UNIVERSITÄT
BERN

Experimental evidence on behavior in organizations and markets

Inauguraldissertation zur Erlangung der Würde eines

Doctor rerum oeconomicarum

der Wirtschafts- und Sozialwissenschaftlichen Fakultät

der Universität Bern

vorgelegt von

Ann-Kathrin Crede

Bern, Juni 2019

Acknowledgements

I would like to thank my first supervisor, Frauke von Bieberstein. Frauke was always open to my research ideas, and supported me in following my own interests. I am impressed by her deep understanding of experimental research and her professional way of dealing with all kind of challenges. I appreciate that Frauke was always available to answer questions and provide feedback. On a personal level, I must thank Frauke for having me brought to Bern, which I consider home now. Finally, I want to thank Frauke for cheering for me at the Grand Prix run every year.

I am also grateful to Christian Zehnder for being my second supervisor. Christian was always a source of inspiration, and gave enriching feedback on my research projects during numerous Summer Schools in Kandersteg and Winter Schools in Champéry.

I would not have enjoyed my 4 years as a PhD student as much as I did without my amazing colleagues. Thanks to Andrea, Jonas, Kathrin, Sebastian, and Steffi for making me feel very comfortable at work, and making the office a really nice place to be. I appreciated the opportunity to always discuss research with them, and to share interesting and enjoyable breaks and free time.

I especially want to thank Jonas for being the coolest office mate and co-author I could imagine. I enjoyed working with him, and appreciated his ideas and perspective on the questions we faced in our research projects. Apart from being a good research team, I enjoyed our weekly TRX trainings and Aare swims.

Finally, I want to thank Adrian for his endless and unconditional support and encouragement, especially in the last, busy weeks, and for much more I would rather tell him in person ☺.

Bern, June 2019
Ann-Kathrin Crede

Contents

List of Figures

List of Tables

Executive Summary

Today, running experiments to uncover causal relationships is a widely acknowledged tool in Managerial and Behavioral Economics. First with the help of laboratory experiments, and later with field and online experiments, the standard neoclassical model was increasingly challenged, and robust deviations from the assumption of a rational and selfish *Homo oeconomicus* were established. In addition, the experimental method is characterized by a broad field of application, and addresses not only economically highly relevant but also multidisciplinary questions.

One common feature of all four essays of this thesis is that they share this multidisciplinarity. Essays 1 to 3 focus on the honesty and morality of individuals in organizations and markets, and thus, are closely related to the fields of Psychology and Philosophy. These projects address the fundamental question of the driving forces of being honest or dishonest, or of behaving according to one or another moral principle. Essay 4 is multidisciplinary in the sense that it focuses on the technical aspect of running interactive experiments online, and thus, is closely related to Computer Sciences.

In essay 1, we investigate the influence of markets on morals. Following philosophy, the morality of an action can be evaluated based on the action or based on the consequences. 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.

In essay 2, I examine the underlying motives of lying aversion. I investigate the role of reputational concerns toward others in the decision not to lie in a die roll experiment. In a between-subject design, I exogenously vary whether the experimenter can observe the outcome of a die roll that determines the payoff. I find that partial lying and full lying disappear when the experimenter can track participants' behavior. This result can

be explained by reputational costs: Participants care about how they are viewed by the experimenter, and thus, abstain from lying when they are tracked.

In essay 3, we adapt the experimental method to family business research. We investigate whether family managers are perceived as more religious by external stakeholders than non-family managers, and how this perception alters stakeholders' decision to behave honestly toward a family manager in comparison to a non-family manager. By running a survey and an experiment, we show that family managers are perceived as more religious than non-family managers. In addition, we find that external stakeholders behave more honestly toward family managers than non-family managers, and that this positive effect is driven by the family managers' attribute of being religious.

In essay 4, we focus on a technical aspect of running interactive experiments online. We illustrate the implementation of websockets in oTree to allow for real-time interactions. As the first application, we run a continuous double auction market online to validate the functionality of our tool. We find that the number of trades and the market price converge toward the predicted equilibrium, as found in many laboratory experiments.

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 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).

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).

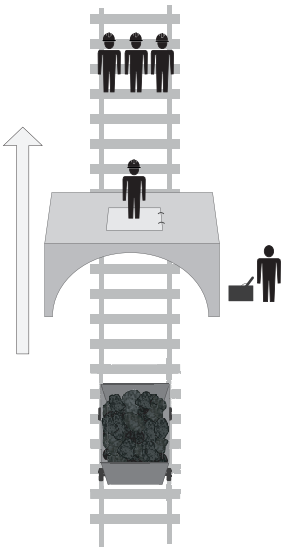


FIGURE 2.1: The boxcar dilemma (own illustration)

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 $n = 100$.

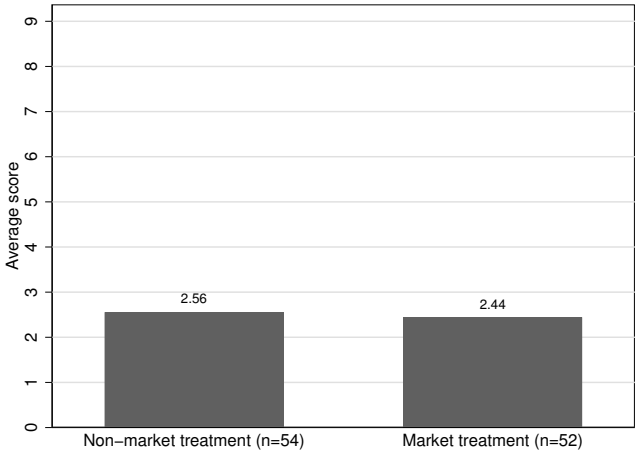


FIGURE 2.2: Results of the manipulation check

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.

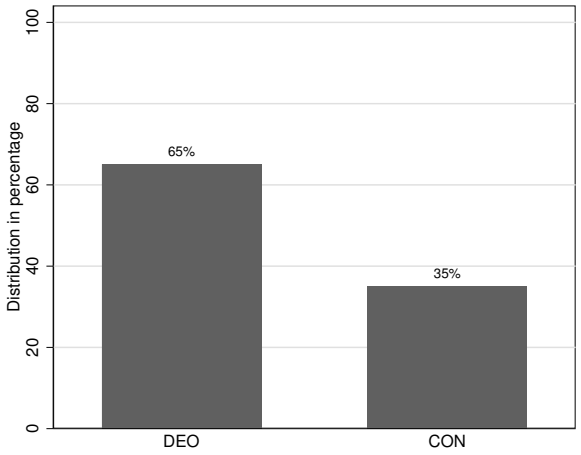


FIGURE 2.3: Distribution of decisions in the baseline
(DEO: deontological, CON: consequentialist)

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.

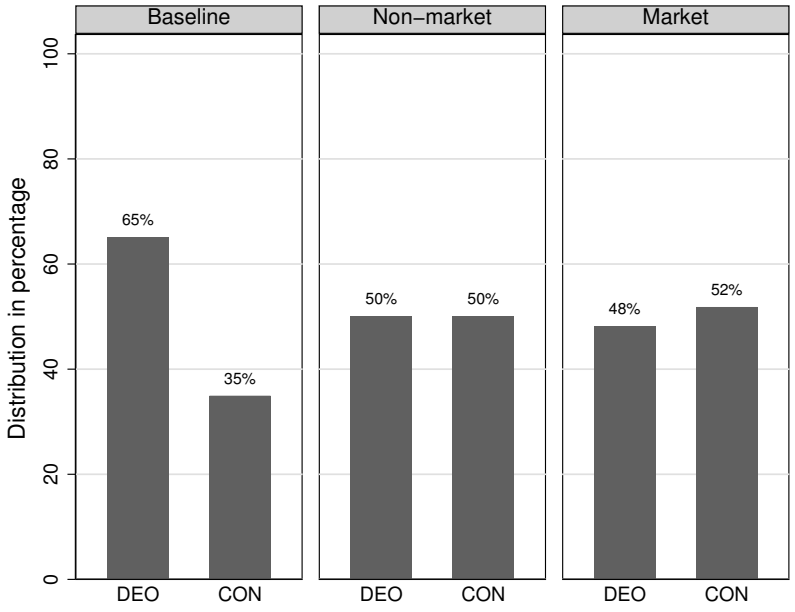


FIGURE 2.4: Distribution of decisions in the baseline versus the non-market and market treatments (DEO: deontological, CON: consequentialist)

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

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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

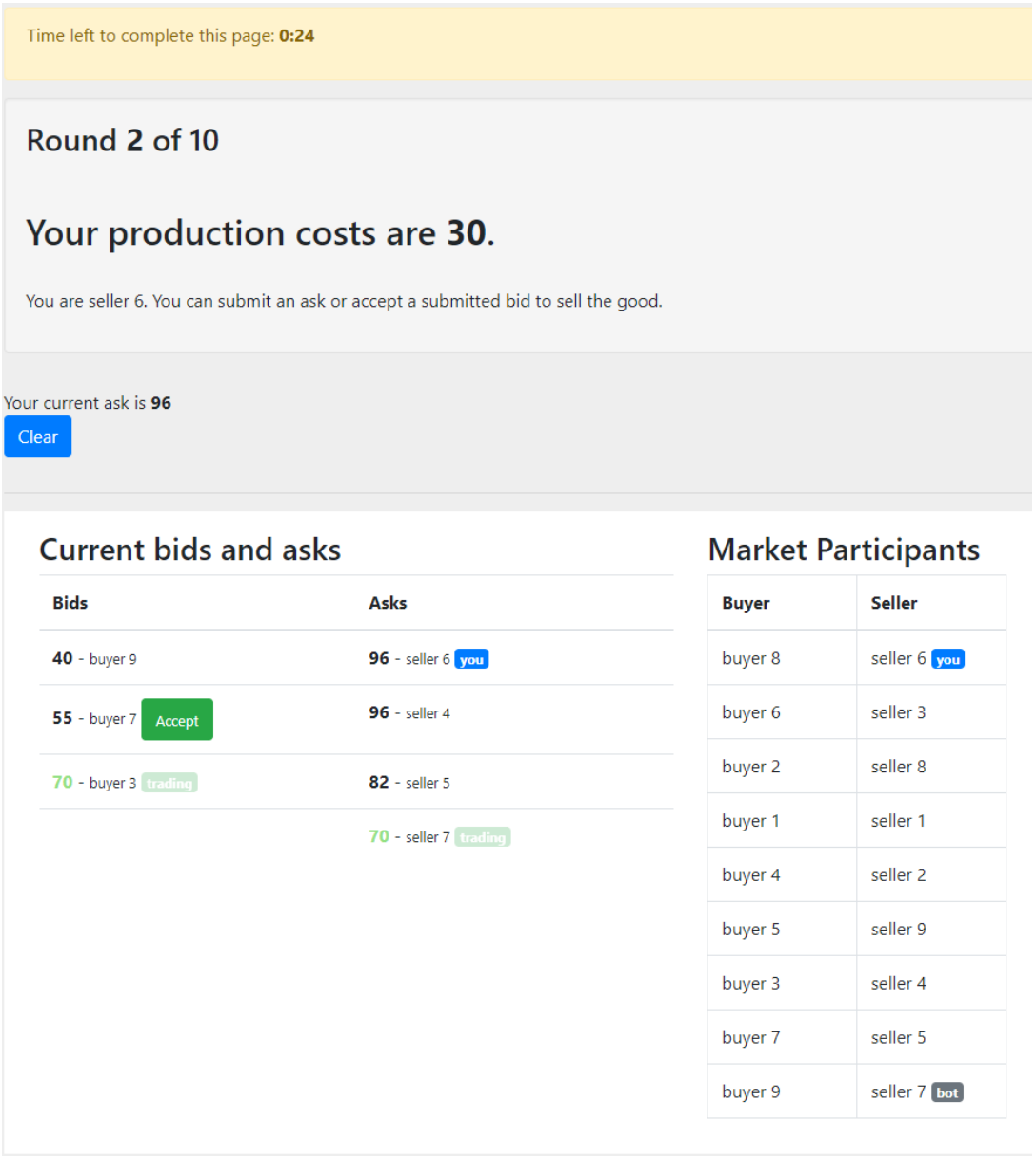


FIGURE 2.5: The Graphical User Interface of the DA market

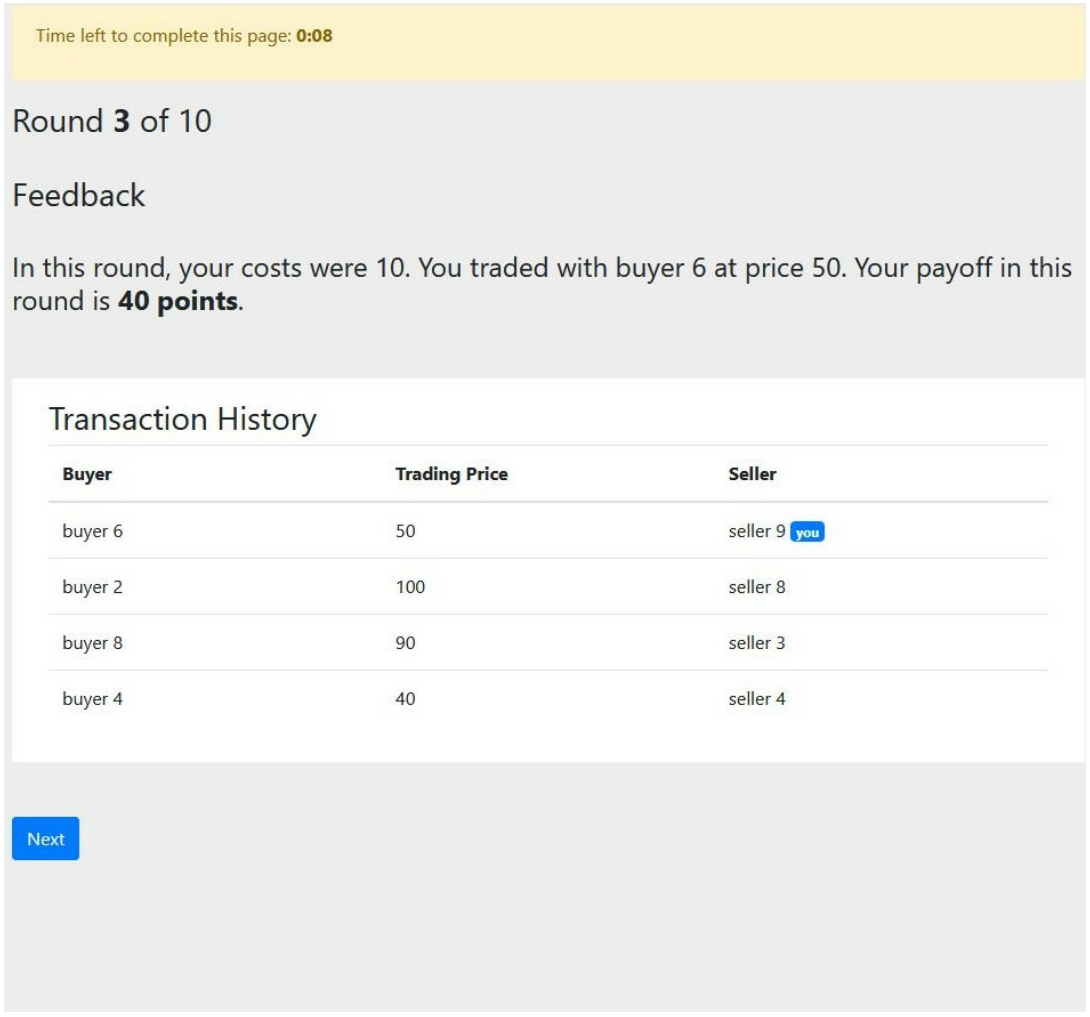


FIGURE 2.6: The feedback screen of the DA market

Appendix B: Word completion task

TABLE 2.3: Full list of word fragments and the corresponding market and non-market solutions. Words 10–14 served as control and could only be completed with a neutral meaning. Note that there might be additional 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

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: Uncovering underlying motives for lying aversion: Evidence from a laboratory experiment

Ann-Kathrin Crede [*]

Abstract

This paper investigates underlying motives for lying aversion. Although there is vast experimental evidence that people do not always lie fully even if this would maximize their monetary payoff, less is known about the driving forces for the aversion to lie. I investigate the role of reputational concerns toward others in the decision not to lie in a die roll experiment. In a between-subject design, I exogenously vary whether the experimenter can observe the outcome of a die roll that determines the payoff: In a control treatment, participants privately roll a six-sided die and report the outcome, which cannot be observed by the experimenter. In a digital die treatment, the outcome of the die roll is determined randomly by the computer, which the experimenter can verify ex post, and which the participants know. I find that partial lying and full lying disappear when the experimenter can track participants' behavior in the digital die treatment. This result can be explained by reputational costs: Participants in the digital die treatment care about how they are viewed by the experimenter, and thus, abstain from lying. These findings have important implications: In an age of increasing digitization, the possibility to observe and track behavior might bring benefits through more compliance with norms such as honesty.

Keywords: laboratory experiment, dishonesty, lying aversion, digital die, reputation, social image

[*] Ann-Kathrin Crede: ann-kathrin.crede@iop.unibe.ch, Institute for Organization and Human Resource Management, University of Bern. I thank Nicolas Hafner for excellent research assistance, Kathrin Friedrich and Lucian Hunger for running the experimental sessions, and René Fahr and Behnud Djawadi for providing the laboratory at the University of Paderborn. I am grateful to Johannes Abeler for providing valuable feedback at the CUSO Winter School (Champéry, 2019), and to Sebastian Berger, Andrea Essl, Kathrin Friedrich, Jonas Gehrlein, Stefanie Jaussi, Elisa Matthewes, and Frauke von Bieberstein for providing very helpful comments.

3.1 Introduction

Dishonest behavior is omnipresent in daily life (DePaulo et al. 1996, Mazar and Ariely 2006, Gino 2015). Even if tiny acts of dishonesty, such as taking office supplies from work or not paying for a bus ticket, might seem negligible and ordinary from an individual perspective, the resulting costs for organizations and society are tremendous (Weber et al. 2003, Kalish and Dermody 2004, Speights and Hilinski 2005). Following a cost-benefit perspective from standard economics, people will always engage in dishonest behavior when the utility from the dishonest act outweighs the potential punishment in case of detection (Becker 1968, Allingham and Sandmo 1972, Lewicki 1984).

However, numerous empirical studies have found that not everyone behaves dishonestly even if dishonesty has no consequences, and people are rarely dishonest to the maximum extent (Jacobsen et al. 2018). Instead, people have a preference for honesty (e.g., Kartik et al. 2007, Matsushima 2008, Kartik et al. 2014). Empirical evidence shows, for instance, that tax compliance is higher than theoretically predicted (Andreoni et al. 1998), managers do report truthfully (Evans III et al. 2001), people hand in full wallets to police stations (West 2005) and customers pay for a newspaper that they buy out of a box on the street (Pruckner and Sausgruber 2013). There is a growing body of literature showing that people renounce financial benefits and behave honestly, even if dishonesty does not involve long-term consequences. In a meta-analysis, Abeler et al. (2019) find the robust result that people forgo, on average, about three-fourth of the potential gains from lying. What are the driving forces that keep people from taking advantage of lies for financial benefits?

One widely used method to experimentally study lying behavior in a non-strategic environment is the die roll paradigm introduced by Fischbacher and Föllmi-Heusi (2013) (henceforth F & FH). The experimental design is simple: Participants privately roll a six-sided die and report the outcome, which determines their payoff. The higher the number they report, the higher the payoff, unless it is a six, which yields a payoff of zero. Because the experimenter cannot observe the outcome, participants have an incentive to report a higher number than they actually rolled (except six) to maximize their payoff. Although it is impossible for the experimenter to tell whether a particular individual lied or not, lying can be detected on an aggregate level, as the underlying true distribution of a six-sided die roll is known. F & FH identify three different behavioral types: There are honest participants who report truthfully, income maximizers who report the highest payoff, and partial liars who report a high, but not the maximum, payoff.

Why do some participants lie only partially? Two suggested motives are the desire to maintain a favorable self-image and reputational concerns.[8] F & FH run an additional double anonymous treatment to address these two motives. In this treatment, the experimenter does not know who *rolled* which number, as is the case in the baseline die roll treatment. In addition, he or she does not learn who *reported* which number. Thus, participants can be sure that the experimenter cannot reveal decisions on an individual level, which should exclude reputational concerns. F & FH do not find a significant change in behavior between the baseline and the double anonymous treatment; that is, they still find truthful reporting, partial lying and full lying. Therefore, they suggest that partial lying might be explained by the desire to maintain a favorable self-image, but not by reputational concerns toward the experimenter, as the experimenter cannot draw conclusions from individual behavior in the double anonymous treatment. Accordingly, some participants lie only partially to benefit from a higher payoff, but do not lie to the full extent, to still be able to maintain a positive image of themselves. However, with their design F & FH cannot address the question whether reputational concerns might play a role if the experimenter can actually observe individual behavior, which was ruled out in both treatments. It seems that participants do not react to the removed observability of their reporting behavior. An open question is whether they would react to the observability of their individual die roll.

In this study, I address this question, and investigate the influence of reputational costs on the decision not to lie when individual behavior can actually be tracked. Therefore, I run a die roll experiment based on the design by F & FH, and exogenously vary whether the experimenter can observe the outcome of a six-sided die roll determining the payoff: In the control treatment, participants privately roll a six-sided die and report the outcome, which cannot be observed by the experimenter. In the digital die treatment, the die roll is executed by the computer, which the experimenter can observe and verify ex post, and which participants know. The results show that partial lying and full lying disappear in the digital die treatment. Because the only difference between the control and the digital die treatment is the possibility of being tracked by the experimenter, the underlying motive for the observed lying aversion must be reputational costs, as self-image concerns are constant across treatments. Participants in the digital die treatment seem to care about how they are viewed by the experimenter, and abstain from lying when they are observed. These reputational costs can be due to an intrinsic preference for appearing honest, or due to the fear of actual consequences, such as being excluded from payment or future experiments.

This paper complements a very young strand of theoretical and empirical literature on the underlying motives for the aversion to lie. Gneezy et al. (2018), Abeler et al. (2019),

[8] Throughout the paper, I use the terms reputational concerns and reputational costs interchangeably.

and Khalmetski and Sliwka (2019) address reputational costs theoretically, and present models that incorporate a desire to appear honest in the utility function. In a different approach, Dufwenberg and Dufwenberg (2018) assume that individuals care about the beliefs others have about the *extent* of lying (instead of the chance of being a liar at all). In additional experiments, Gneezy et al. (2018) and Abeler et al. (2019) introduce the observability of a random draw to determine a payoff by the experimenter, comparable to my design. Whereas Gneezy et al. (2018) find that partial lying disappears while full lying remains when the experimenter can observe the outcome of the random draw, Abeler et al. (2019) find that partial lying and full lying disappear once observability is introduced, similar to my results. I discuss additional similarities and differences in the experimental design and the results throughout the paper.

This paper contributes to the existing literature by introducing observability by the experimenter in the die roll paradigm, and presents additional evidence for the influence of reputational costs on the aversion to lie. This finding has important implications: In an age of increasing digitization, behavior can be tracked more and more easily, and people leave digital marks. By making this potential observability more salient, one could take advantage of the power of reputational costs to keep people from engaging in dishonest behavior.

The paper proceeds as follows: In section 3.2, I briefly review the related literature. In section 3.3, I present the experimental design and procedure. In section 3.4, I show the results. In section 3.5, I discuss the results and conclude.

3.2 Related Literature

Recently, a large body of experimental studies investigating dishonest behavior has emerged (see Rosenbaum et al. 2014 and Jacobsen et al. 2018 for two literature reviews and Abeler et al. 2019 and Gerlach et al. 2019 for two meta-analyses). Dishonest behavior can be assessed in different ways. Four of the most widely used paradigms in the experimental literature are the deception game (Gneezy 2005), the coin flip task (Bucciol and Piovesan 2011), the die roll task (F & FH), and the matrix task (Mazar et al. 2008). Table 3.1 provides an overview of the most important characteristics of each paradigm.

TABLE 3.1: Overview of paradigms to study dishonesty (Gerlach et al. 2019)

	Deception game	Coin flip task	Die roll task	Matrix task
Dishonest behavior	sending a false message to receiver	misreporting a randomly generated outcome	misreporting a randomly generated outcome	misreporting one's performance in a task
Measuring scale	dichotomous	dichotomous	continuous	continuous
Measuring level	individual	aggregate	aggregate	individual/aggregate
Victim identifiable	yes	no	no	no

The deception game (Gneezy 2005) is a two-player game with a sender and a receiver. The sender has private information about the payoffs of two monetary distributions, where one distribution yields a higher payoff for the sender than the other (and vice versa for the receiver). The sender can send a true or wrong message to the receiver, stating "Option 1 will earn you more money," or "Option 2 will earn you more money." After reading the message, the receiver decides which option is implemented without knowing the true payoffs. The advantage of the deception game is that the experimenter can observe cheating on the individual level. However, as the victim of the lie can be identified, this game makes studying dishonest behavior complex, as the sender's social preferences and strategic considerations might enter his or her decision.

In the coin flip task (Bucciol and Piovesan 2011), participants are asked to flip a fair coin in private to determine their payoff. Reporting one outcome yields a prize, whereas reporting the other outcome yields zero payoff. The experimenter cannot observe the outcome of the coin flip and thus, cannot identify dishonest behavior on the individual level. However, as the underlying true distribution of a coin flip is known, lying can be detected on the aggregate level. As in the deception game, the decision whether to lie or not is binary. In contrast to the deception game, participants interact with the experimenter, and not with another participant. Thus, misreporting hurts the experimenter's budget.

In the original die roll task (F & FH), participants are asked to roll a fair six-sided die in private to determine their payoff. They receive CHF 1, 2, 3, 4, or 5 if they report a 1, 2, 3, 4, or 5, respectively, and CHF 0 if they report a 6. As in the coin flip task, the experimenter can identify dishonest behavior only on the aggregate level, and lying affects the experimenter, not another participant. In contrast to the deception game and the coin flip task, in the die roll task lying can be measured on a continuous scale, which makes this task suitable for studying partial lying.

In the matrix task (Mazar et al. 2008), participants have to solve math-based matrices, and are paid per correctly solved matrix. In the control group, the experimenter verifies and counts the number of correctly solved matrices. This performance is compared to a treatment group, in which participants count the number of correctly solved answers

themselves, and thus, can inflate their performance. As in the die roll task, the measurement of dishonesty is continuous. In contrast to the other three paradigms, lying is also related to the desire to appear competent.

Although the four paradigms differ in some important features, and do not yield the same results (see Gerlach et al. 2019 for a detailed comparison), they share the overall finding of the literature on dishonesty, that people *do* cheat in experiments, but not to the extent that a classical cost-benefit analysis predicts. When participants get the opportunity to increase their monetary gain by behaving dishonestly, many resist this option or make only partial use of it (Jacobsen et al. 2018). This result holds for experiments in the laboratory (e.g., Shalvi et al. 2011, Erat and Gneezy 2012, Conrads et al. 2014), as well as for studies conducted in the field (e.g., Levitt 2006, Abeler et al. 2014, Jacobsen and Piovesan 2016). In a recent meta-analysis, Abeler et al. (2019) quantify the forgone payoff by not cheating or not cheating fully to three-fourth of the potential gains, confirming how strong lying aversion overall is.

The literature discusses various motives for the aversion to lie. Following the recent classification by Abeler et al. (2019), three classes of models can be distinguished: First, people experience direct lying costs; that is, not telling the truth involves a utility loss (e.g., Kartik 2009, Gibson et al. 2013, López-Pérez and Spiegelman 2013). Second, people might experience reputational costs associated with how their (potentially) dishonest behavior is perceived by others (e.g., Gneezy et al. 2018, Dufwenberg and Dufwenberg 2018, Abeler et al. 2019, Khalmetski and Sliwka 2019). Third, people's decision whether to behave dishonestly might be influenced by social norms and social comparisons (e.g., Weibull and Villa 2005, Rauhut 2013, Dieckmann et al. 2016). For instance, groups are more inclined to behave dishonestly than individuals, which is driven by communication and learning about norm compliance among group members (Kocher et al. 2017).

Why do people experience reputational costs? The literature on *social image* suggests that many people care about how they are viewed by people in their surroundings, and that social image concerns can strongly affect economic behavior (see Bursztyn and Jensen 2017 for an overview). For example, people vote to be able to signal to others they voted (DellaVigna et al. 2016), work more for social recognition (Kosfeld and Neckermann 2011), and prefer donating in public than in private (Ariely et al. 2009). With respect to the decision whether to lie in a laboratory experiment, participants might pay more attention to their social image when they have the feeling the experimenter can observe their behavior. Some studies support this idea, showing that people are more likely to behave according to a social norm when an audience is present (e.g., Kurzban et al. 2007, Andreoni and Bernheim 2009, Cohn et al. 2018).

Two recent experimental studies (Gneezy et al. 2018, Abeler et al. 2019) examine the link between reputational costs and lying aversion. As both designs are very close to my experiment, I present them in detail. Gneezy et al. (2018) run an experiment in which participants randomly pick a number from a known distribution, and then report this (or any) number to receive a payoff that is solely based on this report. In their experiment, they vary whether the experimenter can observe the participants' outcome or not. In the *unobserved* treatment, participants get sealed envelopes containing 10 folded pieces of paper with numbers from 1 to 10. Participants are asked to take out one piece of paper, observe the number, put the paper back in the envelope, and report this (or any) number to the experimenter to determine the payoff. The payoff in € is equal to the number reported. In this treatment, the outcome of the random pick cannot be observed, and thus, cannot be verified. In the *observed* treatment, participants have to click on one of 10 boxes on a computer screen showing numbers from 1 to 10 and reveal the outcome. After seeing the number, they report this (or any) number to the experimenter and receive a payoff in € equal to the reported number. The participants' clicking behavior can be tracked and verified ex post, and thus this treatment allows observation of potential lying behavior. Importantly, the instructions do not explicitly mention that the experimenter can verify the real outcome ex post. Participants might, of course, form beliefs, and assume that clicking on a box on the screen can be tracked afterward, but the observability by the experimenter is not made common knowledge explicitly.

Gneezy et al. (2018) find less overall lying in the observed compared to the unobserved treatment. More importantly, they find that partial lying vanishes, whereas full lying remains when observability by the experimenter is introduced. Gneezy et al. (2018), therefore, conclude that partial lies can be primarily explained by reputational concerns,[9] and not by a desire to maintain a favorable self-image (as suggested by Mazar et al. 2008). When not being observed, participants might signal honesty toward the experimenter by not lying fully but by reporting a high number (partial lying). Accordingly, in the observed treatment, participants abstain from partial lying if the experimenter can observe and identify lying. Interestingly, the fraction of full liars (reporting a 10) does not change across treatments, suggesting that reputational concerns are heterogeneous across individuals. For those who do not care (or care only a little) about how they are viewed by others, and who also do not have self-image concerns, it does not make a difference whether the experimenter can track lying or not. Thus, they lie to the maximum extent in both treatments.

[9] In their paper, Gneezy et al. (2018) do not talk about reputational concerns, but about social identity, which they define as the way others perceive an individual. I think of the same concept when talking about reputational concerns or costs.

In a very similar setting, Abeler et al. (2019) vary whether the experimenter can observe the outcome of a random draw that determines the payoff. In the *unobservable* treatment, participants receive 50 chips with numbers from 1 to 10 (five chips of each number), and an envelope. They are asked to place all chips in the envelope, shake it a few times, and randomly draw one chip. After placing the chip drawn back in the envelope, participants write down the number of the chip on a payment sheet, and are paid by the experimenter outside the lab. The payoff in £ is equal to the number of the chip that was drawn. In this treatment, the experimenter cannot observe the true outcome. In the *observable* treatment, the random draw is carried out at the computer; that is, the process of drawing a chip from an envelope is simulated on the screen. An important difference from Gneezy et al. (2018) is that Abeler et al. (2019) employ a double-blind payment procedure for the observable treatment.[10] This payment procedure makes it impossible for the experimenter to link a reported number to an individual. Still, the experimenter can identify lying behavior by linking the number of coins left on the desk to the number of the drawn chip on the corresponding computer. As in Gneezy et al. (2018), the instructions for the observable treatment do not explicitly mention that the outcome of the random draw can be verified.

Abeler et al. (2019) observe less overall lying when the experimenter is able to track behavior, which is in line with Gneezy et al. (2018). An interesting difference from Gneezy et al. (2018) is that not only partial lying, but also full lying, disappears. That means in Abeler et al. (2019), even full liars abstain from lying when detecting lying is introduced, suggesting that reputational concerns are even stronger than in Gneezy et al. (2018). This finding is especially surprising, as Abeler et al. (2019) additionally employ a double-blind payment procedure in the observable treatment, making it even easier for participants to disguise their potential lie. Again, these findings cannot be explained by the desire to maintain a favorable self-image, but by reputational concerns. Participants care about how they are viewed by the experimenter, and abstain from lying when doing so could be detected. I come back to the studies by Gneezy et al. (2018) and Abeler et al. (2019) when I present my own design and the results.

3.3 Experimental Design

My experiment consists of a short questionnaire (see Appendix A), the rolling of a die, and the (self-)reporting of the payoff. Following F & FH, I let participants answer four questions to justify the payoff they could earn. They learned that the payoff for

[10] Before the start of the experiment, the experimenter placed envelopes with 10 coins of £ 1 each on the participants' desks. Instead of writing the drawn number on the payment sheet and handing it over to the experimenter, participants were asked to take as many coins from the envelope as determined by their drawn number. After taking the money, the participants could anonymously leave the laboratory.

completing this questionnaire would be different for every participant, and that a die roll served to determine the exact size of this payoff. Paying participants different amounts for filling out the same questionnaire might seem questionable. However, this way might be more credible than mentioning nothing about the purpose of the die roll to avoid an experimenter demand effect. The payoff amounted to € 1, 2, 3, 4, or 5 if participants rolled a 1, 2, 3, 4, or 5, respectively, and € 0 if they rolled a 6.

3.3.1 Treatments

In the control treatment, I gave participants a cup and a six-sided die, and asked them to roll the die to determine their payoff. The instructions (see Appendix A) explicitly explained that they should remember, and report, the outcome of the first die roll, but that they could roll the die several times to verify that the die was fair. After rolling the die once or more often, participants had to enter the outcome of their first die roll and the resulting payoff they would get on the computer. This treatment is very similar to the unobserved treatments in Gneezy et al. (2018) and Abeler et al. (2019), except the medium used to determine the random draw.

In the digital die treatment, I asked participants to determine the outcome of a die roll on the computer. I told them that the outcome of the die roll would be electronically saved, but not checked by the experimenter, for the payment procedure after the experiment. This is different from Gneezy et al. (2018) and Abeler et al. (2019), who did not explicitly mention that the experimenter was able to track outcomes ex post. I added this sentence to prevent participants from forming different beliefs about this possibility. Participants saw a screen with a button labeled *Rolling die*. By clicking on this button, a number between 1 and 6 was randomly generated and displayed next to the button (see Appendix B for a screenshot). As in the baseline treatment, I explained to participants that they should remember, and enter, the outcome of the first die roll, but that they could click the button more often to verify that the random generator worked properly. Similar to Gneezy et al. (2018), but unlike Abeler et al. (2019), I did not employ a double-blind payment procedure for the digital die treatment.

3.3.2 Procedure

I preregistered the experiment in the American Economic Association's registry for randomized controlled trials.[11] I conducted the experiment at the Business and Economic Research Laboratory at the University of Paderborn in December 2017. The experiment

[11] https://www.socialscienceregistry.org/trials/2607/history/23742

was implemented in z-Tree (Fischbacher 2007), and participants were recruited with ORSEE (Greiner 2015). Given that the experiment took only a very short time, I added it to the sessions of another experiment. After participants finished the first experiment, they learned that there would be a second, very short experiment that was unrelated to the first one. At the end of the session, participants received their payoff from the first experiment and the payoff for this experiment. In total, 140 participants took part in this study. I ran two sessions with the control treatment (n = 56) and three sessions with the digital die treatment (n = 84). Because of potential spillover effects from hearing other participants rolling a real die, it was not possible to run a session that included both treatments. I show later that the session does not seem to influence behavior.

3.4 Results

I first present the main variable of interest; that is, the payoff participants reported. Participants were asked to report a payoff of € 1, 2, 3, 4, or 5 if their die showed a 1, 2, 3, 4, or 5, respectively, and a payoff of € 0 if their die showed a 6. Figure 3.1 shows the distribution of the reported payoffs in percentage in the control treatment.

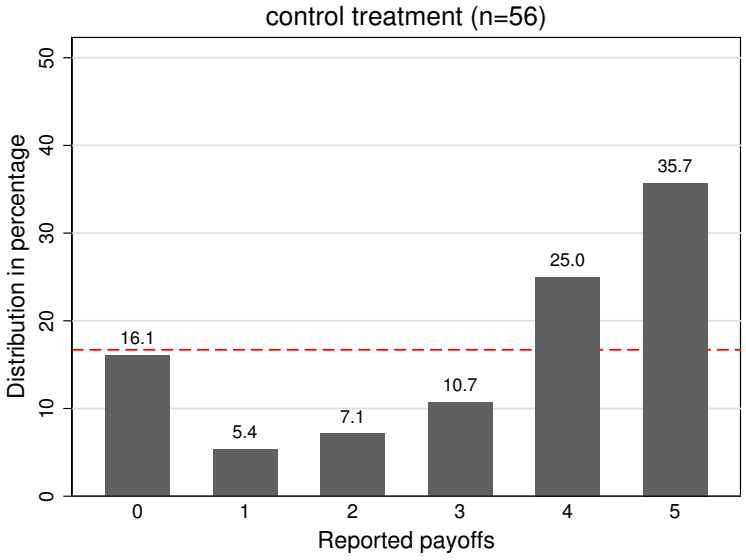


FIGURE 3.1: Distribution of reported payoffs in percentage in the control treatment. The red dashed line shows the uniform distribution.

As Figure 3.1 shows, the distribution is not uniform. I run two-sided binomial tests to check whether the observed percentage for each reported payoff differs from the theoretically predicted percentage of 16.7% (1/6) that would result in the case of full honesty (see the red dashed line). The observed percentage differs significantly from

16.7% for 1 (p=0.019) and 5 (p=0.001), differs weakly significantly for 2 (p=0.070) and 4 (p=0.106), and does not differ significantly for 0 (p=1.00) and 3 (p=0.284). These results are in line with the three types identified by F & FH. There are honest subjects (there is a positive fraction of participants who report 0), there are income-maximizing subjects (the fraction of participants reporting 5 is higher than 16.7%), and there seem to be partial liars (the fraction of participants reporting 4 is higher than 16.7%, though only weakly statistically significant). The main difference between my data and the F & FH results is that I find 16.1% of participants report 0, compared to 6.4% in their study. Thus, I find more honesty, and cannot fully replicate F & FH's monotonically increasing distribution. These results are in line with those of Gneezy et al. (2018) and Abeler et al. (2019), who identify partial lying and full lying in their unobserved treatments.

In the next step, I consider how participants in the digital die treatment behaved. Figure 3.2 shows the distribution of reported payoffs in percentage by participants who rolled the die digitally on the computer.

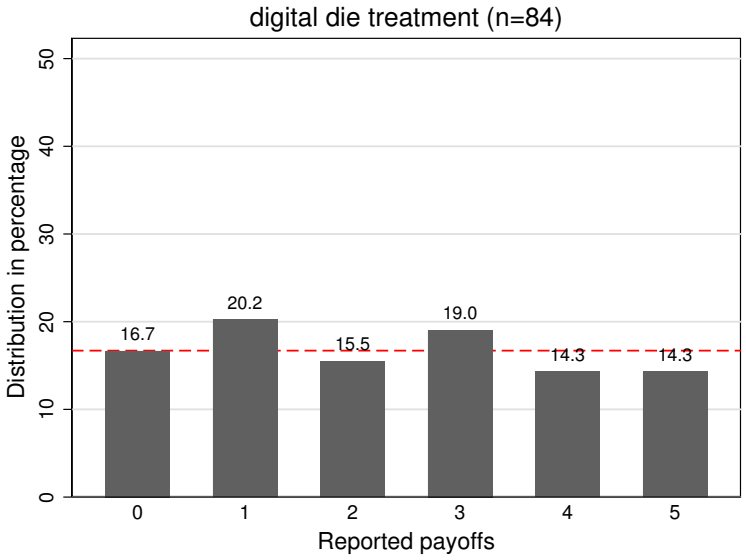


FIGURE 3.2: Distribution of reported payoffs in percentage in the digital die treatment. The red dashed line shows the uniform distribution.

As Figure 3.2 shows, the distribution is very close to uniform. Two-sided binomial tests for the observed percentage for each reported payoff reveal that none differs significantly from 16.7%. Thus, replacing a standard die with a digital die seems to prevent participants from lying, more specifically from partial lying and full lying. I discuss the potential underlying motives in the next section. This result is in line with that of Abeler et al. (2019), who also observe that partial lying and full lying disappear once observability by the experimenter is introduced. Compared to Gneezy et al. (2018), I do not find that full lying remains when the outcome of the random draw can be tracked.

In an additional step, I compare the distribution of reported payoffs in the control treatment with the distribution of reported payoffs in the digital die treatment. Table 3.2 shows the shares of participants reporting the corresponding payoff.

TABLE 3.2: Shares of participants in percentage reporting the corresponding payoff.

	0	1	2	3	4	5	Fisher exact test
Control treatment (n=56)	16.1	5.4	7.1	10.7	25	35.7	p=0.004
Digital die treatment (n=84)	16.7	20.2	15.5	19	14.3	14.3	

As Table 3.2 shows, a Fisher exact test comparing the distributions of payoffs reported in the control treatment and the digital die treatment reveals a statistically significant difference (p=0.004). Thus, replacing the standard die with a digital die leads to a significant change in behavior: The fraction of income maximizers and partial liars vanishes, and a uniform distribution of the die roll evolves, indicating that participants report the result of their die roll truthfully when their reporting behavior can be verified.

To confirm this result, I can compare the outcome of the digital die roll and the reported payoff, as the outcome of the digital die was saved. Of the 84 participants in the digital die treatment, 78 participants (93%) truthfully reported the first number that the digital die showed. Correspondingly, only 6 participants entered a result that did not coincide with the true outcome of the digital die.[12] This finding supports the result that replacing the standard die with a digital die leads to more honesty, and not simply to a uniform distribution due to potentially other (opposing) changes in behavior.

Finally, I run an ordinary least squares (OLS) regression with the reported payoff as the dependent variable and the *digital die* treatment, the amount of *money* gained in the preceding experiment, the *session*, *age*, and being *male* as explanatory variables. Table 3 provides the results.

[12]Out of these 6 participants, 3 participants lied upward, and reported a higher outcome than the digital die actually showed (a 3 instead of a 1, a 5 instead of 2, and a 2 instead of a 1). Surprisingly, 3 participants entered a number that was lower than the actual outcome of the die roll. This might indicate downward lying. However, 2 of these 3 participants entered the outcome of the second die roll. Thus, it seems more plausible that they did not lie downward on purpose, but that they remembered only the second die roll.

TABLE 3.3: OLS regression with the reported payoff as the dependent variable

	Model 1	Model 2
digital die	-0.935***	-1.040***
	(0.306)	(0.308)
money		-0.104
		(0.083)
session		-0.099
		(0.107)
age		0.043
		(0.032)
male		0.280
		(0.318)
_cons	3.304***	4.877***
	(0.245)	(1.843)
N	140	140
R^2	0.065	0.102

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

As the results show, only the treatment dummy has a significant effect; that is, being in the digital die treatment compared to the control treatment reduces the payoff by € 1.04. The amount of money gained in the experiment before, age, being male, and the session do not have a significant influence on the reported payoff.

3.5 Discussion and Conclusion

Summing up, I run a die roll experiment to investigate the influence of reputational concerns on the decision not to lie. Therefore, I exogenously vary whether the experimenter can observe the outcome of a die roll that determines the payoff: In the control treatment, participants privately roll a six-sided die and report the outcome, which the experimenter cannot observe. In the digital die treatment, the die roll is executed on the computer, meaning that the experimenter can verify the outcome ex post. This is common knowledge to participants. I find that partial lying and full lying disappear when the experimenter is able to track participants' behavior in the digital die treatment.

I suggest that this change in behavior can be explained by reputational concerns toward others, in this case toward the experimenter.

What could these reputational costs involve? First, participants could *intrinsically* dislike to be perceived as liars, and thus, abstain from misreporting. This is in line with studies on social image showing that people are more likely to behave according to a social norm when an audience is present. Second, participants could doubt that the outcome of the digital die is only electronically saved, but not checked before payment. Thus, they could be afraid of immediate real consequences in the sense that they forgo their payment when a potential lie is detected. Third, participants could be afraid of long-term consequences, such as being excluded from future experiments, when the experimenter detects misreporting ex post. At this point, I cannot disentangle the different explanations that I subsume under the term *reputational costs*. For further research, it would be interesting to employ a double-blind payment procedure to exclude that participants fear real consequences. In addition, a questionnaire after the experiment could help to learn more about participants' beliefs regarding the consequences of being observed and their motives for the aversion to lie.

Comparing my results to the current experimental literature, I find very similar patterns. Compared to Gneezy et al. (2018), I also find that partial lying disappears when the experimenter is able to observe the outcome of the random draw. In contrast, I do not find that full lying remains. The reason for this difference could be that I made the observability by the experimenter more salient, which might have increased the reputational concerns of those who usually care less about how they are perceived, and who still lied when they were observed in Gneezy et al. (2018). Similar to my design, Gneezy et al. (2018) did not employ a double-blind payment procedure in the observed treatment, so they could not exclude participants' belief of facing real consequences, either. Compared to Abeler et al. (2019), I also find that partial lying and full lying vanish once observability by the experimenter is introduced. This similarity is interesting, as in their study, the observability by the experimenter was not explicitly mentioned, and a double-blind payment procedure was employed in the observable treatment. Thus, participants could be sure that they could not be identified, and thus, did not face any consequences, such as being excluded from payment or future experiments. Still, Abeler et al. (2019) do not find any significant lying in their observable treatment. This finding suggests that fearing real consequences might play a minor role in the decision not to lie, and confirms the powerful impact of *intrinsic* reputational concerns.

I contribute to the existing literature in several ways. First, I replicate the baseline treatment by F & FH, and find support for the three types identified in their study: There are honest participants, income maximizers, and partial liars. Given the current

replication crisis in economics and other disciplines, it seems important to not only build on but also replicate existing findings (Maniadis et al. 2015, Camerer et al. 2016, Christensen and Miguel 2018, Verschuere et al. 2018). Second, I further develop the double anonymous treatment by F & FH by not focusing on the observability of reporting behavior, but by introducing observability of the die roll itself, to study the role of reputational concerns. Thus, I complement experimental findings recently published by other authors (Gneezy et al. 2018, Abeler et al. 2019) who identified similar effects of reputational concerns on the decision not to lie in slightly different settings. Third, I am the first to implement a digital die as a medium to introduce observability by the experimenter. Gneezy et al. (2018) and Abeler et al. (2019) used digital boxes or chips, but I tried to stick as closely as possible to the original design by F & FH, and simulated a die digitally. It seems that the medium used does not influence behavior, which further strengthens the finding for the influence of reputational concerns on the aversion to lie.

This study has several limitations: The sample size is small and the subject pool consisted of only students. For further research, it would be interesting to run similar experiments on a large scale with a mixed subject pool to verify the results. Abeler et al. (2014) investigated lying costs by calling a representative sample of the German population at home, and found that aggregate reporting behavior was close to the expected truthful distribution. Thus, investigating lying behavior in the laboratory might overestimate dishonesty, and calls for experiments conducted in an applied and representative context. Another drawback might be the task itself, as the die roll is artificial and not close to what people encounter in their daily life when it comes to the decision whether to lie or not. It would be interesting to study situations that are closer to the daily dilemmas people face. Next, I let participants make a one shot decision, and thus, could not study how behavior evolves over time. It would be interesting to implement a digital die treatment with repeated decisions to examine whether reputational concerns decline over time. F & FH ran a repeated version of their baseline die roll treatment, and found that participants reported higher payoffs when they participated a second time. This finding might suggest that reputational concerns extenuate over time. Finally, I focused on the observability by the experimenter, and thus, could investigate only reputational concerns toward him or her. In real life, people encounter observability by many different groups of people, such as family members, neighbors, or colleagues at work. It would be interesting to study the influence of the *type* of observer on lying behavior, and see whether reputational concerns differ depending on who is observing.

What are the policy implications? The results of this paper generally seem promising, as they support the finding that people have an aversion to lie, and do not cheat as much as they can, even if they could do so without being detected. However, relying only on voluntary honesty might not be sufficient, as there are always deviations from

average behavior. I suggest taking into account that people care about how they are viewed by others. In the age of digitization, behavior can increasingly be observed and tracked, and people leave digital marks. By making this potential observability more salient, one could take advantage of the power of reputational costs to keep people from engaging in dishonest behavior. At the same time, being observed by humans still seems more powerful than being observed by machines: In a laboratory experiment, Cohn et al. (2018) find that participants cheat significantly more when interacting with a machine than with a person, suggesting that human interaction is key to mitigating dishonesty.

References

Abeler, J., Becker, A., and Falk, A. (2014). Representative evidence on lying costs. *Journal of Public Economics*, 113:96–104.

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

Allingham, M. G. and Sandmo, A. (1972). Income tax evasion: A theoretical analysis. *Journal of Public Economics*, 1(3-4):323–338.

Andreoni, J. and Bernheim, B. D. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, 77(5):1607–1636.

Andreoni, J., Erard, B., and Feinstein, J. (1998). Tax compliance. *Journal of Economic Literature*, 36(2):818–860.

Ariely, D., Bracha, A., and Meier, S. (2009). Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. *American Economic Review*, 99(1):544–555.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, 76(2):169–217.

Bucciol, A. and Piovesan, M. (2011). Luck or cheating? A field experiment on honesty with children. *Journal of Economic Psychology*, 32(1):73–78.

Bursztyn, L. and Jensen, R. (2017). Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure. *Annual Review of Economics*, 9:131–153.

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.

Christensen, G. and Miguel, E. (2018). Transparency, reproducibility, and the credibility of economics research. *Journal of Economic Literature*, 56(3):920–80.

Cohn, A., Gesche, T., and Maréchal, M. A. (2018). Honesty in the digital age. *CESifo Working Paper Series No. 6996*.

Conrads, J., Irlenbusch, B., Rilke, R. M., Schielke, A., and Walkowitz, G. (2014). Honesty in tournaments. *Economics Letters*, 123(1):90–93.

DellaVigna, S., List, J. A., Malmendier, U., and Rao, G. (2016). Voting to tell others. *The Review of Economic Studies*, 84(1):143–181.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., and Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology*, 70(5):979–995.

Dieckmann, A., Grimm, V., Unfried, M., Utikal, V., and Valmasoni, L. (2016). On trust in honesty and volunteering among Europeans: Cross-country evidence on perceptions and behavior. *European Economic Review*, 90:225–253.

Dufwenberg, M. and Dufwenberg, M. A. (2018). Lies in disguise–A theoretical analysis of cheating. *Journal of Economic Theory*, 175:248–264.

Erat, S. and Gneezy, U. (2012). White lies. *Management Science*, 58(4):723–733.

Evans III, J. H., Hannan, R. L., Krishnan, R., and Moser, D. V. (2001). Honesty in managerial reporting. *The Accounting Review*, 76(4):537–559.

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

Fischbacher, U. and Föllmi-Heusi, F. (2013). Lies in disguise–An experimental study on cheating. *Journal of the European Economic Association*, 11(3):525–547.

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

Gerlitz, J.-Y. and Schupp, J. (2005). Zur Erhebung der Big-Five-basierten Persönlichkeitsmerkmale im SOEP. *DIW Research Notes*, 4.

Gibson, R., Tanner, C., and Wagner, A. F. (2013). Preferences for truthfulness: Heterogeneity among and within individuals. *American Economic Review*, 103(1):532–548.

Gino, F. (2015). Understanding ordinary unethical behavior: Why people who value morality act immorally. *Current Opinion in Behavioral Sciences*, 3:107–111.

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, 95(1):384–394.

Gneezy, U., Kajackaite, A., and Sobel, J. (2018). Lying aversion and the size of the lie. *American Economic Review*, 108(2):419–53.

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

Hahn, E., Gottschling, J., and Spinath, F. M. (2012). Short measurements of personality–validity and reliability of the GSOEP Big Five Inventory (BFI-S). *Journal of Research in Personality*, 46(3):355–359.

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.

Jacobsen, C. and Piovesan, M. (2016). Tax me if you can: An artifactual field experiment on dishonesty. *Journal of Economic Behavior & Organization*, 124:7–14.

Kalish, N. and Dermody, C. (2004). How honest are you? *Reader's Digests*, 164(981):114.

Kartik, N. (2009). Strategic communication with lying costs. *The Review of Economic Studies*, 76(4):1359–1395.

Kartik, N., Ottaviani, M., and Squintani, F. (2007). Credulity, lies, and costly talk. *Journal of Economic Theory*, 134(1):93–116.

Kartik, N., Tercieux, O., and Holden, R. (2014). Simple mechanisms and preferences for honesty. *Games and Economic Behavior*, 83:284–290.

Khalmetski, K. and Sliwka, D. (2019). Disguising lies–Image concerns and partial lying in cheating games. *American Economic Journal: Microeconomics (forthcoming)*.

Kocher, M. G., Schudy, S., and Spantig, L. (2017). I lie? We lie! Why? Experimental evidence on a dishonesty shift in groups. *Management Science*, 64(9):3995–4008.

Kosfeld, M. and Neckermann, S. (2011). Getting more work for nothing? Symbolic awards and worker performance. *American Economic Journal: Microeconomics*, 3(3):86–99.

Kurzban, R., DeScioli, P., and O'Brien, E. (2007). Audience effects on moralistic punishment. *Evolution and Human Behavior*, 28(2):75–84.

Levitt, S. D. (2006). White-collar crime writ small: A case study of bagels, donuts, and the honor system. *American Economic Review*, 96(2):290–294.

Lewicki, R. J. (1984). Lying and deception: A behavioral model. *Negotiation in Organizations*, 96(2):68–90.

López-Pérez, R. and Spiegelman, E. (2013). Why do people tell the truth? Experimental evidence for pure lie aversion. *Experimental Economics*, 16(3):233–247.

Maniadis, Z., Tufano, F., and List, J. A. (2015). How to make experimental economics research more reproducible: Lessons from other disciplines and a new proposal. In *Replication in Experimental Economics*, pages 215–230. Emerald Group Publishing Limited.

Matsushima, H. (2008). Role of honesty in full implementation. *Journal of Economic Theory*, 139(1):353–359.

Mazar, N., Amir, O., and Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, 45(6):633–644.

Mazar, N. and Ariely, D. (2006). Dishonesty in everyday life and its policy implications. *Journal of Public Policy & Marketing*, 25(1):117–126.

Pruckner, G. J. and Sausgruber, R. (2013). Honesty on the streets: A field study on newspaper purchasing. *Journal of the European Economic Association*, 11(3):661–679.

Rauhut, H. (2013). Beliefs about lying and spreading of dishonesty: Undetected lies and their constructive and destructive social dynamics in dice experiments. *PloS one*, 8(11):1–8.

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.

Shalvi, S., Dana, J., Handgraaf, M. J., and De Dreu, C. K. (2011). Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes*, 115(2):181–190.

Speights, D. and Hilinski, M. (2005). Return fraud and abuse: How to protect profits. *Retailing Issues Letter*, 17(1):1–6.

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.

Weber, J., Kurke, L. B., and Pentico, D. W. (2003). Why do employees steal? Assessing differences in ethical and unethical employee behavior using ethical work climates. *Business & Society*, 42(3):359–380.

Weibull, J. W. and Villa, E. (2005). Crime, punishment and social norms. *SSE/EFI Working Paper Series in Economics and Finance No. 610*.

West, M. D. (2005). *Law in everyday Japan: Sex, Sumo, Suicide, and Statutes*. University of Chicago Press.

Appendix A: Instructions

Questionnaire

I see myself as someone who

1. ... tends to be lazy.
2. ... values aesthetic experiences.
3. ... tends to be disorganized.
4. ... is reserved, quiet.[13]

Answers given on a 7-point Likert scale: 1 = totally disagree, ..., 7 = totally agree

Die roll game

Control treatment

For the questionnaire, you will receive a small additional payoff. However, this payoff is not the same for every participant. You determine your payoff by rolling the die twice as soon as you are asked to. The first roll determines how much you receive. If you roll a 1, 2, 3, 4, or 5, you receive one euro per number rolled (i.e. if you roll a 1, you receive € 1 if you roll a 2, you receive € 2, etc.). If you roll a 6, you receive nothing. The second die roll only serves to make sure that the die is working properly. If you wish, you can roll the die more than twice. However, only the first die roll counts.

If you have any questions, please raise your hand.

If you are ready, please press OK.

[13] Questions are taken from the German Socio Economic Panel (GSOEP) Big Five Inventory (see e.g. Gerlitz and Schupp 2005, Hahn et al. 2012).

Digital die treatment

For the questionnaire, you will receive a small additional payoff. However, this payoff is not the same for every participant. You determine your payoff by rolling the digital die on your computer twice as soon you are asked to. Therefore, you click the button "Rolling die" and will see the result of your die roll. The result of your die roll is electronically stored, but will not be verified for the payment procedure. The first roll determines how much you receive. If you roll a 1, 2, 3, 4, or 5, you receive one euro per number rolled (i.e. if you roll a 1, you receive € 1, if you roll a 2, you receive € 2, etc.). If you roll a 6, you receive nothing. The second die roll only serves to make sure that the die is working properly. If you wish, you can roll the die more than twice. However, only the first die roll counts.

If you have any questions, please raise your hand.

If you are ready, please press OK.

Appendix B: Screenshot digital die treatment

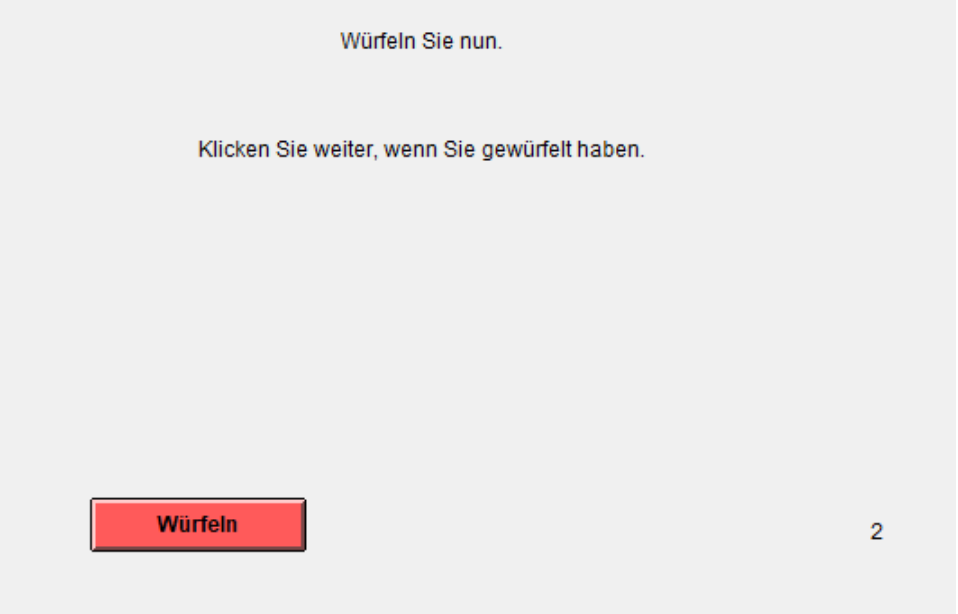


FIGURE 3.3: Implementation of the digital die in z-Tree with the button "Rolling die" on the left side and the outcome of the die roll on the right side. Participants could click on the button several times to check that the digital die works properly.

Essay 3: On family firm stereotypes: Are family managers perceived as religious and why should they care?

Ann-Kathrin Crede, Andrea Essl, Andreas Hack, Frauke von Bieberstein [*]

Abstract

Family firms and their family managers are associated with a variety of positive attributes inducing a distinct and strong reputation toward their stakeholders. Whereas the existing literature gives first hints that family firm values may be steeled by a deep sense of religious identification, little is known about how family managers are actually perceived with respect to religiosity and what the consequences of this perception might be. We close this gap by investigating whether family managers are perceived as more religious by external stakeholders than non-family managers, and how this perception alters stakeholders' decision to behave honestly toward the family manager in comparison to a non-family manager. By conducting a survey ($n = 600$) and an economic experiment ($n = 448$), we can show that family managers are indeed perceived as more religious than non-family managers. In addition, we find that external stakeholders behave more honestly toward family managers than non-family managers and that this positive effect is driven by the family managers' attribute of being religious. Summarizing, our research suggests that family firms can benefit by publicly expressing their family firm image and their family managers' affiliation to religiosity.

Keywords: family firm, reputation, stereotypes, religiosity, experimental methods, honesty

[*]Ann-Kathrin Crede: ann-kathrin.crede@iop.unibe.ch, Institute for Organization and Human Resource Management, University of Bern; Andrea Essl: andrea.essl@iop.unibe.ch, Institute for Organization and Human Resource Management, University of Bern; Andreas Hack: andreas.hack@iop.unibe.ch, Institute for Organization and Human Resource Management, University of Bern; Frauke von Bieberstein: frauke.vonbieberstein@iop.unibe.ch, Institute for Organization and Human Resource Management, University of Bern. We thank Nicolas Hafner, Maruscha Heyner, and Annika Wyss for excellent research assistance, Stefan Palan and the Max-Jung-Labor for providing the laboratory infrastructure, and Kathrin Friedrich, Jonas Gehrlein, Julia de Groote, Thomas Köllen, Alexandra Michel, and Sabrina Schell for valuable comments.

4.1 Introduction

For more than 125 years, the Swiss family firm Victorinox, famous for producing the genuine Swiss army knife, has been doing business based on strong Christian values (Südostschweiz 2014). Not only does the firm act according to Christian values, religious symbolism also plays a role in its professional conduct. For instance, at the start of every year, the firm organizes a church service for all employees to build up strength for the next year. Moreover, Carl Elsener Jr., owner CEO and Supervisory Board President, does not get tired of emphasizing in public that Victorinox is heavily influenced by Christian values.

Victorinox is just one example of many family firms demonstrating their religious affiliation. More than 400 family firms in Switzerland are, for example, members of the "Vereinigung Christlicher Unternehmer Schweiz" (Association of Christian Entrepreneurs Switzerland). Similar associations exist in countries like Germany ("Arbeitskreis evangelischer Unternehmer in Deutschland e.V." or the "Bund katholischer Unternehmer e.V."), France ("Entrepreneurs et dirigeants chrétiens"), Italy ("Unione cristiana imprenditori dirigenti"), or on an international level ("Christian Entrepreneurs Association").

For many family firms, being religious is part of their internal identity, which is the shared perceptions of organizational members regarding who they are as an organization (Brown et al. 2006). Moreover, they also portray their religiosity to their stakeholders to build a strong religious family firm image, which can be viewed as the mental associations that organizational members want others to hold about the organization (Brown et al. 2006). Apart from identity and image, there is also the concept of reputation. Reputation is how outsiders perceive an organization (Dyer and Whetten 2006), including the combined information and assumptions that stakeholders have about it (Brown et al. 2006). According to Binz Astrachan et al. (2018), a family firm's reputation is, amongst other things, the general perception that a family firm's diverse stakeholders have of family firms as a distinct class of economic actors, in comparison to non-family firms.

This view is based on stereotyping considerations (Allport 1954, Ashmore and Del Boca 1981, Tajfel 1969) such that individuals have a set of beliefs about the attributes of a group (e.g. the family firm) in general, irrespective of what the particular individual organization belonging to this group (e.g. Victorinox) in fact shows (Judd and Park 1993). Important to note is that stereotypes need not be negative and need not be inaccurate. In general, stereotypical attributes distinguish a particular group from other groups (Brigham 1971, McCauley et al. 1980).

Many researchers have tried to shed light on the general attributes stakeholders typically ascribe to family firms. Apart from a few exceptions (e.g. Keplinger and Feldbauer-Durstmüller 2012, Othman et al. 2011), most of the associations are positive and relate to the social capital of family firms (Sageder et al. 2018). To the best of our knowledge, there is no single study referring to religiosity as a distinguishing attribute. The first aim of our study is, thus, to find out if being religious constitutes a stereotypical belief about family managers, regarded as the personification of the family firm. According to the F-PEC scale (Astrachan et al. 2002), apart from voting rights, a family's power derives from active family management. Family managers influence the family firm through the family's values and related goals (Rau et al. 2018). Therefore, a family manager could be viewed as the embodiment of the family firm and the main representative of the family firm toward its stakeholders. Thus, the perception of a family manager as being religious might equally influence the reputation of the family firm as a whole.

While it is insightful to have a more comprehensive understanding of family firm reputation, investigating the possible consequences of being perceived as religious is probably even more important. Generally, the strategic value of a firm's reputation is undisputed (Rindova et al. 2005, Roberts and Dowling 2002). This is important for many stakeholder groups such as consumers (Walsh and Beatty 2007) who prefer to choose products from businesses with a good reputation and are willing to buy their products at higher prices (Fombrun and Shanley 1990, Walsh et al. 2009), or also for potential employees who feel more attracted toward firms with a good reputation (Turban and Cable 2003).

These generalized consequences of a good reputation have also been demonstrated for family firms as a category. In an experimental study, Schellong et al. (2018), for example, show that family firms are perceived by consumers as acting more supportive toward internal and external stakeholders and that these perceptions in turn result in higher consumer happiness. For job seekers, Hauswald et al. (2016), applying a conjoint experiment, find evidence that applicants are more attracted to firms with a strong family influence. Even for investors, family firm stereotypes play a role as has recently been established by Lude and Prügl (2019) in a choice-based experiment. Results show that stereotypes shift investors' preferences toward the high-risk alternative.

A positive reputational effect has also been found for religiosity. Clifford and Gaskins (2016) and Hout and Fischer (2002) show that religious people are perceived as being more moral and Bailey and Young (1986) find that religious individuals are rated more likeable than non-religious individuals. Furthermore, people who attend church services regularly are perceived more positively than those who do not attend church regularly (Isaac et al. 1995). In a business context, Bailey and Doriot (1985) show that students perceive professionals who actively express religious beliefs to be more liked, more intelligent,

more trustworthy, more moral, and more personally adjusted than professionals who do not express any kind of religious affiliation.

While there is first evidence on the value of family firm reputation and religiosity reputation, we are not aware of any study investigating whether the reputational effects of family firms and religiosity are interrelated and whether there is a combined and positive influence on stakeholder behavior like honesty. The stereotype of family firms and family managers as being trustworthy and socially responsible, for example, could be reinforced by additional cues that come along with religiosity. This reinforcement of held beliefs, also called stereotype confirmation, could be interpreted as a desire to perceive the world accurately and to perpetuate one's stereotypes (Sherman et al. 2012). If family managers are perceived to act religiously, this cue could strengthen the positive attributes resulting from the family firm reputation. In addition, this cue could activate values and concepts from religiosity (like e.g. "you should not lie" or "you should not steal") and thereby evoke more honesty of stakeholders that deal with the family manager. Honesty could thus be increased in two ways: Indirectly by appealing to the family firm reputation and thereby fostering trusting behavior, for example, and directly by pointing to the values of religiosity.

In this paper, we are especially interested in the combined reputational effect of being a family manager and being religious on stakeholder honesty. Especially in economic settings characterized by asymmetric information such as procurement (Akerlof 1978), job interviews (Stigler 1962), or customer complaint processes (Jacobsen et al. 2018), stakeholder honesty is evidently a crucial factor as it reduces overstating product quality, lying in the CV, or returning used goods as new. Given that cheating, deception, organizational misconduct and other forms of unethical behavior are a big challenge for firms (Gino 2015, Mazar et al. 2008), interacting with honest and ethical stakeholders is of essential importance.

Summarizing, we first investigate whether family managers are perceived as more religious compared to non-family managers. Second, we test whether family managers who deliberately communicate their religiosity will benefit in the marketplace through the combined effect of a family firm reputation, a religiosity reputation, and the confirmation effect religiosity has on family firm reputation. This enhanced reputation might then trigger honest behavior of a family firm's stakeholders. We empirically test our hypotheses with a survey (n = 600) and an economic experiment (n = 448). The survey examines how family and non-family managers are perceived with respect to religiosity. In the experiment, participants play a lying game (Gneezy et al. 2013) which measures honest behavior toward another player. This other player is, according to one of four possible

treatments, a family or non-family manager who is presented as religious or not (between-subject design).

We find that family managers are indeed perceived as more religious compared to their non-family counterparts. Furthermore, our experiment provides first evidence that stakeholders are more honest toward family managers compared to non-family managers. Interestingly, increased honesty toward family managers is contingent on the family manager's religiosity. This gives rise to the assumption that stereotypes trigger stakeholder behavior especially if strengthened by additional cues.

Our study makes several contributions to the literature. First, by showing that religiosity is part of family managers' and therefore also family firms' reputation, we add to a more comprehensive picture of which attributes family firms are associated with in the minds of the general public. This is important as reputation plays a crucial role in the behavioral actions and reactions of family firms' stakeholders. Second, we link the two research fields of family firms and religiosity. While there are first considerations that religious beliefs and behaviors might be crucial to a family firm's social identity (Paterson et al. 2013, Mitchell et al. 2013, Fang et al. 2013), research lacks empirical evidence regarding the influence a family firm's religiosity might have on its stakeholders. Our research takes an outsider perspective proposing that religiosity reputation can alter the behavior of the family firm's stakeholders like consumers, suppliers, or employees. Third, we are one of the first to introduce economic experiments into the family business research. Economic experiments are a very relevant method in research on behavioral reactions of stakeholders because they allow to uncover causal effects (Kagel and Roth 2016). Despite Evert et al.'s (2016) demand for more experimental studies in family business research, we know of only two studies applying this methodology (Hatak and Roessl 2015, Lude and Prügl 2019). Although a methodological issue, we believe that family firm research can benefit from more robust research methods to analyze behavior-oriented research questions.

The paper proceeds as follows: We first outline the theoretical background for our considerations. In doing so, we briefly summarize the main propositions of stereotyping theory. Second, we develop the hypotheses regarding family firm reputation and the impact of family firm and religiosity reputation. Third, we describe the applied methods, procedures, and measures of our statistical analysis. Fourth, we present the empirical results and robustness tests. Last, we discuss our findings and acknowledge limitations of our study while pointing to avenues of future research.

4.2 Theory and Hypotheses Development

4.2.1 Stereotype theory

Throughout the history of psychology research, there has been a strong interest in understanding stereotyping (Brewer and Brown 1998, Dovidio 2001, Duckitt 1992, Fiske 1998). This interest is shared by many disciplines like sociology, political science, or business administration (Ford and Stangor 1992). Especially in business, the practical implications of stereotyping are widely recognized (e.g. Brief et al. 2000, Gao and Mattila 2017).

Lippmann (2017) was the first to introduce the term "stereotype" to refer to the typical picture that comes to mind when thinking about a particular social group. Stereotypes are cognitive schemas used by social perceivers to process information about others (Hilton and Von Hippel 1996). Stereotypes encompass, among other things, beliefs about the traits of the typical group members, their qualities or behavior. This implies having a substantial amount of information about group members beyond their apparent surface traits and qualities.

Individuals rely on stereotypes, although these are often incorrect, rigid, oversimplified, and biased generalizations (English and English 1958), due to the limitations of the human capacity for processing information. Stereotyping provides receivers with tentatively useful information at relatively little cost in cognitive effort and time (Tajfel 1981). Stereotypical perceptions therefore have priority over individuating processes, the latter of which are used only if the correct evaluation of a target's properties is of particular relevance (Fiske and Neuberg 1989). According to the theory of person perception proposed by Fiske and Neuberg (1990), stereotypes in fact represent the default mode of person perception and reactions to others.

Fiske et al. (2002) develop a stereotype content model proposing that there are two fundamental dimensions of stereotypes. Depending on where the perceiver locates an individual within each dimension, special emotions will be elicited, which in turn trigger behavioral reactions. Warm (associated with "cooperation") and competent (associated with "high-status") groups elicit pride and admiration. Warm but incompetent groups produce pity and sympathy; stereotypically cold but competent groups elicit envy and jealousy; and stereotypically cold and incompetent groups generate disgust, anger, and resentment. In this regard, family firms might be seen as warm and competent and thus elicit positive emotions and reactions like benevolence or honesty, as many recent papers show (e.g. Beck and Prügl 2018, Schellong et al. 2018).

How stereotypes of and attitudes toward social groups form, develop, and preserve is intensely discussed. There are basically two types of models explaining stereotyping: motivational and cognitive models. The underlying assumption of motivational models is that stereotypes are the result of the individual's desire to satisfy his or her own basic motivational needs (Ford and Stangor 1992). An example is to maintain positive self-regard through the process of downward comparison with members of social outgroups (e.g. Crocker and Major 1989, Ehrlich 1973, Wills 1981), or to be accepted by significant others, from whom stereotypes are learnt and internalized (e.g. Glock et al. 1975).

Motivational models are particularly concerned with negative stereotypes and are less likely to add to the understanding of neutral or even positive stereotypes. Especially with regard to family firm stereotypes, which are mostly perceived as positive, cognitive models gain importance. The cognitive view states that cognitive processes like social perception and information processing lead to stereotyping (e.g. Fiske 1980). The central assumption of the cognitive approach is that social perception is complicated, and that the social perceiver cannot possibly process all of the social information to which he or she is exposed. Because people do not have the cognitive capacity to deal with all of the information that is available to them, some methods of simplifying the social situation are required (Allport 1954, Ashmore and Del Boca 1981, Fiske 1980, Tajfel 1969).

For example, as information can come from many different sources, including both direct observations of group members but also second-hand information passed down by significant others, especially by media (Mastro 2015) or by social media (Dobson and Knezevic 2018), people focus on the information viewed as the most informative. That means the information is generally high on communicability, contributing to persistent use (Schaller et al. 2002), is accessible (Higgins et al. 1982), has relatively high salience or uniqueness in a given context (Fiske 1980, McGuire et al. 1978), and is of affective importance to a perceiver (Erber and Fiske 1984, Hansen and Hansen 1988).

Information about a group's religiosity might be viewed as informative, since religiosity is mostly perceived as relatively stable (König 2015, Ashton and Lee 2014), is of high salience as it is easily understood (Clifford and Gaskins 2016), triggers affective reactions like feelings of trust in the minds of perceivers (Zaller 1992), and, especially if communicated in a business context, is of high uniqueness.

4.2.2 Hypotheses

Recently, numerous researchers have investigated the specific attributes that stakeholders ascribe to family managers and their respective family firms in comparison to non-family firms and external managers. Sageder et al. (2018) provide a review of the literature and

find that most of these associations relate to the social capital of family firms. Family firms are perceived as socially responsible (e.g. Blodgett et al. 2011, Binz et al. 2013, Beck and Kenning 2015), stable (Micelotta and Raynard 2011, Krappe et al. 2011) and honest (Keplinger and Feldbauer-Durstmüller 2012). The strongest empirical support to date is the perception of family firms as being trustworthy (Binz et al. 2013, Krappe et al. 2011, Orth and Green 2009).

Similar attributes have been found for religious people. Customers' degree of religiosity is positively correlated with their support for corporate social responsibility (Ramasamy et al. 2010) and religious executives tend to show more socially responsible business conduct (Graafland et al. 2007). In addition, religious people tend to favor values of stability that promote the conservation of traditions (Schwartz and Huismans 1995, Saroglou et al. 2004). There is also ample evidence of a positive relationship between religiosity and honesty, where religious priming (Mazar et al. 2008, Norenzayan and Shariff 2008, Stavrova and Siegers 2014) and religiosity (Bloodgood et al. 2008) reduce cheating and religious people are less accepting of unethical economic behavior like tax evasion (Kirchmaier et al. 2018). In addition, religious people are more trustworthy compared to non-religious people (Tan and Vogel 2008).

Research on similarity heuristics (Medin et al. 1993) shows that people's perceptual and conceptual processes have evolved such that they believe that similar things may behave in similar ways, and, more importantly for our context, that they judge the likelihood that an individual is a member of one category rather than another by the degree to which it is similar to others in that category (Tversky and Kahneman 1981, Shafffi et al. 1990). In this regard, it is reasonable to propose that if family managers are perceived as more trustworthy and honest, and thus more similar to religious people, they are more likely judged to be religious as well.

Aside from the proposed similarity effect, research shows that stereotypes are reinforced by second-hand information provided through the media (Mastro 2015). Given that most of the attributes of family firms and family managers are positive, these firms are increasingly communicating their family nature to their stakeholders around the world (Lude and Prügl 2019). This communication can constitute the general nature of being a family firm (e.g. SC Johnson – A Family Company) or it can highlight specific attributes of the family firm. One attribute that many family firms and family managers promote to internal as well as to external stakeholders is their religiosity. For instance, Philipp Pfenniger, the owner of Swiss consumer products company Trisa Group, wrote in the preface of the Annual Report 2010, at the height of the global economic crisis, that he trusted in God. The prevalence of associations of Christian entrepreneurs in Western

countries also shows this importance of religiosity for family firms and the willingness to show this religiosity to their stakeholders.

In sum, it is suggested that the similarity in attributes between family firms and religiosity as well as the communication of their religiosity to diverse stakeholders leads to the perception of family managers being more religious than non-family managers.

H1: Family managers are perceived as more religious than non-family managers.

The generally positive attributes that stakeholders assign to family firms (Sageder et al. 2018) are the building block for the firm's reputation. They might stem from first-hand and second-hand information (Hauswald and Hack 2013) about family firms' prosocial behavior in the market. Many empirical studies provide evidence about this behavior. Block (2010), for example, found that family firms engage less in downsizing. Berrone et al. (2010) showed that controlling families adopt environment-friendly strategies more frequently than non-family firms in polluting industries. On a more generalized level, Vazquez (2018), in his literature review on family firm ethics, states that family firms are described as having higher ethical focus compared to non-family firms. First-hand information, thus, supports the good reputation of family firms. Furthermore, family firms engage heavily in CSR communication, put strong emphasis on the philanthropic dimension and frame CSR as part of their corporate culture (Block et al. 2015) and, therefore, additionally convey positive second-hand information.

This positive reputation, in turn, has been shown to have a positive impact on family firms' financial and non-financial performance. Actions that support the positive reputation like investments in CSR have a positive effect on financial success (Block and Wagner 2014, Kashmiri and Mahajan 2010) and in a choice-based experiment, investors are more willing to invest in a high-risk family firm rather than a high-risk non-family firm (Lude and Prügl 2019). In addition, in a conjoint experiment, job seekers are more willing to join a firm when they receive information regarding the firm's family ownership and control (Hauswald et al. 2016).

The positive attributes that stakeholders assign to the family firm and family managers can influence stakeholders' actions toward the firm through different factors. One important factor is reciprocity (Falk and Fischbacher 2006), which describes the tendency that people reward kind actions (positive reciprocity) and punish unkind ones (negative reciprocity). Reciprocity is embedded in many social interactions (Fehr and Gächter 2000). For instance, small gifts can make people donate more money to charity (Falk 2007) or buy a product after having received a free sample (Cialdini 1993). In the laboratory, people reciprocate high investments by the first mover in a trust game by

sending back higher amounts (Berg et al. 1995) and punish low offers in the ultimatum game by rejecting the offer even at a personal cost (Güth et al. 1982). In addition, first movers' investment in the trust game increases with the perceived religiosity of the second mover. Interacting with a family manager who is perceived to act in a kind and prosocial way can thus induce a kind and prosocial reply.

Another factor to explain stakeholder actions is imitative behavior. Imitative behavior has been found in many settings (Bikhchandani et al. 1998): Television channels imitate shows (Kennedy 2002), firms in the chemicals industry follow the investment activities of others (Gilbert and Lieberman 1987), and the decision to commit crime is positively influenced by observing criminal behavior (Kahan 1997). Imitation is deeply rooted in human beings' nature allowing individuals to benefit from the knowledge of others (Bikhchandani et al. 1998). Imitating a family manager who is supposed to act honestly thus triggers honest behavior of a decision maker.

H2a: Stakeholders are more honest toward family managers than to non-family managers.

In addition to the factors that influence stakeholders' actions described above, there is one further important factor to consider, especially when it comes to ethical actions of stakeholders, which is religiosity. Prior research has shown a strong link between religion and morality, because religions prescribe morality and religious people often see religion as the source of morality (Geyer and Baumeister 2005). Reminding people of religious moral standards (e.g. in the form of the Ten Commandments) reduces subsequent lying behavior in an incentivized experimental study (Mazar et al. 2008). The authors argue that the reminder increases the awareness of participants‘ own moral standards and thus increases the likelihood of behavior consistent with these standards. In our setting, mentioning the religiosity of the interaction partner can serve as such a reminder of moral standards.

Although family managers are perceived as more religious in general (see the arguments leading to H1), the information of dealing with a family manager might not be strong enough to produce the reminding effect of moral standards, because religiosity is less salient in this case. In addition, while stereotypes represent the default mode of person perception and reactions to others (Fiske and Neuberg 1990), because of their ease of use, such judgements can prove erroneous (Macrae et al. 1994). Conversely, forming more accurate individuated perceptions requires attention and cognitive resources. People may invest in individuation processes if the correct evaluation of a target's properties is of particular relevance (Fiske and Neuberg 1989), as in the case of our experiment, where own actions have a strong influence on the interaction partner's (hypothetical) financial outcomes (see methodology section). However, in some situations it is not

possible to gain more individuated information, for instance, as in our case where there is no personal interaction. Here, the decision maker requires additional cues to substantiate the initial stereotypical assessment.

When presented with two or more cues, perceivers are prone to a confirmation bias, meaning that they are reluctant to revise initial expectations (Darley and Gross 1983). For instance, supporters and opponents of the death penalty considered evidence in favor of their opinion as more conclusive and credible than evidence against their opinion. Thus, the same two-sided evidence let both groups feel more convinced of their initial opinion (Lord et al. 1979). The confirmation bias can also explain why some beliefs persist when the initial evidence for them is removed. This has been shown in experiments where participants were first given fake evidence and later debriefed. Some of the initial belief based on the fake evidence remained even after debriefing (Ross et al. 1975), and even when it was clear that participants understood the debriefing (Anderson et al. 1980). Thus, given the confirmation bias, the effect of religiosity information can only work if it is in line with the other cues given, i.e., when dealing with a family manager who is perceived as religious per se, as opposed to a non-family manager who produces no initial perceptions of religiosity.

Taken together, when participants in our experiment are presented with the cue of religiosity, it supports their initial perception of the family manager as being religious. This in turn serves as a reminder of moral standards and increases the likelihood of behavior consistent with these standards.

H2b: The positive effect of more honest behavior toward family managers than to non-family managers is driven by family managers' reputation of being religious.

4.3 Methodology

To test hypothesis 1, we ran an online survey on Clickworker (similar to Amazon Mechanical Turk).[14] To examine our hypotheses 2a and 2b, we chose an experimental approach, more exactly a lying game. Running controlled experiments allows the clear identification of causality and is therefore ideal to uncover the effect of perceived family managers' religiosity on the relationship with their stakeholders. We ran our online experiment after the survey on Clickworker, and, as a robustness check, we additionally conducted an incentivized laboratory experiment. We explain both the survey and the lying game in more detail in the following parts.

[14] Clickworker is a platform that provides a workforce to carry out simple tasks and studies such as surveys: https://www.clickworker.com/.

4.3.1 Survey

The survey consists of two short texts, one describing a family manager and the other a non-family manager of a (random) company. We use the F-PEC scale by Astrachan et al. (2002) as used in Schellong et al. (2018) to design our company descriptions. The text on the family manager emphasizes that the person described owns the company and that the family already owned the majority of shares of this company in the last generation. The text on the non-family manager underlines that diverse private and institutional shareholders own the majority of shares of the company.[15] We present both texts to each participant in a within-subject design and determine the order of the two texts randomly. After each text, participants assess several statements on how they perceive the (non-)family manager's degree of religiosity and other personal traits. In total, 600 people participated in the online survey.

Participants assess the attributes of a non-family and a family manager (independent variable). Our dependent variable is the perceived religiosity of non-family and family managers. We use an adoption of the Centrality of Religiosity Scale (CRS-5) by Huber and Huber (2012) to measure religiosity perception. Participants assess these statements on a 5-point Likert scale ranging from "strongly agree" (coded as 5) to "strongly disagree" (coded as 1). We build the CRS-5 score by taking the average of all five coded answers. The CRS-5 yields a Cronbach's alpha of 0.92 for family managers (RMSEA=0.026; CFI=0.99; TLI=0.99) and 0.93 for non-family managers (RMSEA=0.086; CFI=0.99; TLI=0.98).

For our robustness test, we want to know how non-family and family managers are perceived with respect to fundamental personal traits like trustworthiness, social responsibility, innovativeness, and selfishness. We let participants assess selected statements that are based on the Corporate Character Scale by Davies et al. (2004) using the same 5-point Likert scale. We use the coded answer for each statement as outcome variable.

4.3.2 Lying game

Lying game basics

We base our experiment on the lying game introduced by Gneezy et al. (2013). The game works as follows: There are pairs of two players, named A and B. Each pair is randomly assigned a state s, which is an integer number $1 \leq s \leq 6$. Only A learns the number assigned and sends a message about this number to B. The message reads as

[15] The original descriptions of the family manager and the non-family manager as used in our experiment can be found in Appendix A.

"The assigned number is r", with $r \in \{1, 2, 3, 4, 5, 6\}$. After receiving this message, B decides whether to follow the message or not. A's payoff is a linear function of r and independent of s and B's decision:

$$\pi_A = 100 + 2r$$

Thus, the higher the message r sent by A, the higher A's payoff. B's payoff depends on his or her own decision whether to follow the message or not and in case of following whether the number sent by A corresponds to the true state s:

$$\pi_B = \begin{cases} 100 & \text{if B follows and } r = s \\ 0 & \text{if B follows and } r \neq s \\ 30 & \text{if B does not follow} \end{cases}$$

Thus, it pays for B to follow if A has sent the true message, otherwise it pays not to follow. Our focus lies on the decisions made by players A since these measure the degree of honest behavior. Therefore, we assign all participants to the role of player A. We elicit decisions with the strategy method to get the full profile of player A's honesty, i.e. for every possible state participants decide which message they would send to player B (resulting in six decisions in total).[16]

Lying game treatments

We randomly assign participants to one of four possible treatments in a between-subject design. The treatment variation is based on the attributes of player B. Before participants (all in the role of player A) read the instructions of the game, they receive the description of their assigned player B. They are either playing (1) with a non-family manager whose degree of religiosity is unknown, or (2) a religious non-family manager, or (3) a family manager whose degree of religiosity is unknown or (4) a religious family manager. While our main interest lies in the influence of religiosity, we use the dimension of unknown religiosity as a benchmark level.[17] The comparison between these two dimensions measures the treatment effect of religiosity. Table 4.1 summarizes the 2 x 2 design.

[16] The instructions of the lying game as used in our experiment can be found in Appendix B.

[17] To keep the length of the descriptions as equal as possible, we describe the non-family manager and the family manager whose degree of religiosity is unknown as being athletic. This characteristic serves as a neutral reference that does not allow drawing any conclusion about the degree of religiosity.

TABLE 4.1: Treatments in the lying game

Player B	unknown religiosity	known religiosity
non-family manager	(1) non-family manager unknown	(2) non-family manager religious
family manager	(3) family manager unknown	(4) family manager religious

Experimental procedures

We ran the online experiment on Clickworker with the same participants that completed our survey. After excluding participants who did not fully understand the game, our final sample for the analysis consisted of 458 participants. In the online experiment, participants played a hypothetical version of the lying game. They learned that they would not gain additional money and that they would play the game with another player, which we then described in detail according to the treatment variations. The online experiment concluded with a short questionnaire. Each participant received a fixed payment of € 2.30.

To test the robustness of the lying game in our online experiment, we ran a similar version of the experiment with students of the University of Graz at the Max-Jung-laboratory for economic experiments.[18] The experiment was paper-pencil based and consisted of 8 sessions, with one session having on average 19-20 participants. We recruited 157 subjects using ORSEE (Greiner 2015)[19]. We excluded seven participants from the data analysis because they failed the comprehension check. In contrast to the online experiment, we ran an incentivized version of the lying game and thus paid participants according to their decisions. One randomly chosen participant per session was paid based on her actual decision and the decision of a real (non-)family manager. These managers took part in the experiment beforehand and actually had the attributes as described in the respective treatment.[20] Average payment from the lying game amounted to € 11.70, including a show-up fee of € 6.00.

Lying game measures

Independent variable: In the lying game, participants play with one of four possible manager types in a between-subject design, i.e. we randomly assign them to one of four possible treatments. In the analysis, we use the dummy variable *Family Manager* (FM)

[18] maxjunglab.uni-graz.at

[19] ORSEE (Online Recruitment System for Economic Experiments) is a software to recruit subjects for economic laboratory experiments.

[20] We contacted companies and recruited eight persons who fulfilled the requirements (either being a non-family manager or a family manager together with being either athletic or religious) to participate in our experiment. Thereby, we could match each manager with one of our 8 sessions.

that takes the value 1 if player B is a FM and 0 if he or she is a non-family manager (NFM) as well as a vector consisting of treatment dummy variables indicating each of the four different player B types (NFM unknown, NFM religious, FM unknown, FM religious).

Dependent variables: In the lying game, participants decide for every possible state s which message r they want to send to player B. To measure the degree of honesty, we count the number of true messages that are sent in states 1–5.[21] A message is true if the number sent to player B is identical to the assigned state. This variable is at minimum zero and at maximum five with high numbers indicating more honesty compared to low numbers.

Control variables: In the post-experimental questionnaire, we elicit how participants perceive player B's level of religiosity on a 5-point Likert scale ranging from "very religious" (coded as 5) to "not religious at all" (coded as 1). We use the coded answer as outcome variable. The same outcome variable is used to measure participants' self-assessed own level of religiosity. The questions on socio-demographics yield participants' age (in years) and gender (as dummy variable with males coded as 1 and females as 0) as outcome variables.

Lying game manipulation and comprehension checks

To test whether the manipulation of religiosity in the different treatments of the lying game works, we let participants assess how they perceive the (non-)family manager's level of religiosity in the post experimental questionnaire. As expected, participants in the two religious treatments perceived the manager as more religious (M=3.63, SD=1.23) than participants in the treatments in which the manager's degree of religiosity was unknown (M=2.37, SD=0.91, $p<0.001$). Thus, the manipulation of religiosity was successful.

To check whether participants understand the rules of the lying game correctly, they have to answer two multiple-choice questions. All participants finally got the questions right (those participants who did not answer the questions correctly at the first attempt received some hints that allowed them to get the right answer afterwards).

State 6 in the lying game is an additional comprehension check, as sending another number than the true number 6 does not make any sense from an economic point

[21] Since lying only makes sense when it is economically beneficial, there is no reason to report a different number than 6 when the assigned state is actually 6. We therefore only counted the number of true messages for states 1–5.

of view.[22] Out of 600 participants in the online experiment (157 in the incentivized laboratory experiment), 152 (7) did not choose the true number in state 6. Since we cannot assume that these individuals completely understood the nature of the game, we excluded them from our further analysis.

4.4 Results

4.4.1 Perceived religiosity

The first objective was to test whether family managers (FM) are perceived as more religious than non-family managers (NFM). As each participant assessed the degree of religiosity of both the FM and NFM, our comparisons are within treatment. We therefore used a paired t-test for empirically testing our first hypothesis.[23] The results show that FMs are perceived as significantly more religious (M=1.86, SD=0.84) than NFMs (M=1.34, SD=0.84; $p < 0.001$). Thus, hypothesis 1 is supported.[24]

4.4.2 Honesty toward family and non-family managers

In this section, we examine the number of true messages sent in the lying game across treatments. As our comparisons are between treatments, we use t-tests for independent groups. Table 4.2 presents means and standard deviations.

[22] Sending a lower number than 6 when the true state is 6 means that own earnings will be lower than with truthful reporting and the other player will receive 0 if he or she follows the message because it was a lie.

[23] As all our hypotheses are directional, we use one-sided t-tests for our analysis.

[24] We further checked whether the order of presentation of the description of the FM and the NFM has an effect on our findings. Our results reveal that there are no significant order effects.

TABLE 4.2: Descriptive statistics of the number of true messages

	Number of true messages	
	Mean	SD
Panel A		
NFM (n=227)	3.23	1.94
FM (n=221)	3.58	1.85
Panel B		
NFM unknown (n=112)	3.09	1.90
NFM religious (n=115)	3.37	1.97
FM unknown (n=113)	3.42	1.89
FM religious (n=108)	3.74	1.79

Notes: The table shows means and standard deviations for the number of true messages received by non-family managers (NFM) and family mangers (FM) in the online experiment.

Panel A of Table 4.2 indicates that participants with a FM as a counterpart report significantly more true messages than individuals playing with an NFM (p=0.026). This result is consistent with Hypothesis 2a.

As we proposed in Hypothesis 2b, the reputation of being religious might drive the family manager effect. Thus, we examine whether the combination of family firm and religiosity reputation influences participants' behavior in the lying game. Descriptive statistics for the different treatment groups are presented in Panel B of Table 4.2. Religious FMs receive on average the highest number of true messages, whereas NFMs without any religiosity reputation obtain on average the true message least frequently. A t-test shows that participants send significantly more true messages to religious FMs than to FMs and NFMs whose religiosity is unknown (p=0.0695, or, p=0.005, respectively). All other group comparisons are not significant. Furthermore, when pooling the two NFM groups, we find a significant difference in the number of true messages between religious FM and NFM (p=0.011), but not between FM whose religiosity is unknown and NFM.

The question of whether participants are more honest toward FMs than toward NFMs is also addressed in Table 4.3.

TABLE 4.3: The effect of family firm and religiosity reputation on the number of true messages

	1	2	3	4	5	6	7	8
FM	0.350*	0.350*						
	(0.179)	(0.179)						
FM religious			0.512**	0.511**			0.651***	0.621**
			(0.215)	(0.215)			(0.249)	(0.250)
FM unknown					0.196	0.206	0.335	0.333
					(0.219)	(0.218)	(0.253)	(0.252)
NFM religious							0.276	0.245
							(0.257)	(0.261)
Male		-0.232		-0.061		-0.164		-0.218
		(0.179)		(0.209)		(0.209)		(0.179)
Age		-0.006		0.002		-0.009		-0.005
		(0.008)		(0.009)		(0.009)		(0.008)
Own religiosity		0.105		0.126		0.045		0.099
		(0.072)		(0.081)		(0.084)		(0.072)
Constant	3.229***	3.312***	3.229***	2.920***	3.229***	3.519***	3.089***	3.181***
	(0.129)	(0.354)	(0.129)	(0.395)	(0.129)	(0.395)	(0.180)	(0.374)
N	448	448	335	335	340	340	448	448
R2	0.009	0.018	0.016	0.023	0.002	0.008	0.015	0.023

Notes: The table shows estimates from an OLS regression model. Robust standard errors are in parentheses. Significance level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We apply an OLS regression model with robust standard errors.[25] In specification 1, we regress the number of true messages on the dummy variable FM which takes the value 1 if player B is a FM and 0 if he or she is a NFM. The results of specification 1 confirm the findings of the descriptive analysis and show that participants send more true messages to FMs than to NFMs. The question of whether stakeholders are more honest to religious FMs is addressed in specifications 3–8. In specification 3, we examine whether religious FMs receive more true messages than NFMs, whereas in specification 5, we investigate whether participants also report more truthful messages to FM whose religiosity is unknown. While there are no significant results regarding FMs without any religiosity reputation, our results show that the coefficient of the dummy variable FM religious, which is 1 for religious FMs and 0 for NFMs, is positive and significant. In other words, this finding indicates that family firm in combination with religiosity reputation has a significant positive effect on truthful reporting. In specification 7, we regress the number of true messages on a vector consisting of treatment dummy variables indicating each of the different player B types (NFM unknown, NFM religious, FM unknown, FM religious). The group "NFM unknown" is omitted from the model and serves as the reference category. Results show that participants send true messages significantly more often to religious FMs than to NFMs whose religiosity is unknown. According to Wald

[25] Following Gneezy et al. (2013), we also estimate the likelihood of sending the true number by means of a probit regression (see Appendix C). The results remain qualitatively the same as those obtained from the OLS regression.

tests, the differences between other groups are not significant. In specification 2, 4, 6, and 8, we include different control variables. In particular, we control for demographics like gender and age, and a self-assessment of one's own religiosity. Adding the set of control variables further supports our findings as results show that neither of these variables alter our main effects. Note that the low R-squared are common for behavioral experimental studies in economics (e.g., see recent papers by Buser and Dreber 2015, Cohn et al. 2017, Garofalo and Rott 2017 and papers on lying and religiosity, e.g. by Conrads et al. 2013, Tan and Vogel 2008) and in RCTs (randomized controlled trials) in general. The reason is that due to the randomization, covariates are orthogonal to the treatment indicator.

4.4.3 Robustness test

We ran several robustness checks to our data. First, besides asking our participants about their religiosity evaluation of family firms, we integrated questions regarding their assessment of other attributes such as, for example, trustworthiness, social responsibility, innovativeness and selfishness. Literature has shown that these attributes significantly separate family from non-family firms (for an overview see Sageder et al. 2018). Table 4.4 presents the means and standard deviations for FMs and NFMs.

TABLE 4.4: Descriptive statistics of family managers' attributes

	Family manager (n=600)		Non-family manager (n=600)		p-values (paired t-tests)
	Mean	SD	Mean	SD	
Trustworthiness	2.90	0.78	2.40	0.85	<0.001
Social responsibility	2.79	0.83	2.28	0.92	<0.001
Innovativeness	2.36	0.89	2.61	0.89	<0.001
Selfishness	2.29	0.92	2.53	0.39	<0.001

Notes: The table shows means and standard deviations of (non-)family managers' attributes. The last column reports p-values of paired t-tests.

In line with previous studies, we find that FMs are perceived as significantly more trustworthy and socially responsible and significantly less innovative and selfish than NFMs. As our participants report similar evaluations in previous studies, they can be viewed as a reliable sample.

Second, we further conducted an incentivized laboratory experiment, as behavioral economists usually propose that only incentivized experiments are reliable. 150 individuals participated in the incentivized lying game, as presented in the methodology section. Table 4.5 reports the descriptive statistics for both the non-incentivized online experiment and the incentivized laboratory experiment.

TABLE 4.5: Descriptive statistics of number of true messages for the online and the incentivized laboratory experiment

	Non-incentivized online experiment		Incentivized laboratory experiment		p-values (unpaired t-tests)
	Mean	SD	Mean	SD	
NFM (n=227/n=74)	3.23	1.94	3.07	2.01	0.538
FM (n=221/n=76)	3.58	1.85	3.57	1.91	0.957
NFM unknown (n=112/n=37)	3.09	1.90	2.86	2.00	0.540
NFM religious (n=115/n=37)	3.37	1.97	3.27	2.02	0.800
FM unknown (n=113/n=39)	3.42	1.89	3.44	1.91	0.975
FM religious (n=108/n=37)	3.74	1.79	3.70	1.91	0.913

Notes: The table shows means and standard deviations of (non-)family managers' attributes in the non-incentivized online experiment and the incentivized laboratory experiment. The last column report p-values of unpaired t-tests.

As shown in Table 4.5, there are no significant mean differences between both samples, indicating that the behavior in both experiments is comparable. Focusing on the incentivized laboratory experiment, a t-test for independent groups indicates that participants send more truthful messages to FMs than to NFMs (p=0.061). This result is consistent with the findings of the online experiment. In addition, and again similar to the online experiment, a t-test for independent groups reveals significant differences in the number of true messages received by religious FMs compared to the NFMs whose religiosity is unknown (p=0.035). Finally, we can also show that the mean comparison between religious FMs and NFMs is significant for the sample from the incentivized laboratory experiment (p=0.057).

Third, to further assess the robustness of our results we pooled the sample of the online experiment with the sample of the incentivized laboratory experiment. Table 4.6 shows the means and standard deviations for the pooled sample.

TABLE 4.6: Descriptive statistics of the number of true messages for the pooled sample

	Number of true messages pooled sample	
	Mean	SD
NFM (n=301)	3.19	1.95
FM (n=297)	3.58	1.86
NFM unknown (n=149)	3.03	1.92
FM religious (n=152)	3.34	1.98
FM unknown (n=152)	3.43	1.89
FM religious (n=145)	3.73	1.82

Notes: The table shows means and standard deviations for the number of true messages received by non-family managers (NFM) and family mangers (FM) for the pooled sample.

In line with our previous results, a t-test indicates that participants send more truthful messages to FMs than to NFMs (p=0.007). Furthermore, our analysis shows that participants report more truthful messages to both types of FMs than to NFMs whose religiosity is unknown (p=0.001 for religious FM and p=0.037 for FM whose religiosity is unknown). However, the effect size is greater for religious FMs. Furthermore, we find that compared with religious NFMs, religious FMs receive significantly more true messages (p=0.040). Finally, results show that religious FMs (NFMs) receive more true messages than FMs (NFMs) without any religiosity reputation (FM religious vs FM unknown: p=0.080; NFM religious vs NFM unknown: p=0.086).

4.5 Discussion

Our study shows that stakeholders perceive family managers as more religious compared to non-family managers. This finding is important to further complete the list of attributes that are associated with family firms as a distinct group. Second, we can show that stakeholders behave more honestly, when the interacting party is a family manager in comparison to a non-family manager. This result supports evidence on the positive effect of family firm reputation. Our third finding shows that the increase in honest behavior by external parties toward family managers is driven by the fact that this family manager is presented as religious. We can thus show that the combination of family firm and religiosity reputation results in a positive reaction by external parties in the sense that these behave more honestly when interacting with a religious family manager.

Our study has some limitations that we want to carefully discuss and evaluate in the following. One limitation that is generally linked to the use of economic experiments is the issue of external validity. While experiments have the big advantage of having control over the decision environment and thereby allowing causal inferences, applying the so gained findings to the real world is sometimes difficult. The lying game might be able to represent a very general relationship between a family manager and a stakeholder, but it does not explicitly frame a buying process or the execution of a contract. However, due to the high internal validity, investigating the lying game is an important step before investigating a more realistic and complex scenario, like for example a field setting with real stakeholders. We can also call into question if participants of a laboratory experiment or a study on Clickworker are representative for our target group. To address this issue, it would be interesting to invite real stakeholders to participate in our study. Another important issue is the time factor: In an experiment, interactions are of limited duration and in our setting only occurred once. In a business setting, relationships are more long-lasting and occur repeatedly, thereby allowing strategic considerations to play an important role. It might be of interest to run a study with more than one

round or to collect data at several points in time. A potential drawback unrelated to the method used is that in our study we let participants interact with a person but not with an organization as a whole. Hauswald and Hack (2013) however show that category beliefs lose importance if stakeholders do not interact with individuals (representing an organization) but with the (abstract) organization itself. This could mean that the positive reputational effect of religiosity vanishes when individuals are replaced by firms. While the choice of individuals instead of organizations as interacting parties was necessary for our experimental approach, it would be interesting to run other types of studies with organizations as interacting partners and not individual representatives of a firm. Another restriction is that we limited our data collection to German speaking countries.

Beyond these limitations, there are additional avenues for further research on family firm reputation. First of all, while we have shown a positive reputation effect of religiosity, this effect might, to a large degree, depend on the general appreciation of religion and religiosity in a given region (e.g., rural versus urban regions, collective versus individualistic societies), for a given religion (e.g. Christianity, Islam, Hinduism, Judaism), or specific groups of people (e.g., gender, education, personality). Thus, our study should be replicated in different geographical, cultural, and group settings. This is also important, as culture plays an important role with respect to fundamental behaviors like trust (e.g. Henrich et al. 2001). Second, while we have focused on the influence of family firm and religiosity reputation on stakeholder behavior, there might be another influence worth considering. For family members it is often natural to identify with the family firm (Sharma et al. 2012). According to social identity theory (Tajfel and Turner 1986), group members have the tendency to achieve and maintain positive concepts of themselves and the group. Following Miller and Le Breton-Miller (2003), who suggest that when good deeds are ascribed to family members, a virtuous circle may establish. It might therefore well be that positive stakeholder reactions to family firm status and religiosity enforce family members' family firm and religiosity identity. This strengthened identity might subsequently invigorate family firm and religiosity reputation and again trigger honest stakeholder reactions. Researching these cyclical influences might be a fruitful direction to enrich (family firm) identity theory. Third, our study sheds light on a generalized reaction (i.e., honesty) toward family managers who are perceived as religious. It might be fruitful to conduct empirical studies in more specific contexts like consumers' harmful word-of-mouth behavior, employees' retaliation behavior, or suppliers' product quality violations.

Our study has important implications for family managers. They should not only inform their stakeholders about their family firm nature, as suggested by prior research (Binz et al. 2013, Lude and Prügl 2019), but in addition should highlight their personal religiosity.

Given the prevailing general perception that family managers are more religious and the similarity of the cues of family firms and religiosity, both cues can confirm and reinforce each other. There are two positive effects that such an informational campaign can produce. First, there is the general reputational effect of religiosity. As religiosity is associated with positive attributes, it can further strengthen the positive reputation of family firms. This, in turn, can result in positive outcomes like higher consumer happiness and better recruiting opportunities. Second, and more specific to our study, the moral reminder effect of religiosity can induce stakeholders to behave more honestly. Unethical behavior is a big challenge for firms and society in general today, and interacting with honest and ethical stakeholders is of essential importance for firms. Given that all types of stakeholders regularly face ethical decisions (e.g., employee theft, supplier quality, honest customer word-of-mouth), family managers who highlight their religiosity can achieve a competitive advantage when dealing with stakeholders in these situations.

References

Akerlof, G. A. (1978). The market for "lemons": Quality uncertainty and the market mechanism. In *Uncertainty in Economics*, pages 235–251. Elsevier.

Allport, G. W. (1954). *The nature of prejudice*. Addison-Wesley.

Anderson, C. A., Lepper, M. R., and Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology*, 39(6):1037–1049.

Ashmore, R. D. and Del Boca, F. K. (1981). Conceptual approaches to stereotypes and stereotyping. In *Cognitive Processes in Stereotyping and Intergroup Behavior*, pages 1–35. Psychology Press New York.

Ashton, M. C. and Lee, K. (2014). Personality and Religiousness. In *Religion, Personality, and Social Behavior*, pages 31–46. Psychology Press New York.

Astrachan, J. H., Klein, S. B., and Smyrnios, K. X. (2002). The F-PEC scale of family influence: A proposal for solving the family business definition problem. *Family Business Review*, 15(1):45–58.

Bailey, R. C. and Doriot, P. D. (1985). Perceptions of professionals who express religious beliefs. *Social Behavior and Personality: An International Journal*, 13(2):167–170.

Bailey, R. C. and Young, M. D. (1986). The value and vulnerability of perceived religious involvement. *The Journal of Social Psychology*, 126(5):693–694.

Beck, S. and Kenning, P. (2015). The influence of retailers' family firm image on new product acceptance: An empirical investigation in the German FMCG market. *International Journal of Retail & Distribution Management*, 43(12):1126–1143.

Beck, S. and Prügl, R. (2018). Family firm reputation and humanization: Consumers and the trust advantage of family firms under different conditions of brand familiarity. *Family Business Review*, 31(4):460–482.

Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, 10(1):122–142.

Berrone, P., Cruz, C., Gomez-Mejia, L. R., and Larraza-Kintana, M. (2010). Socioemotional wealth and corporate responses to institutional pressures: Do family-controlled firms pollute less? *Administrative Science Quarterly*, 55(1):82–113.

Bikhchandani, S., Hirshleifer, D., and Welch, I. (1998). Learning from the behavior of others: Conformity, fads, and informational cascades. *Journal of Economic Perspectives*, 12(3):151–170.

Binz, C., Hair Jr, J. F., Pieper, T. M., and Baldauf, A. (2013). Exploring the effect of distinct family firm reputation on consumers' preferences. *Journal of Family Business Strategy*, 4(1):3–11.

Binz Astrachan, C., Botero, I., Astrachan, J. H., and Prügl, R. (2018). Branding the family firm: A review, integrative framework proposal, and research agenda. *Journal of Family Business Strategy*, 9(1):3–15.

Block, J. (2010). Family management, family ownership, and downsizing: Evidence from S&P 500 firms. *Family Business Review*, 23(2):109–130.

Block, J., Stiglbauer, M., Kühn, A.-L., and Wagner, D. (2015). Corporate social responsibility communication of German family firms: A content analysis. *UWF UmweltWirtschaftsForum*, 23(4):251–257.

Block, J. and Wagner, M. (2014). Ownership versus management effects on corporate social responsibility concerns in large family and founder firms. *Journal of Family Business Strategy*, 5(4):339–346.

Blodgett, M. S., Dumas, C., and Zanzi, A. (2011). Emerging trends in global ethics: A comparative study of us and international family business values. *Journal of Business Ethics*, 99(1):29–38.

Bloodgood, J. M., Turnley, W. H., and Mudrack, P. (2008). The influence of ethics instruction, religiosity, and intelligence on cheating behavior. *Journal of Business Ethics*, 82(3):557–571.

Brewer, M. B. and Brown, R. J. (1998). Intergroup relations. In *The Handbook of Social Psychology*, pages 554–594. McGraw-Hill.

Brief, A. P., Dietz, J., Cohen, R. R., Pugh, S. D., and Vaslow, J. B. (2000). Just doing business: Modern racism and obedience to authority as explanations for employment discrimination. *Organizational Behavior and Human Decision Processes*, 81(1):72–97.

Brigham, J. C. (1971). Ethnic stereotypes. *Psychological Bulletin*, 76(1):15–38.

Brown, T. J., Dacin, P. A., Pratt, M. G., and Whetten, D. A. (2006). Identity, intended image, construed image, and reputation: An interdisciplinary framework and suggested terminology. *Journal of the Academy of Marketing Science*, 34(2):99–106.

Buser, T. and Dreber, A. (2015). The flipside of comparative payment schemes. *Management Science*, 62(9):2626–2638.

Cialdini, R. B. (1993). *Influence: Science and practice*. HarperCollins College Publishers.

Clifford, S. and Gaskins, B. (2016). Trust me, I believe in God: Candidate religiousness as a signal of trustworthiness. *American Politics Research*, 44(6):1066–1097.

Cohn, A., Fehr, E., and Maréchal, M. A. (2017). Do professional norms in the banking industry favor risk-taking? *The Review of Financial Studies*, 30(11):3801–3823.

Conrads, J., Irlenbusch, B., Rilke, R. M., and Walkowitz, G. (2013). Lying and team incentives. *Journal of Economic Psychology*, 34:1–7.

Crocker, J. and Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review*, 96(4):608–630.

Darley, J. M. and Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, 44(1):20–33.

Davies, G., Chun, R., da Silva, R. V., and Roper, S. (2004). A corporate character scale to assess employee and customer views of organization reputation. *Corporate Reputation Review*, 7(2):125–146.

Dobson, K. and Knezevic, I. (2018). "Ain't Nobody Got Time for That!": Framing and Stereotyping in Legacy and Social Media. *Canadian Journal of Communication*, 43(3):381–397.

Dovidio, J. F. (2001). On the nature of contemporary prejudice: The third wave. *Journal of Social Issues*, 57(4):829–849.

Duckitt, J. H. (1992). *The Social Psychology of Prejudice.* Praeger Publishers/Greenwood Publishing Group.

Dyer, W. G. and Whetten, D. A. (2006). Family firms and social responsibility: Preliminary evidence from the S&P 500. *Entrepreneurship Theory and Practice*, 30(6):785–802.

Ehrlich, H. J. (1973). *The Social Psychology of Prejudice: A Systematic Theoretical Review and Propositional Inventory of the American Social Psychological Study of Prejudice.* John Wiley & Sons.

English, H. B. and English, A. C. (1958). *A Comprehensive Dictionary of Psychological and Psychoanalytical Terms: A Guide to Usage.* Longmans Green.

Erber, R. and Fiske, S. T. (1984). Outcome dependency and attention to inconsistent information. *Journal of Personality and Social Psychology*, 47(4):709–726.

Evert, R. E., Martin, J. A., McLeod, M. S., and Payne, G. T. (2016). Empirics in family business research: Progress, challenges, and the path ahead. *Family Business Review*, 29(1):17–43.

Falk, A. (2007). Gift exchange in the field. *Econometrica*, 75(5):1501–1511.

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

Fang, H., Randolph, R. V. d. G., Chrisman, J. J., and Barnett, T. (2013). Firm religiosity, bounded stakeholder salience, and stakeholder relationships in family firms. *Journal of Management, Spirituality & Religion*, 10(3):253–270.

Fehr, E. and Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives*, 14(3):159–181.

Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of Personality and Social Psychology*, 38(6):889–906.

Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In *The Handbook of Social Psychology*, pages 357–411. McGraw-Hill.

Fiske, S. T., Cuddy, A. J., Glick, P., and Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82(6):878–902.

Fiske, S. T. and Neuberg, S. L. (1989). Category-based and individuating processes as a function of information and motivation: Evidence from our laboratory. In *Stereotyping and Prejudice*, pages 83–103. Springer.

Fiske, S. T. and Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In *Advances in Experimental Social Psychology*, pages 1–74. Elsevier.

Fombrun, C. and Shanley, M. (1990). What's in a name? Reputation building and corporate strategy. *Academy of Management Journal*, 33(2):233–258.

Ford, T. E. and Stangor, C. (1992). The role of diagnosticity in stereotype formation: Perceiving group means and variances. *Journal of Personality and Social Psychology*, 63(3):356–367.

Gao, Y. L. and Mattila, A. S. (2017). The impact of stereotyping on consumers' food choices. *Journal of Business Research*, 81:80–85.

Garofalo, O. and Rott, C. (2017). Shifting blame? Experimental evidence of delegating communication. *Management Science*, 64(8):3911–3925.

Geyer, A. L. and Baumeister, R. F. (2005). *Religion, Morality, and Self-Control: Values, Virtues, and Vices.* Guilford Press.

Gilbert, R. J. and Lieberman, M. (1987). Investment and coordination in oligopolistic industries. *The Rand Journal of Economics*, 18(1):17–33.

Gino, F. (2015). Understanding ordinary unethical behavior: Why people who value morality act immorally. *Current Opinion in Behavioral Sciences*, 3:107–111.

Glock, C. Y., Wuthnow, R., Piliavin, J., and Spencer, M. (1975). *Adolescent Prejudice.* Harper & Row.

Gneezy, U., Rockenbach, B., and Serra-Garcia, M. (2013). Measuring lying aversion. *Journal of Economic Behavior & Organization*, 93:293–300.

Graafland, J., Kaptein, M., and Mazereeuw-van der Duijn Schouten, C. (2007). Conceptions of god, normative convictions, and socially responsible business conduct: An explorative study among executives. *Business & Society*, 46(3):331–368.

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

Güth, W., Schmittberger, R., and Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, 3(4):367–388.

Hansen, C. H. and Hansen, R. D. (1988). Finding the face in the crowd: An anger superiority effect. *Journal of Personality and Social Psychology*, 54(6):917–924.

Hatak, I. R. and Roessl, D. (2015). Relational competence-based knowledge transfer within intrafamily succession: An experimental study. *Family Business Review*, 28(1):10–25.

Hauswald, H. and Hack, A. (2013). Impact of family control/influence on stakeholders' perceptions of benevolence. *Family Business Review*, 26(4):356–373.

Hauswald, H., Hack, A., Kellermanns, F. W., and Patzelt, H. (2016). Attracting new talent to family firms: Who is attracted and under what conditions? *Entrepreneurship Theory and Practice*, 40(5):963–989.

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.

Higgins, E. T., King, G. A., and Mavin, G. H. (1982). Individual construct accessibility and subjective impressions and recall. *Journal of Personality and Social Psychology*, 43(1):35–47.

Hilton, J. L. and Von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology*, 47(1):237–271.

Hout, M. and Fischer, C. S. (2002). Why more Americans have no religious preference: Politics and generations. *American Sociological Review*, 67(2):165–190.

Huber, S. and Huber, O. W. (2012). The centrality of religiosity scale (CRS). *Religions*, 3(3):710–724.

Isaac, S. V., Bailey, R. C., and Isaac, W. L. (1995). Perceptions of religious and nonreligious targets who participate in premarital sex. *Social Behavior and Personality: An International Journal*, 23(3):229–233.

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.

Judd, C. M. and Park, B. (1993). Definition and assessment of accuracy in social stereotypes. *Psychological Review*, 100(1):109–128.

Kagel, J. H. and Roth, A. E. (2016). *The Handbook of Experimental Economics*, volume 2. Princeton University Press.

Kahan, D. (1997). Social influence, social meaning, and deterrence. *Virginia Law Review*, 83(2):349–395.

Kashmiri, S. and Mahajan, V. (2010). What's in a name? An analysis of the strategic behavior of family firms. *International Journal of Research in Marketing*, 27(3):271–280.

Kennedy, R. E. (2002). Strategy fads and competitive convergence: An empirical test for herd behavior in prime-time television programming. *The Journal of Industrial Economics*, 50(1):57–84.

Keplinger, K. and Feldbauer-Durstmüller, B. (2012). Representations of family businesses in Russian popular culture. *European Journal of Management*, 12(1):1–10.

Kirchmaier, I., Prüfer, J., and Trautmann, S. T. (2018). Religion, moral attitudes and economic behavior. *Journal of Economic Behavior & Organization*, 148:282–300.

König, L. B. (2015). Change and stability in religiousness and spirituality in emerging adulthood. *The Journal of Genetic Psychology*, 176(6):369–385.

Krappe, A., Goutas, L., and von Schlippe, A. (2011). The "family business brand": An enquiry into the construction of the image of family businesses. *Journal of Family Business Management*, 1(1):37–46.

Lippmann, W. (2017). *Public Opinion*. Routledge.

Lord, C. G., Ross, L., and Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37(11):2098–2109.

Lude, M. and Prügl, R. (2019). Risky decisions and the family firm bias: An experimental study based on prospect theory. *Entrepreneurship Theory and Practice*, 43(2):386–408.

Macrae, C. N., Milne, A. B., and Bodenhausen, G. V. (1994). Stereotypes as energy-saving devices: A peek inside the cognitive toolbox. *Journal of Personality and Social Psychology*, 66(1):37–47.

Mastro, D. (2015). Why the media's role in issues of race and ethnicity should be in the spotlight. *Journal of Social Issues*, 71(1):1–16.

Mazar, N., Amir, O., and Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, 45(6):633–644.

McCauley, C., Stitt, C. L., and Segal, M. (1980). Stereotyping: From prejudice to prediction. *Psychological Bulletin*, 87(1):195.

McGuire, W. J., McGuire, C. V., Child, P., and Fujioka, T. (1978). Salience of ethnicity in the spontaneous self-concept as a function of one's ethnic distinctiveness in the social environment. *Journal of Personality and Social Psychology*, 36(5):511–520.

Medin, D. L., Goldstone, R. L., and Gentner, D. (1993). Respects for similarity. *Psychological Review*, 100(2):254–278.

Micelotta, E. R. and Raynard, M. (2011). Concealing or revealing the family? Corporate brand identity strategies in family firms. *Family Business Review*, 24(3):197–216.

Miller, D. and Le Breton-Miller, I. (2003). Challenge versus advantage in family business. *Strategic Organization*, 1(1):127–134.

Mitchell, R. K., Robinson, R. E., Marin, A., Lee, J. H., and Randolph, A. F. (2013). Spiritual identity, stakeholder attributes, and family business workplace spirituality stakeholder salience. *Journal of Management, Spirituality & Religion*, 10(3):215–252.

Norenzayan, A. and Shariff, A. F. (2008). The origin and evolution of religious prosociality. *Science*, 322(5898):58–62.

Orth, U. R. and Green, M. T. (2009). Consumer loyalty to family versus non-family business: The roles of store image, trust and satisfaction. *Journal of Retailing and Consumer Services*, 16(4):248–259.

Othman, S., Darus, F., and Arshad, R. (2011). The influence of coercive isomorphism on corporate social responsibility reporting and reputation. *Social Responsibility Journal*, 7(1):119–135.

Paterson, T. A., Specht, D., and Duchon, D. (2013). Exploring costs and consequences of religious expression in family businesses. *Journal of Management, Spirituality & Religion*, 10(2):138–158.

Ramasamy, B., Yeung, M. C., and Au, A. K. (2010). Consumer support for corporate social responsibility (CSR): The role of religion and values. *Journal of Business Ethics*, 91(1):61–72.

Rau, S. B., Astrachan, J. H., and Smyrnios, K. X. (2018). The F-PEC revisited: From the family business definition dilemma to foundation of theory. *Family Business Review*, 31(2):200–213.

Rindova, V. P., Williamson, I. O., Petkova, A. P., and Sever, J. M. (2005). Being good or being known: An empirical examination of the dimensions, antecedents, and consequences of organizational reputation. *Academy of Management Journal*, 48(6):1033–1049.

Roberts, P. W. and Dowling, G. R. (2002). Corporate reputation and sustained superior financial performance. *Strategic Management Journal*, 23(12):1077–1093.

Ross, L., Lepper, M. R., and Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology*, 32(5):880–892.

Sageder, M., Mitter, C., and Feldbauer-Durstmüller, B. (2018). Image and reputation of family firms: A systematic literature review of the state of research. *Review of Managerial Science*, 12(1):335–377.

Saroglou, V., Delpierre, V., and Dernelle, R. (2004). Values and religiosity: A meta-analysis of studies using Schwartz's model. *Personality and Individual Differences*, 37(4):721–734.

Schaller, M., Conway III, L. G., and Tanchuk, T. L. (2002). Selective pressures on the once and future contents of ethnic stereotypes: Effects of the communicability of traits. *Journal of Personality and Social Psychology*, 82(6):861–877.

Schellong, M., Kraiczy, N. D., Malär, L., and Hack, A. (2018). Family firm brands, perceptions of doing good, and consumer happiness. *Entrepreneurship Theory and Practice*, https://doi.org/10.1177/1042258717754202.

Schwartz, S. H. and Huismans, S. (1995). Value priorities and religiosity in four western religions. *Social Psychology Quarterly*, 58(2):88–107.

Shafffi, E. B., Smith, E. E., and Osherson, D. N. (1990). Typicality and reasoning fallacies. *Memory & Cognition*, 18(3):229–239.

Sharma, P., Chrisman, J. J., and Gersick, K. E. (2012). 25 years of family business review: Reflections on the past and perspectives for the future. *Family Business Review*, 25(1):5–15.

Sherman, J. W., Allen, T. J., and Sacchi, D. L. (2012). *Stereotype Confirmation and Disconfirmation.* Guilford Press.

Stavrova, O. and Siegers, P. (2014). Religious prosociality and morality across cultures: How social enforcement of religion shapes the effects of personal religiosity on prosocial and moral attitudes and behaviors. *Personality and Social Psychology Bulletin*, 40(3):315–333.

Stigler, G. J. (1962). Information in the labor market. *Journal of Political Economy*, 70(5):94–105.

Südostschweiz (2014). Victorinox-CEO Carl Elsener plädiert für christliche Werte im Wirtschaftsalltag. *https://www.suedostschweiz.ch/zeitung/victorinox-ceo-carl-elsener-plaediert-fuer-christliche-werte-im-wirtschaftsalltag.*

Tajfel, H. (1969). Cognitive aspects of prejudice. *Journal of Biosocial Science*, 1(S1):173–191.

Tajfel, H. (1981). *Human Groups and Social Categories: Studies in Social Psychology.* Cambridge University Press.

Tajfel, H. and Turner, J. (1986). *The Social Identity Theory of Intergroup Behavior.* Chicago Nelson-Hall.

Tan, J. H. and Vogel, C. (2008). Religion and trust: An experimental study. *Journal of Economic Psychology*, 29(6):832–848.

Turban, D. B. and Cable, D. M. (2003). Firm reputation and applicant pool characteristics. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, 24(6):733–751.

Tversky, A. and Kahneman, D. (1981). *Judgments of and by representativeness*. Stanford University.

Vazquez, P. (2018). Family business ethics: At the crossroads of business ethics and family business. *Journal of Business Ethics*, 150(3):691–709.

Walsh, G. and Beatty, S. E. (2007). Customer-based corporate reputation of a service firm: Scale development and validation. *Journal of the Academy of Marketing Science*, 35(1):127–143.

Walsh, G., Mitchell, V.-W., Jackson, P. R., and Beatty, S. E. (2009). Examining the antecedents and consequences of corporate reputation: A customer perspective. *British Journal of Management*, 20(2):187–203.

Wills, T. A. (1981). Downward comparison principles in social psychology. *Psychological Bulletin*, 90(2):245–271.

Zaller, J. R. (1992). *The nature and origins of mass opinion*. Cambridge University Press.

Appendix A: Descriptions non-family and family manager

Description of the family manager (non-family manager) in the survey:

Imagine a family manager (manager) holding a management position in his own company (in a company). The majority of the company's shares and voting rights of the company have been in the family's possession since the previous generation (have been held in free float by various private and institutional investors for decades.)

Description of the family and non-family manager according to the treatments in the lying game:

Religious family manager (family manager whose degree of religiosity is unknown)

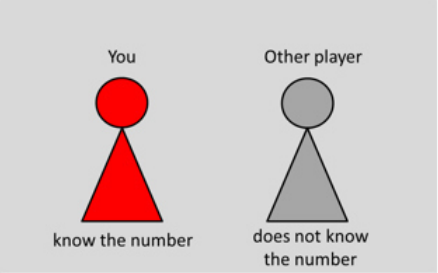
Your fellow player is a family manager who holds a management position in his own company. The majority of the company's shares and voting rights have been in the family's possession since the previous generation. Your fellow player is an active Christian and regularly attends church services. (Your fellow player is an active sportsman and regularly participates in sports events.)

Religious non-family manager (non-family manager whose degree of religiosity is unknown)

Your fellow player is a manager who holds a management position in a company. The majority of the company's shares and voting rights of the company have been held in free float by various private and institutional investors for decades. Your fellow player is an active Christian and regularly attends church services. (Your fellow player is an active sportsman and regularly participates in sports events.)

Appendix B: Instructions lying game

1) Start

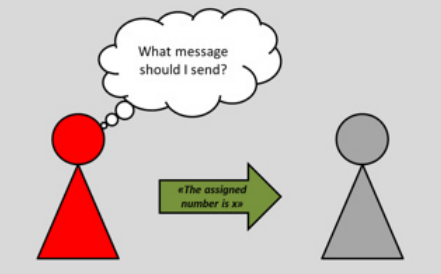


There are 2 players: You and the other player. You are the red player.

You are playing with the person from the previous description.

A number between 1 and 6 (1, 2, 3, 4, 5 or 6) is randomly assigned to you and the other player. You **are** informed about this assigned number, the other player **is not.**

2) Your decision

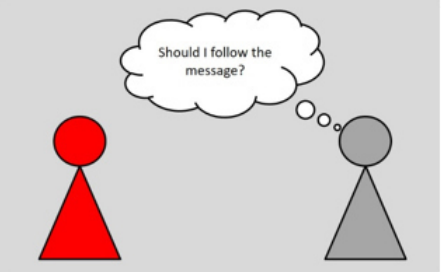


You can send a message about the assigned number to the other player. The message does not have to contain the actually assigned number.

You receive € 100 **plus** twenty times the amount of the number sent in the message in €.

For example: You receive € 120 if you send the message that the assigned number is 1; you receive € 140 if you send the message that the assigned number is 2 etc.

3) Decision of the other player



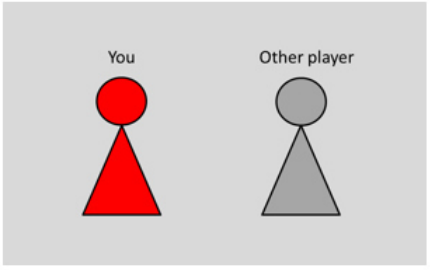
The other player must decide whether to follow the message or not.

If he **follows** the message and the message contains the actually assigned number, he receives € 100. If the message **does not** contain the actually assigned number, he receives nothing.

If the other player **does not follow** the message, he receives € 30.

Example:

1) Start

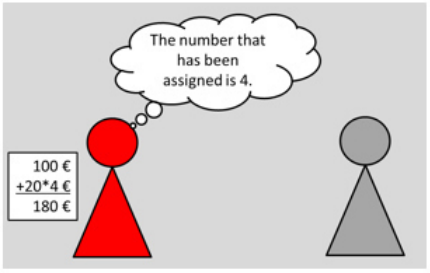


You and the other player are randomly assigned a number between 1 and 6.

You are informed about this assigned number, the other player is not.

Suppose the number 3 is randomly assigned to you and the other player.

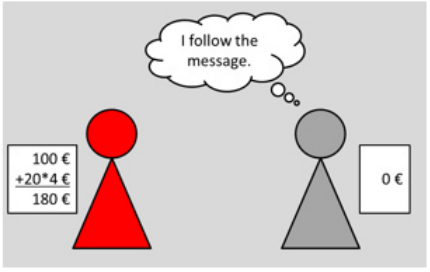
2) Your decision



You send the message "The assigned number is 4" to the other player.

You receive € 100 plus twenty times the amount of the number sent in the message in €. Therefore, you receive € 180 (100 € + 20 € * 4).

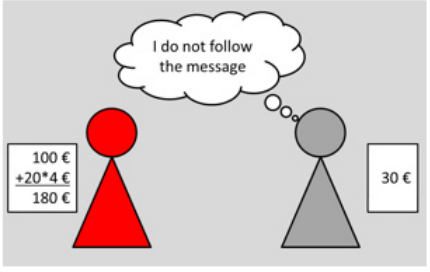
3) Decision of the other player



The other player must decide whether to follow the message or not.

Scenario 1:

Suppose he follows the message. Then he receives € 0 because the message does not contain the actually assigned number.



Scenario 2:

Suppose he does not follow the message. Then he receives € 30.

You receive € 180 in both cases.

Comprehension questions

To ensure that we have explained the game in a comprehensible way, please answer the following questions briefly.

Assume that:

- The number 5 is randomly assigned to you and the other player.
- You send the message "The assigned number is 5" to the other player.
- The other player decides to follow the message.

Please tick the correct answer.

How many € do you have at the end of the game?

☐ € 100

☐ € 200

☐ € 300

How many € does the other player have at the end of the game?

☐ € 0

☐ € 30

☐ € 100

Your decision

Since this is a hypothetical game, please indicate what message you would send to the other player for each possible number assigned to you. The message does not have to match the actually assigned number. While making these decisions, please imagine that you are playing for real money, even if your decisions are hypothetical and will not be executed. If you wish to review the instructions, please click here (link to the instructions).

(Please tick a number between 1 and 6 in each of the 6 gray boxes.)

Suppose that the actually assigned number is:	What message do you then send to the other player?
1	***The assigned number is*** ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6
2	***The assigned number is*** ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6
3	***The assigned number is*** ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6
4	***The assigned number is*** ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6
5	***The assigned number is*** ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6
6	***The assigned number is*** ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

Appendix C: Probit regression

TABLE 4.7: The effect of family firm and religiosity reputation on the likelihood of sending the true number

	1	2	3	4	5	6	7	8
FM	0.191*	0.194**						
	(0.098)	(0.098)						
State (s)	0.249***	0.250***	0.251***	0.252***	0.256***	0.257***	0.250***	0.251***
	(0.013)	(0.013)	(0.015)	(0.015)	(0.015)	(0.015)	(0.013)	(0.013)
FM religious			0.288**	0.291**			0.358**	0.346**
			(0.126)	(0.126)			(0.140)	(0.140)
FM unknown					0.104	0.113	0.174	0.176
					(0.119)	(0.118)	(0.133)	(0.133)
NFM religious							0.141	0.125
							(0.134)	(0.135)
Male		-0.129		-0.034		-0.088		-0.120
		(0.098)		(0.114)		(0.111)		(0.098)
Age		-0.003		0.001		-0.005		-0.003
		(0.004)		(0.005)		(0.005)		(0.004)
Own religiosity		0.059		0.073		0.024		0.057
		(0.041)		(0.046)		(0.045)		(0.041)
Constant	-0.280***	-0.238	-0.286***	-0.473**	-0.303***	-0.153	-0.353***	-0.309
	(0.086)	(0.198)	(0.092)	(0.224)	(0.091)	(0.214)	(0.106)	(0.206)
N	2688	2688	2010	2010	2040	2040	2688	2688
Pseudo-R2	0.080	0.084	0.084	0.087	0.081	0.084	0.082	0.087

Notes: The table shows the estimates of a probit regression with clustered standard errors on an individual level in parentheses. The dependent variable is a dummy variable that takes the value 1 if the individual sent the true number and 0 otherwise. Significance level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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.

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: 96).

Our novel implementation relies on websockets[26] 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)[27] to test the functionality of the DA market. We find that the module proved to work smoothly (see section 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. 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

[26] 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.

[27] 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.

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[28] and EconPort[29] 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: 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 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

[28] See http://veconlab.econ.virginia.edu
[29] See http://econport.org/econport/request?page=web_home

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).

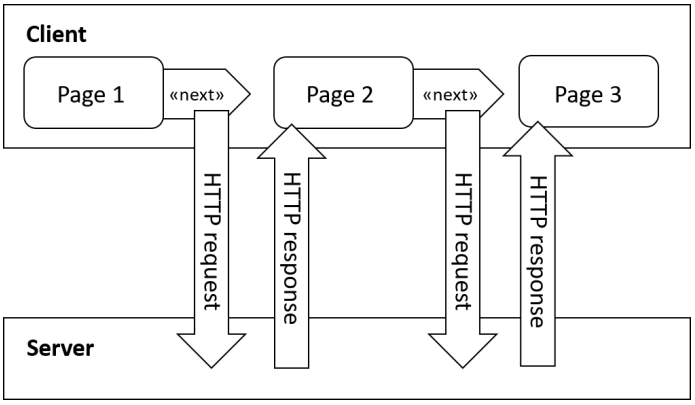


FIGURE 5.1: The current communication of an oTree application is based on HTTP requests

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.[30] 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

[30] 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.

messages and terminate its connection. Multiple channels can be grouped to send messages to multiple clients. Figure 5.2 illustrates this architecture.

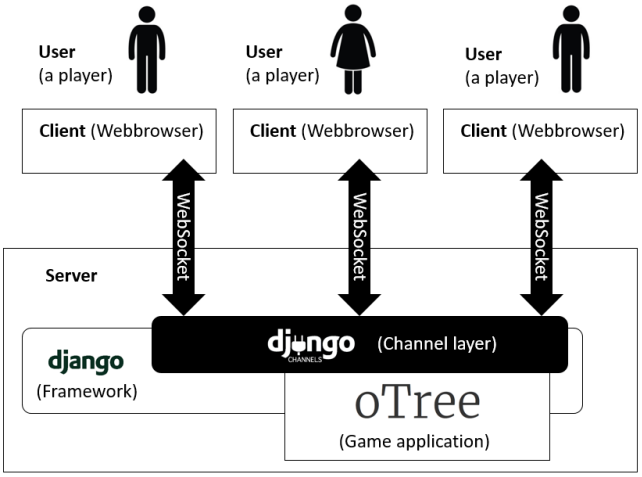


FIGURE 5.2: The websockets architecture within oTree using Django channels

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.

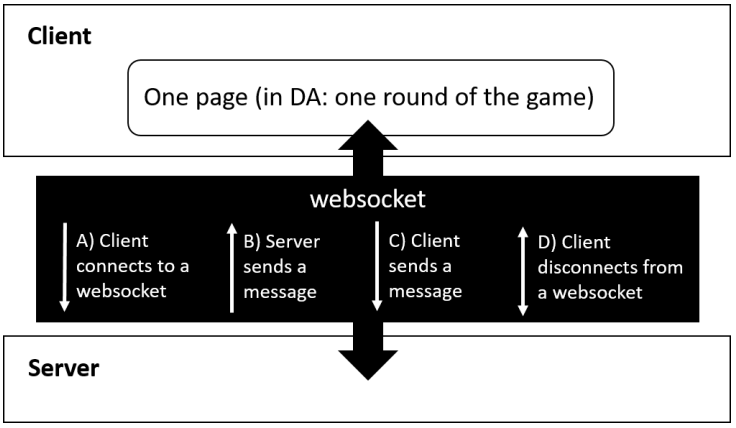


FIGURE 5.3: Websockets form a permanent communication channel between the client and the server

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[31]. 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.[32] 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.[33] 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.

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

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

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

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 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.

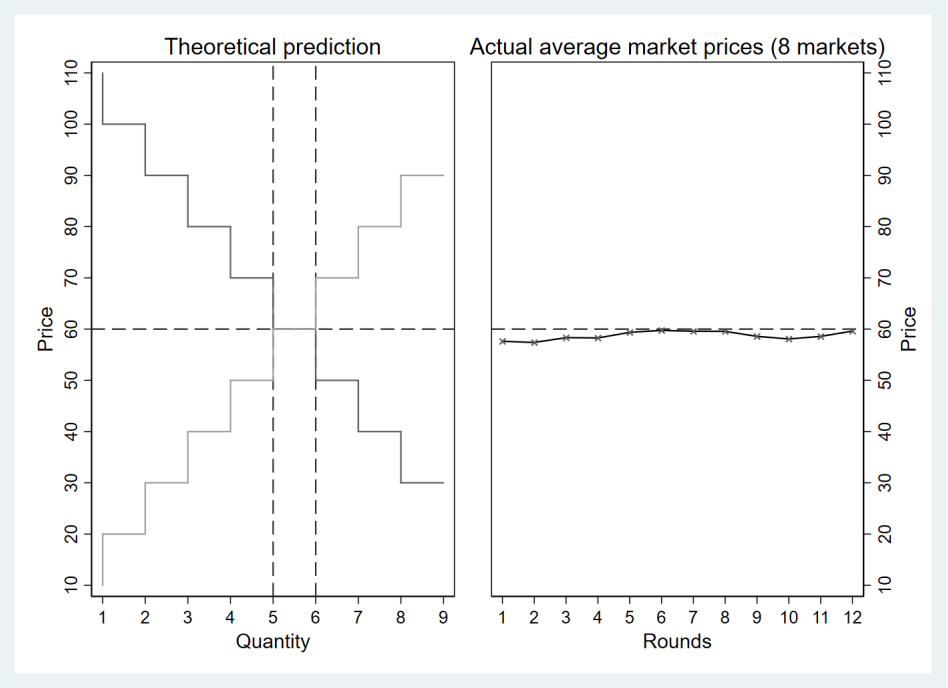


FIGURE 5.4: Theoretical prediction (left) and actual average market prices (right)

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.[34] 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.[35] 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 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

[34] 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.

[35] 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.

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

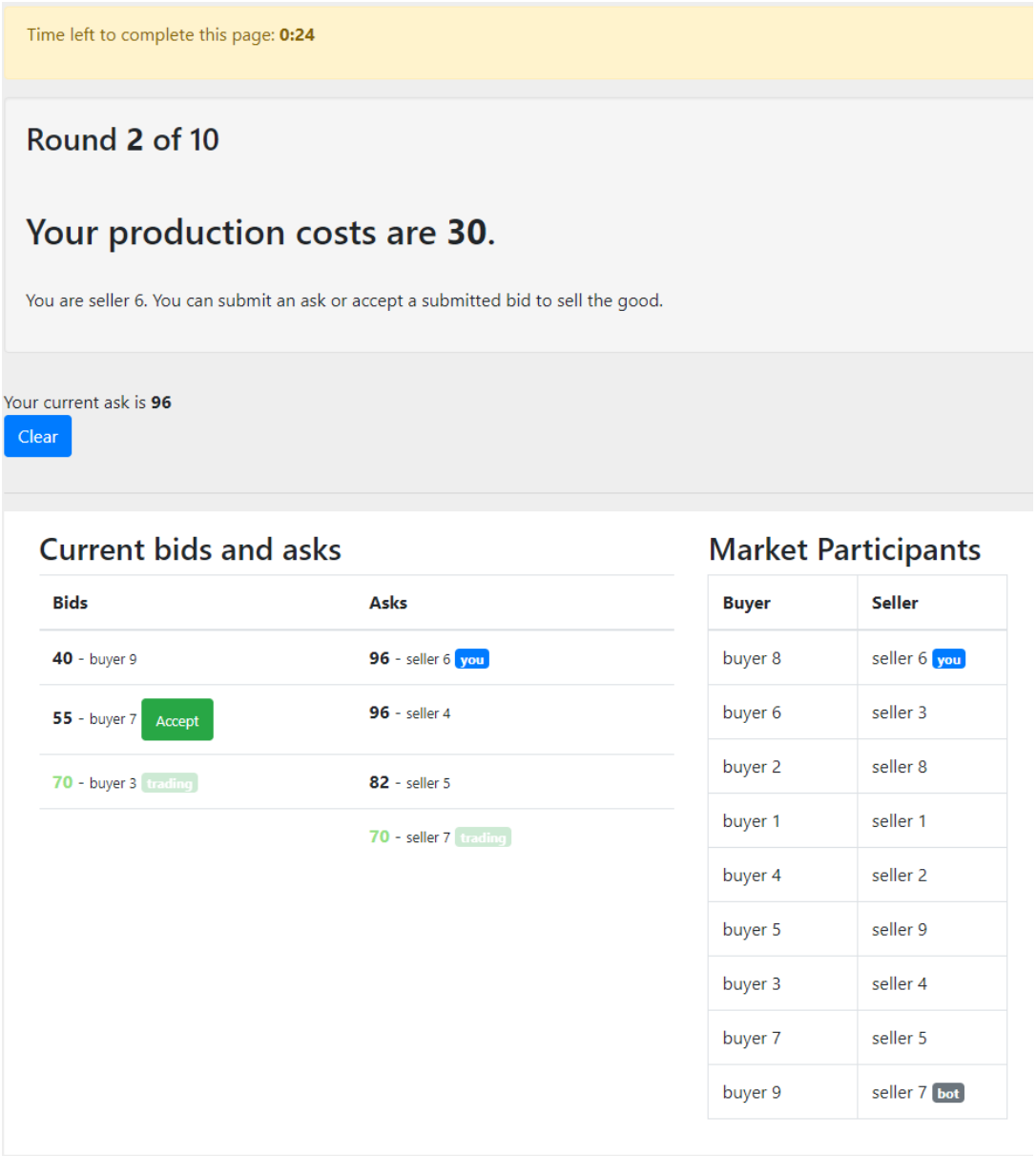


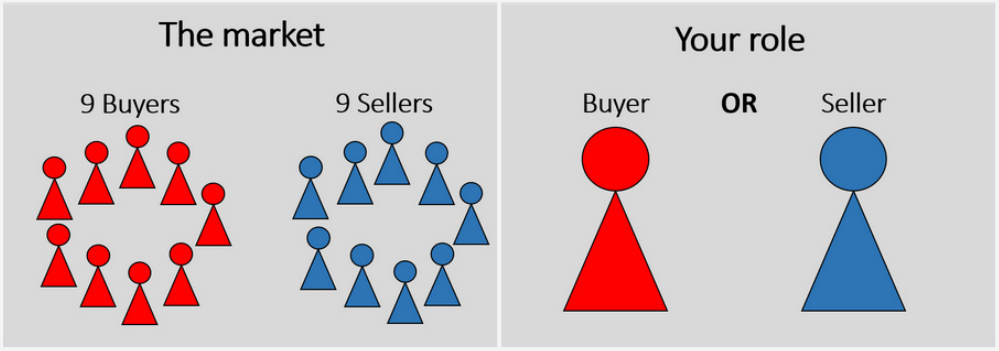
FIGURE 5.5: The Graphical User Interface of the DA market

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.

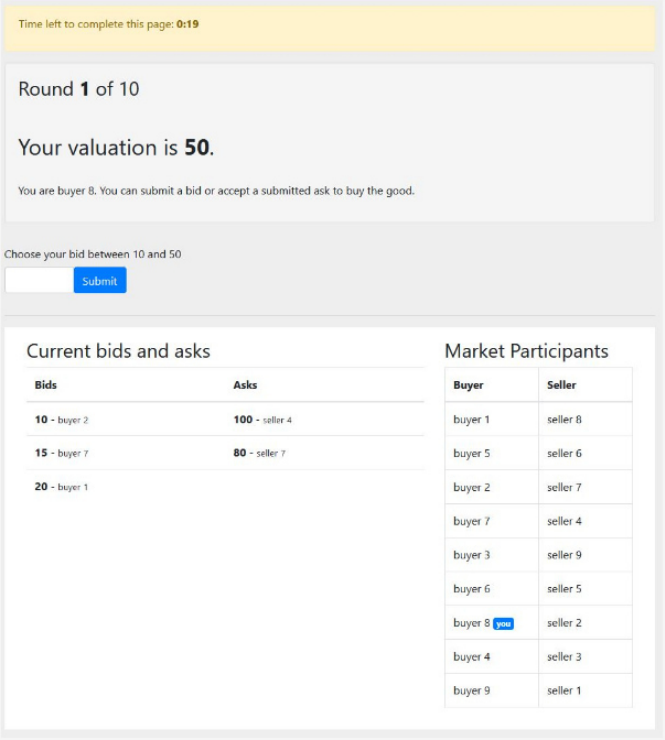
How does trading work for the buyer?

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.

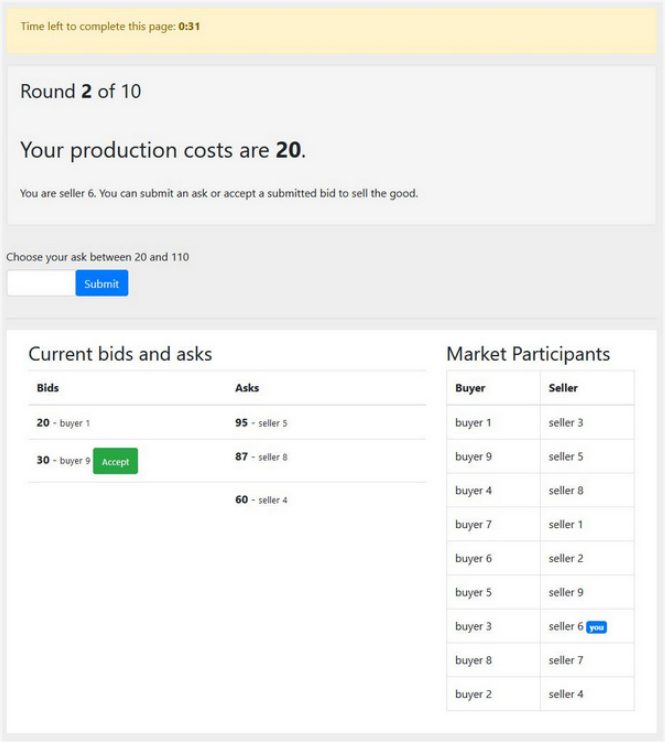
How does trading work for the seller?

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: A. Crede

Date: 07.06.2019