dictator game with punishment and disguise

The computer again randomly presents two of the four options to the dictator. The dictator then chooses one of the available options. Afterward she can disguise her intentions to some extent by framing the punishment context for two of the three receivers. More specifically, she can choose which option to present alongside the actual chosen option. Thus, if she faced a choice between Options 2 and 3, and chose the unfair option, she might present the very unfair Option 4 along with Option 3 to the uninformed receivers. These uninformed receivers then decide about punishing the presented options, without knowing whether they reflect the dictator's true choice set. Finally, the receivers state their belief whether the presented choice set resembles the original one faced by the dictator. Meanwhile, the third receiver learns about the actual two alternatives. We call this player an *informed receiver*. This informed receiver also knows whether the dictator has manipulated the presentation of the available options to the uninformed receivers.[24] We chose this ratio of dictator to informed or uninformed receivers because we want to model situations in which a decision maker faces a reasonable chance not to be detected with a lie. The punishment proceeds as described above. The computer randomly selects one of the three receivers, and implements her punishment decision for the option that has been chosen by the dictator. The costs of that punishment decision are shared among the receivers, and all other decisions become irrelevant.

4.2.4 Game order, group composition, and determination of payoffs

After all of the participants played Game I, the sequence of the remaining two games (Game II and Game III) is randomized. In the case of multiple decisions within Game I, we randomize the sequence of the choice sets presented. In all games, the presentation of the two options in a choice set is randomized (i.e., which one is left or right). Furthermore, there is no feedback between the games about the actual implemented option (remember, receivers punish as if either of the two options is implemented), to guarantee the independence of the choices of confounding wealth effects or income targets. This means that punishment has no impact on the dictator's choice in the respective next game.

[24] In case of manipulation, the informed receiver reads the following message: "Player A decided to replace the alternative, which has not been chosen, with another one." Otherwise, she reads: "Player A decided against replacing the alternative, which has not been chosen."

Additionally, the group composition changes between the games, using a perfect stranger-matching procedure. At the end of the game, subjects receive the information about all decisions in the three games. A throw of a die determines the game that became relevant for the payoff. Another throw then determines the specific relevant decisions in that game.

4.2.5 Procedure

To collect the experimental data, we conducted nine sessions in June and July 2015 with a total of 248 students. They were recruited via ORSEE (Greiner 2015), and invited to the Lakelab, the experimental laboratory of the University of Konstanz. z-Tree software (Fischbacher 2007) was used to program the experiment. Subjects earned €17, on average, across sessions, with a maximum of €48 and a minimum of €3 (the show-up fee). At the beginning of the session, every subject was assigned randomly to one of the PC terminals, where they were told to read the instructions carefully. Before each session started, the subjects answered control questions, to ensure that they understood every feature of the experiment properly. Sessions lasted an average of about 50 minutes. At the end, each subject completed a questionnaire that requested some socioeconomic data, and provided the opportunity to comment on the experiment. Finally, subjects were called one by one to exit the laboratory. Using this method, we made sure that the payoff was known only to the corresponding subject.

4.3 Behavioral Predictions

The theoretical literature provides different predictions in the context of this experiment, which we present as competing hypotheses. Note that the experiment consists of three separate dictator games. Because of the deferred feedback about the decisions and the random reassignment of groups and decision sequences, we can rule out reputation and reciprocity concerns across the games. Therefore, we can study them as independent decisions.

4.3.1 Model of rationality

Assuming common knowledge about rational and purely selfish agents leads to the following benchmark predictions.

Hypothesis 1 (Pure selfishness)

i) Dictators will choose the option that maximizes their own payoff.

ii) As retribution is costly, receivers will not punish dictators for any decision in Game II and Game III.

iii) In the absence of punishment, dictators are indifferent in their decision to disguise their true choice sets in Game III.

4.3.2 Outcome-oriented social preferences

Theoretical contributions as those by Fehr and Schmidt (1999) or Bolton and Ockenfels (2000) suggest that people dislike unequal outcomes, in particular, if the individuals are worse off than the person with whom they compare themselves. The experimental design ensures that a receiver compares her outcome with the dictator's payoff, because all receivers get the same payment under any condition. Such outcome-oriented preferences have two direct implications. First, receivers have a motive to punish any unfair distribution, because punishment reduces the gap in payoffs between the receivers and the dictator. Note that the pure focus on closing the income gap implies that punishment of a certain option does not depend on the parameters of the alternative option that is available to the dictator. Second, dictators may refrain from making choices that maximize their own payoffs. In Game I, dictators do so if they dislike being better off than the receivers. In Games II and III, dictators additionally consider the receiver's potential punishment.

Hypothesis 2 (Inequality aversion)

i) Dictators will choose the unfair options less often in Games II and III than in Game I.

ii) Receivers will punish dictators in Games II and III for choosing Option 3 or 4.

iii) The punishment for choosing Option 4 exceeds the punishment for choosing Option 3.

iv) Receivers focus only on the outcome; thus, for the dictator's given choices, the punishment does not differ between Games II and III.

4.3.3 Intention-based models of reciprocity

Another set of theories (Dufwenberg and Kirchsteiger 2004, Falk and Fischbacher 2006) argues that people also assess the relative kindness of a person when they consider punishing that person. A critical factor in this assessment is the question whether an alternative feasible outcome would have been more desirable, or could have been even worse. In this context, similar to the outcome-oriented theories mentioned above, the relationship between the payoffs still matters. Falk and Fischbacher (2006, p. 297) note that "it is not reasonable to demand that the other person is fair to me if this implies that (relative to me) she puts herself in a disadvantageous position." The implications of these kindness considerations are fairly obvious in Game II, when dictators cannot disguise their intentions.

Hypothesis 3a (Punishment by receivers in Game II)

i) Receivers in Game II punish the choice of Option 3 more severely if the dictator's alternative choice was the egalitarian Option 2, rather than Option 4 or the altruistic Option 1.

ii) Receivers in Game II punish the choice of Option 4 more severely if the dictators' alternative choice was the more egalitarian Option 2 or Option 3.

iii) Punishment for the very unfair option is minimized when Option 1 was the alternative.

The same predictions also hold for the *informed receivers* in Game III. However, the informed receivers also learn whether the dictator has disguised her decisions for the uninformed receivers. In the case of an attempt to disguise, Brandts and Charness (2003) suggest increased retaliation, as the disguise attempt reveals bad intentions.

Hypothesis 3b (Reciprocity by informed receivers in Game III)

i) Informed receivers in Game III punish any choice more strongly if the dictator has disguised her true decision alternatives.

Punishment decisions by the *uninformed receivers* in Game III might depend on their beliefs about whether the dictator has disguised her intentions. This requires considerations about dictator's specific disguise strategy. In this context, it is reasonable to assume that the dictator wants to frame her actual choice as the relatively kind one. For example, if she has chosen the unfair Option 3, she will display the particularly unfair Option 4 alongside. If she has chosen Option 4, she will display the altruistic Option 1 alongside, because the choice of 0 payoff appears to be unreasonable, as the quotation by Falk and Fischbacher (2006) suggests. Thus, the disguising dictator will not display the egalitarian Option 2 if it has not been chosen. Table 4.2 shows, for any potential choice by the dictator, the option that minimizes the unkindness of the chosen option. The dictator will always display Option 4, either because she has selected it or because any alternative choice appears as the kind one in comparison.

Table 4.2: Actual choices and alternatives that minimize perceived unkindness

Actual Choice	Alternative that minimizes perceived unkindness	
Option 1 (0, 8, 8, 8)	Option 4 (18, 2, 2, 2)	Disguising unnecessary because choices are unlikely to be punished
Option 2 (6, 6, 6, 6)	Option 4 (18, 2, 2, 2)	
Option 3 (12, 4, 4, 4)	Option 4 (18, 2, 2, 2)	Option 4 makes receivers even worse off
Option 4 (18, 2, 2, 2)	Option 1 (0, 8, 8, 8)	Option 1 would eliminate the dictator's income

Hypotheses about differences in the punishment behavior of uninformed receivers between Game II and Game III is unclear. In the presence of the possibility to disguise, they might assume bad intentions when presented with a suspicious choice set (i.e., containing Option 4 or Option 1). However, we assume that uninformed receivers act according to the principle *in dubio pro reo* (doubts benefits the accused). This means that bad intentions are not presumed, and thus, dictators benefit from disguising their true choice set.

Hypothesis 3c (Punishment by uninformed receivers in Game III)

i) For a given choice set, uninformed receivers do not increase the punishment in Game III relative to Game II.

ii) Dictators benefit from disguising their true choice set.

4.4 Results

We have shown that different theoretical perspectives lead to competing hypotheses. In this section, we describe the results, and provide the relevant statistical tests. If not stated otherwise, the reported p values derive from one-sided Mann-Whitney U tests, because most of the hypotheses are directional. Then, we evaluate the competing hypotheses in light of the results, and discuss the findings.

4.4.1 Dictator behavior

First, we focus on the dictator's decisions. Table 4.3 shows the relevant decisions in the three games. Recall that in Game I we let all players decide as dictators, using the strategy method. In that simple dictator game, we observe that about half of the players choose the more fair option if it implied a positive payoff for themselves. They opt overwhelmingly for selfish options only if the altruistic Option 1 is the alternative (columns 1/3 and 1/4).

TABLE 4.3: Dictator behavior in Game I through III

Available Options	1/3		1/4		2/3		2/4		3/4	
Choice	1	3	1	4	2	3	2	4	3	4
Game I:	6.5%	93.5%	16.5%	83.5%	50.4%	49.6%	50.8%	49.2%	53.6%	46.4%
N	248		248		248		248		248	
Game II:	0%	100%	9.1%	90.9%	41.2%	58.8%	33.3%	66.7%	58.3%	41.7%
N	13		11		17		9		12	
Game III:	0%	100%	0%	100%	33.3%	66.7%	61.5%	38.5%	46.2%	53.8%
N	17		10		9		13		13	

Interestingly, the implementation of the punishment mechanism in these games does not lead to more pro-social choices. Table 4.4 shows how dictators modify their decision between Game I and the subsequent games. The decisions are remarkably stable, and do not become kinder.[25]

Table 4.4: Changes in dictator behavior relative to Game I

	Game II	Game III
Same Choice (kind)	12	12
Same Choice (unkind)	35	42
More kind	6	3
Less kind	9	5
$\sum$	62	62

We now focus on how dictators disguise their intentions in Game III. Figure 4.1 shows which option dictators display alongside the allocation actually chosen.

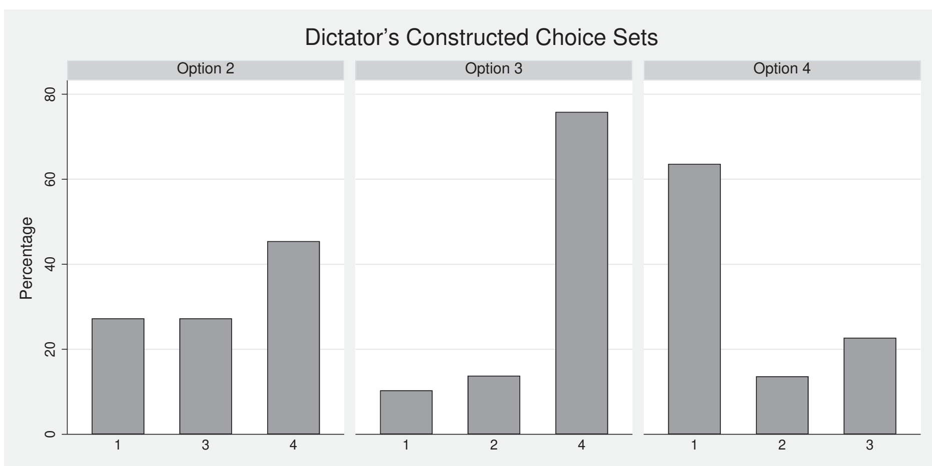


Figure 4.1: Alternative chosen to be displayed alongside the implemented option.

Most dictators choose the option that minimizes their perceived unkindness. In particular, they display Option 4 alongside Option 3, while displaying Option 1 alongside Option 4. Generally, dictators are ready to display the actual alternative with the option that casts the best light on their own distribution decision. This result is in line with that of

[25] Based on Table 4.4, we can calculate the number of kind choices in Game I through III. A comparison of proportions (chi-square) yields no statistically significant difference ($p = 0.5663$ and $p = 0.2306$) between Game I and either of the latter games.

Schächtele et al. (2011), where most players used the possibility to deceive their recipients. Only about 31% of dictators display their actual choice set, mostly if it already puts their choice in the kindest context. Thus, dictators consider the perception of intentions in their decision. Result 1 summarizes the dictator decisions.

Result 1 (Dictator behavior)

i) About half of the dictators choose the more fair option if that allows for positive payoffs. The other subjects maximize their own payoff.

ii) The threat of punishment does not substantially increase fairness.

iii) Most dictators disguise their intentions; that is, dictators present their choice in the most favorable choice set.

4.4.2 Outcome-oriented punishment

In Figure 4.2, the behavior of receivers (i.e., punishment decisions) is compared across Game II and Game III. In Figure 4.2, the punishment for one option is aggregated across the different available alternatives.

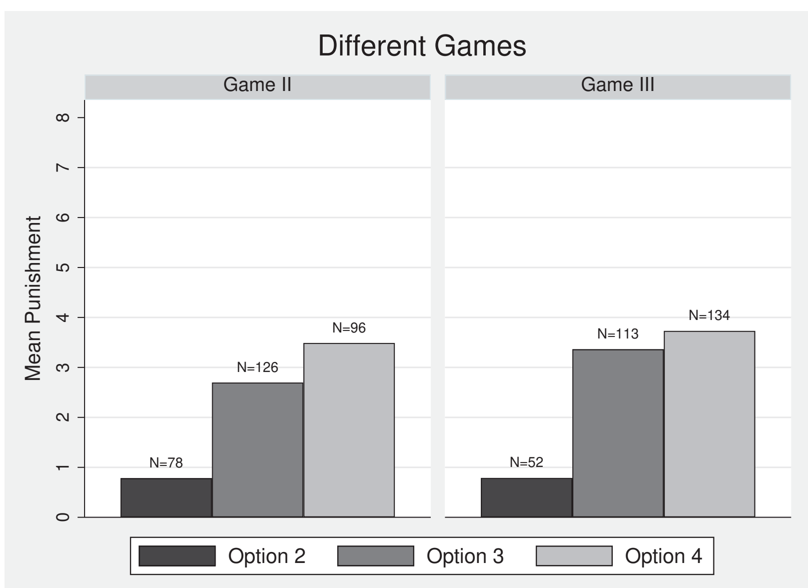


FIGURE 4.2: Punishment (in points) aggregated over each possible option across Game II and Game III.

The punishment generally increases in the inequality of an allocation. The increase from Option 3 to Option 4 is statistically significant in Game II (one-sided t-test, $p = 0.0818$), whereas in Game III the increase is not (one-sided t-test, $p = 0.2549$). The increase from Option 2 to any other option in both games is highly statistically significant.[26]

4.4.3 Intention-based punishment

Now we focus on the role of (perceived) intentions. First, we look at Game II, in which the dictators cannot disguise their intentions. Figure 4.3 provides the punishment decisions for an option in the context of the displayed alternative.

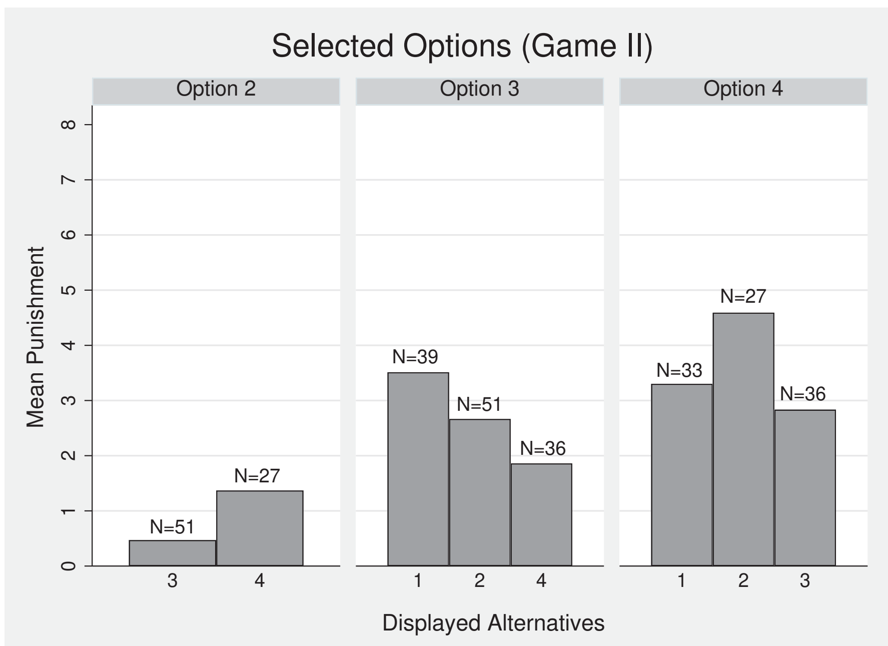


FIGURE 4.3: Punishment (in points) in Game II for each option separated by alternatives.

The data shows that the displayed alternative has some impact on the punishment of a specific choice. Take the punishment for Option 3, which is the second cell in Figure 4.3. Punishment for an unfair dictator whose alternative is the egalitarian allocation (Option 2) is slightly larger than for dictators who face the very unfair option as the alternative ($p = 0.097$). Interestingly, having only the altruistic option as the alternative does not reduce the punishment. The punishment for dictators who implement the very unfair

[26]Note that we find some punishment for the fair Option 2. This punishment might be explained by a preference for spite, as investigated by Levine (1998). However, the actual impact of that punishment is negligible.

Option 4 is significantly higher if receivers see the fair option rather than the unfair option or the altruistic option ($p = 0.066$ and $p = 0.098$, respectively). Despite the lower punishment with the altruistic alternative, compared to the egalitarian alternative, the punishment is not significantly reduced, compared to Alternative 3. This result suggests that the altruistic option is not a reasonable candidate for disguise. Result 2 summarizes the insights.

Result 2 (Receiver behavior on intention-based decisions)

i) Punishment is based not only on the outcome but also on the available alternative. Intentions matter.

ii) Receivers punish the implementation of the unfair option less if the very unfair option is the alternative.

iii) Punishment for the very unfair option is highest if dictators forgo the possibility to make the egalitarian choice.

iv) In the absence of the possibility to disguise, the altruistic option does not deter punishment of unfair options.

4.4.4 Intention-based punishment with disguise

Now we study Game III, the dictator game with punishment, and the possibility to disguise intentions. Figure 4.4 presents the punishment decisions of uninformed receivers. We differentiate the choice of both *unfair* options with every feasible alternative. Recall that dictators have constructed the choice sets displayed to the (uninformed) receivers.

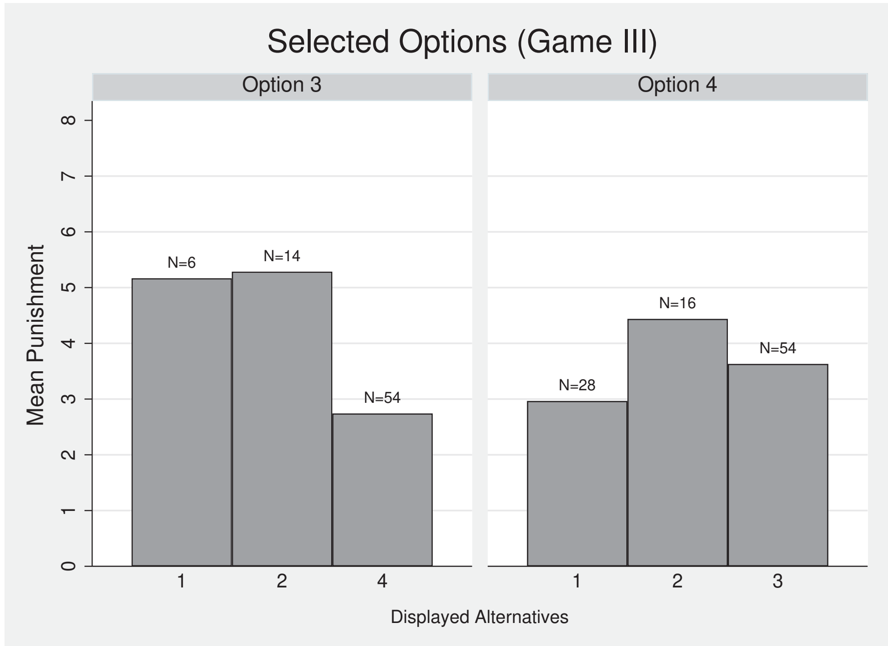


Figure 4.4: Punishment (in points) in Game III of uninformed receivers for unfair options separated by alternatives.

The results show that uninformed receivers respond to the displayed alternatives. They reduce their punishment if an unfair choice is placed in a relatively kind context. This result holds, in particular, for the choice of the unfair Option 3, which triggers much less punishment if displayed alongside the very unfair option rather than the egalitarian option ($p = 0.033$). After the very unfair option is chosen, the mean punishment does not differ significantly by varying the alternatives. Thus, once dictators opt for the utmost unfavorable choice, they cannot minimize the expected punishment by presenting an alternative which is very unfavorable to themselves. Notably, the altruistic option, once again, fails to be a suitable candidate for disguise.

Now we study the punishment behavior of receivers who learned that the dictator had altered the presentation of the choice set. These informed receivers see the dictator's original choice set together with a message that the dictator had exchanged one option for another. In Figure 4.5, we compare the punishment in that context to that in Game II, where disguise was impossible.

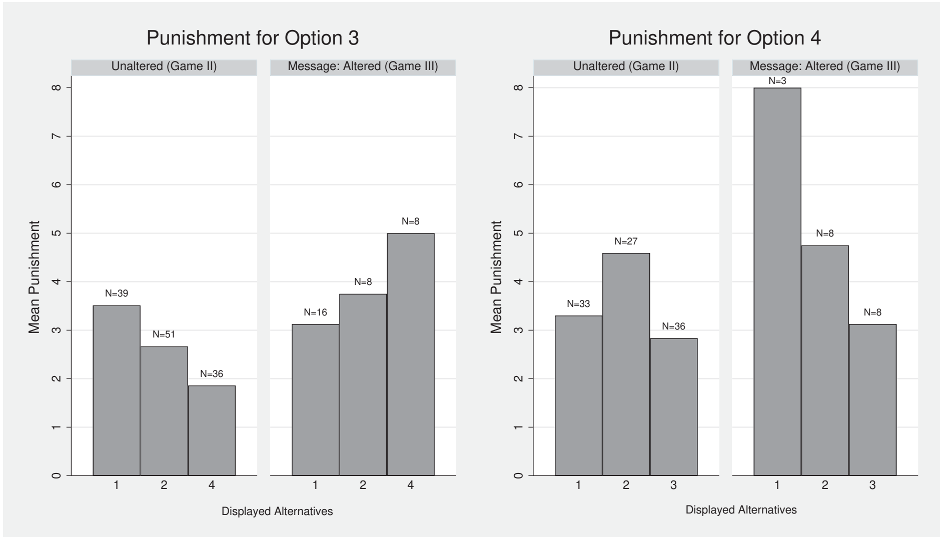


FIGURE 4.5: Punishment (in points) for unfair options separated by alternatives for unaltered choice sets (represented by Game II punishment decisions, equal to those presented in Figure 4.3) and choice sets with an "altered" message. The punishment for the egalitarian option shows no difference, and therefore, is omitted.

Due to the low number of observations, caused by the endogeneity of the design, we do not find clear statistical evidence that a dictator's attempted disguise systematically increases punishment. However, the punishment for the unfair Option 3, which has Option 4 as the alternative, is higher once the receiver is informed about the disguise ($p = 0.096$). Nevertheless, the threat of the expected punishment after being caught window dressing is small, and cannot outweigh the expected gain. The anticipation of this lack of sanctions could be one possible explanation for the vast majority of dictators altering the choice set. To summarize the findings, Table 4.5 shows the results for a regression in which we combine the punishment decisions of all receivers across Games II and III. To compare the punishment decisions of receivers among different possible options, the outcome variable is defined as the "share of punishment". This variable normalizes the actual punishment to the respective possible punishment. As the reference choice set, we chose the extreme case of forgoing the highest personal benefit, and choosing the egalitarian option (Choice 2, Alternative 4). The player type is derived from the choices of the dictator game without punishment (equality concerned serves as the reference group) and as the information set, Game II serves as the reference. We also control for an order effect of Game II and Game III.

Table 4.5: Ordinary least squares (OLS) regression of the share of punishment dependent on the choice set

	Share of punishment
Choice 3/Alternative 1	0.132^{*}
	(0.0721)
Choice 3/Alternative 2	0.110^{*}
	(0.0560)
Choice 3/Alternative 4	0.0292
	(0.0540)
Choice 4/Alternative 1	0.110^{*}
	(0.0625)
Choice 4/Alternative 2	0.186^{***}
	(0.0575)
Choice 4/Alternative 3	0.102^{*}
	(0.0571)
Game III/uninformed	0.0470
	(0.0285)
Game III/informed	0.0361
	(0.0375)
Undefined	0.0206
	(0.0546)
Homo oeconomicus	-0.0848^{*}
	(0.0474)
Game II first	0.0156
	(0.0435)
Constant	0.166^{***}
	(0.0616)
Observations	605
R^2	0.071

Notes: Choice 2, Alternative 4 serves as the reference choice set (fair choices are included in the regression but omitted in the table). Game II serves as the baseline for the information set, where every receiver is informed. The player type is derived by Game I (reference group: equality concerned). Robust standard errors, clustered on the individual level, in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We find clear evidence for intention-based punishment. In particular, the very unfair option is punished statistically significantly more if the alternative was the fair option. We find little effect of the information set of the receivers.[27] This means that the attempt to disguise, either detectable or undetectable, does not significantly influence the punishment decisions of the receivers. Thus, dictators benefit from disguising their

[27] The belief of an uninformed receiver about whether the choice set was altered or not does not influence the punishment decision.

intentions. This result also holds when we aggregate all punishment decisions in Game III (informed and uninformed), and analyze the dictators' expected payoff.

4.4.5 Dictators' payoff

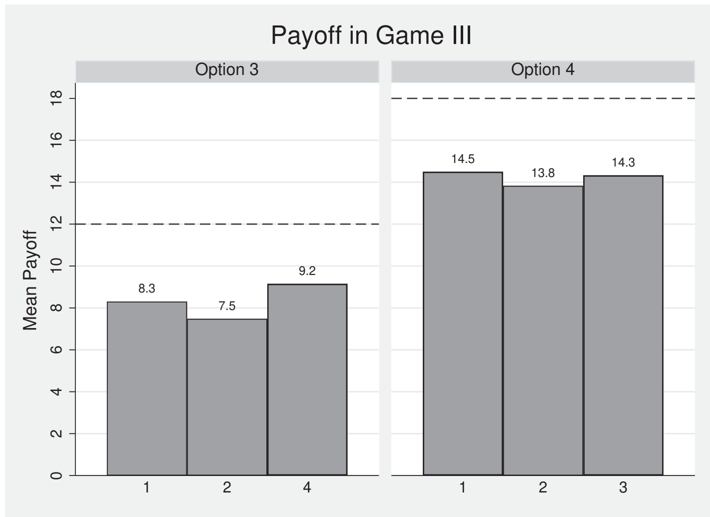


FIGURE 4.6: Payoff (in points) of dictators in Game III. Dashed reference lines indicate the maximum payoff of the respective option.

Figure 4.6 shows that after the punishment the payoff is maximized by choosing Option 4. Although this is strongly dependent on the punishment mechanism and parameters, the dictator especially benefits from disguising her intentions in the case when they opt for the unfair Option 3. By putting that option in light of the more unfavorable Option 4, the payoff increases significantly (one-sided t-test, $p = 0.0451$). We can conclude the results as follows:

Result 3 (Does it pay to play spin doctor?)

i) There is no clear statistical evidence that an observed attempt to disguise unkind intentions leads to higher punishment.

ii) Dictators can disguise unkind intentions.

iii) Punishment is minimized by disguising unkind intentions; thus, the payoff is maximized in expectation.

4.5 Conclusion

We used variations of a mini-dictator game to study how people try to disguise bad intentions and the punishment of such behavior. We presented three competing sets of hypotheses. The standard *Homo oeconomicus* predictions in Hypothesis 1 do not capture the observed behavior well. Dictators are sometimes fair, and receivers punish. In line with purely outcome-oriented predictions (Hypothesis 2), punishment increases with the inequality of the outcome. However, we also see some differentiation in punishment for a given monetary outcome. To some extent, we can explain these differences with intention-based models of reciprocity (as predicted in Hypotheses 3a and 3b), as the punishment tends to decrease with the unfairness of the available (Game II) or the displayed alternative (Game III). Crucially, dictators are quite sensitive to the role of intentions, and typically, present their unfair decisions in the kindest context. This is interesting, as the threat of punishment per se does not induce them to change their behavior very much (see the comparison between Game I and the other two games). However, some results, for example, the punishment behavior of informed receivers in Game III, are hard to reconcile with intention-based reciprocity models, and require more attention in future studies. Future studies with larger samples should focus on discerning inter-individual differences in punishment motives that might help address this concern. In addition, because dictators overwhelmingly opt to disguise intentions, we advice abstaining from an endogenous choice design, as presented here, as certain (interesting) corner cases are only rarely observed. The difference between the sensitive dictators and the comparatively blunt punishment decisions of the recipients provides some insights for our understanding of communication processes, for example, in business or politics. It shows a certain information asymmetry about intentions, because disguising dictators know their motives better than the recipients. The manager knows that she fires people to increase her own payoff, while the workers may attribute their firing to an exogenous shock. Differences in cost considerations between these different agents also play a role in this context. It is easy for a manager to spin a story about the reasons behind job cuts, while the workers find it difficult to punish her for this action. However, the observation also helps explain why some populists do not lose touch with voters, even when they are obviously dishonest about their intentions. The punishment of unfair and disguising politicians is a subtle process, whose effects are drowned in statistical noise. Many subjects may not consider such behavior to be punishable. Politicians may overinvest in spin doctoring, if they are not aware of this heterogeneity, and have lost touch with their voters.

References

Abeler, J., Nosenzo, D., and Raymond, C. (2019). Preferences for truth-telling. *Econometrica (forthcoming).*

Alempaki, D., Doğan, G., and Saccardo, S. (2016). Deception and reciprocity. *Experimental Economics*, pages 1–22.

Bolton, G. E. and Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, 90(1):166–193.

Brandts, J. and Charness, G. (2003). Truth or consequences: An experiment. *Management Science*, 49(1):116–130.

Charness, G. and Levine, D. I. (2003). The road to hell: An experimental study of intentions. *Mimeo*.

Charness, G. and Levine, D. I. (2007). Intention and stochastic outcomes: An experimental study. *The Economic Journal*, 117(522):1051–1072.

Cox, J. C., Friedman, D., and Gjerstad, S. (2007). A tractable model of reciprocity and fairness. *Games and Economic Behavior*, 59(1):17–45.

Cushman, F., Dreber, A., Wang, Y., and Costa, J. (2009). Accidental outcomes guide punishment in a "trembling hand" game. *PloS one*, 4(8):e6699.

Dufwenberg, M. and Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior*, 47(2):268–298.

Falk, A., Fehr, E., and Fischbacher, U. (2008). Testing theories of fairness—Intentions matter. *Games and Economic Behavior*, 62(1):287–303.

Falk, A. and Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, 54(2):293–315.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.

Friehe, T. and Utikal, V. (2018). Intentions under cover–Hiding intentions is considered unfair. *Journal of Behavioral and Experimental Economics*, 73:11–21.

Gerlach, P., Teodorescu, K., and Hertwig, R. (2019). The truth about lies: A meta-analysis on dishonest behavior. *Psychological Bulletin*, 145(1):1–44.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1):114–125.

Jacobsen, C., Fosgaard, T. R., and Pascual-Ezama, D. (2018). Why do we lie? A practical guide to the dishonesty literature. *Journal of Economic Surveys*, 32(2):357–387.

Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics*, 1(3):593–622.

Lundquist, T., Ellingsen, T., Gribbe, E., and Johannesson, M. (2009). The aversion to lying. *Journal of Economic Behavior & Organization*, 70(1-2):81–92.

Orhun, A. Y. (2018). Perceived motives and reciprocity. *Games and Economic Behavior*, 109:436–451.

Rand, D. G., Fudenberg, D., and Dreber, A. (2015). It's the thought that counts: The role of intentions in noisy repeated games. *Journal of Economic Behavior & Organization*, 116:481–499.

Rosenbaum, S. M., Billinger, S., and Stieglitz, N. (2014). Let's be honest: A review of experimental evidence of honesty and truth-telling. *Journal of Economic Psychology*, 45:181–196.

Sánchez-Pagés, S. and Vorsatz, M. (2007). An experimental study of truth-telling in a sender–receiver game. *Games and Economic Behavior*, 61(1):86–112.

Sánchez-Pagés, S. and Vorsatz, M. (2009). Enjoy the silence: An experiment on truth-telling. *Experimental Economics*, 12(2):220–241.

Schächtele, S., Gerstenberg, T., and Lagnado, D. (2011). Beyond outcomes: The influence of intentions and deception. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 33.

Toussaert, S. (2017). Intention-based reciprocity and signaling of intentions. *Journal of Economic Behavior & Organization*, 137:132–144.

Appendix A: Instructions

We welcome you to this economic experiment.[28]

In this experiment your decisions as well as the decisions of other participants influence your payoff. Therefore, it is very important that you read through these instructions carefully. Whenever you have questions, please contact us **before** the experiment starts.

During the experiment it is forbidden to communicate with other participants.

If you break this rule we have to exclude you from the experiment and your payoff. During the experiment we talk about points rather than Euros. Your payoff will first be computed in points. The points you gained during the experiment are translated into Euros at the end of the experiment at:

$$1\ Point = 2.5\ Euro$$

In addition, you receive 3 Euro for you showing up on time. Your payoff will be handed in cash to you after the experiment. The following pages give you the instructions for the experiment.

Procedure

The experiment is divided into three parts. In every part you play in a group of four other participants. After every part the groups are reallocated. Before we explain the differences between the parts we will first stress the commonalities of the different parts.

Commonalities of the three parts

In every group there is a Player A and three Players B. Player A can distribute 24 points between him and the other players. He will have the choice between **two** of the following four options. Which options Player A will face is decided randomly by the computer.

[28] Translated from German.

Note that the computer will never give a Player A the choice between Option 1 and Option 2.

Points for	Option 1	Option 2	Option 3	Option 4
A	0	6	12	18
B1	8	6	4	2
B2	8	6	4	2
B3	8	6	4	2

Now we will explain to you in more detail the differences between the three parts of the experiment. **Note that the order of Part II and III can be exchanged during the experiment.**

Part I

In Part I no participant knows whether he is Player A or Player B. All participants decide for the case where they actually are Player A. You will have the choice between two of the four options presented above. You can then always choose one to be implemented for the group. After you decided for one pair you get another one etc. Therefore, every player will make five decisions in this part.
In the case where Part I will be the one which is paid, you will be informed **after** the experiment whether you have been chosen to be Player A. In that case **the computer** decides randomly which of the five decisions will be paid. In that case Player B do not have an influence on the payoff.

Part II

At the beginning of Part II you will be informed whether you are Player A or Player B. Player A will get the choice between two options **once**. Which options he can choose from is randomly determined by the computer. Note once again that there will never be the choice between Option 1 and Option 2.
The Players B receive knowledge about which options were presented to Player A and then decide independently from each other how many points to deduct from Player A for each choice he can make.

Part III

In the case where Part III comes after Part II you keep your role. In the other case you will be informed whether you are Player A or Player B before the part starts.
Player A again has the choice between two of the options presented above. Afterwards he *can* decide to replace the option which he did not take by any other option **before** Players B see the options.
In the case where Player A decided to replace the not chosen option, two of the three Players B will see the choice of Player A together with the option he decided to present with his choice. The two Players B do not know which option Player A decided to implement for the group. The third Player B receives the information about the initial pair of options and a short information text whether Player A decided to replace the not chosen option or not. He also does not know which option Player A decided to implement for the group. All Players B decide independently if and how many points they want to deduct from Player A if he chose to implement the one or the other option for the group.

Example Part III

Player A had the choice between Option 3 and 4 where he chose Option 3. Then he decides to replace Option 4 with Option 1. Two Players B now see Option 1 and Option 3 as potential options of Player A, whereas the third Player B sees that Player A had the choice between Option 3 and 4. This Player B also learns that Player A decided to replace Option 4.

Deducting points

In Part II and III Players B can decide if and how much points they want to deduct from Player A. Only one of these decisions will be implemented at the end of the experiment. Which one that is, is defined by a dice. Once one decision is implemented, **all** Players B join in on the costs for the deduction. In order to deduct 1 point from Player A the group of Players B has to pay 0.5 points. But because all Players B bear the cost for the deduction at the same rate, in order to deduct 1 point from A this costs one Player B $\frac{1}{6}$ points effectively. Note that you can deduct maximally the amount Player A receives through the option or you can afford to pay by implementing that option.

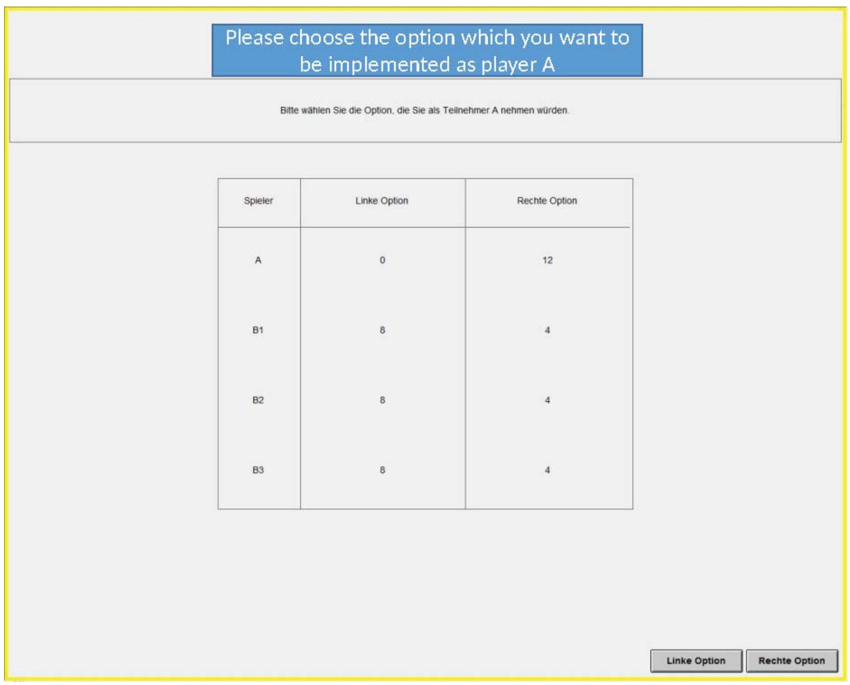
Example: Player A implements the option for the group where every player receives 6 points. One player decides to deduct 6 points from Player A. The group of Players B has to give up $0.5 * 6 = 3$ points. Since every Player B bears the same amount of costs, every Player B has to give up 1 point. This results in a payoff of 0 points for Player A and 5 points for every Player B.

Ordering of the different parts

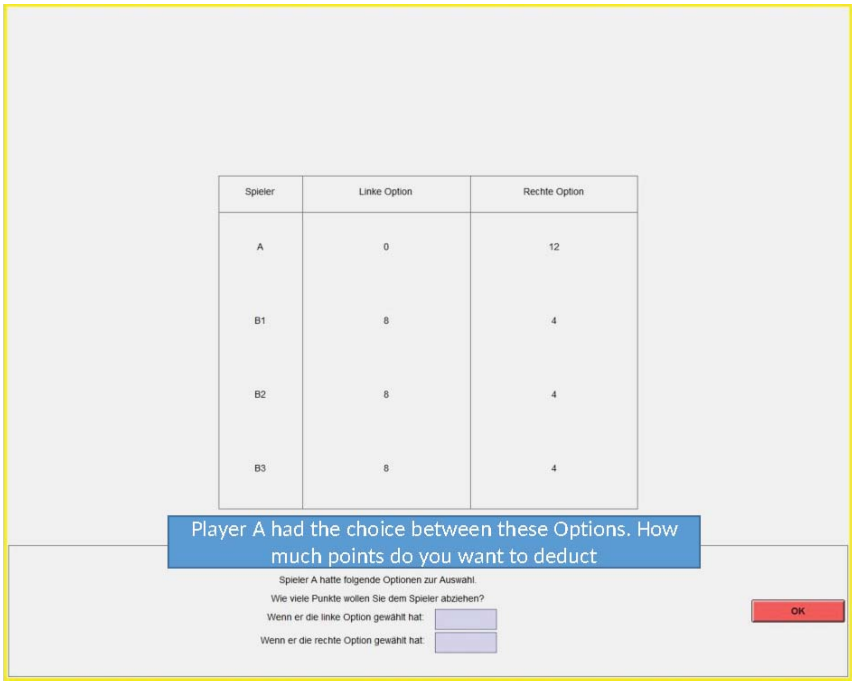
The following ordering of the parts can happen during the experiment:
Case 1: Part I → Part II → Part III
Case 2: Part I → Part III → Part II

Procedure on the PC-Screen

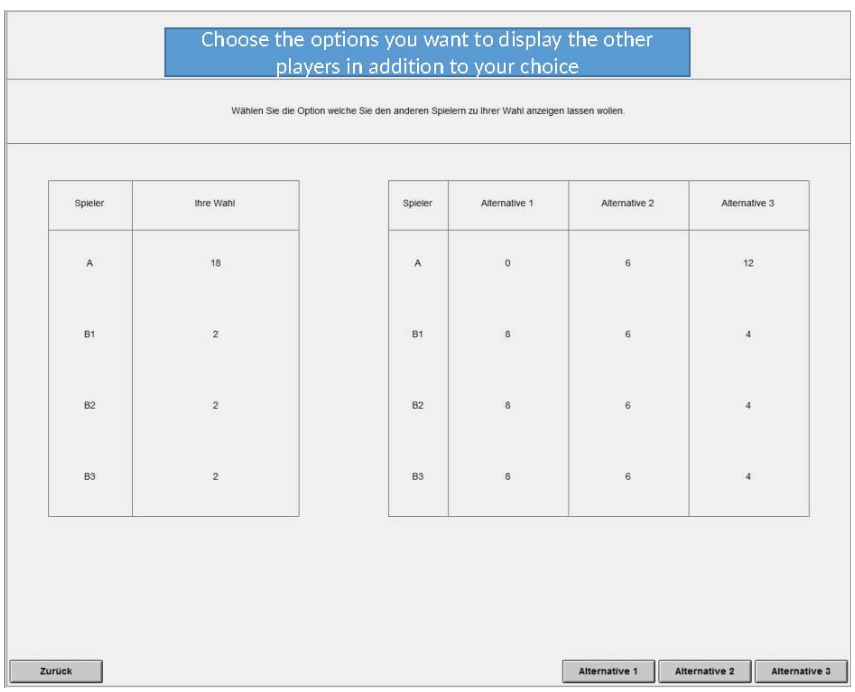
Part I: All players decide as Player A and choose one of the two options presented on the screen. This decision is repeated for every pair with a total of five decisions.



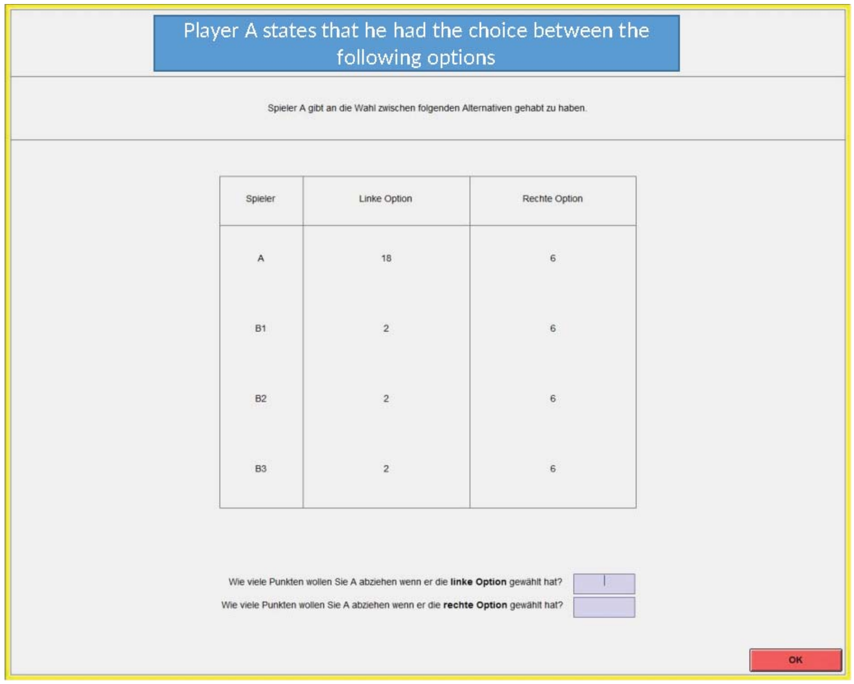
Part II (can also come after Part III): Player A chooses one of the presented options. Players B can decide independently from each other if and how much points to deduct from Player A if he chooses the one or the other.



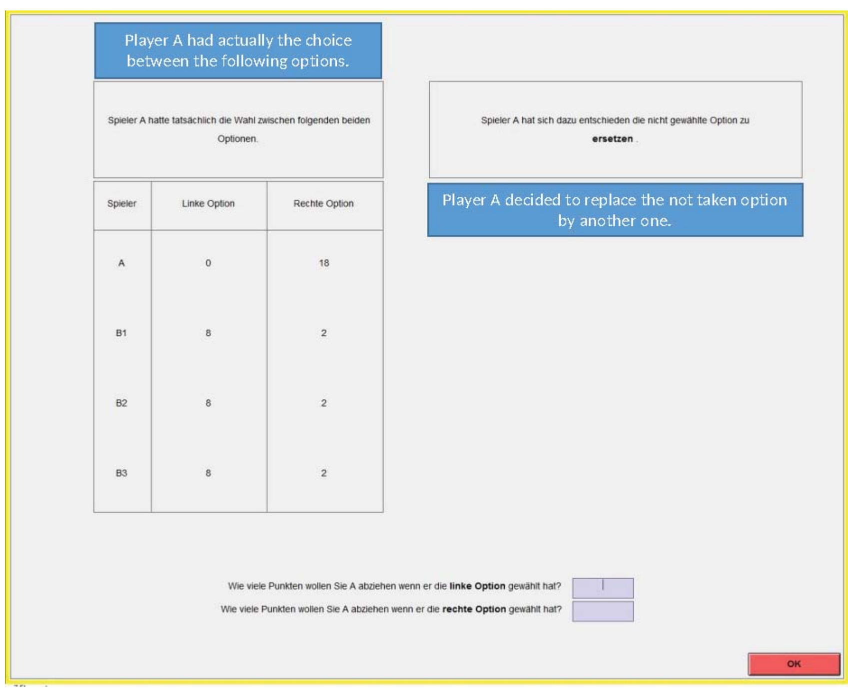
Part III (can also come before Part II): Player A chooses one of the presented options. Afterwards he decides which other option he wants to present to two other Players B on the screen.



Two Players B decide independently if and how much points to deduct from Player A.

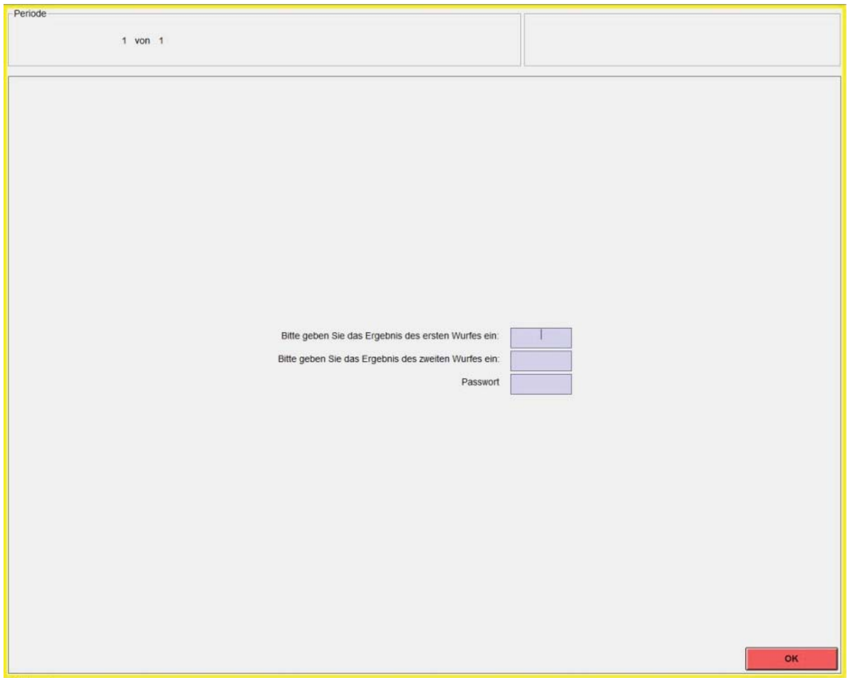


The third Player B decides accordingly to the initial two options if and how much points to deduct.



Determination of the Payoff: At the end of the experiment the participant at the computer assigned with the number 1 roles two dice. The first throw decides which game is chosen to be paid. A throw of 1 or 2 leads to Part I, 3 or 4 to Part II and 5 or 6 to Part

III. A further throw will help the computer to randomly decide which point deduction decision in that certain part will be implemented. Afterwards you will be informed about your payoff.



End of Instructions

In addition to the instructions, we provided a table, on a separate piece of paper, for a better understanding of how the punishment works.

Deduction Points for Player A	0	1	2	3	4	5	6	7	8	9	10	11	12
Costs for Players B in total	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6
Costs for one Player B	0	0.17	0.33	0.5	0.67	0.83	1	1.17	1.33	1.5	1.67	1.83	2

Essay 4: oTree: Implementing websockets to allow for real-time interactions: A continuous double auction market as first application

Ann-Kathrin Crede, Jan Dietrich, Jonas Gehrlein, Oliver Neumann, Matthias Stürmer, Frauke von Bieberstein*

Abstract

This article illustrates the implementation of websockets in oTree (Chen et al. 2016) to allow for real-time interactions. Whereas oTree generally allows to overcome the need for participants to be at the same location to interact with each other, a real-time module in the sense that the user interface responds within milliseconds to actions from other participants is currently not available. We address this gap and further develop oTree by making real-time interactions between a large number of players with immediate updates possible. As a first application, we run a continuous double auction market on Amazon Mechanical Turk to validate its functionality. This ready-to-use software is of special interest for the research of large (online) markets and for teaching purposes. We provide the code open-source on GitHub, thereby encouraging users to develop it further and applying the websocket technology to other real-time settings.

Keywords: experimental economics, oTree, double auction market, websockets, open-source

*Ann-Kathrin Crede: ann-kathrin.crede@iop.unibe.ch, Institute for Organization and Human Resource Management, University of Bern; Jan Dietrich: jan.dietrich@tecpose.com, Tecpose; Jonas Gehrlein: jonas.gehrlein@iop.unibe.ch, Institute for Organization and Human Resource Management, University of Bern; Oliver Neumann: oliver.neumann@inf.unibe.ch, Research Center for Digital Sustainability, University of Bern; Matthias Stürmer: matthias.stuermer@inf.unibe.ch, Research Center for Digital Sustainability, University of Bern; Frauke von Bieberstein: frauke.vonbieberstein@iop.unibe.ch, Institute for Organization and Human Resource Management, University of Bern. We thank Christoph Zangger, Roden Safar, Annika Wyss, and Nicolas Hafner for excellent research assistance. We are grateful to Ilija Braun and Andrea Essl for very valuable comments. All remaining errors are ours.

5.1 Introduction

The use of oTree as a software to run experiments has a number of advantages that have been documented and tested since its introduction (e.g., Chen et al. 2016, Holzmeister and Pfurtscheller 2016, Holzmeister 2017). These advantages are especially visible in comparison to the widely used z-Tree (Fischbacher 2007). The main benefit is that only a web browser is necessary, thereby allowing that participants need not be at the same location to interact with each other. Especially for large-scale experiments where interaction between many subjects is necessary, oTree offers clear advantages. The native integration with Amazon Mechanical Turk (AMT) provides an additional beneficial feature to run experiments online with non-standard subject pools. While oTree obviously offers new possibilities for running experiments, one important feature is not available so far: It is missing a real-time component in the sense that the user interface does not update within milliseconds to actions from other participants (Chen et al. 2016, p. 96).

Our novel implementation relies on websockets[29] allowing that the input of any kind of information of one participant is immediately transmitted to other participants without the need to refresh the current page. This automatic update allows participants to interact rapidly and in real-time. With respect to our application to a double auction (DA) market, the real-time component creates a realistic and authentic environment of a marketplace, where trading occurs fast and requires quick decision-making. In this paper, we evaluate data gathered by a related project (Adrian et al. 2019)[30] to test the functionality of the DA market. We find that the module proved to work smoothly (see section 5.4).

We chose a DA market as first application to meet economists' ongoing interest in understanding and investigating markets and focus on market clearance through pure competition (for other mechanisms see Plott and Sunder 1988). In the 1960s, Vernon Smith was the first to run an oral DA market experiment with his students. In the basic version, participants are randomly assigned to a group of buyers or a group of sellers.

[29]We use the term "websockets" to describe the technology in general, although more precisely the term "WebSockets" describes the protocol. The implementation of the websocket technology in oTree is done by Django channels (https://github.com/andrewgodwin), see section 5.2 for details.

[30]The study by Adrian et al. (2019) investigates the influence of markets on moral decisions. In their experiment, the data collected on the market is not analyzed, as it is not part of their research question. The DA only serves as stimulus for a subsequent moral decision.

Each buyer privately learns the maximum price he is willing to pay for one unit of the good. Each seller privately learns the minimum price at which he is willing to give away the good. Each buyer and each seller can trade once per round over several rounds in total. Despite his intention to reject competitive market theory, Smith found that the realized price under the DA market mechanism instead converged toward the competitive equilibrium even though basic assumptions (e.g., infinite number of buyers and sellers) were not met (Smith 1962). His work initiated a long wave of research and contributed to the promotion of laboratory experiments as a research method in economics (e.g., Smith 1976, Miller et al. 1977, Plott and Smith 1978, Williams 1980, Ketcham et al. 1984, Smith et al. 1988).

Today, investigating markets experimentally is common practice, especially in the context of financial markets (see Nuzzo and Morone 2017 for a recent overview). Most of the studies rely on some form of computer software providing the market infrastructure. While there is a large range of software for experimental markets, most researchers develop their own, custom-built solutions that are often not publically available (Palan 2015). However, there is a number of exceptions: VeconLab[31] and EconPort[32] provide online platforms that allow to (partly) configure and run different kinds of market games. GIMS (Graz-Innsbruck Market System) by Palan (2015) is a publically available market software, but is technically restricted to the features of the z-Tree application. ConG (Continuous Games) by Pettit et al. (2014) provides the fundament for running experiments with real-time interactions, but is not designed to run two-sided market institutions such as a continuous DA market (Pettit et al. 2014, p. 647). nodeGame by Balietti (2017) allows to run real-time experiments online with participants only needing a web browser access but does not provide a module for a DA market yet. Together, an experimental market software that is suitable for a wide range of research applications, easily usable and customizable and provided open-source is not available so far. We close this gap by offering an easy to implement and customizable DA market module for use within oTree. By providing the code open-source on GitHub, we want to encourage users to develop it further and apply the websocket technology to other real-time settings.

This paper proceeds as follows: Section 5.2 gives technical details on the use of websockets within oTree. Section 5.3 continues with the explanation of how to set up and use the

[31] See http://veconlab.econ.virginia.edu

[32] See http://econport.org/econport/request?page=web_home

DA market module. Section 5.4 shows the experimental design of the DA market, the procedural details and first results. Section 5.5 shortly concludes.

5.2 Real-time Interactions in oTree with Websockets

Real-time interactions are currently not provided as default option in oTree. In the following section, we illustrate how the current framework operates and why this complicates the implementation of real-time interactions. We then illustrate an approach to solve this issue with examples from the DA market code.

5.2.1 Current limitations of oTree

The current oTree architecture is not suitable for real-time interactions as part of a DA market, since such applications need a persistent connection with uninterrupted communication between the clients and the server. Currently, the user requests a page from the oTree server, makes some inputs and clicks on the "next" button to save the state and advance to the next page (see Figure 5.1).

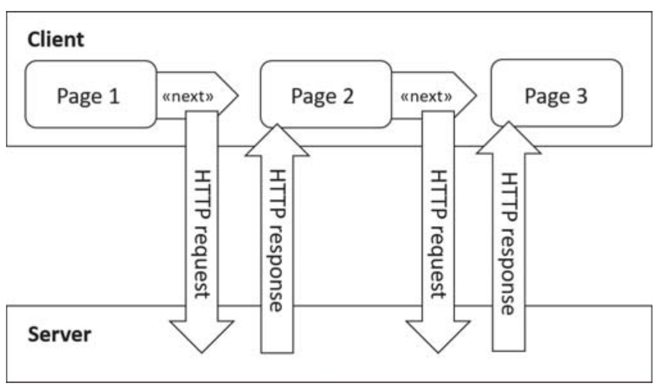


FIGURE 5.1: The current communication of an oTree application is based on HTTP requests.

Technically, only the "client" can initiate each step with a new HTTP GET or POST request to the server in order to load an entirely new HTML page. It is therefore necessary to use a technical protocol that allows the client to simultaneously send and receive messages from the server.

5.2.2 Using websockets in oTree

The DA market module relies on the application of websockets to handle interactions among players. A websocket is a continuous connection between a client and a server. In the case of oTree, this means that the participant's web browser keeps an open internet connection to the oTree service where messages can be sent and received simultaneously.[33] A way to cope with missing websockets is offered by the underlying Django framework which provides an abstraction of websockets, called "Django channels". Such channels represent an active websocket connection to a client that can be used to send or receive messages and terminate its connection. Multiple channels can be grouped to send messages to multiple clients. Figure 5.2 illustrates this architecture.

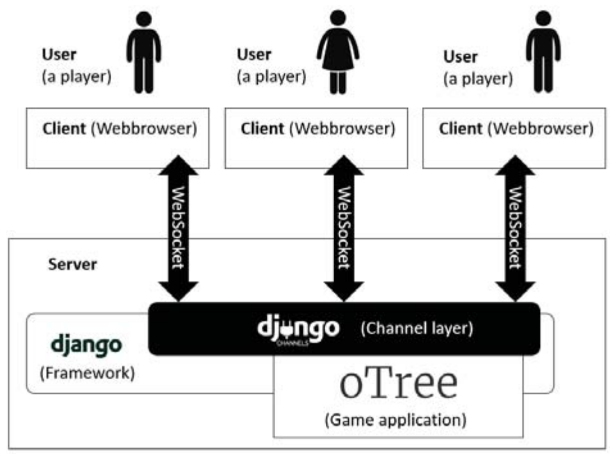


Figure 5.2: The websockets architecture within oTree using Django channels.

We show the use of websockets for four different forms of interactions: (1) connecting to a websocket, (2) sending a message, (3) receiving a message, and (4) disconnecting from the websocket. These four interactions, which are illustrated in Figure 5.3, require different implementations on the server- and client-side. For the server-side implementation, we provide various code snippets for further illustration using the DA market example.

[33]Websockets are already used in oTree to auto-advance players initiated by a process called "oTree worker": https://otree.readthedocs.io/en/latest/timeouts.html. However, applications to games are currently limited.

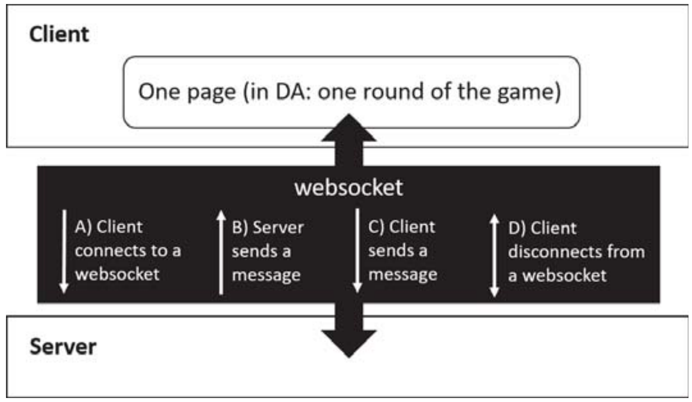


Figure 5.3: Websockets form a permanent communication channel between the client and the server.

5.2.3 Server-side implementation

The following section focuses on the server-side implementation and provides (pseudo) code written in Python as an illustration. In the following examples a websocket consumer class is used to handle the server-side interaction.

A) Connecting to a websocket: Various functions can be defined on the server which specify the actions of a client once it connects to the websocket. In our application, the client receives all information regarding the current state of the market, i.e. all bids and asks with the corresponding contractors, as multiple messages. Additionally, the websockets of all players of a DA market are grouped together in order to be able to broadcast group messages which are relevant for all market participants. A sample code is:

```
def connect(connection, **connection_params):
    """
    client connects to websocket and is automatically added to
    group_channel
    """
    session_code, player, group_channel = get_infos(connection_params)
    self.send(current_game_state)
```

B) Sending a message: As mentioned above, either Django channels or Django channel groups can be used to send messages, such as “New bid of 40 of buyer 6”. At the beginning of a round the connection is established and when a player changes his bid or ask, all other players need to be immediately informed about this change. The server informs all DA market players about the following actions:

- A player has updated his bid/ask,
- A player has deleted his bid/ask,
- A player has disconnected and is replaced by a bot,
- A player has reconnected and replaced the bot.

A sample code is:

```
# This sends a message to all DA market participants' websocket
group_send(group_channel, {
    "type": "match",
    "buyer": buyer_id,
    "seller": seller_id,
    "value": value
})
```

C) Receiving a message: The server collects all messages of user actions, calls the respective functions and replies with corresponding messages. In the DA market, there are multiple types of actions that a player can perform. The user can update or clear his entry or accept an offer/a bid from the "Current bids and asks" table. The server receives the message, evaluates the type of action, verifies its permissibility and acts accordingly, e.g. saves the changes to the database and informs the other players about the updated state. In case a message from a client is invalid, the server ignores it. The sample code is:

```
def receive(message, **conection_params):
    """
    When a message arrives the market state is updated according
    to the message.action_type
    """
    session_code, player, group_channel = get_infos(connection_params)
    if message.action_type == "clear":
        clear_bet(player.id)
    elif message.action_type == "seller" or action_type == "buyer":
        create_bid(message)
    ...
    endif
```

D) Disconnecting from the websocket: In case a client loses or closes the connection to the server, several game-specific functions can be implemented. The DA market module supports automated players that can replace a user who dropped the connection. If the bot service is enabled (see section on bots), all other players are informed about that dropout by labeling that player visually as a bot. The code for disconnecting is:

```
def disconnect(connection, **connection_params):
    """
    When a player disconnects he is automatically removed from
    group_channel
    """
    session_code, player, group_channel = get_infos(connection_params)
    replace_player_with_bot(player)
```

Once a connection is dropped, the websocket is discarded from the Django channel group.

5.2.4 Client-side implementation

As the client-side interaction is handled in the user's web browser, the corresponding websockets are implemented with built-in JavaScript functions. Possible actions correspond to those described for the server-side implementation.

A) Connecting with a websocket: After the page has loaded in the web browser, the client opens a websocket connection which stays open until the browser tab is closed or a new page is loaded. In our application, no actions are required on the client-side as the server handles the entire interaction.

B) Sending a message: If the user makes an input on the current page, a message will be sent through the websocket connection to inform the server of the change. In our application, a user can send three different types of messages:

- Update of a bid/ask
- Clearance of a bid/ask
- Accepting a bid/ask from another player

C) Receiving a message: An incoming message through the websocket leads to an update of the corresponding part of the webpage of the client. For example, a new message could be reading “New bid of 40 of buyer 6”. Accordingly, the “Current bids and asks” table is updated to show the new state of the market. Additionally, current values are analyzed to find suitable bids or asks and to show a corresponding “Accept” button.

D) Disconnecting from the websocket: When the connection is interrupted, the client is automatically disconnected from the websocket. In this case, the server is informed about the interruption. In our implementation, no specific action is required on the client-side as an irregular end of a connection (e.g. by closing the web browser) is handled by the server. The disconnected player can be replaced by a bot. It is also possible to disable the bot service in the settings. In this case, a player that drops out is labeled as “inactive”.

Bots

One option to handle drop-outs is to enable the bot service: When a player drops out of the DA market (due to attrition or technical issues), i.e. the websocket is closed, he is immediately replaced by a bot and all other players are notified of that change. Those players receive the label “bot” which is added to the player ID in the table “Market participants” (see Appendix A for a screenshot). A bot is programmed to place a bid/ask equal to his valuation/production costs at a randomized point of time (relative to when the websocket closed). To achieve this, a scheduled task is added to the “oTree worker”. Missing players can always return and instantaneously replace “their” bots by restoring the web session.

5.3 Setup and Usage

To use the DA market module, the experimenter is required to install oTree, Python and Redis[34]. Participants of the DA market only require an active internet connection and a common web browser on a computer or mobile device such as a smartphone or tablet. A preconfigured version of the DA market can be downloaded from GitHub.[35]

[34] https://redis.io. Redis is a message broker that handles schedules tasks and that is necessary to make use of the bots.

[35] https://github.com/IOP-Experiments

It consists of the oTree project setup, a copy of the DA market (which can be found in the subfolder double_auction) and a readme file with additional detailed information on the setup. Once the code is cloned, the command `otree devserver` starts the web application from a terminal (e.g. windows command prompt) with the oTree admin interface.[36] A session can be created by clicking on the game and the oTree web interface provides the links to start the experiment. To customize the DA market for individual needs, important variables can be either changed through the user interface or configured within the `settings.py` file:

```
bot_enable #Enables the bot service
delay_before_market_opens #Countdown before round starts (in seconds)
market_size #Maximum number of players in each market
num_of_test_rounds #Number of test rounds
production_costs_increments #Production costs increments of the seller
production_costs_max #Maximum production costs of seller
production_costs_min #Minimum production costs of seller
time_per_round #Time of one round (in seconds)
valuation_increments #Valuation increments of the buyer
valuation_max #Maximum valuation of buyer
valuation_min #Minimum valuation of buyer
participation_fee #Fixed compensation for participation
real_world_currency_per_point #Variable compensation for participation
```

The series of valuations (production costs) is created as following: The minimum valuation (production costs) is incremented by the specified parameter until the maximum valuation (production costs) is reached. The thereby generated values are then randomly assigned to the players. Each value is only assigned once among sellers and among buyers. When the number of players exceeds the number of values, the additional players receive a draw from another series, which is generated as described above. Due to the current structure of oTree, the number of rounds cannot be adjusted from the web interface but rather has to be changed in `models.py` by changing the `num_rounds` variable. Generally, users of the DA market module are advised to read the provided readme file in the repository.

5.4 The Double Auction Market

In the following, we describe the DA market as conducted for the first time by Adrian et al. (2019). While in their paper, the DA market is only needed to expose participants

[36]https://otree.readthedocs.io/en/latest/tutorial/intro.html

to a market environment to examine its impact on moral decisions, here we focus on the DA market data. We describe the game with predetermined values for the particular parameters. However, these can be easily changed and adapted for individual purposes (see section 5.3).

5.4.1 Experimental design

In our experiment, we ran a continuous DA market consisting of 9 buyers and 9 sellers over 10 rounds (with 2 additional, non-incentivized test rounds). We assign subjects randomly to either the role of a buyer or seller. Subjects keep their role for the entire 12 rounds. In every round, they can trade at most once a fictional good for 60 seconds. At the beginning of each round, buyers privately learn their valuation of the good and sellers privately learn their production costs of the good. Valuations and costs are randomly drawn from the sets $v \in \{30, 40, 50, \ldots, 120\}$ and $c \in \{10, 20, 30, \ldots, 90\}$. In each round, every value can only appear once among the buyers and sellers. While the distribution of demand and supply is common knowledge, the realization of v (for a buyer) or c (for a seller) is private knowledge to each market participant.

Sellers can sell and buyers can buy one unit of the fictional good in each round. Once the market opens, sellers can submit asks, i.e. the price at which they are willing to sell the product. Buyers can submit bids, i.e. the price at which they are willing to buy the product. All asks and bids appear in the table "Current bids and asks" and are visible to all market participants (see Appendix A for a screenshot). A trade occurs if a seller makes an ask that is lower than a current bid or if a buyer makes a bid that is higher than a current ask. The trade is closed at the price (of the bid or the ask) that was posted first. A trade is also possible by directly accepting a bid or ask that appears in the table. Sellers and buyers can modify their asks and bids until the market closes, as long as they did not trade yet. If a trade occurs, the payoff is $\pi_S = price - production\ costs$ for the sellers and $\pi_B = valuation - price$ for the buyers. Production costs only occur when trading, which means that it is not possible that a seller produces the good at a personal cost but cannot sell it on the market. Furthermore, buyers and sellers cannot make bids or asks that would lead to a negative payoff. After each round, sellers and buyers receive feedback and see a table with all trades and prices. The complete instructions can be found in Appendix B. Competitive equilibrium theory predicts an average trading price of 60 with a frequency of trades between 5 and 6 per round (see Figure 5.4).

5.4.2 Procedure

In the first test of conducting the DA market (Adrian et al. 2019), we ran 8 markets on AMT. After publishing the Human Intelligence Task (HIT), we let participants queue in a waiting room until all 18 spots were filled. This took on average less than 5 minutes. After that, subjects were provided with the instructions and control questions. To facilitate the formation of a group of 18 participants in a reasonable time, the evening before running sessions we posted the starting times on platforms such as turkerhub.com, as done by Suri and Watts (2011). Participation in the whole experiment (DA market plus two additional parts) took approximately 40 minutes and participants earned on average \$5.40 (\$3.00 participation fee plus the payment from one randomly selected round, where one point was transferred to \$0.15). The sessions were conducted between November 2nd and November 7th, 2018.

5.4.3 Results

Overall, $N_p = 116$ (out of 144 potential) participants in $N_m = 8$ markets completed the DA market, i.e. out of 18 players, we had on average 3–4 bots in each market. We analyze the number of trades and the actual average market price. Table 5.1 provides an overview.

TABLE 5.1: Aggregated market history

Round	Average number of trades	Average market price
test 1	5.63	57.62
test 2	6.19	57.39
1	6.00	58.33
2	6.00	58.27
3	5.94	59.37
4	6.13	59.76
5	6.00	59.58
6	6.25	59.56
7	6.25	58.58
8	6.25	58.08
9	5.88	58.57
10	5.88	59.62

As the table shows, the number of trades fluctuates around the theoretically predicted number of trades of 5–6 starting from the first round and the average market price converges rapidly toward the theoretically predicted (market-clearing) price of 60. Figure 5.4 illustrates the evolution of market prices graphically.

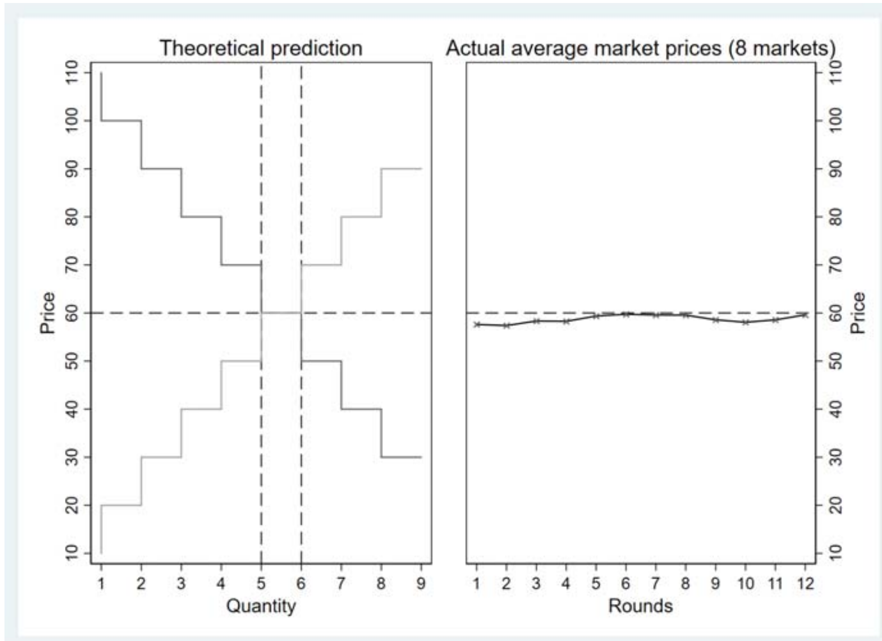


Figure 5.4: Theoretical prediction (left) and actual average market prices (right).

A comparison of our results to the existing literature shows a similar pattern of convergence with respect to the number of trades and the market price: In his first experimental markets, Smith (1962) shows that the predicted quantity and the predicted price arise within the first rounds of trading. Later studies consistently replicated that the DA mechanism is characterized by a fast convergence toward the theoretically predicted market price and highly efficient outcomes, especially in comparison to other market institutions such as a posted-offer market (e.g. Smith 1976, Smith et al. 1982, Ketcham et al. 1984). In addition, we find a comparably low variation of market prices as in Smith (1962). By running the DA market online on AMT and replicating results similar to previous laboratory experiments, we can further validate online subject pools as a reliable data source (for the discussion of the reliability of data from online experiments see e.g. Suri and Watts 2011, Mason and Suri 2012, Arechar et al. 2018).

5.4.4 Attrition and bots

One major challenge of conducting interactive games online on platforms such as AMT is that participants might leave the experiment at an early stage (even if this means that they are excluded from all payments), which the experimenter cannot influence. To cope with this problem of attrition, we implemented computerized players, so called bots.[37] These bots substitute human players once they close their web session. Bots are programmed such that they submit an ask equal to their production costs (as a seller) or a bid equal to their valuation (as a buyer) at a random point of time during the remaining seconds of the round. To comply with the rule of no-deception of study participants, the existence and trading strategy of bots is common knowledge to all participants and bots are indicated as such. By the use of bots, we can ensure that the experiment continues despite participant dropouts and allow the remaining players to continue without interruption. The code, however, allows to disable the bot service and thereby provides maximal flexibility for users who want to cope with attrition differently.[38] In this case, players who dropped out are labeled as "inactive" to other participants. Additionally, participants only received payment when they completed the whole experiment and we paid participants only for one randomly selected round to keep the importance of participating in every round high.

In the following, we examine attrition in our subject pool in greater detail. We can distinguish between two types of attrition: Attrition *before* the game, i.e. while participants are in the waiting room or reading the instructions, and attrition *during* the game, i.e. while participants are trading. Although participants were told at the beginning that the experiment would start after at most 15 minutes, most attrition (26 subjects or 18.06%) occurred during the time allocated to wait or read the instructions. The frequency of dropouts during the game is rather negligible. 116 of 118 participants (98.3%) participants who started the first round of the DA market also completed the last round. Of those 116 participants, only 4 participants were absent for some rounds

[37] oTree commonly uses the term "bot" to express automated testing of the experiment by inputs done by computerized players. We use the term for automated players who are playing in real-time with human players.

[38] Including bots might lead to a higher number of trades than theoretically predicted, as the bots are programmed such that they do not make a surplus by trading but only make bids/asks equal to their valuations/production costs. Switching off the bot service therefore provides a tool to exclude this dynamic.

during the game but reconnected before the last round. This suggests that subjects do not make noticeable use of dis- and reconnecting to the experiment in a strategic way.

As we experienced a varying influx of participants upon publishing the HIT, we chose to group participants into a market before giving access to the instructions. We wanted to avoid that after reading the instructions we could not start the market with sufficient participants and either pay for lost observations or deny the remaining participants payment. Based on our experience with low attrition during the game, a group formation after presenting the instructions might be an interesting alternative for future users to reach a high number of (human) participants.

5.5 Conclusion

In this paper, we illustrate the implementation of websockets in oTree that allow for real-time interactions. By running a DA market on AMT, we test both the replicability of basic results from the existing literature and the technical functionality of the DA market module. We find that participants from AMT behave comparably to participants in the laboratory, as we find similar convergence patterns in the DA market. In addition, we could verify that the technology works smoothly. We provide the code for the DA market module open-source on GitHub and encourage researchers and teachers to use and improve it to custom needs. For example, the extension to a multi-unit market would be a valuable contribution. In addition, we suggest to use websockets in oTree to investigate further synchronous, real-time games. Integrating websockets in games such as a dynamic public goods game or an asset market allows to run large-scale experiments and can thereby enrich behavioral experimental research.

References

Adrian, N., Crede, A.-K., and Gehrlein, J. (2019). Do markets foster consequentialist decisions? Evidence from an online experiment. *Working Paper*.

Arechar, A. A., Gächter, S., and Molleman, L. (2018). Conducting interactive experiments online. *Experimental Economics*, 21(1):99–131.

Balietti, S. (2017). nodeGame: Real-time, synchronous, online experiments in the browser. *Behavior Research Methods*, 49(5):1696–1715.

Chen, D. L., Schonger, M., and Wickens, C. (2016). oTree–An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.

Holzmeister, F. (2017). oTree: Ready-made apps for risk preference elicitation methods. *Journal of Behavioral and Experimental Finance*, 16:33–38.

Holzmeister, F. and Pfurtscheller, A. (2016). oTree: The "bomb" risk elicitation task. *Journal of Behavioral and Experimental Finance*, 10:105–108.

Ketcham, J., Smith, V. L., and Williams, A. W. (1984). A comparison of posted-offer and double-auction pricing institutions. *The Review of Economic Studies*, 51(4):595–614.

Mason, W. and Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, 44(1):1–23.

Miller, R. M., Plott, C. R., and Smith, V. L. (1977). Intertemporal competitive equilibrium: An empirical study of speculation. *The Quarterly Journal of Economics*, 91(4):599–624.

Nuzzo, S. and Morone, A. (2017). Asset markets in the lab: A literature review. *Journal of Behavioral and Experimental Finance*, 13:42–50.

Palan, S. (2015). GIMS–Software for asset market experiments. *Journal of Behavioral and Experimental Finance*, 5:1–14.

Pettit, J., Friedman, D., Kephart, C., and Oprea, R. (2014). Software for continuous game experiments. *Experimental Economics*, 17(4):631–648.

Plott, C. R. and Smith, V. L. (1978). An experimental examination of two exchange institutions. *The Review of Economic Studies*, 45(1):133–153.

Plott, C. R. and Sunder, S. (1988). Rational expectations and the aggregation of diverse information in laboratory security markets. *Econometrica*, 56(5):1085–1118.

Smith, V. L. (1962). An experimental study of competitive market behavior. *Journal of Political Economy*, 70(2):111–137.

Smith, V. L. (1976). Bidding and auctioning institutions: Experimental results. In *Bidding and Auctioning for Procurement and Allocation*, pages 43–64. New York University Press.

Smith, V. L., Suchanek, G. L., and Williams, A. W. (1988). Bubbles, crashes, and endogenous expectations in experimental spot asset markets. *Econometrica*, 56(5):1119–1151.

Smith, V. L., Williams, A. W., Bratton, W. K., and Vannoni, M. G. (1982). Competitive market institutions: Double auctions vs. sealed bid-offer auctions. *The American Economic Review*, 72(1):58–77.

Suri, S. and Watts, D. J. (2011). s. *PloS one*, 6(3):e16836.

Williams, A. W. (1980). Computerized double-auction markets: Some initial experimental results. *Journal of Business*, 53(3):235–258.

Appendix A: Screenshot DA market game

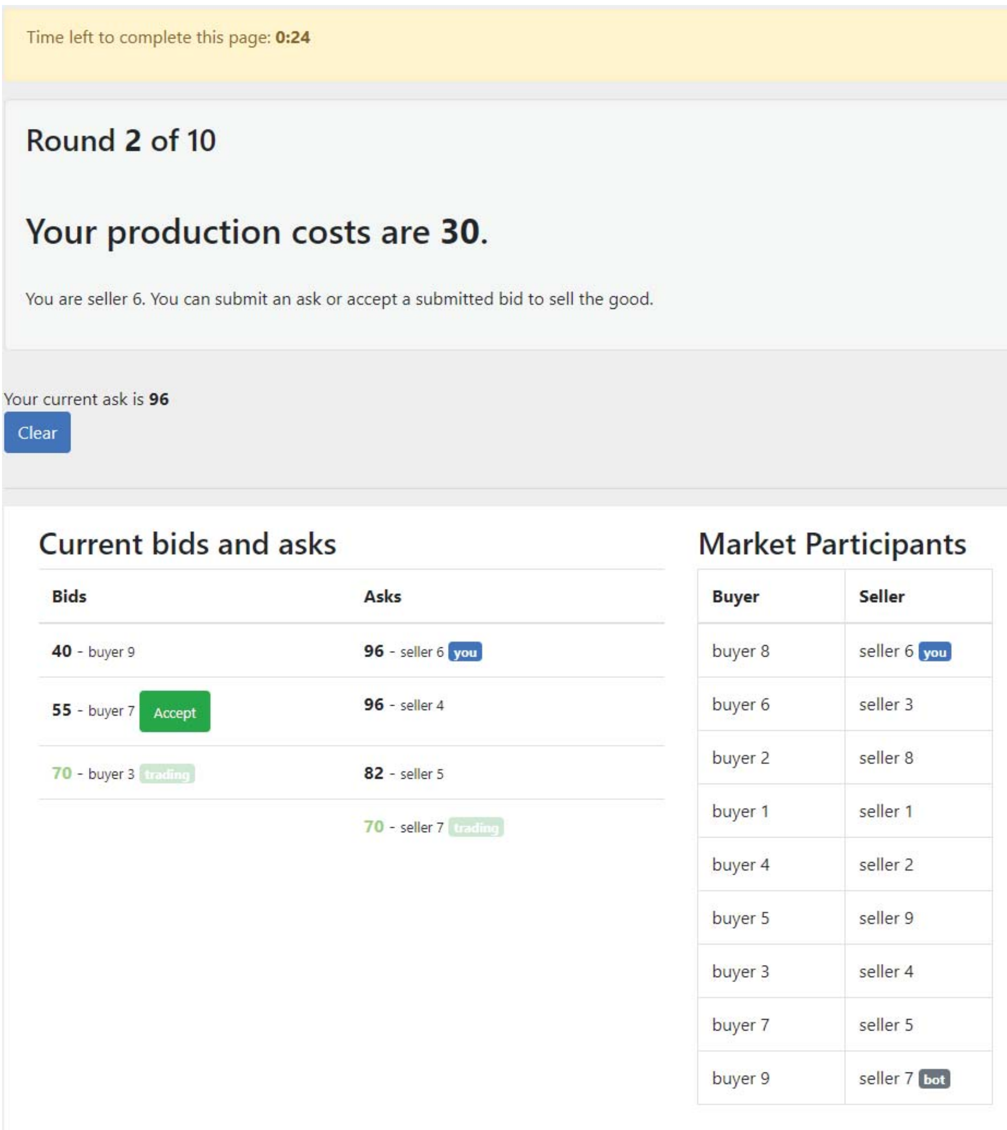


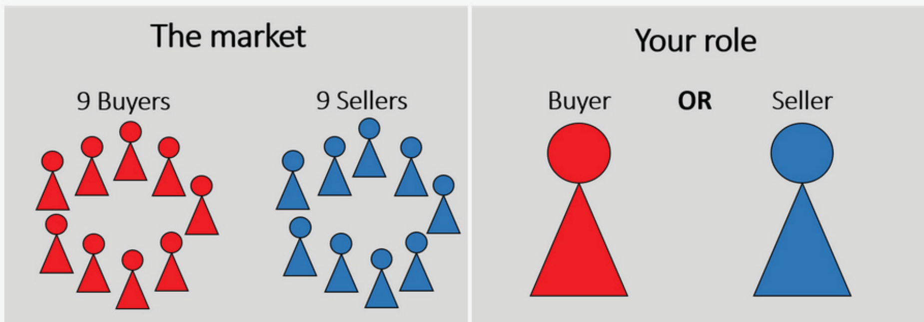
FIGURE 5.5: The Graphical User Interface of the DA market.

Appendix B: Instructions DA market game

General rules

In this part, you will be interacting in an online **market** consisting of 1 buyers and 1 sellers. These are real people interacting in real-time. You will be randomly assigned to the role of a buyer or the role of a seller. You will keep this role throughout the entire duration of the game. You will learn your role after reading the instructions.

There will be 10 trading rounds in which you can earn points by trading. One of these 10 rounds will be randomly chosen at the end of the study to count for your payment. In each of the 10 trading rounds, the market opens for 60 seconds, during which trading between buyers and sellers is possible.



What can a buyer do?

In each trading round, each buyer can buy **one unit** of a fictional good. By buying and hence owning this good, buyers receive a benefit in terms of a valuation. At the beginning of each trading round, each buyer learns how much the good is worth to him, i.e. he learns his own valuation. These valuations are different for each buyer and measured in points. The valuations will be randomly assigned to the buyers in each round and can be 30, 40, 50, 60, 70, 80, 90, 100 or 110 points. Among the buyers, each number is assigned only once within a round, i.e. one buyer is assigned a valuation of 30 points, another buyer is assigned a valuation of 40 points, yet another buyer is assigned a valuation of 50 points and so on.

What can a buyer earn?

A buyer can earn points by trading, i.e. by buying the good from a seller. If a trade occurs, a buyer gets the valuation (measured in points) minus the price (measured in points):

Buyer's earnings in points = Valuation - price

If no trade occurs, a buyer earns 0 points.

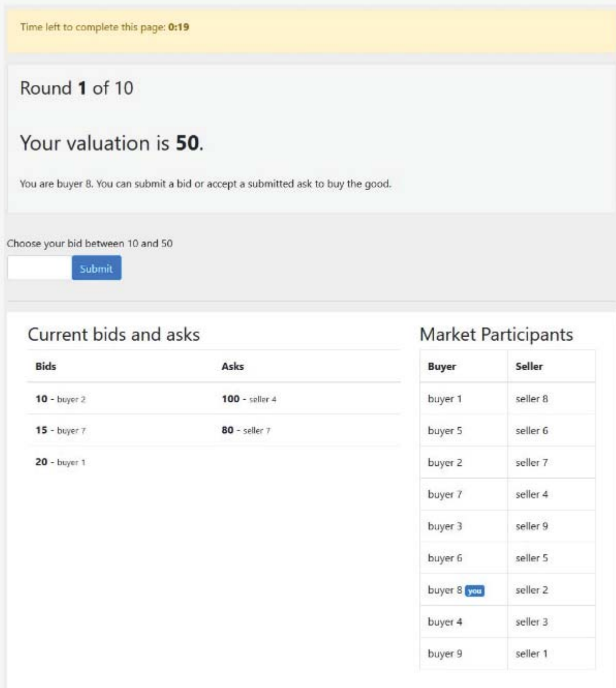
How does trading work for the buyer?

Trading is done on an online market platform. A buyer can trade in two possible ways:

1. He can accept an ask that has been submitted by a seller. The trade then occurs at the price of the ask.
2. Alternatively, he can submit a bid, i.e. the price at which he is willing to buy. If a seller accepts this bid or submits a lower ask, the trade occurs at the price of this bid.

The two possible ways of trading will be explained in more detail later on the screen.

The following screenshot shows how the online market platform looks like:



In each trading round, buyers are numbered consecutively from 1 to 9. The numbers change each round such that no buyer can be identified. In the example, the buyer has number 8. The valuation of the buyer in this round is 50, as you can see from the message on the screen "*Your valuation is 50*". You see a list with all market participants at the right side of the screen. Bids and asks of the buyers and sellers are displayed in the table "*Current bids and asks*".

At the beginning of each round, there is a countdown of 10 seconds during which each buyer learns his valuation. Then the market opens for 60 seconds. While the market is open, each buyer can trade one unit of the good by accepting an ask of a seller or by submitting a bid (these are the two possible ways of trading shortly described before):

1. **Each buyer can accept an ask** from the table "*Current bids and asks*". He does so by clicking on the accept button that shows up next to the lowest ask in the table. The good then trades for the price of the ask.

2. Alternatively, **each buyer can submit a bid**, i.e. a price at which he is willing to buy the good. In order to do so, he can enter a value and click on *Submit*. The bid then appears in the table "*Current bids and asks*" and is visible to all sellers and buyers. Within a trading round, a buyer can revise his bid as many times as he likes and replace it by a new one. If a seller accepts the bid of the buyer, trade occurs at the price of the bid. To avoid a loss, a buyer can only submit bids that are equal to or lower than his valuation.

   If a buyer submits a bid and there are lower asks in the table, trade occurs at the price of the lowest ask. In principle, it is the same as if the buyer had directly accepted the lowest (and thus currently best) ask in the table.

When the market closes, each buyer receives feedback about his payoff and all trades from that round.

What can a seller do?

In each trading round, each seller can produce **one unit** of a fictional good that he can sell in the market. At the beginning of each trading round, each seller learns how much it costs for him to produce this good, i.e. he learns his own production costs. These production costs are measured in points. They will be randomly assigned to the sellers in each round and can be 10, 20, 30, 40, 50, 60, 70, 80 or 90 points. Among the sellers, each number is assigned only once within a round, i.e. one seller is assigned production costs of 10 points, another seller is assigned production costs of 20 points, yet another seller is assigned production costs of 30 points and so on.

What can a seller earn?

A seller can earn points by trading, i.e. by selling the good to a buyer. If a trade occurs, a seller gets the price (measured in points) minus the production costs (measured in points):

Seller's earnings in points = Price – production costs

If no trade occurs, the good is not produced, i.e. the seller does not pay the production costs. Thus, if no trade occurs, a seller earns 0 points.

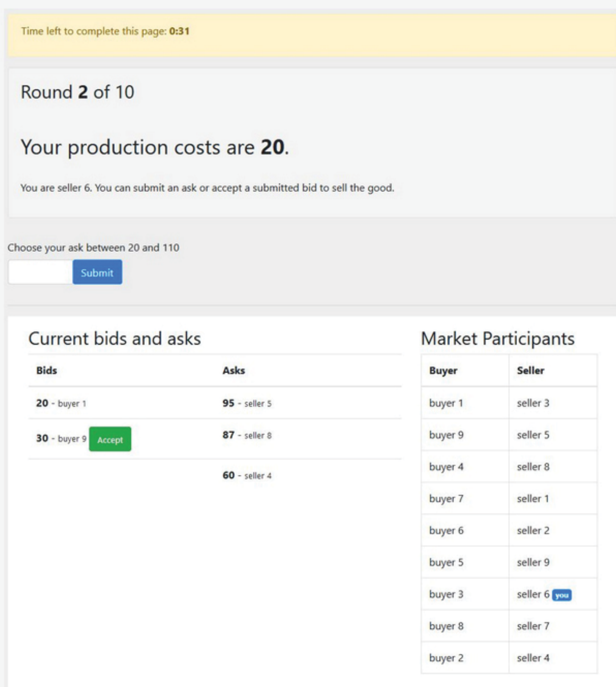
How does trading work for the seller?

Trading is done on an online market platform. A seller can trade in two possible ways:

1. He can accept a bid that has been submitted by a buyer. The trade then occurs at the price of this bid.
2. Alternatively, he can submit an ask, i.e. the price at which he is willing to sell. If a buyer accepts this ask or submits a higher bid, the trade occurs at the price of this ask.

The two possible ways of trading will be explained in more detail later on the screen.

The following screenshot shows how the online market platform looks like:



In each trading round, sellers are numbered consecutively from 1 to 9. The numbers change each round such that no seller can be identified. In the example, the seller has number 6. The production costs of the seller in this round are 20, as you can see from the message on the screen "*Your production costs are 20*". You see a list with all market participants at the right side of the screen. Bids and asks of the buyers and sellers are displayed in the table "*Current bids and asks*".

At the beginning of each round, there is a countdown of 10 seconds during which each seller learns his production costs. Then the market opens for 60 seconds. While the market is open, each seller can trade one unit of the good by accepting a bid of a buyer or by submitting an ask:

1. **Each seller can accept a bid** from the table "*Current bids and asks*". He does so by clicking on the accept button that shows up next to the highest bid in the table. The good trades at the price of the bid.

2. Alternatively, **each seller can submit an ask**, i.e. a price at which he is willing to sell the good. In order to do so, he can enter a value and click on *Submit*. The ask then appears in the table "*Current bids and asks*" and is visible to all sellers and buyers. Within a trading round, a seller can revise his ask as many times as he likes and replace it by a new one. If a buyer accepts the ask of the seller, trade occurs at the price of the ask. To avoid a loss, a seller can only submit asks that are equal to or above his production costs.

   If a seller submits an ask and there are higher bids in the table, trade occurs at the price of the highest bid. In principle, it is the same as if the seller had directly accepted the highest bid in the table.

When the market closes, each seller receives feedback about his payoff and all trades from that round.

Quiz

Please answer the following questions to make sure you understood the rules of the game correctly.

You are a buyer. Your valuation for the good is 50 points. You submit a bid of 40 points and a seller accepts this bid. What are your earnings (in points)?

You are a seller. Your production costs for the good are 20 points. You submit an ask of 25 points and a buyer accepts this ask. What are your earnings (in points)?

You are a buyer. Your valuation for the good is 40 points. Is it possible to submit a bid of 60 points?

○ Yes ○ No

Selbstständigkeitserklärung

Ich erkläre hiermit, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Alle Koautorenschaften sowie alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Mir ist bekannt, dass andernfalls der Senat gemäss Artikel 36 Absatz 1 Buchstabe o des Gesetzes vom 5. September 1996 über die Universität zum Entzug des aufgrund dieser Arbeit verliehenen Titels berechtigt ist.

Signed:

Date: 27.06.2019