Economics of Ethical and Prosocial Behavior: Experiments on Social Responsibility, Ignorance, and Cheating

Inauguraldissertation
zur
Erlangung des Doktorgrades
der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der
Universität zu Köln

2016

vorgelegt
von

Agne Kajackaite, M.Sc.
aus
Sakiai

Referent: Prof. Dr. Dirk Sliwka

Korreferent: Prof. Dr. Bernd Irlenbusch

Tag der Promotion: 30.05.2016

Acknowledgments

Over the past four years I have received encouragement from a great number of individuals. First, I would like to express my sincere gratitude to my advisor Dirk Sliwka for the continuous support during my PhD studies. His guidance helped me in learning how to conduct research and, importantly, to appreciate theory. One could not wish for a better advisor.

I would also like to thank the other members of my thesis committee, Bernd Irlenbusch and Axel Ockenfels, for their insightful comments and encouragement.

My sincere thanks goes to Uri Gneezy, who is my coauthor, teacher and friend. We started collaborating on a number of projects during the last couple of years, and it changed the way I think about research questions and how I design experiments.

My sincere thanks also goes to my coauthors outside of this thesis: Steffen Andersen, Gary Charness, Catherine Eckel, Julie Marx, Joel Sobel and Peter Werner. I have learned a lot from these collaborations and have grown as a researcher.

I am also thankful for my dear colleagues with whom I had long discussions about the research questions – mine, theirs, new joint ideas – or who helped me with the research in some other ways. Among others, those people are Kristina Bott, Suparee Boonmanunt, Lea Cassar, Tobias Danzeisen, Patrick Kampkötter, Andrew Kinder, Pedro Rey-Biel, Silvia Saccardo, Anne Schielke, Uta Schier, Marina Schröder, Marta Serra-Garcia, Shaul Shalvi, Charlie Sprenger, and Marcin Waligora.

I also thank the German Science Foundation (DFG) for the financial support through the research unit "Design & Behavior" (FOR 1371), which made my experiments possible.

I dedicate this thesis to my family – my parents and my sisters, who supported and encouraged me with their best wishes.

Contents

List of Figures

List of Tables

1 Introduction

In this thesis I experimentally investigate prosocial and ethical behavior in economic interactions. The thesis consists of three experimental research papers that have a broad range of research questions on social responsibility, ignorance and cheating. With these experiments I aim to better understand when and why people behave *good* and *bad* and which consequences it has on their own and other players' payoffs, and on overall efficiency. By acting *good*, I mean two kinds of behavior. First, one might conform to the social norms, i.e. telling the truth, even if lying has a higher financial benefit (*weak goodness; ethical behavior*). Second, one might go above just conforming to the social norms and also voluntarily benefit others, for instance, by donating to a charity or working for a good cause (*strong goodness; prosocial behavior*). Under acting *bad*, on the other hand, I understand *unethical behavior*, i.e. not conforming to the social norms, for instance, misreporting of private information in order to receive a higher payoff or working for a bad cause.

Chapter 2 looks at prosocial behavior, whereas chapters 3 and 4 address (un)ethical behavior, which together underscore that the focus of the thesis is on both ethical and prosocial behavior. The approach in the thesis centers on using the experimental method in the laboratory, where one can observe how people make decisions in a completely controlled environment (for the method, see Roth, 2005).

Prosocial behavior, also called other-regarding preferences, has been extensively studied in experimental and behavioral economics over the last decades (see, Cooper and Kagel, forthcoming, for an overview). The experiments have shown that people are willing to voluntarily forego some money and give it to the others (see ultimatum and trust games), and it cannot be solely explained by strategic concerns (see dictator games). The non-strategic prosocial behavior is argued to be driven by pure or warm glow altruism (see Andreoni, 1989, and Andreoni, 1990), inequity concerns (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) and efficiency concerns (Charness and

Rabin, 2002). The experimental evidence also suggests that experimental participants are willing to treat someone nicely if this person treated them (or others) nicely before, which suggests that some people behave reciprocally. This has been widely shown in the so-called gift-exchange literature, where higher wages tend to be reciprocated with higher efforts by the employees (see, for instance, Charness, 2004).

Chapter 2 of this thesis focuses on one specific research question with regard to prosocial behavior, with which it contributes to this broad research area (or even maybe the most extensive research area by date in experimental economics). In this chapter, which is called "Social Responsibility and Incentives in the Lab: Why Do Agents Exert More Effort when Principals Donate?" and is joint work with Dirk Sliwka, we experimentally and theoretically analyze whether and why principals' charitable giving has a positive effect on agents' effort, i.e. we use a variation of a gift-exchange game.

The motivation for this research paper is that many companies invest high amounts of money in corporate social responsibility (CSR) activities, have CSR departments and make their prosocial activities public through CSR reports. Since companies spend a share of their profits for the social causes, does it mean that they want to do good? And maybe it pays to do good? In this paper, we study one particular channel through which a firm's social behavior might affect its financial performance, namely why and under what conditions CSR has a positive effect on employees' performance.

In order to study how donating decisions of the principal affect employees, we use a laboratory experiment with a simple principal-agent setting, where a principal first decides whether or not to donate a fixed amount to a charity and, in the next step, an agent chooses his effort level. We argue that there are three potential mechanisms that can trigger a positive effect of an employer's charitable giving on an employee's effort in such a setting: (i) distributional concerns, (ii) reciprocal altruism and the signaling of prosocial motives, and (iii) shared warm glow utility. As expected, we find that agents choose higher efforts when principals had donated to a charity. We disentangle the different mechanisms by varying the intentionality and size of the donation as well as the wealth of the principals. We find robust evidence for reciprocal altruism as a driver for agents' performance reactions to employers' donating decisions, while distributional concerns and warm glow motives are pronounced to a weaker extent. Additionally, we test for whether donating decisions are strategic (think of the companies anticipating that employees reward donations with higher effort and donating for this reason). We find that a high fraction of principals donate strategically in order to induce higher efforts, but agents reward charitable acts irrespective of the extent to which these acts are driven by strategic motives.

In chapter 3, I move thematically from what I call *strong goodness* to *unethical behavior*. In this chapter ("If I close my eyes, nobody will get hurt: The effect of ignorance on performance in a real-effort experiment"; single-authored; published in "Journal of Economic Behavior & Organization", 2015), I analyze whether experimental subjects choose to stay ignorant about the negative consequences of one's own actions and if so how it affects their further decisions. With this chapter, I contribute to a relatively new and not yet extensively researched area in behavioral economics – ignorance (see experiments by Dana et al., 2006; Dana et al., 2007; Larson and Capra, 2009; Fong and Oberholzer-Gee, 2011; Matthey and Regner, 2011; Conrads and Irlenbusch, 2013; Grossman and van der Weele, 2013; Bartling et al., 2014; Grossman, 2014; van der Weele, 2014).

In the CSR experiment, participants could act prosocially by donating to a charity or transferring points to the principal or not take these actions. In this chapter, on the other hand, the default situation is negatively framed. In most of the treatments of this study, participants can act unethically by donating or not to a negatively perceived charity. That is, there is no way to act prosocially, but one can decide on whether one's actions are ethical. Thus, in contrast to the previous chapter, the focus of this study is on ethical behavior, and in particular its interaction with ignorance.

The motivation for the experiment in this chapter is the observation that people often willfully decide to avoid information. Especially when it comes to ignorance in organizations, a number of examples of willful ignorance are available, as for instance, huge corporate scandals like Enron. In most of the cases, ignorance increases monetary gain by maintaining the illusion of acting in a fair way. For instance, consider an employee who does not know the negative effect of her work on the economy. She can maintain her positive self-image while still earning a high income. I simulate such an environment, in which work output increases one's own payoff but might have a negative externality, and am interested in observing how choosing to be ignorant about the negative consequences of one's own work affects work output.

The game in this experiment was the following: The participants worked on a tedious real-effort task, and their effort either increased only their own payoff or also increased the donation to a negatively perceived charity (National Rifle Association, a US gun lobby). I manipulated the information between conditions: In some conditions, participants knew that their effort benefits the NRA, and in some they did not. I introduced ignorance by letting agents decide whether to learn if their effort benefitted the NRA. As expected, agents exerted significantly higher efforts if they knew the NRA would receive no benefits. Yet when given the choice, almost a third of the agents chose to stay ignorant and exert significantly more effort than agents who knew their effort would benefit the charity. Importantly, if uncertainty about the donation to the charity was introduced exogenously, agents exerted lower effort than ignorant agents, which suggests that not having information about the consequences of one's own actions alone does not lead to self-interested behavior, but rather the sorting of social agents of a low type into ignorance drives self-interested behavior of ignorant agents.

Finally, in chapter 4 ("Lying Costs and Incentives", joint work with Uri Gneezy), I move from ignorance to another topic in behavioral economics and ethics – lying. Lying behavior has gained attention of many behavioral economists over the last ten years (see experiments by Gneezy, 2005; Mazar, Amir, and Ariely, 2008; Sutter, 2008; Dreber and Johannesson, 2008; Erat and Gneezy, 2012; Fischbacher and Föllmi-Heusi, 2013). These experiments show that for some people lying is intrinsically costly – they would prefer an outcome if they achieve it honestly over achieving it by lying. Although the economics literature has extensively studied lying, many questions regarding individual lying behavior remain unanswered, for instance, the structure of the intrinsic cost of lying. Given the important economic consequences of cheating (think of insurance fraud or recent cheating examples, such as the one involving VW), understanding the factors that influence the decision to cheat is important for understanding many economic behaviors. Contributing to the knowledge about the structure of the intrinsic lying cost is the aim of chapter 4.

In this study, Uri Gneezy and I concentrate on “intrinsic” costs of lying and use two experiments to better understand the structure of these costs. In the first experiment, we replicate the finding in the cheating game literature that lying does not increase with incentives and show that this insensitivity is not a characteristic of the intrinsic lying costs, but rather a result of concern about being exposed as a liar. When we modify the cheating game into a “mind” game in which this concern is reduced, we find that people lie more in general and react to the incentives associated with lying. The second experiment allows us to look deeper into the structure of the intrinsic cost of lying. In this experiment, we test the functional form of the lying costs. Our results reject the common assumption of “small lies” that are told because of a convex cost of lying. In contrast, our data are consistent with a fixed intrinsic cost of lying: When our participants lie, they do so to the full extent, whereas partial lying is rare. Taken together, our results show that for many participants, the decision to lie follows a simple cost-benefit analysis: They compare the intrinsic cost of lying with the incentives to lie. Once the incentives are higher than the cost, they switch from telling the truth to lying to the full extent.

Together, the results from the three experimental studies presented in this thesis contribute to our understanding and knowledge about when and why people behave prosocially and/or ethically. In particular, the experimental evidence suggests that (i) donations to charity by employees are rewarded in an experimental setting, and the effect is driven by reciprocal concerns; (ii) there is a significant fraction of people who decide not to know about negative consequences of own actions, and the sorting of social agents of a low type into ignorance drives self-interested behavior of ignorant agents; and (iii) if the possibility of being exposed as a liar is small, the tendency to lie increases with incentives, indicating that some people have positive and finite costs of lying. Furthermore, when the participants lie, they lie to the full extent, which suggests that the intrinsic cost of lying is fixed.

In the following chapters, I present the three experimental studies in detail.

2 Social Responsibility and Incentives in the Lab: Why Do Agents Exert More Effort when Principals Donate?

2.1 Introduction

Milton Friedman (1970) famously claimed that “the social responsibility of business is to increase its profits.” On the other hand, we observe that companies spend high amounts of money on Corporate Social Responsibility (CSR) activities. Many of them have CSR departments and make these socially-responsible activities public through CSR reports (for example, think of such companies as Google, BMW or Disney). Since companies spend parts of their profits to social causes, does it mean that they want to do good? And does it pay to do good? While there are a large number of observational studies (see, for instance, Stanwick and Stanwick, 1998, and Margolis et al., 2007 for a meta-analysis) showing that the financial performance of a firm is (weakly) positively correlated with social responsibility activities, there is still a lack of evidence on the causal impact of an employer’s social behavior on financial performance. In this paper, we study one particular channel through which a firm’s social behavior may affect its financial performance, namely we investigate why and under what conditions a charitable act by a principal has a positive effect on an agent’s motivation to exert effort that is beneficial for this principal.

A number of recent lab experiments (Charness et al., 2014; Imas, 2014; Koppel and Regner, 2014; Cassar, 2015, or Tonin and Vlassopoulus 2015) have established that charitable giving can raise employee incentives. Our main aim is to understand the behavioral mechanisms driving an agent’s reaction to observable prosocial acts by a principal in the lab. In particular, we consider a baseline setting in which principals first decide on a donation to a charity and agents then exert effort that affect principals' profits. We argue that there are three different channels through which a principal's social activity can raise employees' motivation: (i) reciprocal altruism (see Levine, 1998), (ii) shared warm glow motives (see Andreoni, 1989, and Andreoni, 1990)[1] and (iii) distributional concerns (see Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002). By varying characteristics of our baseline setting we aim to disentangle these different behavioral channels.

We study a very simple setting in which a principal is matched to an agent, and prior to the agent's effort choice there is either a donation to a charity from the principal's account or not. In the second stage of the game, the agent observes the donation decision and decides on his effort. There are no performance-contingent incentives, meaning that higher efforts only increase the principal's payoffs and reduce the agent's. When the principal decides on the donation and players have full information about the game, the agents' effort can be affected through all three mechanisms, i.e., reciprocal altruism, "warm glow" utility and distributional concerns. Reciprocal altruism comes into play since a donating principal signals that she is a prosocial type, which may lead to higher efforts by agents who are reciprocal altruists. "Warm glow" utility can affect efforts when donations trigger the possibility that agents can "buy into" these donations by exerting effort. Note that we here extend the notion of "warm glow" proposed by Andreoni (1989) and Andreoni (1990). The idea is that an agent may feel that he "finances" (part of) the donation when exerting effort for the donating principal even when the donation decision has been made before the agent decides how much effort he exerts. Finally, distributional concerns can lead to a positive reaction to a donation because donations raise the effort level at which the principal's and agent's payoffs from the interaction are equalized. We illustrate these mechanisms in a formal model (reported in the Appendix A) and experimentally disentangle motives by conducting several treatment variations, which differ with respect to the intentionality of the principal's decision, the size of the donation and the principal's initial endowment.

[1] See, for instance, Crumpler and Grossman (2008) or Tonin and Vlassopoulus (2010) for experimental evidence on warm glow giving.

In the baseline treatment of our experiment, principals choose whether or not to donate to a charity and are aware that agents will learn about the decision prior to choosing their effort levels. In our non-intentional condition, on the other hand, similarly to Blount (1995), Offerman (2002), Charness and Levine (2007) and Falk et al. (2008), a random device determines whether the principal donates or not. This treatment variation allows us to study the role of reciprocal altruism as a driving force, since the decision is randomly assigned and, hence, the principal cannot signal his type. If reciprocal altruism matters, an agent's reaction to a donation decision should be diminished in the non-intentional condition when compared to the baseline condition, but warm glow motives and distributional concerns may still play a role.

We conduct further treatments to better understand the role of warm glow motives and distributional concerns. First, we vary the size of the potential donation. Both warm glow and distributional motives can positively affect efforts when the donation's size increases since a larger donation raises the scope to "buy into" the warm glow of the donation and shifts the effort level at which equity is attained upwards. Second, we run a treatment in which the principal is wealthier than the agent. If distributional concerns matter, raising the principal's wealth should reduce the agent's efforts as higher efforts in this case would increase inequity even further. However, if warm glow motives drive agents' behavior, we should still observe a positive effort reaction to a donation in this treatment even if the donation is exogenously assigned.

In additional treatments, we test how possible strategic motives behind a donation affect agents' effort by conducting sessions in which principals are unaware that agents make a subsequent effort decision and agents know that principals were unaware when they made the choice. Donations in these treatments can thus be a more "pure" signal of the principal's prosocial inclinations.

We find that in seven out of the eight treatments agents choose significantly higher efforts when principals donate.[2] In line with the importance of intentions and reciprocal altruism, the effect of donating is mitigated when donations are exogenously

[2] The only treatment where agents do not exert more effort after a donation is the one in which donations are exogenously assigned and the principal is richer than the agent.

assigned. In the non-intentional conditions, average efforts are (slightly) higher when nothing is donated and lower after a donation.[3]

We find that high donations do not trigger higher effort levels than those observed after lower donations. Increasing the principal's wealth, however, substantially reduces agents' effort choices after a donation, in particular when the donation is exogenously assigned. But when principals donated intentionally agents choose significantly higher efforts even when the principal is substantially wealthier than the agent. Hence, reciprocal altruism is a robust channel through which observable prosocial behavior by principals affect agents' efforts.

As we noted above, there are several recent experimental studies that analyze the impact of social incentives on employees' performance.[4] Tonin and Vlassopoulus (2015) analyze how social incentives (i.e., a piece rate or a fixed donation transferred to a charity) motivate subjects in a real effort experiment. They show that social incentives lead to a rise in productivity, regardless of their form (fixed vs. piece rate) or strength, but are not as effective as monetary incentives in motivating workers. Imas (2014) also compares monetary and social incentives (in a form of a piece rate to a charity) in a real effort experiment and finds that while social incentives work better when the stakes are low, social incentives are not more effective than monetary incentives when the stakes increase. Similarly, Charness et al. (2014) find that social incentives work better when stakes are low, and monetary incentives are more effective in motivating employees when stakes are high. Koppel and Regner (2014) analyze the effect of CSR on the agents' effort in an experimental gift-exchange game. In their experiment, the principal simultaneously decides on the wage and a share of profits transferred to a charity and the agent decides on his effort. They find that agents' efforts increase according to the share of profit that principals donate to the charity. Similarly, Cassar (2015) also shows that in a gift-exchange game agents are motivated by a principal's piece-rate donation and exert more effort. We build on these studies and aim to identify and disentangle the behavioral

---

[3] The positive reciprocal responses are significant in the rich principals' treatments but not when the principal is "poor," while negative reciprocal responses are significant only in poor principals' treatments.

[4] There are also a growing number of experiments analyzing how social responsibility of firms' affects consumers' behavior. See, for instance, Bartling et al. (2015) and Pigors and Rockenbach (forthcoming) for recent evidence.

mechanisms driving the link between charitable giving by the principal and employees' effort. Note that in our setting, the charitable act has already been carried out when the agents choose their efforts and these efforts do not affect the size of the donation. Hence, our setting aims at studying how a charitable act by a principal changes the relationship between principal and agent excluding the direct social incentives for an agent to raise charitable giving by his efforts.

The paper is organized as follows: In section 2.2, we describe the experimental design and experimental procedure. In section 2.3, we describe the theoretical predictions. Section 2.4 reports the results. We discuss our findings and conclude in section 2.5.

2.2 Experimental Design

We conducted a laboratory experiment with eight treatments using a simple principal-agent game. The game consisted of two stages and was played only once. Two participants were matched, one in the role of participant A (the principal) and one in the role of participant B (the agent). We randomly assigned the roles prior to the experiment.[5]

We conducted the main treatments to analyze intentional, distributional, and warm glow effects of social incentives on agents' performance. In the baseline condition *Intentions 20* (denoted by I20), each principal initially received 50 ECU (10 ECU = 1 EUR) in her experimental account. In the first stage of the game, principals could decide whether to donate 0 or 20 ECU to the charity "Médecins sans Frontières", which is a renowned charity and received the Nobel Peace Prize in 1999.[6]

In the second stage of the game, the agent was informed about the donation by the principal whom he was assigned to. The agent could decide how many points of his endowment of 100 ECU to transfer to the principal, where every integer between 0 and

[5] In the instructions (see Appendix for translations from German), we used neutral formulations. In particular, we did not use words such as principal or agent.

[6] In the experiment by Koppel and Regner (2014), in which subjects can choose which charity they want to donate to, the majority chooses "Médecins sans Frontières".

100 was feasible. We doubled each transferred point, and thus the payoff functions are $\pi_P = 50 - d + 2e$, $\pi_A = 100 - e$ and $\pi_C = d$ for the principal, agent and charity, respectively, whereas d stands for the donation and e for effort.[7]

The second treatment *No Intentions 20* (NI20) differed from the baseline condition only in the nature of the donation decision. In this treatment, we ruled out intentionality of the principal's action, i.e., we used a random device to make the donation decision. The experimenter rolled a six-sided die in front of every agent. If the die showed 1, 2 or 3, no donation was made, and if the number was 4, 5 or 6, we transferred a donation of 20 ECU from the principal's account.

Further treatments *Intentions 40* (I40) and *No Intentions 40* (NI40) differed from I20 and NI20 in the size of the donation. Here, the donation that the principal (or the random device) could choose was 40 ECU, which is equal to 80% of the principal's endowment. These treatments give us insights into the importance of warm glow and distributional concerns since there is a larger scope for a warm glow "buy in" (the donation that can be "financed" by the agent is higher) and the effort level that generates equitable payoffs is shifted upwards.

In the treatments *Intentions 20 with rich principals* (I20rich) and *No Intentions 20 with rich principals* (NI20rich), we increased the endowment of the principal up to 120 ECU, such that $\pi_P = 120 - d + 2e$. This treatment helps us to get further insights on warm glow motives and distributional concerns. While warm glow would still predict positive effort after a donation, the distributional concerns should lead to no (or very low) effort, given that the principal is richer than the agent and higher efforts only increase inequity further.

Finally, we conducted the last two treatments, *Non-strategic 20* (NS20) and *Non-strategic Online 20* (NSO20), to investigate agents' reactions to non-instrumental donations, where the I20 condition served as the reference for these treatments. While in all the treatments described above both the agent and the principal had complete information about the game, in NS20 agents and principals were informed about the existence of the second stage of the game, but the content of the stage was withheld until

[7] The payoff functions are similar to those used by Falk and Kosfeld (2006).

the end of the first stage. Not knowing about the agent's possible reciprocal response should diminish strategic concerns when making a donation. However, since some principals could still expect that the payoffs in the second stage might depend on the donation made in the first stage, we conducted a further treatment, NSO20, which ruled out all strategic concerns. Here, we let the principals decide whether to donate 0 ECU or 20 ECU in a single-stage online donation experiment. Thus, the donations made by the principals should be completely non-strategic. Later we invited subjects who were assigned to the role of agents to the lab. Each agent observed the donation of a principal that we assigned to her from the online experiment and decided on her effort. We subsequently transferred the money generated for the principals via bank transfer.

We used the strategy method (Selten, 1967) for the agents' choices in all the treatments.[8] Specifically, we asked the agents to state their effort levels for both possible principals' decisions (donation and non-donation) before they found out the actual decision. Table 2.1 summarizes the experimental treatments.

Table 2.1: Treatments

Treatment name	Intentional decision	High donation	Wealthy principal	Strategic concerns	N of agents
I20	X			X	30
NI20				X	29
I40	X	X		X	30
NI40		X		X	30
I20rich	X		X	X	30
NI20rich			X	X	30
NS20	X				30
NSO20	X				30

[8] According to Brandts and Charness (2011), most of the experimental studies do not find significant differences between results gathered using the strategy method and the direct-response method.

We conducted our experiment in April 2013 - April 2015 using the experimental software zTree (Fischbacher 2007). We recruited the participants via ORSEE (Greiner 2004). We ran sixteen sessions (two sessions for each treatment condition), fifteen in the lab and one session online using the survey software Unipark. Altogether 478 subjects participated in our experiment, half of them were agents and the other half principals. No subject participated in more than one of our sessions. Thus, we collected roughly 30 independent observations for both agents' and principals' decisions in each treatment condition.[9]

Instructions for the experiment were displayed on the computer screen. We allowed participants to ask questions privately. After they read the instructions, participants had to answer a number of questions in order to guarantee that they understood the setting. Then the actual experiment started. After the experiment was finished, we asked the subjects to complete a post-experiment questionnaire including questions on gender, age, field of study and motives behind their decisions. At the end, they privately received their payoffs in cash and left the laboratory. In the online condition, the participants received their payments via bank transfer.[10]

Each lab session lasted approximately 50 minutes. The average payoff was 9.42 Euros and 11.65 Euros for principals and agents, respectively, including a show-up fee of 2.50 Euros. The overall revenue generated for the charity "Médecins sans Frontières" was 366 Euros. We made the transfers to "Médecins sans Frontières" directly after each session. To ensure that the donations are credible, we wrote in the instructions that the subjects can stay after the experiment and observe how we make the donation. In the online experiment, participants could give us their email address to receive the proof of the donation. Only one participant stayed in the lab after the experiment and observed how we transferred the money to the charity. In the online experiment, approximately half of the subjects stated that they wanted to receive proof of the donation via email.

[9] There are 30 independent observations for principals' and agents' behavior in all the treatments except NI20, which has 29 independent observations, since two subjects did not show up to the session.

[10] We transferred the payment together with the doubled agents' transfer from the second stage of the game. The principals received an email, in which we explained why they may receive the additional money.

2.3 Predictions

We analyze our experimental game in a formal signaling model, from which we derive predictions for the agents' and principals' behavior (see Appendix A for the formal model). Our model builds on Levine's (1998) model of reciprocal altruism where agents are characterized by their degree of "prosociality" and an agent cares more for another individual when he considers this other individual to be a more prosocial person. We incorporate two further motives into this framework. First, we allow for the possibility that, as in Charness and Rabin (2002), these prosocial concerns matter as long as the payoff from the interaction does not leave an agent worse off than her interaction partner. Second, we allow for the possibility that the agent receives an indirect "warm glow" utility when exerting effort for a principal who made a donation as long as the resulting profit is smaller than the donation made ex-ante. The idea is that agents may reap psychological benefits from "buying into" the donation. In the following sections, we describe the resulting predictions.

Overall Effect

We expect agents in all of our treatment variations to exert more effort after a donation than after a non-donation. In the baseline condition I20, all three mechanisms can come into play and lead to higher efforts after a donation. First, the donation made by the principal alters the income distribution in favor of the agent, and thus an agent who has distributional concerns will exert more effort if the principal donates. Second, through charitable actions an employer signals that she is a prosocial type, in turn increasing the agent's expectation about the principal's type and triggering a stronger identification on the part of the agent with the employer's outcomes. Finally, if indeed agents care for indirect warm glow, by making a donation an employer creates the possibility that the employee "buys into" the warm glow of this donation by exerting effort.

Intentions and Signaling

In the next step, we explore the signaling motive by comparing I20 with the NI20 condition. Donations in the latter do not reveal any information on the principal's type as they are exogenously assigned. We thereby exclude reciprocal responses, while donations still might affect effort due to distributional and warm glow concerns. In NI20, we expect average efforts to be lower after donations and higher when nothing is donated when compared to I20.

Donation Size

A comparison of the distribution of efforts between the high (d=40 in Treatments I40/NI40) and low donation (d=20 in I20/NI20) treatments allows us to gain further insights into the role of warm glow and distributional effects and study the agents' sophistication in understanding the signaling mechanism.

First, note that both warm glow and distributional concerns predict higher efforts after a donation. In particular, the model predicts the existence of "spikes" in the distribution of efforts: At an effort level $e = \frac{d}{2}$ the effort fully "finances" the donation, and at this point any potential "warm glow" utility should break off. Hence, an agent with a strong warm glow motive can at most be driven to choose an effort level of $e = \frac{d}{2}$. Furthermore, agents with strong distributional concerns choose the payoff equalizing effort $e = \frac{50+d}{3}$. Note that these spikes should vary with the size of the donation *d*.

Concerning the signaling mechanism, the model shows that size of the donation may affect the strength of the reciprocal reactions due to its possible effect on the signal strength. In this context, it is instructive to use Bénabou and Tirole's (2012) notion of "normal" and "admirable" prosocial acts. If the donation is small, many people donate and the donation becomes "normal." In this case, a donation is not a particularly strong positive signal of prosociality, but a non-donation is a negative signal (generating "stigma"). On the other hand, if the donation is high, fewer people may donate and the donation becomes "admirable," i.e., it acts as a stronger positive signal about the principal's prosociality (generating “honor”). Thus, if we observe that many principals donate, efforts after a donation should only be weakly larger in the baseline I20 treatment than in NI20 (where donations are exogenously assigned). When there is no donation, efforts should be reduced to a stronger extent in I20 compared to NI20. If, however, donations are less frequent, as we expect for the high donation conditions, we should observe the opposite pattern.

Wealth of the Principal

In a further treatment variation, we raise the wealth of principals such that they are always richer than the agents. This variation helps us to distinguish between warm glow and distributional concerns as drivers of agents' effort choices. If the agents receive warm glow utility from “buying into” the principal's donation, we should still observe positive efforts after a non-intentional (i.e., exogenously assigned) donation by a rich principal. However, if this indirect warm glow motive does not matter (or is not very strong) but instead distributional concerns drive behavior in our baseline treatments, an agent would not exert effort even after a donation as higher efforts only further increase payoff inequality in our treatments. As Charness and Rabin (2002) have shown, social concerns towards another player tend to vanish when this other player is wealthier than oneself.

Strategic or Non-Strategic Donations

Finally, we argue that in non-strategic donation conditions a lower number of principals will donate and expected efforts after both a donation and a non-donation will then be larger. As the principal does not know about the second stage of the game and thus extrinsic benefits are not taken into account (i.e., the instrumental value that a donation creates higher efforts), the overall incentives to donate are weaker. In turn, fewer principals donate and those who do are more prosocial, leading to larger efforts after a donation. However, because fewer principals donate, those who do not donate should also be more prosocial on average. A non-donation is less of a negative signal on the principal's type, and agents should react to this with higher effort than in the baseline.

2.4 Results

We now present the results concerning these predictions, starting with the baseline treatment, and then move on to the other treatment variations to disentangle the underlying behavioral mechanisms step by step.

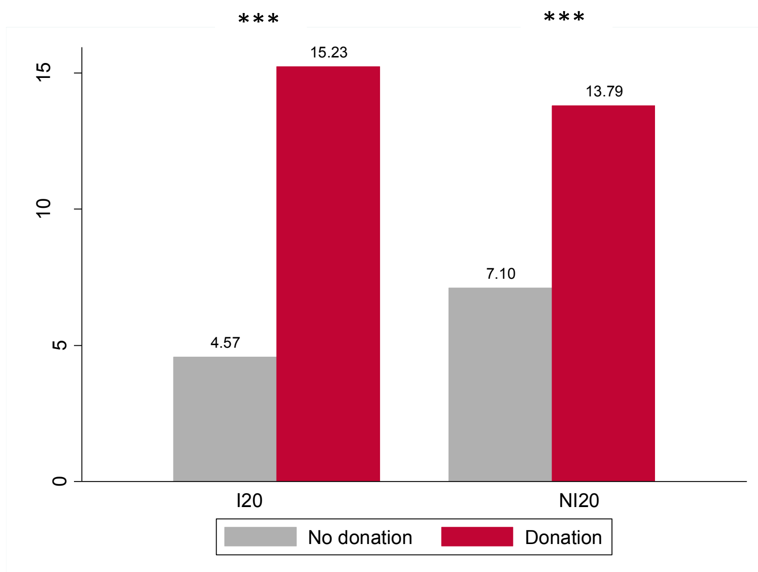
2.4.1 Do donations increase efforts?

In the baseline condition I20, a substantial fraction of principals decides to donate: 76.67% of the principals (23 out of 30) transfer 20 ECU to the charity. Thus, in the baseline setting, most of the principals are either sufficiently prosocial or believe that charitable giving has an instrumental value by raising agents' efforts.

When it comes to the agents' behavior, we find that, although effort is costly, the average transfer levels in I20 are positive (see Figure 2.1 for the average effort levels). Importantly, in line with our predictions, agents transfer substantially more to a principal when he has given to the charity: In I20, the average effort amounts to 4.57 after a non-donation and increases to 15.23 if the principal donated. The difference is highly statistically significant, with $p<0.001$ (two-sided Wilcoxon Matched Pairs Signed Ranks test (WMPSR)).[11] In this treatment, 22 out of 30 agents reacted positively to a donation, while 7 were neutral (i.e., chose the same effort level irrespective of the donation) and 1 subject reacted negatively. At this point, the positive reaction to a donation in I20 may be driven by all of the defined mechanisms: agents' distributional concerns, indirect warm glow giving, or reciprocal altruism. In the following sections, we aim to disentangle these mechanisms.

[11] All of the tests we report in this study are two-sided if not otherwise noted.

Figure 2.1: Average effort in lower donation treatments



Note: *** indicates significance on the 1%-level.

2.4.2 Do intentions matter?

To see whether and to what extent intentions matter for the agents' reactions, we compare efforts in I20 to NI20. We find that in the non-intentional condition NI20 the donation leads to a significantly higher effort than a non-donation ($p<0.001$, WMPSR). Hence, reciprocal altruism alone cannot explain positive agent's reactions in I20. The impact of the principal's decision is reduced when the donation cannot signal information about his type. First, we find that in NI20 agents transfer slightly less in the event of a donation than in I20 (13.79 instead of 15.23; or 9.5% lower), but the difference is not significant ($p=0.591$ in a two-sided Mann Whitney U test (MWU)). Second, after a non-donation, at 7.10 the average effort in NI20 is 55.36% larger than in I20. This difference is significant on the 5% level ($p=0.030$, MWU). Hence, in I20 principals are punished for their selfish decision and slightly rewarded, if they decide to donate. The result is in line with the experiments by Offerman (2002), in which he also finds more evidence for negative reciprocal responses than positive ones.

It follows that intentionality matters for understanding why charitable giving affects employee incentives. An intentional donation choice reveals information about the principal's type, and thus the agent's reaction is different than in exogenously-assigned donation conditions, even though the payoff consequences are identical.

2.4.3 (How) does the size of the donation affect behavior?

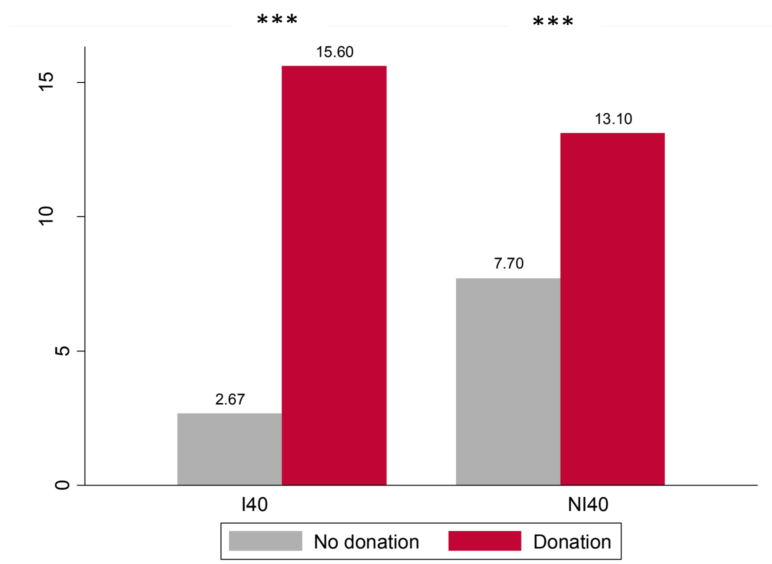
In a next step, we study the I40/NI40 treatments, in which the size of the donation is increased to 40. First, in the I40/NI40 treatments we observe very similar effort patterns as in the baseline I20/NI20 condition (see Figure 2.2). A donation leads to a higher effort than a non-donation, and the within-treatment differences are highly statistically significant, with p<0.001, WMPSR, in both I40 and NI40. Second, agents again react more negatively to a non-donation when the decision is endogenous: Efforts after a non-donation are 65.32% lower in I40 than NI40 (2.67 instead of 7.70; p=0.039, MWU). Third, intentional donations are not substantially rewarded. In I40, the effort after a donation is 15.60 and thereby 16.03% higher than in NI40, where the average effort is 13.10, but the difference is not significant (p=0.321, MWU).

In the next step, we compare efforts between the low and high donation conditions. Surprisingly, we find that effort in the higher donation treatments does not differ from the effort exerted in the lower donation conditions. In I40 after a non-donation, the average agent exerts an effort of 2.67, which is not significantly different from the 4.57 in I20 (p=0.905, MWU). And if the principal donates in I40, agents exert an average effort of 15.60, which is very close to the average effort of 15.23 in I20 (p=0.739, MWU).[12] We also do not observe any significant differences between efforts in NI20 and NI40 either after a non-donation or after a donation (p>0.1, MWU), where the average efforts amount to 7.10 (13.79) and 7.70 (13.10) after a non-donation (donation) in NI20 and NI40, respectively.[13]

[12] Note that if, on average, the agents react equally to a donation of 20 and 40, overall efficiency is the same in both treatments.

[13] Tonin and Vlassopoulus (2015) and Imas (2014), also find no effect of the size of the social incentive on agents' performance. But note that in their settings the size of the piece rate paid to the social cause is varied rather than the size of the lump sum payment.

Figure 2.2: Average effort in higher donation treatments



Note: *** indicates significance at the 1%-level.

Concerning the principals' behavior, we find that in the I40 condition the fraction of donors is slightly reduced to 66.67 % as compared to the 76.67% in I20, indicating that fewer principals donate when donating is expensive. Yet the difference is not significant (p=0.567, Fisher's exact two-sided test). The signaling value of donating is apparently not strongly affected by the size of the donation, resulting in a similarity of effort reaction that is consistent with the similarity of the signaling value.[14]

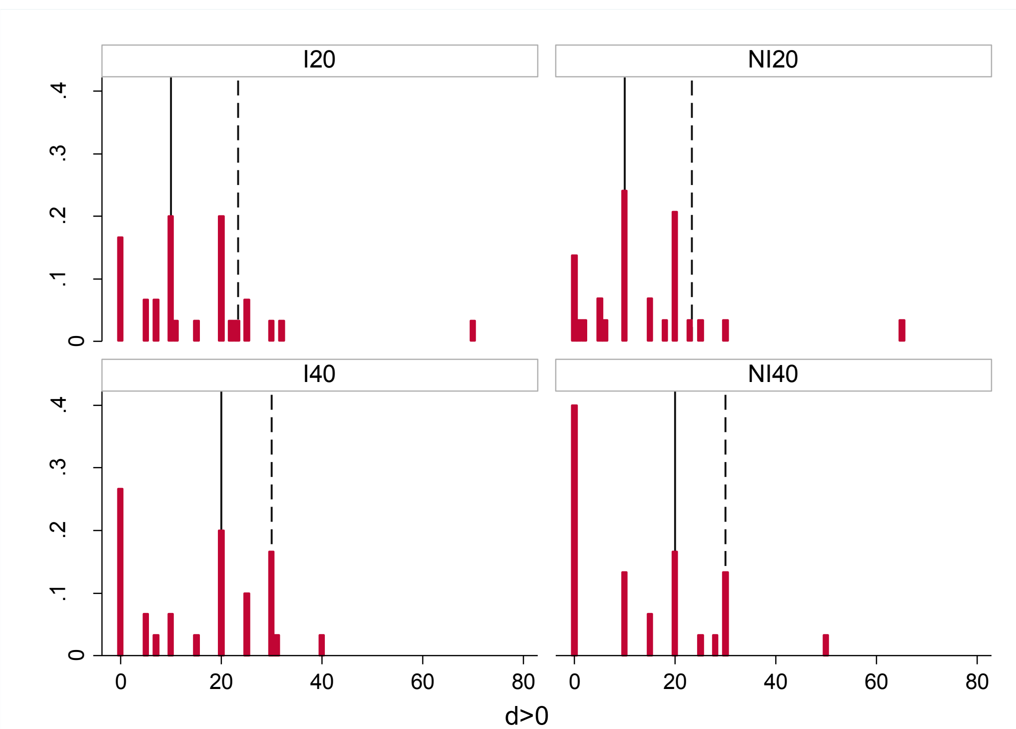
The result is more puzzling from the perspective of the warm glow and distributional concerns motives. In our model, the size of the donation affects the warm glow motive as it raises the scope to "buy into" the warm glow of the donation. It also affects distributional concerns by shifting the effort level at which equity is attained upwards. If either motive is sufficiently strong, we should see an upward shift in the

[14] In the model, we show that a separating equilibrium exists, in which a principal donates if and only if her "degree of prosocial orientation" is larger than a certain cut-off value. If the cut-off value is large, fewer people donate, and donations are a stronger positive signal of prosocial preferences. In other words, the fraction of donators and the signaling value of a donation are directly linked. If the fraction of donating principals does not differ to a strong extent, there can be no substantial difference in the signaling value of a donation.

level of effort. However, when looking at the aggregate results, we do not observe this, which could indicate that these motives are present but of weaker importance. In our model, for instance, the warm glow maximizing level of effort may be well below the "donation-financing" effort, even with a donation of 20. In that case, higher donations should not lead to higher efforts from this motivation.

It is therefore important to investigate whether we observe evidence for these motives when we compare the distribution of efforts between the high and low donation treatments. Figure 2.3 shows histograms for the efforts chosen in these treatments after a donation has been made. Our theoretical model predicted that the distribution of efforts should have three mass points: at 0 (for the very selfish agents), at $e = \frac{d}{2}$ (for those agents with strong warm glow preferences who "fully finance" the donation with their own efforts), and at the payoff equalizing effort $e = \frac{50+d}{3}$. These hypothesized mass points are marked with dotted ("equitable effort") and solid ("donation financing effort") vertical lines in the histograms.

Figure 2.3: Distribution of efforts



It is noteworthy that the donation financing efforts are indeed prominent in both the low donation treatment (where it is equal to 10) and the high donation treatments (where it is 20), a result well in line with the idea that warm glow motives matter. With respect to distributional concerns, while there is a mass point at the equitable effort for the high donation treatments, this equitable effort is not prominent in the low donation treatment. However, this may to some extent be due to the fact that the equitable effort is exactly 30 in the high-donation treatment but a less salient 23.33 in the low donation treatment, which may have led participants to pick 20 in the latter. Interestingly, the choice of zero is more frequent in the high donation treatments (p=0.032, Fisher's exact test), which explains why efforts after d=40 do not exceed efforts after d=20, even though both the equitable and the donation-financing efforts are higher and also frequently chosen by agents. One interpretation for this pattern is that agents apply a boundedly rational heuristic, such that they revert to an effort choice of 0 when the choice of both focal points is perceived as being too expensive in d=40 instead of choosing an effort level in between, as our model would predict.

Hence, we find some evidence of warm glow and distributional motives since we observe peaks at (or close to) both the donation-financing and the payoff-equalizing levels. However, these motives are not sufficiently strong to boost effort reactions when the donations are increased.[15] In a next step, we provide another test on the role of distributional and warm glow motives.

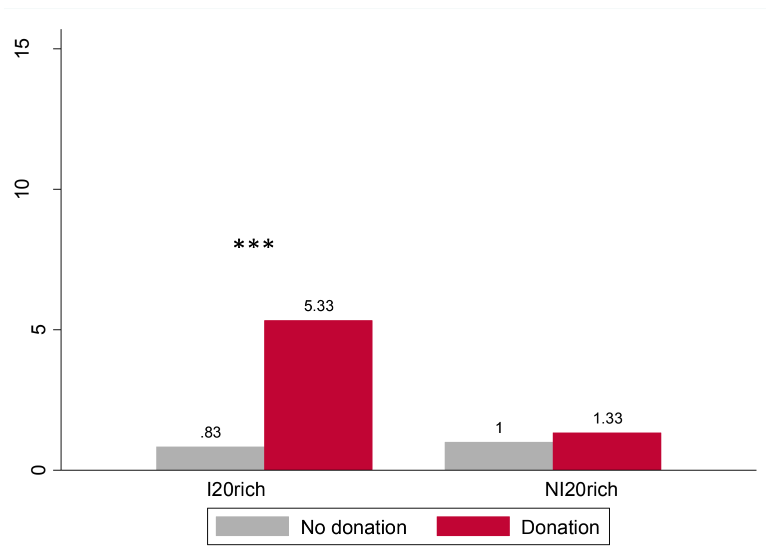
2.4.4 What happens when we "turn off" distributional concerns?

Note that in our treatments so far the agent was always better off than the principal at an effort level of zero. In turn, fairness concerns could motivate the agent to exert positive effort. Furthermore, donations create a stronger inequality to the disadvantage of the principals and thus should trigger higher efforts, if the agent is motivated to reduce

[15] Note that the observation that efforts do not substantially increase with the donation is consistent with the model if warm glow motives are not very strong (low η), the disutility of inequity $h(x_i - x_j)$ is a linear function and there are not many subjects with very strong prosocial inclinations θ_A. In this case, only the first case in the first order condition (3) matters, predicting positive efforts after a donation that do not vary with the size of this donation.

inequality. However, this is no longer the case when the principal has a higher initial wealth. Therefore, we conducted the I20rich and NI20rich treatments, in which the principal has an initial endowment that makes her wealthier than the agent even when the effort level is zero. In this case, equity concerns are eliminated as a driving force of higher efforts – higher efforts then only would increase inequity to the further disadvantage of the agent (see also subsection 2.6.3 in the formal model). On the other hand, warm glow motives may still play a role since agents could still benefit from warm glow utility when buying into the principal's donation irrespective of the principal's wealth.

Figure 2.4: Average effort in rich principals' treatments



Note: *** indicates significance at the 1%-level.

We find that agents exert virtually no effort if there are no donations or if the donation decision has been assigned exogenously, as Figure 2.4 shows. When there is no donation, we observe an average effort of 1.00 in the "non-intentional" NI20rich condition, which is not significantly different from the 0.83 in the "intentional" I20rich condition ($p=0.986$, MWU) and also not significantly different from the average effort of 1.33 when a donation was exogenously assigned. Efforts are therefore significantly lower in NI20rich than NI20 in both cases after a donation and non-donation, with p<0.001, MWU. Hence, in line with the role of distributional concerns an increase in the principal's wealth to a level where she is always richer than (or at least as rich as) the agent eliminates the agent's incentives to exert effort when the donation is exogenously assigned. Moreover, the low effort in NI20rich also shows that the aim to "buy into" the warm glow of giving cannot be a strong motive since this motive should be unaffected by the principal's wealth.[16]

But intentional donations still have a significant effect on the agents' motivation. After an intentional donation, agents on average choose an effort level of 5.33, which is significantly higher than both the effort when there is no donation (p<0.001, WMPSR) and the effort when the donation is non-intentional (p=0.023, MWU). It follows that the most robust motive for higher efforts after a donation is reciprocal altruism/indirect reciprocity: Agents are willing to exert higher efforts to generate profits for principals even when these principals are richer than themselves (and become more so with every unit of effort exerted) if these principals have made a costly donation decision on their own volition.[17]

Finally, note that the absence of distributional concerns also substantially reduces overall efforts in the intentional condition. The effort of 5.33 exerted in I20rich after a donation is substantially lower than the corresponding effort level of 15.23 in I20, where the principal was poor (p<0.001, MWU).

Looking at the principals' behavior, we find that in I20rich a substantial fraction[18] of 83.33% of the principals donate, which is slightly but not significantly higher than in I20 and I40 (p>0.1, Fisher's exact test).[19] Again, as the majority of people

---

[16] Another interpretation (not captured by our model) is that warm glow motives matter but are outweighed by envy, i.e., marginal returns to a warm glow buy in may exist but are smaller than the marginal utility loss from increasing disadvantageous inequity.

[17] Reciprocal altruism thus matters even towards wealthier individuals.

[18] 58.62 %, 60 % and 50 % of principals donate in NI20, NI40 and NI20rich, respectively. The fractions of donors do not differ significantly between random and endogenous donation treatments in two-sided Fisher's exact tests (p>0.1).

[19] We asked agents after they made their effort choice what fraction of principals do they think donates. The beliefs were not incentivized. The beliefs are 53.27%, 37.40% and 62.33% in I20, I40 and I20rich, respectively, whereas the difference in the beliefs is slightly significant between I20 and I40 (p=0.068, MWU), significant between I40 and I20rich (p=0.002, MWU) and not significant between I20 and I20rich (p=0.24, MWU).

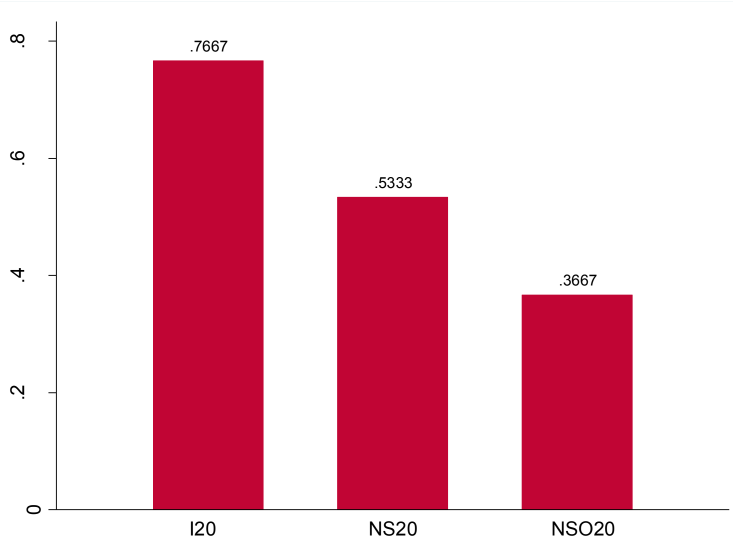
donates, donations are rather “normal” and therefore should not be a particularly strong positive signal of prosociality. It is important to note that even if non-donations thus should lead to a stronger “stigma”, here it cannot lead to a more pronounced effort reduction as agents already choose an effort level close to zero when the donation is exogenously assigned.

2.4.5 Are non-strategic donations rewarded?

In this final section, we present our results from control treatments in which we study how agents react to non-instrumental donations. We argued that more principals should donate in the baseline condition than in the non-strategic treatments since in the baseline treatment – where principals are aware that agents learn about their donation prior to the effort choice – principals also have an instrumental motive to donate in order to raise agents’ efforts. In contrast, in the non-strategic treatments, strategic concerns are ruled out to a stronger (NSO20 – here principals made their donation decisions in a prior one-shot experiment in which they only made a donation) or weaker (NS20 – here principals were informed that there was a second stage of the game but did not know what happened at this stage) extent. Also, if principals know that agents observe their decision, image motivation may lead to a positive donation. Similarly, Ariely et al. (2009) show that subjects make significantly higher donations to a renowned charity when this is publicly observed. However, this channel is less relevant in our experimental setting since the relationship between the principal and agent is anonymous.

As expected, we find that fewer principals donate in the non-strategic treatments, with the fraction of donors decreasing from 76.67 % in I20 to 53.33 % in NS20 and to 36.67 % in NSO20 (see Figure 2.5). When we compare the differences between the fractions in all three conditions with a Jonckheere-Terpstra test, we find that they have a significant descending ordering (p=0.001, one-sided). Comparing the treatments pairwise, we find no statistically significant difference in principals’ behavior between I20 and NS20 (p=0.103, a two-sided Fisher’s exact test) or NS20 and NSO20 (p=0.299, Fisher’s exact test) but a highly significant difference between I20 and NSO20 (p=0.004, Fisher's exact test). Hence, principals anticipate that a donation may pay off and thus have an instrumental motive to trigger a high effort by choosing a donation. Approximately half of the donating principals in I20 thus donate strategically. While in NSO20 only 11 out of 30 donate, in I20 this number increases to 23 out of 30.

Figure 2.5: Fraction of donors in the baseline and non-strategic conditions



We also predicted that agents react to a stronger extent to a non-strategic donation because it should be a more "pure" and thus more credible signal about the principal's prosociality. However, this hypothesis is not supported by the data: We observe that in NS20 (NSO20) agents choose an average effort of 3.10 (2.87) after a non-donation and 14.10 (12.40) after a donation. Thus, while there is again a large within-treatment difference between agents' reactions to a donation and non-donation ($p<0.001$, WMPSR in NS20 and NSO20), we do not find any differences between the strategic and the non-strategic treatments (all $p>0.1$, MWU). Hence, while a substantial fraction of principals donates strategically, agents do not change their behavior when moving from the baseline to the non-strategic conditions.

2.5 Discussion and Concluding Remarks

We conducted a set of laboratory experiments to study the effects of principals' charitable giving on agents' effort in a principal-agent game and to disentangle the motives driving agents' reactions. Overall, we find that agents choose significantly higher effort levels after a donation. This effect is robust against treatment variations, such as changing the instrumentality or the size of a donation. The positive performance effect of donations disappears only when the principal is wealthy and the donation is not intentionally made by the principals.

Regarding the mechanisms driving agents' behavior, we find robust evidence for the importance of reciprocal altruism. Specifically, an intentional choice not to donate leads to significantly lower efforts (as compared to a situation where this is assigned exogenously) when the principal is not wealthy, and an intentional donation led to significantly higher efforts when the principal is wealthy. Distributional concerns and indirect warm glow motives, on the other hand, are of weaker importance.

Furthermore, when analyzing the instrumentality of the donations, we find that about half of the donating principals donate strategically, i.e. they donate only when they know that their choice is observed by an agent who then can exert efforts that affect the principals' profits. Interestingly, the agents' effort is not affected by the principals' strategic concerns since agents do not react differently regardless of whether the principals' donations had an instrumental value or were purely altruistic. In particular, the latter observation is important for CSR activities in firms: Our results indicate that CSR can affect employee motivation even when the true underlying motives may to some extent be instrumental – as will often be the case in companies driven by profit or shareholder value maximization motives.[20]

On a broader methodological level, our observations hint at the importance of asymmetries in wealth in lab experiments on employment relations. In most real world settings, employers are substantially wealthier than their employees, but in most lab experiments wealth is rather symmetrically distributed (or agents even have a higher

[20] See, for instance, Barron (2007) for a formal analysis of strategic and non-strategic CSR motives from an investor's perspective.

endowment ex-ante). As we have found, this substantially alters patterns of effort reactions. Our results show that positive intentions play a much bigger role when principals are wealthier. In our experiment reciprocal altruism is thus the most robust driving force behind performance effects of charitable acts by a principal. We should therefore expect that this motive is also the strongest driver of positive reactions to social activities by real world employers.

2.6 Appendix A: Theoretical Analysis

2.6.1 The Model

We develop and analyze a formal model as a conceptual framework to capture several mechanisms that may affect the way in which agents react to social activities of their employers. The main idea of the formal model is to apply standard approaches discussed in the behavioral economics literature while at the same time limiting the number of parameters used. The model incorporates reciprocal altruism (as in Levine, 1998), distributional concerns (as in Fehr and Schmidt, 1999, or Charness and Rabin, 2002), and warm glow motives of impure altruism (based on Andreoni, 1989, and Andreoni, 1990). In the experiment we aim to disentangle these motives.

It is instructive to first consider an actor in a two person game who has a utility function of the following form

$$u_i(x_i, x_j) = x_i - \frac{\theta_i + \lambda\theta_j}{1+\lambda} h(x_i - x_j)$$

where x_i and x_j are the material payoffs both players obtain from their interaction and

$$h(\Delta) = 0 \text{ if } \Delta \leq 0,$$
$$h'(\Delta) > 0 \text{ if } \Delta \geq 0 \text{ and}$$
$$h''(\Delta) \geq 0.$$

The degree of prosocial preferences of an actor i is captured by a single parameter θ_i. Note that the model incorporates both reciprocal altruism and distributional preferences: A player is inequity averse as his utility is lower when the other player is worse off. A player is also a reciprocal altruist (as in Levine, 1998) as he cares more for another player when the other player is a more prosocial type (i.e., θ_j is large), which he may infer from this other player's decisions. Note that the model encompasses specific cases of standard behavioral economics approaches: When $h(\Delta)$ is linear and $\lambda = 0$ the model boils down to the linear social welfare preferences model from Charness and Rabin (2002), in which an agent cares for another agent if this other agent is poorer than herself. It is then also equivalent with a Fehr and Schmidt (1999)-type model with $\alpha = 0$ and thus well in line with key observations in the empirical analysis of Charness and Rabin (2002), who summarize that their "*(...) regression analysis indicates that a B who has a higher payoff than A puts great weight on A's payoff. However, if B has a lower payoff than A and no reciprocity is involved, the weight on A's payoff is close to 0. When A has mistreated B, B significantly decreases positive weight or puts negative weight on A's payoff*".[21]

Following Levine (1998), we assume that the θ of the fellow players are initially unknown and follow some continuous distribution with cdf $F(\theta)$ and unlimited support (which allows for spitefulness). In that case, player i infers information about θ_j from player j's actions.

The game we study closely follows the design of the later experiment. An agent A is matched to a principal/firm P. Both have the type of equity & intentions utility function outlined above with (unobservable) preference parameters θ_A and θ_P. The agent receives an (exogenously given) wage w. Ex-ante the firm has a wealth level of W and can give an amount $d \in \{0, D\}$ to a charity, which is publicly observable.[22] After the

---

[21]The difference to Charness and Rabin (2002) is that we capture reciprocity in a different manner. In our model, reciprocity follows Levine's (1998) idea that positively reciprocal behavior occurs because actors make inferences about other person's types from their actions.

[22] In real world settings it may not always be easy for employees to observe social activities of their employers perfectly and thus signals of prosociality may be somewhat noisy. De Haan, T., Offerman, T., & Sloof (2011) theoretically and experimentally investigate noisy signaling and show that separating (and

agent has observed the donation decision d he invests an effort $e \in [e, e_{\max}]$ at linear cost e. The firm receives payoff ke_i, with $k > 1$. We impose one additional restriction on the parameters, namely that $d < w$, i.e., the donation is smaller than the wage.[23] Moreover, we initially assume that *2w>W,* i.e., the principal is not too wealthy (in subsection 3.2 we will discuss the case in which *W* is larger).

Both the principal and the agent may also care for the benefit of the charity. This is captured by treating the charity as a third party that may add a further social welfare utility, whereby we also consider one further motive that can play a role in our context, namely that donations can be driven by "impure altruism" as people enjoy the donation as such beyond its true charitable value. We extend this idea to allow for the possibility that an employee receives utility from "buying into" a donation made by their employer. To be specific, the principal's expected utility is

$$W + k \cdot e - w - d - \frac{\theta_P + \lambda\theta_A}{1+\lambda} h((W + ke - w - d) - (w - e)) + \frac{\theta_P + \lambda\theta_C}{1+\lambda} \cdot g(d) \tag{1}$$

and the agent's is

$$w - e - \frac{\theta_A + \lambda\theta_P}{1+\lambda} h((w - e) - (W + ke - w - d)) + \eta \cdot \frac{\theta_A + \lambda\theta_C}{1+\lambda} g(\min\{ke, d\}). \tag{2}$$

where the last part in both functions captures the potential "warm glow" from participating in the donation (with $g(0) = 0$ and $g'(x) > 0$ and $g''(x) < 0$ for $x \geq 0$) and θ_C measures the "social worth" of the charity. Note that the principal's preference for d may encompass "pure" as well as "impure" motives. But as the agent does not affect the size of d and this already has been determined at the agent's choice, node *g()* entirely captures "impure" warm glow motives. The parameter η measures to what extent agents

thus credible signaling) occurs in the experiment also for low levels of noise where separating equilibria do not exist in their theoretical model.

[23]This in particular implies that the effort that equalizes the principal's and agent's payoffs $\frac{2w-W+d}{1+k}$ is larger than the effort $\frac{d}{k}$ which is chosen when the agent tries to completely compensate the donation with his effort.

care for sharing this "warm glow." Finally, note that we assume that more prosocial agents experience a stronger warm glow.[24]

2.6.2 Equilibrium Analysis

In the model, we analyze how employers' publicly observable social actions can affect agents' behavior. Note that three effects may here come into play: First of all, a donation may simply alter the distributional set-up and thus will affect the agent's fairness concerns. In addition, an employer may try to reveal though charitable actions that she is a prosocial type, in turn triggering a stronger identification by the agent with the employer's outcomes. Finally, when making a donation, an employer creates the possibility that the employee buys into the warm glow of this donation by exerting effort. As the principal's actions may reveal information about her type, the game is a signaling game, and we characterize Perfect Bayesian Equilibria.

We start by analyzing an agent's effort choice for a given donation. The agent's expected utility becomes

$$u_A\left(e,\theta_A,\hat{\theta}_P,d\right)=w-e-\tfrac{\theta_A+\lambda\hat{\theta}_P}{1+\lambda}h\left(2w-W-(1+k)e+d\right)+\eta\tfrac{\theta_A+\lambda\theta_C}{1+\lambda}g\left(\min\{ke,d\}\right)$$

where $\hat{\theta}_P = E\left[\theta_P|d\right]$ denotes the agent's expectation about the principal's type. Note that $e=\frac{2w-W+d}{1+k}$ is the effort level that equalizes the payoffs of principal and agent such that equity concerns do not affect the agent's marginal utility above this point. The first derivative of the agent's objective function with respect to e is[25]

---

[24]Most of the theoretical results do not rely on $\eta > 0$.

[25]The second derivative is

$$\frac{\partial Eu_A}{\partial e}=\begin{cases}-\frac{\theta_i+\lambda E[\theta_P|d]}{1+\lambda}\cdot h''\left(2w-W-(1+k)e+d\right)(1+k)^2+\eta\cdot\frac{\theta_i+\lambda\theta_C}{1+\lambda}g''(ke)k^2 & if \quad e\leq\frac{d}{k} \\ -\frac{\theta_i+\lambda E[\theta_P|d]}{1+\lambda}\cdot h''\left(2w-W-(1+k)e+d\right)(1+k)^2 & if \quad e>\frac{d}{k}\end{cases}.$$

and, hence, the function is strictly concave below $\frac{2w-W+d}{1+k}$.

$$\begin{cases} -1+\frac{\theta_A+\lambda\hat{\theta}_P}{1+\lambda}h'(2w-W-(1+k)e+d)(1+k)+\eta\frac{\theta_A+\lambda\theta_C}{1+\lambda}g'(ke)k & if \quad e\le\frac{d}{k} \\ -1+\frac{\theta_A+\lambda\hat{\theta}_P}{1+\lambda}h'(2w-W-(1+k)e+d)(1+k) & if \quad \frac{d}{k}<e<\frac{2w-W+d}{1+k}. \\ -1 & if \quad e\ge\frac{2w-W+d}{1+k} \end{cases} \qquad (3)$$

The agent thus never exerts an effort beyond the payoff equalizing level of $\frac{2w-W+d}{1+k}$. We can now characterize the agent's reaction function

$$e^*(\theta_A,\hat{\theta}_P,d)=\arg\max_e u_A(e,\theta_A,\hat{\theta}_P,d)$$

and, in turn, the expected efforts $E_{\theta_A}\left[e^*(\theta_A,\hat{\theta}_P,d)\right]$ as a function of the donation d and the agent's belief about the principal's type $\hat{\theta}_P$.

Proposition 1 *Expected efforts* $E_{\theta_A}\left[e^*(\theta_A,\hat{\theta}_P,d)\right]$ *increase according to the donation* d *and the agent's expectation about the principal's prosociality (*$\hat{\theta}_P$*). For a given donation* d *the distribution of efforts has the support* $\left[0,\frac{2w-W+d}{1+k}\right]$ *with three mass points at* $e=0$, $e=\frac{d}{k}$ *and* $e=\frac{2w+W+d}{1+k}$.

Proof: First, we characterize the function

$$e^*(\theta_A,\hat{\theta}_P,d)=\arg\max_e u_A(e,\theta_A,\hat{\theta}_P,d)$$

As $u_A(e,\theta_A,\hat{\theta}_P,d)$ is strictly concave, the optimal effort $e^*>0$ if and only if

$$\frac{\theta_A+\lambda\hat{\theta}_P}{1+\lambda}\cdot h'(2w-W+d)(1+k)+\eta\cdot\frac{\theta_A+\lambda\theta_C}{1+\lambda}g'(0)k>1$$

$$\Leftrightarrow \theta_A>\frac{1+\lambda-\lambda\hat{\theta}_P h'(2w-W+d)(1+k)-\lambda\theta_C\eta g'(0)k}{(h'(2w-W+d)(1+k)+\eta g'(0)k)}\equiv\theta_0^d.$$

and otherwise $e^*=0$. Moreover, agent will choose an effort level of $e^*=\frac{d}{k}$ if the left derivative at this point is positive and the right derivative is negative. The left derivative is positive if

$$\frac{\theta_A + \lambda\hat{\theta}_P}{1+\lambda} \cdot h'\left(2w - W - (1+k)\frac{d}{k} + d\right)(1+k) + \eta \cdot \frac{\theta_A + \lambda\theta_C}{1+\lambda} g'(d)k > 1$$

$$\Leftrightarrow \theta_A > \frac{1 + \lambda - \lambda\hat{\theta}_P h'(2w - W - (1+k)\frac{d}{k} + d)(1+k) - \eta\lambda\theta_C g'(d)k}{h'(2w - W - (1+k)\frac{d}{k} + d)(1+k) + \eta g'(d)k} \equiv \theta_1^d$$

and the right derivative is negative if

$$1 > \frac{\theta_A + \lambda\hat{\theta}_P}{1+\lambda} \cdot h'\left(2w - W - (1+k)\frac{d}{k} + d\right)(1+k)$$

$$\Leftrightarrow \theta_A < \frac{1+\lambda}{h'(2w - W - (1+k)\frac{d}{k} + d)(1+k)} - \lambda\hat{\theta}_P \equiv \theta_2^d.$$

It is straightforward to see that $\theta_1^d < \theta_2^d$ and thus the respective interval is non-empty. An agent will choose an effort level of $e^* = \frac{2w - W + d}{1+k}$ if

$$-1 + \frac{\theta_A + \lambda\hat{\theta}_P}{1+\lambda} \cdot h'\left(2w - W - (1+k)\frac{2w - W + d}{1+k} + d\right)(1+k) > 0$$

$$\Leftrightarrow \theta_A > \frac{1+\lambda}{h'(0)(1+k)} - \lambda\hat{\theta}_P \equiv \theta_3^d.$$

Hence, when $\theta_0^d < \theta_A < \theta_1^d$ the effort level is uniquely characterized by

$$\frac{\theta_A + \lambda\hat{\theta}_P}{1+\lambda} \cdot h'(2w - W - (1+k)e^* + d)(1+k) + \eta \cdot \frac{\theta_A + \lambda\theta_C}{1+\lambda} g'(ke^*)k = 1 \tag{4}$$

and when $\theta_2^d < \theta_A < \theta_3^d$ the effort is characterized by

$$\frac{\theta_A + \lambda\hat{\theta}_P}{1+\lambda} \cdot h'\left(2w - W - (1+k)e^* + d\right)(1+k) = 1. \quad (5)$$

The function $e^*\left(\theta_A, \hat{\theta}_P, d\right)$ is thus defined by

$$\begin{array}{lll}
e^* = 0 & \text{if} & \theta_A \le \theta_0^d \\
\frac{\theta_A+\lambda\hat{\theta}_P}{1+\lambda} h'\left(2w - W - (1+k)e^* + d\right)(1+k) + \eta \frac{\theta_A+\lambda\theta_C}{1+\lambda} g'\left(ke^*\right)k = 1 & \text{if} & \theta_0^d < \theta_A \le \theta_1^d \\
e^* = \frac{d}{k} & \text{if} & \theta_1^d < \theta_A \le \theta_2^d \\
\frac{\theta_A+\lambda\hat{\theta}_P}{1+\lambda} h'\left(2w - W - (1+k)e^* + d\right)(1+k) = 1 & \text{if} & \theta_2^d < \theta_A \le \theta_3^d \\
e^* = \frac{2w-W+d}{1+k} & \text{if} & \theta_A > \theta_3^d
\end{array}$$

The function is continuous as at each interval boundary the left hand side optimum converges to the right hand side optimum: To see this, just note that the interval boundaries are defined such that the first order conditions are equal to zero at $e^* = 0,\ \ e^* = \frac{d}{k}$ or $e^* = \frac{2w-W+d}{1+k}$, respectively.

The left hand sides of the second and fourth condition are both increasing in d, θ_A , and $\hat{\theta}_P$. And both $\frac{d}{k}$ and $\frac{2w+d}{1+k}$ are increasing in d. Hence, $e^*\left(\theta_A, \hat{\theta}_P, d\right)$ is non-decreasing in each of its arguments and, in turn, the expected effort

$$E_{\theta_A}\left[e^*\left(\theta_A, \hat{\theta}_P, d\right)\right] = \int_{-\infty}^{\infty} e^*\left(\theta_A, \hat{\theta}_P, d\right) dF(\theta_A)$$

is also increasing in d and $\hat{\theta}_P$.

□

Hence, even for fixed beliefs about the principal's type, the donation has a direct impact on effort for two reasons: (i) The donation alters the income distribution in favor of the agent, who then raises efforts to increase distributional fairness; and (ii) higher donations increase the scope for warm glow participation by the agent.

Furthermore, note that even though the underlying type distribution is continuous and has unlimited support, the efforts chosen have three mass points (as the same effort level is chosen by all agents who have a θ_A in a certain interval). The first spike is at $e = 0$ as all agents with very low prosocial preferences do not exert any efforts. The next spike is at $e = \frac{d}{k}$ where the effort fully "finances" the donation since at this point the "warm glow" utility breaks off. The third spike is at the payoff equalizing effort $e = \frac{2w-W+d}{1+k}$ as individuals with very strong social preferences are willing to exert effort until the principal and agent are equally well off but never beyond.

Moreover, the expected efforts increase along with the agent's expectation about the principal's type $\hat{\theta}_P = E[\theta_P | d]$, and this creates a third motive for an incentive impact of observable charitable giving: The principal may have an incentive to signal that she is a prosocial type. To study this motive in more detail and to fully characterize Perfect Bayesian Equilibria of the game, we have to analyze the principal's behavior.

First, recall that $e^*(\theta_A, \hat{\theta}_P, d)$ never exceeds the "equitable effort" $\frac{2w-W+d}{1+k}$. Therefore, the principal's utility (1) simplifies to $W + k \cdot e - w - d + \frac{\theta_P + \lambda\theta_C}{1+\lambda} \cdot g(d)$ such that her expected utility is

$$E_{\theta_A}\left[W + k \cdot e^*\left(\theta_A, \hat{\theta}_P^d, d\right) - w - d + \frac{\theta_P + \lambda\theta_C}{1+\lambda} \cdot g(d)\right].$$

We now determine P's optimal choice given A's beliefs and best response. Let $\hat{\theta}_P^D = E_A[\theta_P | d = D]$ and $\hat{\theta}_P^0 = E_A[\theta_P | d = 0]$ the agent's beliefs about the principal's type on the equilibrium path. The principal donates to the charity if

$$E_{\theta_A}\left[W + ke^*\left(\theta_A, \hat{\theta}_P^D, D\right) - w - D + \frac{\theta_P + \lambda\theta_C}{1+\lambda} g(D)\right] > E_{\theta_A}\left[W + ke^*\left(\theta_A, \hat{\theta}_P^0, 0\right) - w\right]$$

but this is equivalent to

$$\theta_P > (1+\lambda)\frac{D - k\left(E_{\theta_A}\left[e^*\left(\theta_A, \hat{\theta}_P^D, D\right)\right] - E_{\theta_A}\left[e^*\left(\theta_A, \hat{\theta}_P^0, 0\right)\right]\right)}{g(D)} - \lambda\theta_C \equiv \overline{\theta}_P. \tag{6}$$

Hence, any Perfect Bayesian Equilibrium has the form such that P donates if θ_P is larger or equal than a cut-off value $\overline{\theta}_P$ and the agent's posterior belief on the principal's type is

$$E_A\left[\theta_P \middle| d = D\right] = E\left[\theta_P \middle| \theta_P \geq \bar{\theta}_P\right] \equiv \hat{\theta}_P^D \text{ and}$$
$$E_A\left[\theta_P \middle| d = 0\right] = E\left[\theta_P \middle| \theta_P < \bar{\theta}_P\right] \equiv \hat{\theta}_P^0.$$

Hence, we have obtained the following result:

Proposition 2 *In any Perfect Bayesian Equilibrium, firms donate to charity if and only if they are sufficiently prosocial (i.e., if* θ_P *larger than a cut-off value* $\bar{\theta}_P$*). The agent's beliefs about the prosociality of the principal and thus the expected efforts are strictly higher when a principal has given to charity.*

Proof: The condition that sufficiently prosocial principals donate in any PBE directly follows from condition (4).

To show that such equilibria always exist, we need to show that the function

$$(1+\lambda)\frac{D-k\left(E_{\theta_A}\left[e^*\left(\theta_A, E\left[\theta_P \middle| \theta_P \geq \bar{\theta}_P\right], D\right)\right] - E_{\theta_A}\left[e^*\left(\theta_A, E\left[\theta_P \middle| \theta_P < \bar{\theta}_P\right], 0\right)\right]\right)}{g(D)} - \lambda\theta_C - \bar{\theta}_P$$

has at least one root. The function is continuous as $e^*\left(\theta_A, \hat{\theta}_P, d\right)$ is continuous (see the proof of Proposition 1). As

$$\begin{aligned}
&\lim_{\bar{\theta}_P \to -\infty} (1+\lambda)\frac{D-k\left(E_{\theta_A}\left[e^*\left(\theta_A, E\left[\theta_P \middle| \theta_P \geq \bar{\theta}_P\right], D\right)\right] - E_{\theta_A}\left[e^*\left(\theta_A, E\left[\theta_P \middle| \theta_P < \bar{\theta}_P\right], 0\right)\right]\right)}{g(D)} - \lambda\theta_C - \bar{\theta}_P \\
&= (1+\lambda)\frac{D-kE_{\theta_A}\left[e^*(\theta_A, E[\theta_P], D)\right]}{g(D)} + \infty \\
&= \infty
\end{aligned}$$

and

$$\begin{aligned}
&\lim_{\bar{\theta}_P \to \infty} (1+\lambda)\frac{D-k\left(E_{\theta_A}\left[e^*\left(\theta_A, E\left[\theta_P \middle| \theta_P \geq \bar{\theta}_P\right], D\right)\right] - E_{\theta_A}\left[e^*\left(\theta_A, E\left[\theta_P \middle| \theta_P < \bar{\theta}_P\right], 0\right)\right]\right)}{g(D)} - \lambda\theta_C - \bar{\theta}_P \\
&= (1+\lambda)\frac{D-k\left(\frac{2w+d}{1+k} - E_{\theta_A}\left[e^*(\theta_A, E[\theta_P], 0)\right]\right)}{g(D)} - \lambda\theta_C - \infty \\
&= -\infty
\end{aligned}$$

existence follows from the intermediate value theorem.

□

Hence, charitable giving is indeed a signal about the principal's type: Only for sufficiently prosocial principals do the benefits of charitable giving exceed the costs. In turn, agents exert higher efforts after observing a donation not only because of distributional concerns or to participate in the warm glow, but also because reciprocal altruists internalize the payoffs of a more prosocial principal to a stronger extent and the fact that a donation represents a credible signal of the principal's social preferences.

2.6.3 Disentangling Motives

In the lab experiment, it is important to disentangle the distributional, warm glow, and intention-based effects when explaining the agents' potential reactions to a donation. To do this, we consider following modifications of the model.

Exogenously Assigned Donations

To eliminate the signaling motive of donations, it is instructive to consider a benchmark case where donations do not reveal information on the principal's type as they are exogenously assigned. The agent's optimal effort choice is here:

$$e^*(\theta_A, E[\theta_P], d).$$

Proposition 1 then directly implies the following result:

Proposition 3 *When donations are exogenously assigned, average efforts are lower after donations and higher when nothing is donated:*

$$\begin{aligned} &E[e^*(\theta_A, E_A[\theta_P | d = D], D)] > E[e^*(\theta_A, E[\theta_P], D)] \\ &> E[e^*(\theta_A, E[\theta_P], 0)] > E[e^*(\theta_A, E_A[\theta_P | d = 0], 0)]. \end{aligned}$$

Hence, without an intentional act, donations still affect efforts due to distributional and warm glow effects. However, when the social activity is endogenously chosen, prosocial acts trigger a stronger positive effect on efforts, and the absence of the social activity has a stronger negative effect because it reveals that the principal has only weak prosocial preferences.

It is therefore important to study how strong the "signaling of intentions" affects beliefs and, in turn, affects efforts in both directions. To do that, it is instructive to consider a specific type distribution. Suppose that θ is normally distributed with mean m and variance σ^2. In that case[26]

$$E\left[\theta_P \middle| d = D\right] = m + \sigma \frac{\phi\left(\frac{m-\bar{\theta}_P}{\sigma}\right)}{\Phi\left(\frac{m-\bar{\theta}_P}{\sigma}\right)} \text{ and } E\left[\theta_P \middle| d = 0\right] = m - \sigma \frac{\phi\left(\frac{\bar{\theta}_P - m}{\sigma}\right)}{\Phi\left(\frac{\bar{\theta}_P - m}{\sigma}\right)}.$$

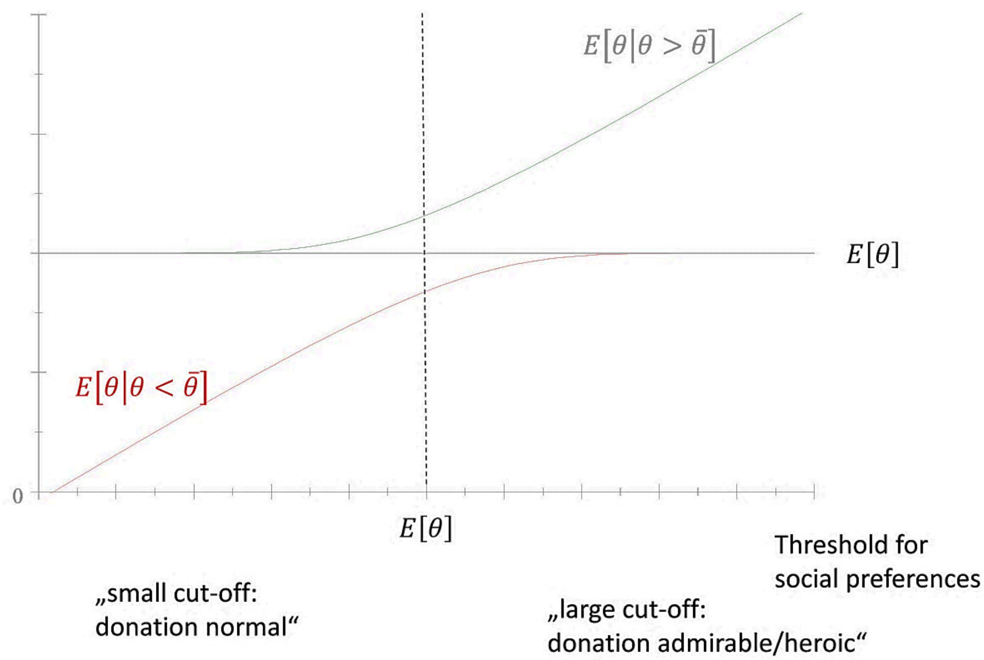
Hence, we can now plot the supervisor's reputation for prosociality as a function of the equilibrium cut-off, which is shown in Figure 2.6. The upper curve shows the reputation of a donating and the lower curve of a non-donating principal. If the cut-off is small, many people donate (in the language of Bénabou and Tirole (2012), the donation is then "normal"). In this case, a donation is not a particularly strong positive signal of prosociality, but a non-donation is a negative signal. On the other hand, if the cut-off is high, few people donate (in the language of Bénabou and Tirole (2012), the donation is then "admirable or heroic"). In that case, a non-donation is not a particularly negative signal, but a donation is a strongly positive signal about the principal's prosociality.

This leads to predictions for the experiment when we compare exogenously-determined with endogenous donations. If we observe that many principals donate, efforts after a donation should only be weakly larger when the donation is endogenous than in the baseline. But when there is no donation, efforts should be reduced to a stronger extent when this non-donation had been endogenously determined as compared to an exogenous imposition of the donation.

If, however, donations are rare, we should observe the opposite pattern. Efforts should be only weakly smaller after endogenous non-donations but larger after endogenous donations.

[26]This can be easily derived from R. 185(i) in Gourieroux and Monfort (1989), p. 526.

Figure 2.6: Reputation for Prosociality



Varying the principal's wealth

As laid out in the above, we vary the principal's wealth in a further experimental treatment. Note that in the model, we can eliminate distributional concerns as a driver of the agent's effort choices if

$$w - e \leq W + ke - w - d$$

even for *e=0* which is equivalent to

$$W \geq 2w + d\,.$$

In this case, the agent is always worse off than the principal, and higher efforts will only increase the disadvantageous inequity. Only warm-glow motives matter, and the first order condition of the agent's objective function is

$$\eta \tfrac{\theta_A + \lambda\theta_C}{1+\lambda} g'(ke)k = 1\,.$$

Hence, if we observe strictly positive effort levels in the experiment even if $W \geq 2w + d$ this would provide evidence for the importance of warm glow motives.

Non-Strategic Donations

We can also compare the outcome to another benchmark, namely when we consider the agent's reaction to purely non-instrumental donations. To do that we make a simple thought experiment supposing that the principal (erroneously) believes that the game is over after the donation. In that case, P will give to the charity if

$$\frac{\theta_P + \lambda\theta_C}{1+\lambda} \cdot g(D) > D$$

or

$$\theta_P > \frac{(1+\lambda)D}{g(D)} - \lambda\theta_C \equiv \breve{\theta}_P. \qquad (7)$$

We can compare this cut-off $\breve{\theta}_P$ to the equilibrium cut-off $\overline{\theta}_P$ as characterized in (4):

Proposition 4 *When the principal's donation decision is non-strategic, a lower number of principals will donate. The expected efforts after both a donation and a non-donation are then larger.*

Proof: The cut-off value $\breve{\theta}_P$ will exceed $\overline{\theta}_P$ if

$$\frac{(1+\lambda)D}{g(D)} - \lambda\theta_C > (1+\lambda)\frac{D-k\left(E_{\theta_A}\left[e^*\left(\theta_A, \hat{\theta}_P^D, D\right)\right]-E_{\theta_A}\left[e^*\left(\theta_A, \hat{\theta}_P^0, 0\right)\right]\right)}{g(D)} - \lambda\theta_C$$

$$\Leftrightarrow E_{\theta_A}\left[e^*\left(\theta_A, \hat{\theta}_P^D, D\right)\right] > E_{\theta_A}\left[e^*\left(\theta_A, \hat{\theta}_P^0, 0\right)\right]$$

which always holds due to Proposition 1. Hence,

$$E\left[\theta_P \middle| \theta_P \geq \breve{\theta}_P\right] > E\left[\theta_P \middle| \theta_P \geq \overline{\theta}_P\right] \text{ and}$$
$$E\left[\theta_P \middle| \theta_P < \breve{\theta}_P\right] > E\left[\theta_P \middle| \theta_P < \overline{\theta}_P\right]$$

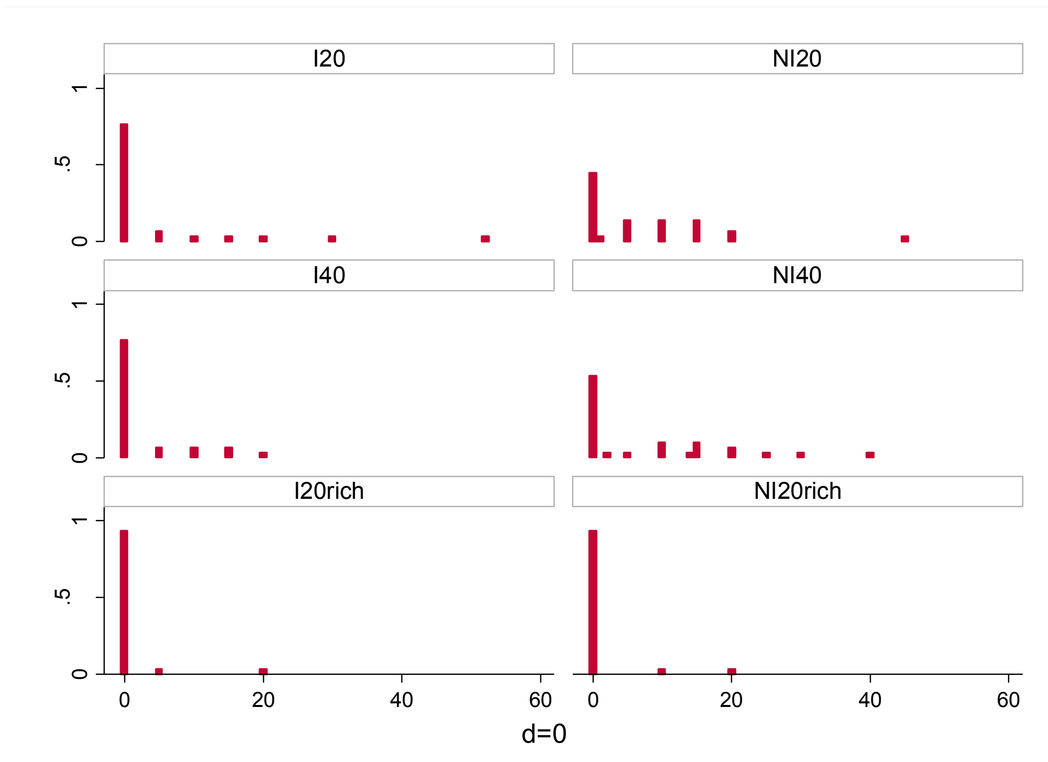
and, as expected, efforts are strictly increasing in the agent's expectation about θ_P the result follows.

□

Hence, a non-strategic donation reveals the "intrinsic motives" of a principal as only the intrinsic concern for the social cause affects the principal's behavior. As the extrinsic benefits are not taken into account (i.e., the instrumental value that a donation creates higher efforts), the overall incentives to donate are weaker. In turn, fewer principals donate and those who do so are more prosocial, leading to larger efforts after a donation. However, as fewer principals donate, those who do not donate should also be more prosocial on average. In this case, a non-donation is less of a negative signal on the principal's type.

2.7 Appendix B: Further Statistics

Figure 2.7: The distribution of agents' efforts after a non-donation



Note: The payoff equalizing level is 16.67 in I20, NI20, I40 and NI40, while in I20rich and NI20rich the payoffs cannot be equalized.

2.8 Appendix C: Experimental Instructions

Below you find translated instructions for participants A and B in I20 treatment. Original instructions were in German and were displayed on the screen.

[Instructions Participant A]

Welcome in our experiment! Please read the instruction carefully. If you have a question, please raise your hand. We will come over to you and answer your question. Communication with other participants is not allowed. If you break this rule, we will have to exclude you from the experiment and you will not receive any payment. Please switch off your mobile phone or any other devices which may disturb the experiment.

Every participant will receive 2.50 Euros for attending, which will be paid out independently of the decisions made in the experiment.

Furthermore, you can get additional payoffs in this experiment. The procedure is described more precisely below. In the experiment, experimental currency units (ECU) are used. The payoff in ECU will be converted into euros and paid in cash. The exchange ratio is:

10 ECU=1 Euro

Neither during the experiment, nor after the experiment will any of the participants be informed about the identity of other participants or about their payoffs.

In this experiment, there are two types of participants: Participants A and participants B. The types were assigned randomly and remain constant for the whole experiment. You are participant A. One of the participants B will be randomly assigned to you for the experiment.

You will receive a starting cash balance of 50 ECU. You can decide whether you donate 0 or 20 ECU to the organization “Ärzte ohne Grenzen e.V.”

The participant B assigned to you will receive 100 ECU starting cash balances. Participant B can decide how many points he transfers to you. Each ECU that participant B transfer to participant A is doubled. It means that each point which B transfers to you increases your payoff by 2 points and reduces B's income by one point. Participant B can transfer to you any amount between 0 and 100 ECU.

Thus, your payoff, the payoff of participant B and the donation to charity are given by:

Your payoff = 50 – donation by A + 2 * transfer by B

Donation to charity = donation by A

Payoff of participant B = 100 – transfer by B

Hence, the experiment consists of two stages:

Stage 1:

In the first stage of the game, you can decide whether you donate 0 or 20 ECU to the organization "Ärzte ohne Grenzen e.V.".

Stage 2:

In the second stage of the game, participant B will be informed about your donation and will decide how many ECU he transfers to you.

Participant B has to decide how many ECU he transfers to you, before he is informed whether you have donated 0 or 20 ECU. Participant B has to make two decisions: First, he has to decide how many ECU he transfers, if you donate 0 ECU. And, he decides, how many ECU he transfers to you, if you donate 20 ECU. The screen that will be shown to the participant B will look like this:

You have 100 ECU.

You can transfer from 0 to 100 ECU to participant A.

Every point that you transfer will be multiplied with a factor of 2.

Case 1: Participant A transfers 0 ECU.

How many ECU do you transfer to participant A? ___

Case 2: Participant A transfers 20 ECU.

How many ECU do you transfer to participant A? ___

It depends on your decision which of these two decisions will be payoff-relevant. If you donate 0 ECU, the decision for the Case 1 will determine your payoff. I you donate 20 ECU, the decision for the Case 2 will determine your payoff.

Subsequent to the experiment, the donations of all participants will be added up and the team of Cologne Laboratory for Economic Research will transfer the donation to „Ärzte ohne Grenzen e.V.“ We will transfer the money right after the experiment and you can stay and watch how we transfer the donation.

If you have any questions, please raise your hand.

[Instructions Participant B]

Welcome in our experiment! Please read the instruction carefully. If you have a question, please raise your hand. We will come over to you and answer your question. Communication with other participants is not allowed. If you break this rule, we will have to exclude you from the experiment and you will not receive any payment. Please switch off your mobile phone or any other devices which may disturb the experiment.

Every participant will receive 2.50 Euros for attending, which will be paid out independently of the decisions made in the experiment.

Furthermore, you can get additional payoffs in this experiment. The procedure is described more precisely below. In the experiment, experimental currency units (ECU) are used. The payoff in ECU will be converted into euros and paid in cash. The exchange ratio is:

10 ECU=1 Euro

Neither during the experiment, nor after the experiment will any of the participants be informed about the identity of other participants or about their payoffs.

In this experiment, there are two types of participants: Participants A and participants B. The types were assigned randomly and remain constant for the whole experiment. You are participant B. One of the participants A will be randomly assigned to you for the experiment.

You will receive a starting cash balance of 100 ECU. You can decide how many ECU you transfer to participant A. Each ECU that you transfer to participant A is doubled. It means that each point which you transfer to participant A increases the payoff of participant A by 2 points and decreases your payoff by 1 ECU. You can transfer any amount between 0 and 100.

Participant A, who was assigned to you, receives 50 ECU starting cash balances. He can decide whether he donates 0 or 20 ECU to the organization “Ärzte ohne Grenzen e.V.”

Thus, your payoff, the payoff of participant A and the donation to charity are given by:

Your payoff = 100 – transfer by B

Donation to charity = donation by A

Payoff of participant A = 50 – donation by A + 2 * transfer by B

Thus, the experiment consists of two stages:

Stage 1:

In the first stage of the game, participant A can decide whether he donates 0 or 20 ECU to the organization "Ärzte ohne Grenzen e.V.".

Stage 2:

In the second stage of the game, you will be informed about the donation of player A and will decide how many ECU you transfer to participant A.

You have to decide how many ECU you transfer to participant A, before you are informed whether participant A has donated 0 or 20 ECU. You have to make two decisions: First, you have to decide how many ECU you transfer to participant A, if he donates 0 ECU. And, you decide, how many ECU you transfers to participant A, if he donates 20 ECU. The screen will look like this:

You have 100 ECU.

You can transfer from 0 to 100 ECU to participant A.

Every point that you transfer will be multiplied with a factor of 2.

Case 1: Participant A transfers 0 ECU.

How many ECU do you transfer to participant A? ___

Case 2: Participant A transfers 20 ECU.

How many ECU do you transfer to participant A? ___

It depends on the decision of participant A which of your two decisions will be payoff-relevant. If participant A donates 0 ECU, the decision for the Case 1 will determine your payoff. If participant A donates 20 ECU, the decision for the Case 2 will determine your payoff.

Subsequent to the experiment, the donations of all participants will be added up and the team of Cologne Laboratory for Economic Research will transfer the donation to „Ärzte ohne Grenzen e.V.“ We will transfer the money right after the experiment and you can stay and watch how we transfer the donation.

If you have any questions, please raise your hand.

3 If I close my eyes, nobody will get hurt: The effect of ignorance on performance in a real-effort experiment

3.1 Introduction

In many situations, people willfully decide to avoid information. For instance, a smoker lights up a cigarette in a bar without asking the person sitting next to him whether it disturbs him, assuming it does not. Or say you suspect your favorite restaurant in town launders money; however, you do not want to know.

When it comes to ignorance in organizations, a number of examples of willful—and mostly profitable—ignorance are available (e.g., McGoey, 2012): starting with hip students who work at Apple stores and do not read about the working conditions in Apple factories; investment bankers who use instruments they do not understand in detail; doctors prescribing their patients drugs that pharmacy companies lobbied them to prescribe instead of thinking about better but less profitable alternatives; employees in marketing departments of tobacco companies avoiding thinking about the danger of their product; or huge corporate scandals as in Watergate or Enron[27]; ending up with one of the largest American drug makers Merck producing a painkiller which increased risk of heart attacks- however, the company claimed not to know about the side effect and refused to redraw the drug from the market for a while.[28]

In most of the aforementioned examples, ignorance increases the monetary gain by maintaining the illusion of acting in a fair way. In this paper we will concentrate on employees' behavior. Consider an employee who does not know the negative effect of her work on the economy, climate, or other domains. She can maintain her positive self-

[27] See Simon (2005) for detailed analysis of Watergate and Enron scandals.
[28] See http://www.nytimes.com/2006/08/22/business/22mistakes.html?pagewanted=all.

image while earning a high monetary payoff. Simulating such an environment in which work output increases one's own payoff but might have a negative externality on a third party, we are interested in how choosing to be ignorant about the negative consequences of one's own work affects work output.

To test the link between ignorance and work output, we conducted a real-effort experiment with a simple task in which agents decoded letters (as in Charness et al., 2014). By exerting effort, agents could increase only their payoff or also the payoff of the National Rifle Association (NRA), which a pre-test confirmed is considered negative on campus (see also Ariely et al., 2009). Some of our conditions provided no payment to the NRA, whereas other conditions provided a piece-rate payoff to it.

We also manipulated the information between conditions, such that participants either knew or not about the payment to the NRA. In the first condition with incomplete information, participants knew the chance that the NRA would be paid was equal to the chance it would not, and they could choose to reveal the actual realization. We call the participants who decided not to know "ignorant agents."

An interesting question is whether selecting to be ignorant is the same as simply not knowing. In the second condition with incomplete information, we tested whether a sorting occurs of social agents of a relatively "low type" into ignorance, whereas a social agent of a "low type" has self-image concerns but acts prosocially only if such behavior comes at a low cost. In the model by Grossman and van der Weele (2013), a social agent of a low type may sort into ignorance because doing so allows her to avoid a situation in which she has to behave unambiguously selfishly. Sorting may thus be the driver of ignorant agents' self-interested behavior. This model predicts that if we eliminate the option to reveal the information, people will be less selfish than those who choose to be ignorant. To test for the sorting, we designed a condition in which the agents did not know the piece rate of the NRA and had no possibility of finding out, which we then compared with the condition in which agents endogenously chose not to know.

As predicted, in the conditions with complete information, agents exerted significantly higher efforts if the NRA received no payment. Considering the role of ignorance, we observe that when agents decided not to know about the piece rate, they exerted a high effort, similar to their behavior in the condition in which the piece rate of the NRA was zero. If, on the other hand, the agents were not informed exogenously, they behaved as if they were in the condition with a positive payoff for NRA, and exerted a low effort. Thus, in line with the model of Grossman and van der Weele (2013), not having information about the consequences of one's own actions does not alone lead to a self-interested behavior, but rather, the sorting of low-type social agents into ignorance drives our results.

Our experiment relates to a number of experimental studies on ignorance. Ignorance was analyzed in dictator games (Dana et al., 2006; Dana et al., 2007; Larson and Capra, 2009; Fong and Oberholzer-Gee, 2011; Matthey and Regner, 2011; Grossman and van der Weele, 2013; Bartling et al., 2014; Grossman, 2014; van der Weele, 2014), in ultimatum games (Conrads and Irlenbusch, 2013), and in trust and moonlighting games (van der Weele et al., 2014). Most of these studies find that ignorant agents behave in a more selfish way, because they use lack of knowledge as an excuse for selfish behavior and may also have the illusion of acting fair. One exception is van der Weele et al. (2014) who find no effect of ignorance on behaving more selfishly, which they argue to be due to the richer moral context in their experiment. However, none of these studies analyzes how ignorance about negative externalities of one's own work might affect performance in a real-effort setting, which distinguishes our study from the previous literature.

The remainder of this paper is organized as follows: In section 3.2, we describe the experimental design. We report our results in section 3.3, and section 3.4 concludes.

3.2 Experimental Design

3.2.1 Set-up

In our experiment, participants worked on a real-effort task similar to the decoding task by Charness et al. (2014), which consisted of decoding letters into two-digit numbers. The table with letters in the first column and numbers in the second column was displayed on the computer screen in zTree (Fischbacher, 2007), whereby only one particular letter in the table had to be decoded with the corresponding number. After the subject decoded the letter, a new table with different numbers and letters combinations appeared. Moreover, the accuracy of entries was checked, because a participant could move to the next decoding task only if he decoded the letter correctly.[29]

We implemented a between-subjects design with four treatments. First, we conducted our baseline condition (denoted by BA) to measure the base pace of work in the piece-rate environment. Here, all the participants acted as agents and received 5 ECU for each letter they coded correctly, whereby 100 ECU were equal to 1 EUR. Thus, the more an agent worked, the more he earned.

The second condition (denoted by NRA) differed from the reference condition insofar as we introduced a negatively perceived charity. We told agents they would get 5 ECU for each decoded letter. At the same time, we informed them that each decoded letter yielded 7 ECU for the National Rifle Association (NRA). The participants received a short description of the NRA and were told that 93% of the subjects who took part in the survey at the campus perceived the NRA negatively.[30]

We conducted the ignorance treatment to investigate if and how ignorance might affect agents' performance. Here, agents could choose whether they wanted to learn the piece rate of NRA, whereby they knew the probability of it being 0 or 7 ECU was 50%. Thus, the performance could either increase only the agent's payoff and have no benefits for the NRA *or* increase the payoffs for both the agent and the NRA. We decided on the piece rate by a flip of a coin just before a session started, and this random mechanism was common knowledge. If a subject decided to become informed about the impact of her effort, the randomly chosen piece rate appeared on the screen, and if she did not want to reveal the piece rate, a question mark appeared. We denote the participants who learned that the piece rate was 0 (7) ECU by NI0 (NI7) in the remainder of the paper.

[29] See instructions in Appendix B for an example of a code table that we used in the experiment.

[30] We ran the survey two weeks before conducting the first sessions (see Appendix B for the survey). A student research assistant asked people passing by on campus to take part in a survey that would take only 1 minute. The participants filled in the survey in private, folded the sheets of paper, and gave them back to the assistant. Overall, 100 subjects participated in the survey, and collecting the data took approximately two hours. Participants received no payment for the survey, and were asked to rate the NRA on a 7-point scale (1=*I find it very bad*, 7=*I find it very good*). We considered answers between 1 and 3 to be negative.

We denote the participants who stayed ignorant by IG. We denote the group of all the subjects in the ignorance treatment (both ignorant and not ignorant) by NI&IG.

Finally, we conducted the uncertainty treatment (denoted by UN) to test whether social agents of a relatively low type sort into ignorance, which may possibly explain what drives the behavior of ignorant agents. In this treatment, agents knew that the piece rate of the NRA is 0 or 7 ECU, with a 50% probability of either rate. However, different from the ignorance treatment, here agents did not have the possibility of finding out the actual piece rate of the NRA. When comparing the output in the UN treatment and the output exerted by IG subjects, we can measure the sorting effect.

Moreover, in all the conditions, agents could choose a timeout by pushing the "Pause" button. If they used the button, the screen locked for 20 seconds. The agents received 4 ECU for each pause, which represents the opportunity cost of working.[31]

3.2.2 Hypotheses

Given the low costs of exerting effort (i.e., the task is simple and requires little thinking) and the piece-rate incentives, we expect agents to exert a positive effort in the baseline condition. With respect to the NRA condition, we expect agents to work less than in the baseline condition, because working on the task not only financially benefits the agent, but also the NRA, which the majority of University of Cologne students perceive negatively. Thus, working on the task creates a trade-off between self-image and monetary payoff.[32]

We expect a substantial fraction of agents to stay ignorant. We base this hypothesis on the previous systematic experimental evidence showing that a substantial number of participants remain ignorant about the consequences of their own decisions.[33] Furthermore, we expect the IG agents to work more than agents who know their work

[31] We use a form of opportunity costs similar to Berger et al. (2013), who used the "Pause" button for a break in their real-effort experiment.
[32] Furthermore, Ariely et al. (2009) showed that subjects produced significantly less output for NRA than Red Cross.
[33] For instance, 44% of dictators are ignorant in the seminal game by Dana et al. (2007), and 53% in the study by Larson and Capra (2009).

will benefit the NRA, which may be driven by the moral wiggle room that protects self-image (as in Dana et al., 2007) and/or the sorting of social agents of a relatively low type into ignorance (as in Grossman and van der Weele, 2013).

As for the sorting effects, we compare performance in exogenous and endogenous information conditions. According to the model by Grossman and van der Weele (2013), selfish and social agents are present, whereas the degree of being social varies. Although a selfish agent will always act in a payoff-maximizing way and a social agent of a high type will always act prosocially, a social agent of a low type has (self-) image concerns and will act prosocially only if doing so comes at a low cost. Importantly, in the case in which acting prosocially is expensive, ignorance allows the lower social type to avoid a situation in which the agent has to pool with selfish types (i.e., because acting prosocially is expensive, the low-social-type agent will exert a high effort, even if doing so benefits NRA), and thus the social agent of a lower type chooses ignorance and then acts in a self-interested way. We expect IG agents to behave more selfishly (i.e., decode more letters) than UN agents, because agents in the ignorance condition sort into ignorance, but agents in the uncertainty treatment do not. On the other hand, agents who choose to inform themselves about the NRA's piece rate will behave more prosocially (i.e., decode fewer letters) than agents who received the information exogenously, because social agents of a higher type choose to know. Finally, we do not expect the output levels to be different in BA and NI0, because all the decisions made in BA or NI0 conditions affect only one's own payoff and not the NRA's payoff, and thus the agents' type should not affect the agents' behavior.

As for the timeouts, we expect agents will not use the pause button if the NRA will not receive any donation or if the participants are ignorant, because on average, working leads to a higher monetary payoff than taking a break.[34] However, if the agents know the piece rate of the NRA is 7 or if the agents were assigned to the UN condition, the pause button may represent a morally good alternative to earning money instead of working for the NRA, and thus agents should push the timeout button more often.

[34] In the BA condition, the average amount of decoded letters is 18.47 per minute. It follows that the average agent earns more than 30 ECU in 20 seconds, which is notably more than 4 ECU for a pause.

3.2.3 Experimental Procedure

We conducted our experiment in March-June 2014 at the Cologne Laboratory for Economic Research using the experimental software zTree (Fischbacher, 2007). We recruited participants via ORSEE (Greiner, 2004). We ran 10 sessions with a total of 267 participants. No subject participated in more than one of our sessions.

At the beginning of the experiment, participants received written instructions for the experiment and were allowed to ask questions privately. To ensure each participant understood the instructions, subjects had to answer comprehension questions that the experimenters examined before the task started. Then each participant had 90 seconds for the trial period of the decoding task.[35] In this stage, they received 5 ECU for every correctly decoded letter and could not choose a timeout, and we introduced this stage as an ability checker. After each participant finished the trial task, the actual decoding task started, which subjects had 10 minutes to complete. Each agent received identical decoding tasks in the same order.[36]

After completing the actual experiment, the participants were asked to complete a post-experiment questionnaire that included questions on gender, age, field of study, the Big Five personality traits, and motives behind the decisions. At the end, participants privately received their payoffs in cash and left the laboratory. Each session lasted approximately one hour. The average payoff was 12.81 Euros (minimum 4.20 Euro, maximum 16.15 Euro), including a show-up fee of 2.50 Euros.

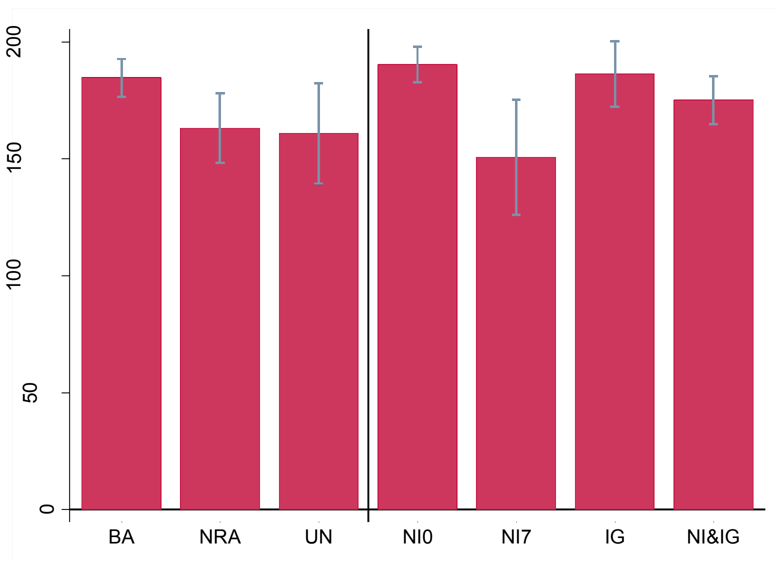
The transfer to the NRA was made after all sessions were finished. To ensure the donation was credible, we told students in the instructions that they could give us their email address if they wanted to receive proof of the donation, and we sent them the proof at a later date.[37]

[35] In the trial period, participants knew already which treatment they are assigned to.
[36] We limited the decoding tasks to 240. Nine out of 267 participants reached the limit at the end of the task.
[37] We wrote, "Please write down your email address, if you want to get a proof of the donation" (see Appendix B for the post-experiment questionnaire). Forty-four out of 219 participants (or 20.09%) wrote down their email addresses.

3.3 Results

In line with our predictions, agents exerted substantial effort in the baseline condition and decoded an average of 184.67 letters in 10 minutes of working on the task (see Figure 3.1 and Figure 3.2). With respect to the NRA condition, we observe that agents worked less when their efforts benefitted the NRA: With 163.15 decoded letters, average performance in this condition was significantly smaller compared to BA (p=0.01093, two-sided Fisher-Pitman permutation test for independent samples).[38] Hence, we observe that agents are less motivated to work if a negatively perceived organization benefits from their efforts, which is in line with our hypothesis and the results by Ariely et al. (2009).

Figure 3.1: Average performance per agent, in number of decoded letters



Notes: Figure 3.1 plots the average amount of decoded letters. Number of subjects amounts to 48, 48, 44, 46, 45, 36, and 127 in BA, NRA, UN, NI0, NI7, IG, and NI&IG, respectively. Error bars are standard errors of the mean.

[38] In this paper, we use the Fisher-Pitman permutation test for independent samples, which is a non-parametric statistical test and provides *p*-values for the difference in means of outcomes in two independent groups. The test is superior to Wilcoxon tests if the observed values are given on at least an interval scale (see Kaiser, 2007). Note that we use Monte Carlo simulations to calculate the p-values (200000 runs). Furthermore, all of our results also hold if we use *t*-test.

Furthermore, we asked subjects in the post-questionnaire whether their decisions in the experiment were fair (on a Likert scale from 1 to 7 with 1=*very fair* and 7=*not fair at all*). If the subjects reported that they felt their decisions were on average more unfair in conditions with positive giving to NRA than in the non-donation condition, we took the report as an indicator that subjects in the experiment perceived donating to the NRA as less desirable than working only for themselves. Whereas the average response amounts to 3.15 in the NRA condition, it is only 2.29 in BA (p=0.02803).[39] Thus, it follows that subjects felt worse about their decisions in the condition in which their efforts benefitted the NRA than in the conditions in which their efforts did not.[40]

In the ignorance treatment, 28.35% of subjects (or 36 out of 127) decided not to learn how their actions would affect others. This share is smaller than, for instance, in the seminal study by Dana et al. (2007), in which 44% of subjects chose to remain ignorant.[41] With respect to the output, we observe that agents who chose to find out how their action would affect others, and learned the NRA piece rate was 0 (NI0), decoded 190.33 letters on average. At the same time, the performance of agents who learned the piece rate was 7 (NI7) decreases to 150.69 letters. The difference between output levels in NI0 and NI7 is highly statistically significant with p=0.00234.

Concerning the impact of ignorance on the output, we observe that among agents who decided not to learn the NRA piece rate, performance increased by 23.64% to 186.31 letters, on average, compared to condition in which agents knew the piece rate was 7 (NI7).[42] This difference is significant with p=0.01812. Thus, as predicted, agents who are ignorant about the consequences of their actions work more than agents who are aware of a negative externality. One interpretation of the result is that not knowing maintains the illusion of acting in a good way. Moreover, when we compare the ignorant

---

[39] See Figure 3.5 in Appendix A for the average results of feeling unfair in all the conditions.

[40] However, as the editor pointed out, the subjects in the experiment did not necessarily think the NRA was a bad charity. Rather, subjects maybe have merely wanted to conform to the social norm that the experimenter conveyed to them.

[41] The difference between our result and theirs might reflect the fact that our setting implemented the opportunity to give to the NRA. In the study by Dana et al. (2007), the *worse* game only means that if the dictator chooses the payoff-maximizing option, the receiver gets less money than in the *better* game. In our setting, on the other hand, the *worse* game leads to a donation to a gun-lobbying organization, which is perceived negatively.

[42] If we pool all the subjects in the ignorance treatment (NI0, NI7, and IG), the average output amounts to 175.14 letters (denoted by NI&IG in Figure 3.1).

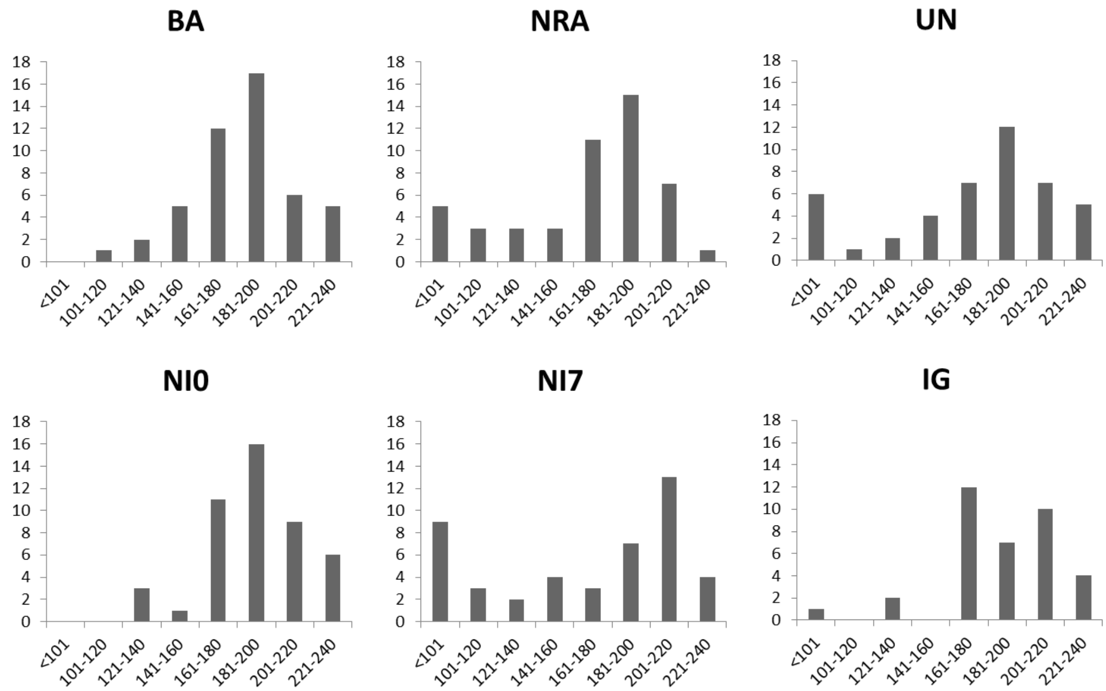
agents with BA or NI0 agents, we see no difference in the output levels (p>0.1), which suggests that ignorant agents behave as if they were in the condition with the zero piece rate for the NRA.

In the next step, we analyze the selection effects of agents. First, we compare agents in the NRA condition with agents in the NI7 condition. As discussed above, agents who choose to know should be more prosocial on average than agents who receive the information exogenously. However, when comparing the output in NI7 and NRA, we find no significant differences in the distributions (p=0.37904).[43] Thus, it follows that choosing to be informed when working for the NRA had no effect on effort.

Second, we expect agents who choose to be ignorant to behave, on average, less prosocially than agents in the uncertainty condition, because social agents of a lower type sort into not knowing. We found that agents in the UN condition decoded 160.91 letters, which is not statistically different from output in NRA or NI7 (p>0.1) and suggests the subjects behaved as if they were in the condition with a positive piece rate for the NRA. Therewith, participants in UN behaved more prosocially on average than agents in IG, and decoded 13.63% fewer letters. The effect is marginally statistically significant with p=0.05757. Thus, it follows that not knowing about the consequences of one's own actions does not necessarily lead to selfish behavior. Rather, in our experimental setting, the selection effect drives the negative effect of ignorance; that is, social agents of a lower type sort into ignorance and behave in a more selfish way. In line with this finding, the results from the post-questionnaires show (see Figure 3.5 in Appendix A) that IG agents feel significantly less unfair about their decisions in the experiment than UN agents, with a p=0.05001 in a Fisher-Pitman test. Finally, the result that IG agents behave more selfishly than exogenously UN agents is in line with the result by Fong and Oberholzer-Gee (2011) that in a dictator game, dictators who choose not to know whether their recipient is disabled or a drug user give significantly less to the recipient than those who were uninformed exogenously.

[43] We find no significant differences in the output distributions between BA and NI0 (p=0.31812). Note that here we do not expect the output levels to be different, as discussed in the Hypotheses section.

Figure 3.2: The distribution of decoded letters (in number of subjects)



Notes: Figure 3.2 plots the distribution of decoded letters. Number of subjects amounts to 48, 48, 44, 46, 45, 36, and 127 in BA, NRA, UN, NI0, NI7, IG, and NI&IG, respectively.

All the results described above still hold if we control for ability of the participants.[44] To derive a measure of output that takes ability into account, we divide the output level in the main part of the experiment by the number of decoded letters in the trial period for every participant.[45] When comparing this measure over the treatments, we find the same effects stay significant (not significant) in a Fisher-Pitman test as when comparing only the output in the main part of the experiment.[46]

[44] The ability of subjects is not significantly different over the conditions, because pairwise comparisons of decoded letters in the trial period between conditions lead to p-values larger than 0.1 in a Fisher-Pitman test. In the trial period, subjects in BA, NRA, UN, NI0, NI7, and IG decode on average 21.29, 20.85, 22.00, 21.93, 22.04, and 22.62 letters, respectively.

[45] We are thankful to an anonymous referee who suggested using this type of measure.

[46] The output in the main part divided by the trial output is on average 8.88 in BA, 8.05 in NRA, 7.31 in UN, 8.95 in NI0, 7.15 in NI7, and 8.57 in IG. The difference between BA and NRA (BA and UN) is statistically significant with p=0.06517 (p=0.00173). The measure is also significantly different between NI0 and NI7, NI7 and IG, and UN and IG with p=0.01051, p=0.06957, and p=0.07544, respectively. The differences between BA and IG, NRA and UN, NRA and NI7, and BA and NI0 stay statistically insignificant with p>0.1.

Finally, with respect to the pauses, we find that agents almost never take a break in the baseline condition. Here, the average number of timeouts is 0.10 (see Figure 3.3 in Appendix A). In the NRA condition, on the other hand, the average number of pauses increases to 3.25. This between-treatment difference is highly statistically significant with p=0.00005. One interpretation of the result is that in the NRA condition, agents take breaks because they believe that earning money is a moral alternative to working to benefit the NRA. In line with this interpretation, agents in NI7 also take significantly more breaks than agents in NI0, whereas the average number of breaks is 5.84 and 0.07 in NI7 and NI0, respectively (p=0.00011). Finally, agents in IG take 1.03 breaks, on average, which is significantly less than in the NI7 condition (p=0.01707). On the other hand, agents who were assigned to the incomplete information condition UN take 3.95 breaks on average, which is almost four times more than agents in the IG condition. However, the difference is only marginally statistically significant (p=0.10782).[47]

All in all, we observe that agents who work to benefit the NRA or are in the UN condition substitute part of their work with incentivized timeouts, whereas agents who work only for themselves or are ignorant work significantly more and take only a few or no breaks. Thus, it follows that agents care about the moral consequences of their actions, even if they are uncertain about whether a negative externality even exists. However, agents who choose to be ignorant behave in a more selfish way by exerting more effort, because, on average, social agents of a lower type sort into ignorance.

3.4 Conclusion

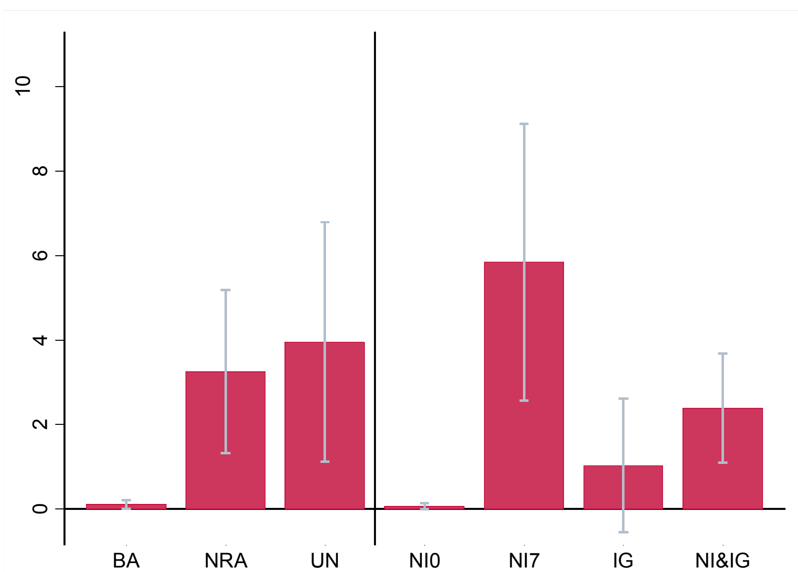
We conducted a real-effort experiment to study the effect of ignorance regarding negative externalities of own work on agents' performance in a real-effort setting. Overall, we find that, as expected, agents work significantly more if their effort does not benefit a recipient that is perceived as socially undesirable. Agents choose to forego a part of their monetary payment to avoid this negatively perceived outcome.

[47] When pooling the data from all the conditions, a Spearman rank correlation coefficient shows a negative and significant correlation between output and timeouts (rho=- 0.5498, p=0.0000).

However, agents who choose to stay ignorant about the negative consequences of their actions exert high effort and behave as in the condition in which the negative effect is zero. We explain the self-interested behavior of the ignorant agents with a selection effect, showing that in the treatment in which agents are not informed about the piece rate to the negative party exogenously, they work less and behave as if they were in the condition with a negative effect. Thus, we conclude that uncertainty about the consequences of own actions does not necessarily lead to a self-interested behavior, but rather, the selection of lower social types into ignorance drives the negative effect of ignorance.

3.5 Appendix A: Further statistics

Figure 3.3: Average amount of timeouts per agent



Notes: Error bars are standard errors of the mean.

Figure 3.4 The distribution of timeouts (in % of subjects)

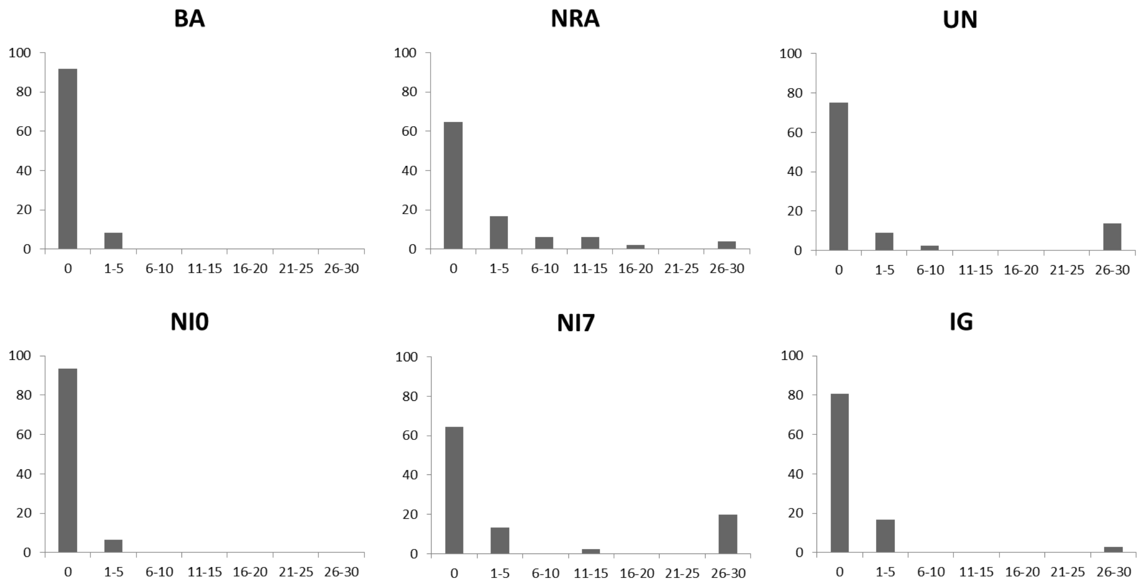
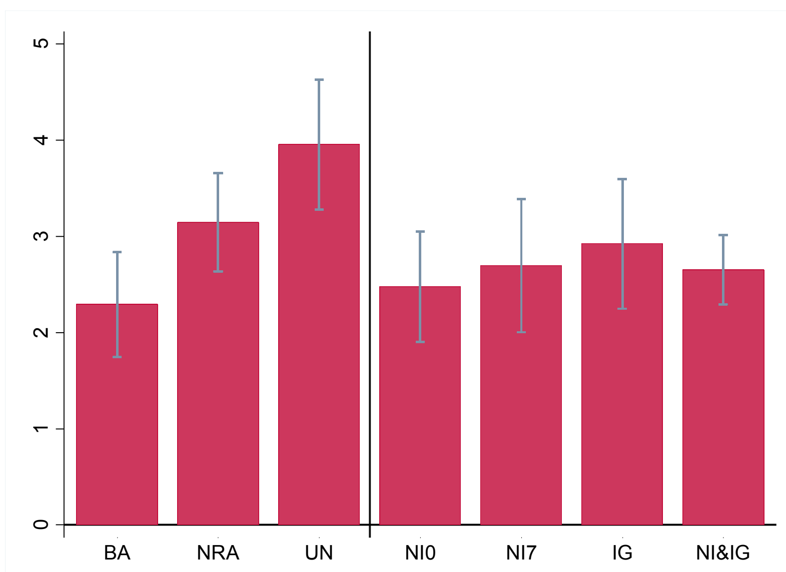


Figure 3.5 Feeling unfair about own decisions in the experiment



Notes: The feeling of being unfair was measured on a Likert scale with 1=*very fair* and 7=*not fair at all*. Error bars are standard errors of the mean.

3.6 Appendix B: Experimental Instructions

Below you find translated instructions. Original instructions were in German.

Welcome to our experiment! Please read the instructions carefully. If you have a question, please raise your hand. We will come over to you and answer your question. Communication with other participants is not allowed. If you break this rule, we will have to exclude you from the experiment and you will not receive any payment.

Every participant will receive 2.50 Euros for attending, which will be paid out independent of the decisions made in the experiment.

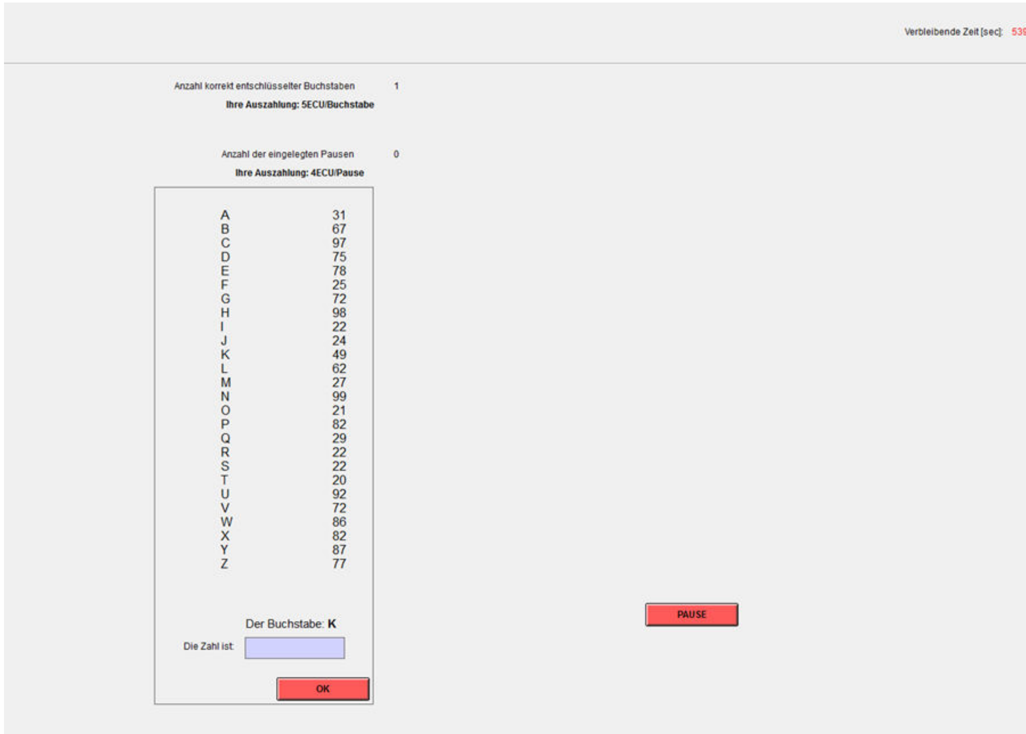
Furthermore, you can get additional payoffs in this experiment. We more precisely describe below how the payoff system works. The experiment uses experimental currency units (ECU). The payoff in ECU will be converted into euros and paid in cash. The exchange ratio is 100 ECU=1 Euro.

Neither during nor after the experiment will we inform any of the participants about the identity of other participants or about their payoffs.

All participants received the same instructions.

The task:

A table with two columns (one column with letters and one column with numbers) will appear on your computer screen. The table looks like the following screenshot:



Your task is to decode the letter, which is denoted in the table. To decode the letter, please find its corresponding number. In the example displayed above, the letter "K" has to be decoded. From the table, it follows that the corresponding number is "49" (the number is placed in the same row as the letter "K"). Therefore, you should enter number "49" in the blue field and press the "OK" button.

If you have decoded the letter correctly, a new input mask with a new table will appear. Now, you can decode a new letter. If you made a mistake while decoding the letter, an error message will appear. In that case, please correct the input and press the "OK" button.

A "Pause" button is also displayed on the computer screen. If you press the "Pause" button, your computer screen will be frozen for 20 seconds. During this time, you will be unable to decode letters. Please note that no pauses can be taken within the last 20 seconds of the task.

Procedure and payoffs:

The experiment consists of two stages:

Stage 1: Trial period. Before the actual experiment starts, you have an opportunity to practice the task. The trial period will last 90 seconds and you will receive 5 ECU for every letter that you decode correctly. The remaining time will be displayed on the screen.

In the trial period, the "Pause" button will not be displayed.

Your payoff in Stage 1 = 5 ECU * number of correctly decoded letters.

Stage 2: Main part of the experiment. In this stage, you have will have 10 minutes for the task. The remaining time will be displayed on the screen.

[Treatment BA]

For every correctly decoded letter, you will receive 5 ECU.

If you press the "Pause" button, the screen will be frozen for 20 seconds. For every pause you take, you will receive 4 ECU.

The payoff in Stage 2 is

Your payoff in Stage 2 = 5 ECU * number of correctly decoded letters +

4 ECU * number of pauses.

If you have any questions, please raise your hand.

[Treatment NRA]

In the main part of the experiment, your effort increases not only your payoff, but also the donation to National Rifle Association (NRA). The NRA is a gun-lobbying organization in the United States that fights for the right for all US citizens to be able to buy, own, carry, pass on, and use a gun. The United States is one of the countries with the highest death rates caused by firearms.

Two weeks ago, we ran a questionnaire about the image of the NRA at the campus of the University of Cologne. Ninety-three percent of the subjects perceived the NRA negatively.

For every correctly decoded letter, you will receive 5 ECU. Additionally, for every correctly decoded letter, 7 ECU will be transferred to the NRA.

If you press the “Pause” button, the screen will be frozen for 20 seconds. For every pause you take, you will receive 4 ECU.

Your payoff and the payoff of NRA are

Your payoff in Stage 2 = 5 ECU * number of correctly decoded letters +

4 ECU * number of pauses.

Your contribution to the NRA = 7 ECU * number of correctly decoded letters.

Subsequent to the experiment, the contributions of all the participants will be combined and transferred to the NRA. We will transfer the money after the experiment. You can give us your email address and we will send you proof of the donation.

If you have any questions, please raise your hand.

[Treatments SI and UN]

In the main part of the experiment, your effort may increase not only your payoff, but also the donation to the National Rifle Association (NRA). The NRA is a gun-lobbying organization in the United States that fights for the right of all US citizens to be able to buy, own, carry, pass on, and use a gun. The United States is one of the countries with the highest death rates caused by firearms. Two weeks ago, we ran a questionnaire about the image of the NRA at the campus of the University of Cologne. Ninety-three percent of the subjects perceived the NRA negatively.

For every correctly decoded letter, you will receive 5 ECU. Additionally, for every correctly decoded letter, 0 ECU or 7 ECU will be transferred to the NRA. A flip of a coin before this experiment determined whether 0 ECU or 7 ECU will be transferred to the NRA for every correctly decoded letter. The probability that the payment to the NRA is 0 ECU or 7 ECU is in both cases 50%.

If you press the "Pause" button, the screen will be frozen for 20 seconds. For every pause you take, you will receive 4 ECU.

Your payoff and the payoff of NRA are:

Your payoff in Stage 2 = 5 ECU * number of correctly decoded letters +

4 ECU * number of pauses.

With a 50% probability, your contribution to the NRA = 0 ECU * number of correctly decoded letters.

With a 50% probability, your contribution to the NRA = 7 ECU * number of correctly decoded letters.

[This paragraph only in the SI treatment] Before the main part of the experiment begins, you can decide if you want to find out whether the NRA gets 0 ECU or 7 ECU for every correctly decoded letter. If you want to find it out, press the "Yes" button. If you do not want to find out about the contribution to NRA, press "No."

Subsequent to the experiment, the contributions of all the participants will be combined and transferred to the NRA. We will transfer the money after the experiment. You can give us your email address and we will send you proof of the donation.

If you have any questions, please raise your hand.

[Post-experiment questionnaire]

Please answer the following questions.

Age: _____

Gender: _____

Major: _____

Semester: _____

Please describe briefly how you made your decisions in this experiment: _____

Please answer on a scale from 1 to 7 (1=*I agree fully*, 7=*I do not agree at all*):

The task in itself was great fun: _____

I was trying very hard when working on the task: _____

Did you use the pause button? _____

If yes, why did you use the pause button? _____

Were your decisions in the experiment fair? (1=*very fair*, 7-=*not fair at all*) _____

Please write down your email address if you want to get proof of the donation: _____

Please write down how much the following statements are true for you. Use a scale from 1 to 5 (1=*strongly disagree*, 5=*strongly agree*):

1. I am reserved ____

2. I am generally trusting ____

3. I tend to be lazy _____

4. I am relaxed, handle stress well _____

5. I have few artistic interests _____

6. I am outgoing, sociable _____

7. I tend to find fault with others _____

8. I do a thorough job _____

9. I get nervous easily _____

10. I have an active imagination _____

Please answer on a scale from 1 to 7 whether you answered the questions truthfully (1=*fully truthfully*, 7=*not truthfully at all*): _____

[Survey about the image of NRA]

Your opinion about the NRA

We are running this survey about National Rifle Association (NRA) for a research project in the Department of Economics.

The NRA is a gun-lobbying organization in the United States that seeks for the right for all US citizens to be able to buy, own, carry, pass on, and use a gun.

Please tell us your opinion about this organization. Please answer on this scale: 1=*I find it very bad*, 7=*I find it very good*.

1	2	3	4	5	6	7
Very bad						Very good

Your age: ______________________________

Your gender: male___ female__

4 Lying Costs and Incentives

4.1 Introduction

Much of the interaction in the marketplace are based on asymmetric information, which makes some people tempted to misreport the private information they have for their advantage. For example, consider VW cheating by installing defeat devices in its cars in order to evade emission tests, or insurance fraud, which, according to the FBI, totals $40 billion a year in the U.S. in the non-health insurance domain.[48] Given the important economic consequences of cheating, understanding the factors that influence the decision to cheat is important in order to understand many economic behaviors.

For some people, lying is costly—they prefer an outcome if they reach it honestly over getting it by lying. In principle, we can distinguish between three types of people according to their cost of lying. Some people may be unwilling to tell a lie, regardless of their benefit from it ("ethical type"). Scholars such as St. Augustine (421) and Kant (1787) advocated such an uncompromising approach to lies. People who are not willing to lie could be described in our approach as having an infinite cost of lying. Other people may have a finite positive cost of lying. These people will lie when the benefit of lying is higher than the associated cost ("finite positive cost type"); at the extreme are people with a zero cost of lying ("economic type").

We concentrate on "intrinsic" costs of lying—those that are not based on strategic considerations. Intrinsic costs could result from a direct distaste of lying, or from reasons that are indirectly related to lying. Such indirect preferences may relate to interdependent preferences in which the agent's utility depends on the utility of others, either because of payoffs (Rabin, 1993; Fehr and Gächter, 2000) or because of guilt aversion (Battigalli and Dufwenberg, 2007; Battigalli, Charness, and Dufwenberg,

[48] See: https://www.fbi.gov/stats-services/publications/insurance-fraud/insurance_fraud.

2013). They may also relate to maintaining a positive social and/or self-image (Bénabou and Tirole, 2006; Andreoni and Bernheim, 2009), to the nature of communication (Lundquist, Ellingsen, Gribbe, and Johannesson, 2009), or to aversion to violating social norms (Elster, 1989; Bicchieri, 2006).

Standard economic models assume "an economic agent" who has no intrinsic cost of lying and considers only strategic aspects of the decision, such as the probability of being caught and the punishment resulting from it. Costs of lying in such models may be associated with strategic concerns about lying, such as reputation and repeated game consideration, but not with intrinsic preferences (e.g., Becker, 1968; Allingham and Sandmo, 1972; Holmstrom, 1979; Crawford and Sobel, 1982). Note that this is just a simplifying assumption of the models, and the intrinsic cost of lying could be added to the standard models (e.g., Kartik, 2009).

In this paper, we use two sets of experiments to better understand the structure of the intrinsic costs of lying. In the first set of experiments ("Incentives Size"), we manipulate the incentives associated with lying and observe the effect on people's tendency to lie. The size of the incentive would not affect people who never lie or people who always lie when doing so benefits them. By contrast, if the cost is positive but finite and the incentives are increased enough, some people will switch from telling the truth to lying. As we will argue below, understanding the structure of the intrinsic cost of lying will help in constructing policies to reduce cheating.

The experimental economics literature offers two main approaches to studying lying behavior, and their results differ regarding the effect of incentives. The first line of research uses deception games, in which a sender has private information regarding payoffs associated with each of the choices that a receiver faces. The sender sends a message to the receiver that could be true or false, and the receiver makes a choice that determines the payoffs for both players. As a two-player game, the deception game includes elements of strategic thinking as well as other-regarding preferences. The results from experiments with this game are that senders are more likely to lie when the incentives to do so are increased (e.g., Gneezy, 2005; Sutter, 2008; Dreber and Johannesson, 2008; Erat and Gneezy, 2012).

The second approach to studying lying in the lab uses some form of a cheating game in which the participant receives private information about the state of the world and reports it to the experimenter; this report, in turn, determines the participant's payoff. This type of game involves no strategic element, because the relation between the participant's report and the payoff is one to one. Furthermore, no social preferences are involved (assuming the participant does not take the experimenter's payoff into account). In contrast to the finding in the deception game, the literature on cheating games reports that higher incentives do not result in a higher fraction of lies (see Mazar et al., 2008; Fischbacher and Föllmi-Heusi, 2013, and the discussion in the literature review below).

In our experiments, we focus on cheating games, with one of the goals being to understand the source of the difference between the lying results in strategic and non-strategic settings in the literature. In the "Incentives Size" experiment, we first replicate the result of the cheating games by testing a game in which a participant rolls a six-sided die in private and then reports the result to the experimenter. If the participant reports a 5, she receives \$X (\$X=\$1, \$5, \$20 and \$50, between participants), and zero otherwise. The size of the incentives we test and the difference between them (a factor of 50) is more extreme than in the literature (e.g., Mazar et al., 2008; Fischbacher and Föllmi-Heusi, 2013). In line with the findings reported above, in this game, we find no increase in cheating as the incentives are raised.

In the deception game, lying is an option suggested in the rules of the game where the sender is asked to decide whether to send a true or false message. This is not the case in the cheating game, in which participants are asked to report the actual outcome without being told about the option to lie. Because the lying option is not in the rules of a cheating game, participants might be afraid of possible consequences of a lie. In addition, participants in the cheating game might question whether their die roll is really private, or whether the experimenter has a way to find out the outcome. When the incentives are raised, concerns may increase regarding the possible consequences of being exposed when lying (see social psychology literature, e.g., Ekman, 1985; Ekman, 1988; Vrij, 2008).

To test the effect of this potential concern, we modified the cheating game with a new design in which we ask participants to think about a number in private, then roll the die in private, and report whether the number that came up is the same as the one they thought of (see Jiang, 2013, and Shalvi and De Dreu, 2014, for similar manipulations). If they report it was the same number, they receive $X as before, and zero otherwise. As predicted, the results of this "mind game" show an overall increase in participants' tendency to lie for every level of X. This result indicates that some participants in the first treatment chose not to lie because of the risk of being exposed, and not because of the intrinsic cost of lying. In addition, and in contrast to the first treatment, the fraction of participants who lied increases with X. That is, in the mind game, participants lied more when the incentives were increased.

After demonstrating that some people are more likely to tell a lie when the incentives are increased, which indicates a positive and finite cost of lying, we try to better understand the structure of their lying cost. In the second experiment ("Cost Structure"), we introduce a novel cheating game in which, unlike in the "Incentives Size" experiment and the existing literature, we can observe the individual lying behavior. In our experiment, we ask participants to click, in private, on one of 10 boxes on a computer screen. The participants are told that behind each box is a different number between 1 and 10. After revealing the number, participants are asked to report it. Payments equal the number they report in Euros. After the experiment, by comparing the actual number with the reported one, we can observe how often and to what extent participants lie.

This design helps us directly test hypotheses regarding the structure of the lying cost function. In particular, previous papers argued (e.g., Mazar et al., 2008) or assumed (e.g., Kartik, 2009) that the cost function is convex—meaning the marginal cost of a lie increases with the magnitude of the lie, such that "small lies" are less painful than "big lies." The size of the lie could be defined on different dimensions—most notably on the payoffs (how much more one earns from the lie) and the reported number (how far the number reported is from the actual number). In the "Cost Structure" experiment, those two dimensions are the same. But in the "Incentives Size" experiment, the size in terms of the reported lie is the same between treatments and increases with X in the payoff dimension.

According to the convex-cost-of-lying hypothesis, if someone is lying in the current game, that person might exaggerate the number he or she reports by a little, but not go all the way to the maximum. An alternative hypothesis is that lying costs are fixed: a person suffers a cost from telling a lie, and this cost does not depend on the size of the lie. According to this hypothesis, if a participant will lie, he or she will lie to the full extent and report a 10.

Our results from the “Cost Structure” experiment offer strong support for the fixed-cost-of-lying hypothesis. We find most of the participants who lie report the maximal outcome of 10. We also find participants reacted to the incentive to lie: the fraction of lies decreased when the actual outcome increased. In other words, when the return on lying increased, people lied more.

Together, our results provide an explanation for the findings in the literature on cheating games and offer support for the hypothesis that the tendency to lie increases with incentives, indicating some of our participants have positive and finite costs of lying. In particular, the evidence suggests that our participants compare the cost and benefit of the lie, where the former does not depend on the size of the lie. When the benefit from lying is higher than the cost of lying, they lie to the full extent.

4.2 Literature review

As discussed in the introduction, the evidence from cheating games suggests increasing the incentives does not increase lying. In the leading paper on the topic in psychology and marketing, Mazar et al. (2008) asked participants to solve a set of 20 matrices in private and report their success to the experimenter later on. The participants received payment based on a piece rate per correctly solved matrix. Payments were increased, between participants, from $.50 to $2 (maximal payoff of $10 to $40). The results of this treatment were compared with a baseline test of ability. The results show participants lie, but there is no difference between the two levels of incentives.

Even stronger than this finding, Mazar et al. (2008) report the results of additional experiments in which they manipulated the amount of payment to each participant per correctly solved matrix ($.10, $.50, $2.50, and $5). They find, compared with the control, a "limited dishonesty" in the $.10 and $.50 treatments, but no lying in the $2.50 and $5 conditions. Mazar et al. (2008) conclude behavior is consistent with a convex cost function: the cost of lying increases faster than the benefits, such that when the incentives to cheat are increased, people are not more likely to lie.

In the leading paper on the topic in economics, Fischbacher and Föllmi-Heusi (2013) introduce the die-roll method, in which participants roll a die in private and then report the outcome to the experimenter. The participants' payoff depends on their report: they receive 1, 2, 3, 4, 5, 0 Swiss Francs for reporting 1, 2, 3, 4, 5, 6, respectively. To test the effect of incentives, in one of the control treatments, all payments were multiplied by a factor of 3 (from a maximal payoff of 5 Swiss Francs to 15 Swiss Francs). Fischbacher and Föllmi-Heusi (2013) argue that by doing so, the incentives were changed in potentially opposing ways: whereas the monetary incentives to lie increase, reporting a higher number "increases the size of the lie."

The results reported by Fischbacher and Föllmi-Heusi (2013) are similar to Mazar et al. (2008): the fraction of participants who lie does not change significantly when the stakes are tripled. The authors conclude, "Either the effect of increasing the benefits of lying is counterbalanced by the negative effect of increased costs of lying when a person earns more through deception, or the concept of lying is not directly related to stakes at all."

Related to this, Weisel and Shalvi (2015) introduce a cheating game with an element of cooperation – two paired players have to report the same outcome sequentially in order to receive a payoff, where the payoff equals the number reported. Weisel and Shalvi (2015) show that a substantial fraction of participants collaborates and lies. However, doubling the incentives does not affect the lying behavior.

By contrast, in the experiment reported in Gibson, Tanner, and Wagner (2013), each participant acted as a CEO and had to report earnings per share for the previous quarter. The earnings had two states, and reporting the false state resulted in the higher payoff for the participant. The incentive to lie was varied such that the payoff for the wrong report was the same (1.5 Swiss Francs), but the payoff for a truthful message differed over the five treatments from 0.3 to 1.5 Swiss Francs. In addition, the authors used a within-subject design and told participants that reporting wrong earnings was legal. Gibson et al. (2013) find that lying increases from 18% up to 79%, when it becomes more beneficial.

Gneezy, Rockenbach, and Serra-Garcia (2013) report the results of a hybrid game combining elements of the deception game with elements of cheating games. In the experiment, participant A observed the state of the world, which was a number between 1 and 6, and reported the number to participant B. Participant's A's payoff increased with the number he or she reported. Participant B earned more when following a true message than not following, but earned nothing if he or she followed a false message. In this design, as in cheating games, the reported number increased the payoff of participant A regardless of the action of participant B. Unlike in cheating games, lying had a negative externality on participant B. Gneezy et al. (2013) find that lying is more frequent when the actual state is low, that is, when the incentives to lie are higher. In addition, the majority of participants who lie do so to the full extent.

Together, the experimental evidence from cheating games suggests no effect of incentive size effect on lying, unless the game is modified such that it is more similar to the deception game. In particular, Gibson et al. (2013) tell the participants that lying is legal—in the deception game, lying is also an option the rules of the game suggest. Gneezy et al. (2013) introduce messages to the second player as well as negative externalities associated with lying; both studies find an incentives effect on lying. By contrast, if one uses a standard cheating game as in Mazar et al. (2008) or Fischbacher and Föllmi-Heusi (2013), the fraction of people who lie does not increase with the incentive to lie and might even decrease.

With respect to the functional form of the lying cost, most of the previous papers on cheating games argue the lying cost function is convex; "small" lies hurt less than "big" lies, and thus people lie just a little bit. In Mazar et al.'s (2008) matrices-solving task, participants could report to have correctly solved up to 20 matrices, but in practice, the average reported number of matrices varies between 2.8 and 9.4 over the treatments with a lying option. They conclude the magnitude of dishonesty per person is relatively low and argue the participants lie only a little so that they can maintain a positive self-image: “A little bit of dishonesty gives a taste of profit without spoiling a positive self-view.”

Fischbacher and Föllmi-Heusi (2013) also report evidence for what they call “partial liars.” In the die-rolling game, 35% of the participants report an outcome with the highest possible payoff, and statistical inference tells us some of them lie to the full extent. However, 27% of the participants report a number resulting in the second highest payoff; thus, some participants lie but not to the full extent. One interpretation of this result is that participants do not want to tell a big lie (they have a convex lying cost); an alternative explanation is that they try to disguise a lie by not reporting the maximum.

Abeler, Becker, and Falk (2014) use a coin-flipping task to measure lying. Participants toss a coin four times and then report the number of tails. In different treatments, participants report the outcome either via a computer or a phone call, and receive 5 Euros for each tail they report. The majority of the participants in the laboratory reported the second highest outcome of three tail flips. That is, many participants lie but not to the full extent. Cohn, Fehr, and Maréchal (2014) use a similar coin-flipping task and find some participants who lie report the maximum amount of successful coins flips, but the majority lie partially.

Overall, the experimental evidence on the size of lies suggests some participants lie to the full extent, and some lie partially by not reporting the maximal possible outcome. The partial lying might be due to trying to disguise a lie in order to maintain a positive social image or because big lies hurt self-image more than small lies.

4.3 Experimental design and procedure

4.3.1 Incentives Size Experiment

In the "Incentives Size" experiment, we concentrate on changing the incentives and measuring the effect of these changes on lying behavior. In the following, we describe four treatment variations we conducted in this set of the experiments.

Cheating game

The cheating game we use is a variation of Fischbacher and Föllmi-Heusi's (2013) experiment described above, in which a participant rolls a six-sided die in private and then reports the outcome to the experimenter. In the game, we told the participant that if she would report a 5, she would receive $X; otherwise, she would earn $0. In a between-participants design, we vary X such that $X=$1, $5, $20, and $50. We use this game to measure how people react to the incentives in a standard cheating game. The procedure of the cheating game is described on the left side of Table 4.1.

Table 4.1: Structure of the cheating and mind games

Cheating Game	**Mind Game**
• Role a die in private • Report the number • If report "5" receive $X; otherwise receive $0	• Think of a number • Role a die in private • Report "Yes" if the number is the same as you thought of; report "No" otherwise • If report "Yes" receive $X; otherwise receive $0

Notes: The structure of the cheating game is described on the left side, and the structure of the mind game on the right side. In both experiments X=$1, $5, $20 and $50 between subject. 75 observations per cell (600 in total).

Extended cheating game

A possible reason people cheat is that they want to punish someone after being mistreated (e.g., Houser, Vetter, and Winter, 2012). In a sense, people might use being treated badly as a justification to lie. Having such an excuse to cheat may increase cheating in general and change the effect of incentives in particular.

To test this prediction, we gave our participants a tedious retyping task before they took part in the same cheating game as in the first treatment. We gave them piles of 200 questionnaire forms and asked them to type in the data in an excel sheet. We asked them to work for 15 minutes. We explicitly told the participants they would not receive payment for the task. The payoffs and the rules of the cheating game were the same as in the first treatment.

Mind game

In the cheating game described above, participants might be concerned about being exposed when lying, and this concern might increase with incentives. Participating in the lab experiment such as the cheating game for $1 is a normal procedure for many students. Participating in this experiment and receiving $50 for reporting the lucky outcome is unusual, and might make participants question what the consequences will be if he or she lies.

To test the effect of this concern about being exposed on one's behavior in the cheating game, we introduce a second game, which we call a "mind game" (see Jiang, 2013, and Shalvi and De Dreu, 2014, for a similar manipulation). In these treatments, the participant is first asked to think about a number between 1 and 6 and remember the number (without reporting it to us). She then rolls a six-sided die. If the die outcome is the same as the number she thought of, the participant is asked to report "Yes" and otherwise to report "No." If the participant reports "Yes," she receives $X as above. If she reports "No," she receives $0. The procedure of the mind game is described on the right side of Table 4.1.

Two-players mind game

As discussed above, one major difference between the deception game and the cheating and mind games is the effect the decision to lie has on another participant's payoffs. In the next treatment, we test the effect of social preferences by introducing a second player whose payoff depends on the participant's decision to lie. In this two-player version of the mind game, if the participant reports "Yes," she receives $X and the other participant assigned to her receives zero; if she reports "No," she receives zero and the other participant receives $X. Having the second player in the game should increase cheating in general, because reporting "No" results in envy for the first player, and reporting "Yes" leads to compassion, where envy is supposed to be stronger than compassion (e.g., Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000).

This treatment also addresses the concern that participants do not cheat, because they care about the payoff of the experimenter. Because the experimenter pays no matter what the participant reports, such a concern is not applicable here.

4.3.2 Cost Structure Experiment

In the second part of our investigation, we concentrate on understanding the form of the lying cost function. In the following, we describe two treatment variations we conducted in the "Cost Structure" experiment.

Observed game

To test for different hypotheses about the functional form of the lying cost, we introduce a variation of a cheating game, in which we can observe the individual lying behavior. In this game, we ask participants to click, in private, on one of 10 boxes on a computer and reveal a number between 1 and 10; each box has a different number. After seeing the number, the participant is asked to report it, where payments are equal to the number the participant reports in Euros. Because we can later observe the actual number the participant saw, in this treatment, we can observe how often and to what extent participants lie, by comparing the actual number with the reported one.

Non-observed game

As we discussed above, in the previous experiments on cheating, some participants lied, but not to the full extent. One explanation for this behavior is that the participants have signaling concerns with respect to the experimenter when reporting the number. To signal he is not a liar, the participant might report a high number (e.g., 9 in our game) but not the highest number (10 in our game) when lying.

To test for the signaling-to-others hypothesis, we conduct a control treatment in which we do not observe the actual outcomes, and therewith create a possibility for such signaling. In this game, we give the participant a sealed envelope with 10 folded pieces of paper that have numbers from 1 and 10 on them. We ask the participant to take out one piece of paper, observe the number he took out, put it back into the envelope, and then report it. As in the observed game, the payments are equal to the number he reported in Euros.

4.3.3 Experimental procedure

We conducted the experiments in January-July 2015 at the Rady Behavioral Lab, UC San Diego (the first set of treatments with the cheating and mind games), and Cologne Laboratory for Economic Research, University of Cologne (the observed and non-observed treatments). We used the experimental software zTree (Fischbacher, 2007) and recruited participants via ORSEE (Greiner, 2015) and Sona Systems. Overall, we recruited 1,693 participants (50.1% female), and none of them participated in more than one session.

In the UC San Diego experiment (N=1200), we collected 75 observations per treatment (type of game and size of incentive). Participants played the lying game at the end of an experimental session. At the beginning of our experiment, participants received written instructions (full instructions for both sets of experiments are reported in Appendix A) and were allowed to ask questions privately. Then each participant rolled the die in private and reported the outcome on a sheet of paper. After completing the actual experiment, participants were asked to complete a post-experiment questionnaire that included questions on gender, age, and field of study. At the end, participants privately received their payoffs in cash and left the laboratory.

We conducted the Cost Structure experiment at the University of Cologne (N=493), with 390 participants in the observed treatment and 103 in the non-observed treatment. Participants played only our game in the experimental session. We displayed instructions on the computer screen, and the participants were allowed to ask questions privately. Then, depending on the treatment, participants either received the envelopes with numbers in it or were asked to click on one of the boxes on the computer screen and reveal the number. In the next stage, participants reported the outcomes on a sheet of paper and filled out a post-experiment questionnaire that included questions on gender, age, field of study, and motives behind the decisions.

Table 4.2 summarizes the experimental treatments and procedure.

Table 4.2: Experimental summary

Experiment	Treatment name	Description	N
"Incentives Size" (X=$1,$5, $20, $50)	Cheating game	Receive $X if report 5; $0 otherwise.	300
	Extended cheating game	Tedious task before cheating game. Receive $X if report 5; $0 otherwise	300
	Mind game	Receive $X if report yes; $0 otherwise	300
	Two-players mind game	Receive $X if report yes, and other player receives $0; $0 otherwise and other player receives $50	300
"Cost Structure"	Observed game	Numbers between 1 and 10 on the computer. Receive payment equal to the number reported in Euros	390
	Non-observed game	Numbers between 1 and 10 in an envelope. Receive payment equal to the number reported in Euros	103

4.4 Hypotheses

In the "Incentives Size" experiment, we are interested in the effect of the monetary incentives to lie on the fraction of participants who choose to do so. In the cheating game, the monetary benefit from telling a lie is $1, $5, $20, or $50. The first hypothesis is based on a simple incentives effect:

Hypothesis 1: Increasing incentives in the cheating game will increase lying.

However, when the incentives are increased, the concern about being exposed is also increasing. The participant might experience concern about being exposed when lying, for example, from standard preferences regarding penalties. This concern may positively depend on the size of the monetary gain from lying. According to the social psychology literature, and in line with intuition, stakes are predicted to increase the fear of getting caught when lying (e.g., Ekman, 1985; Ekman, 1988; Vrij, 2008). If the concern about being exposed when lying is high enough, the incentive effect in the cheating game might even be negative.

By contrast, in the mind game, participants know their lie will not be exposed. This reduction in the concern about being exposed suggests two hypotheses. First, with respect to the overall lying, participants will lie more in the mind game than in the cheating game, because the concern element disappears:

Hypothesis 2: Participants lie more in the mind game than in the cheating game.

Second, because the concern about being exposed is eliminated, we also anticipate a stronger effect of incentives:

Hypothesis 3: Increasing incentives in the mind game will increase lying.

For all the hypotheses stated above, we need to assume the magnitude of the intrinsic lying cost is increasing more slowly than the magnitude of the incentive for lying; otherwise, we would not expect a positive incentive effect. In the "Cost Structure" experiment, we test for the functional form of the lying cost, which in turn also tests the assumption we made while formulating Hypotheses 1-3. We test two opposing hypotheses:

Hypothesis 4a: Fixed cost of lying: If participants lie, they do so to the full extent.

Hypothesis 4b: Convex cost of lying: If participants lie, they do not lie to the full extent.

If we find that in the observed game, the majority of participants report the payoff-maximizing outcome of 10, such a finding supports the fixed-cost-of-lying hypothesis. If, by contrast, participants overreport the actual outcomes only by a bit, such a finding will support the convex-cost-of-lying explanation.

People could be lying not to the full extent for two reasons. First, they might have a convex cost of lying—small lies hurt the self-image less than big lies (signaling to self). Second, they might not want to appear as liars (signaling to others). In the observed game, appearing honest is not possible if the participant lies but reports a number lower than 10, because the experimenter can verify the actual outcome. In the non-observed game, the experimenter does not know what the actual state was; therefore, the participant can signal she is not a liar, by reporting a number that is higher than the number she observed but lower than 10. In the non-observed game, a participant who cares enough about her social image might report a number that is higher than the number she observed but lower than 10. As a consequence, we expect to find more partial lying in the non-observed game than observed game:

Hypothesis 5: More partial lying occurs in the non-observed than in the observed game.

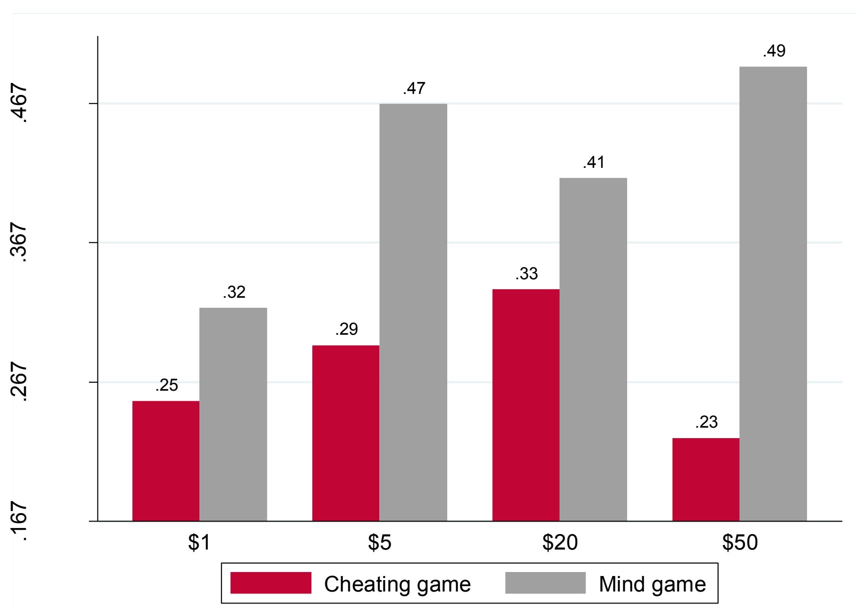
4.5 Results

In what follows, we first examine the incentive effects on lying in the cheating and mind games and show the results are robust to treatment variations, such as adding the second player or letting participants work first on a tedious task without payment. We then report the results from the "Cost Structure" experiment, in which we take a closer look at the lying cost structure and identify the functional form of the lying cost.

4.5.1 Reaction to incentives in cheating and mind games

The left bars of Figure 4.1 display the fraction of participants who reported a die outcome of 5 in the cheating game. When the incentive is $1, 25% of participants claim to have rolled a 5. Slightly more participants report a 5 when the incentive is $5 or $20, with 29% and 33% of positive reports, respectively. When the incentive is increased to $50, the reporting of a 5 declines and amounts to 23%. The pairwise between-treatments differences are not significant in a Fisher test (p>0.1, two-sided). The overall trend is also not significant—when testing for an order (both ascending or descending) in the four-incentives conditions, we find no effect of stakes on the fraction of lies (p>0.1, Jonckheere-Terpstra).

Figure 4.1: Lying in the cheating and mind games



Notes: The left columns represent levels of cheating (above the expected 16.7%) in the cheating game by incentives. The right columns represent these numbers for the mind game.

The expected number of participants who rolled a 5 is 16.67%; hence, in expectation, any value above it is based on lies. We run a binomial test and find participants lie significantly for $1, $5, and $20 with p=0.061, p=0.008, and p<0.001, respectively. In the \$50 treatment, lying becomes insignificant, with p=0.164. This result is in line with the concern about being exposed increasing with incentives, such that participants do not lie when the stakes are increased by a factor of 50.

With respect to the control treatment, we find that in the cheating game with a tedious task beforehand, 32% of participants report a 5 when the incentive is \$1, \$5, and \$20, and 29% report a 5 for \$50. As in the main cheating game, the pairwise between-treatments differences are not significant in a Fisher test (p>0.1). When testing for an ascending order in a Jonckheere-Terpstra test, we find no significant effect of stakes on the fraction of lies (p>0.1). Finally, we compare the theoretical distribution with reported outcomes and find lying is significant for each size of the incentive in a binomial test on a 1% level.

We run a Probit regression to check for the robustness of the incentive effects in the cheating games (Table 4.3 lists the results). In the regression, we also include the dummy *Real-effort stage* to control whether lying is affected in the control cheating game. As can be seen from the first model (columns 1 and 1m), we find the same results in the regression as in the non-parametric tests. The insignificant coefficients of the *\$5, \$20 and \$50* dummies in the model show that in the cheating game the size of incentives has no effect on lying. The regression also shows lying in the cheating game does not change if we introduce the real-effort stage. From the insignificant coefficient of *Real-effort stage* (see column 1) follows that participants do not lie more when they have to work on a tedious task before the cheating game.

Based on this regression result, and the fact that all pairwise non-parametric comparisons between the original cheating game and the cheating game with the real-effort stage are also not significant (p>0.1, Fisher test), we pool the data from the main cheating game and the cheating game with the real-effort stage and use it as the cheating-games data in the following analysis; to be conservative, we also report the analysis of the separate treatments in the text. The above analysis reveals our data reject Hypothesis 1. In particular, we find the following:

Result 1. Increasing incentives in the cheating game does not increase lying.

Table 4.3: Probit regression models for cheating and mind games

	(1)	(1m)	(2)	(2m)
Dependent variable	Lying	Lying	Lying	Lying
Type of the game	Cheating	Cheating	Mind	Mind
Model	Probit	Probit(ME)	Probit	Probit(ME)
$5	0.049	0.017	0.250*	0.098*
	[0.153]	[0.053]	[0.149]	[0.059]
$20	0.124	0.043	0.421***	0.165***
	[0.153]	[0.054]	[0.148]	[0.058]
$50	-0.072	-0.025	0.519***	0.203***
	[0.156]	[0.053]	[0.148]	[0.058]
Real-effort stage	0.145	0.050		
	[0.110]	[0.038]		
Second player			-0.079	-0.031
			[0.105]	[0.041]
Male	0.288***	0.099***	0.152	0.059
	[0.111]	[0.038]	[0.105]	[0.040]
Constant	-0.786***		-0.580***	
	[0.140]		[0.134]	
Observations	599	599	599	599
Pseudo R-squared	0.013	0.013	0.020	0.020

Notes: The models are Probit specifications with the fraction of positive reports ("5" or "Yes") as the dependent variable. Robust standard errors are clustered on the level of the subject. Columns 1m and 2m report marginal effects from Probit models. *, ** and *** denote significance at the 10%, 5% and1% level, respectively. $5, $20, $50, Male, Real-effort stage, and Second player are dichotomous variables equal to 1 if the participant is in the $5 treatment, $20 treatment, $50 treatment, the participant is male, the participant worked on a real-effort task, and we introduced a second player, respectively. The reference group in Model 1 consists of participants in the cheating game. The reference group in Models 2 consists of participants in the mind game.

Moving from the cheating game to the mind game, we conjectured two effects: first, overall lying will increase; second, the fraction of participants who lie will depend positively on incentives.

The right bars in Figure 4.1 present the results from the mind game in terms of the fraction of participants who reported the number they thought of was the same number they rolled afterward (reporting "Yes"). With respect to the first conjecture, we indeed observe that more participants lie in the mind game than in the cheating game. In the mind game, 32%, 47%, 41%, and 49% of participants report a positive outcome for $1, $5, $20, and $50, respectively, whereas in the cheating games, the positive reporting rate varied between only 23% and 33%. When we compare the fractions of positive reports in the pooled cheating-game data and the mind game, we observe that for each size of the stake, participants lie more in the mind game than in the cheating game, and for stakes of $5 and $50, the between-treatments difference is significant in a Fisher test, with p=0.027 and 0.001, respectively. Other pairwise comparisons are not significant (p>0.1).

In particular, when comparing the lying in the cheating game (not pooled) with the mind game, all the effects hold – participants lie more in the mind game than in the cheating game, and for the stakes of $5 and $50, the between-treatments difference is significant in a Fisher test with p=0.043 and 0.001, respectively. Other pairwise comparisons are not significant (p>0.1). In addition, in the mind game, lying is significant for each size of the incentive in a binomial test (p<0.001).

In the two-player mind game, participants also lie more than in the cheating games. Here, 28% of participants report "Yes" for $1, 29% do so for $5, and the fraction of positive reports amounts to 49% for both $20 and $50. In this game, lying is significant in a binomial test for each size of the stakes on the 1% level. When we compare the lying in the two-player mind game with lying in the pooled-cheating game, we find the fractions of positive reports are higher in the two-player mind game, when the stakes are $20 and $50, with p=0.020 and p=0.001 (Fisher test), respectively, and are not different for $1 and $5. When comparing the lying in the cheating game (not pooled) with the two-player mind game, all effects hold. We find the fractions of positive reports are higher in the two-player mind game than in the cheating game, when the stakes are $20 and $50, with p=0.068 and p=0.001 (Fisher test), respectively, and are not different for $1 and $5 (p>0.1).

We run a Probit regression to find further support for lying being higher in mind games than cheating games. Table 4.4 reports the regression results. We include the dummy *Mind game* to measure whether participants lie more in the mind game. As can be seen from column 1m, the marginal effect of dummy *Mind game* is positive, significant, and large in size. The marginal effect of 0.147 means that, on average, participants lie 14.7 percentage points more in the mind game than in the cheating game. That is, making the environment safe such that there is no possibility of being caught when lying makes participants lie significantly more. The above analysis reveals our data do not reject Hypothesis 2. In particular, we find the following:

Result 2. A larger fraction of participants lie in the mind game than in the cheating game.

Table 4.4: Probit regression model for overall lying

	(1)	(1m)
Dependent variable	Lying	Lying
Model	Probit	Probit(ME)
Mind game	0.401***	0.147***
	[0.107]	[0.039]
$5	0.159	0.060
	[0.108]	[0.041]
$20	0.278***	0.105***
	[0.107]	[0.041]
$50	0.238**	0.090**
	[0.107]	[0.041]
Real-effort stage	0.137	0.051
	[0.109]	[0.041]
Second player	-0.072	-0.026
	[0.104]	[0.038]
Male	0.218***	0.080***
	[0.076]	[0.028]
Constant	-0.889***	
	[0.114]	
Observations	1198	1198
Pseudo R-squared	0.022	0.022

Notes: The model is a Probit specification with the fraction of positive reports as the dependent variable. Robust standard errors are clustered on the level of the subject. Column 1m reports marginal effects from the Probit model. *, ** and *** denote significance at the 10%, 5% and 1% level, respectively. $5, $20, $50, Male, Real effort stage, Second player, and Mind game are dichotomous variables equal to 1 if the participant is in the $5 treatment, $20 treatment, $50 treatment, the participant is male, the participant worked on a real-effort task, we introduced a second player, and the participant played the mind game, respectively. The reference group consists of participants in the cheating games.

In the regression models, we also control for gender effects. We use regression models in Table 4.3 to measure these gender effects. We find men lie significantly more in the cheating game—the marginal effect of *Male* is significant, positive, and large (see column 1m, Table 4.3). The marginal effect of 0.099 means that, on average, men lie 9.9 percentage points more than women. This gender effect, however, disappears when we move to the mind game—the insignificant marginal effect of *Male* in column 2m (Table 4.3) shows men do not lie more than women in the mind game.[49]

The result that women and men lie to the same degree in the mind game, but that in the cheating game, women lie less is in line with the gender differences found in the risk-aversion literature, where women are shown to be more risk averse (e.g., see Charness and Gneezy, 2012). In the cheating game, the risk of getting caught keeps women from lying, whereas the risk affects men to a lower extent. As soon as this risk is eliminated, women and men lie to the same degree.

After we show that a larger fraction of participants lie in the mind game than in the cheating game, the second conjecture we have with respect to the mind games is that lying will increase with incentives in the mind game; the data support this conjecture. In the mind game, when the gain from reporting "Yes" is $1, 32% of participants report it. This fraction increases to 47% ($5), 41% ($20), and 49% ($50). The pairwise between-treatments comparisons show a marginally significant difference between lying for $1 and $5 and a significant difference between $1 and $50, with $p=0.094$ and $p=0.046$, respectively. Other pairwise comparisons are not significant ($p>0.1$). When we test for an ascending order in a Jonckheere-Terpstra test over the four conditions, we find a significant increasing effect of stakes on the fraction of lies, with $p=0.034$. That is, participants lie more when the stakes are raised in the mind game.

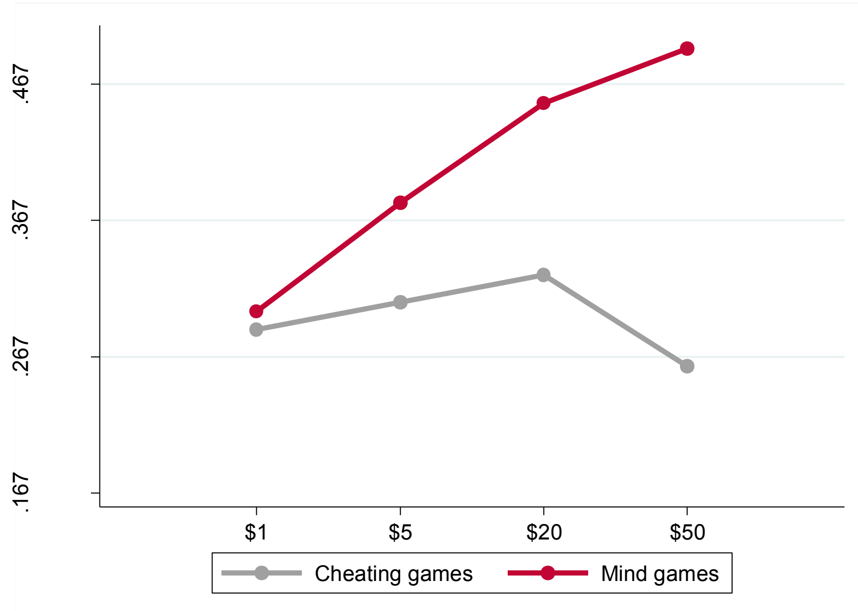
[49] In general, gender differences in lying are still relatively unexplored. Dreber and Johannesson (2008) find that in situations where the sender can benefit at the expenses of the receiver by lying ("selfish black lies"), men are more likely to lie than women. Erat and Gneezy (2012) replicated this result, but found that in situations when the sender loses but the receiver gains in the case of a lie, women were more likely to lie. Conrads and Lotz (2015) find that women are more responsive to changes in communication channels (see Croson and Gneezy, 2009). Muehlheusser et al (2015) used a similar die-rolling paradigm as in Fischbacher and Föllmi-Heusi (2013), dividing participants up into groups of two. They find no gender differences in cheating, but more cheating in male groups and mixed groups than in female groups.

In the two-player mind game, lying also increases with stakes. The fraction of participants reporting “Yes” goes from 28% and 29% for $1 and $5, respectively, up to 49% in the $20 and $50 conditions. The pairwise between-treatments comparisons show a significant difference between lying for $1 and $20, $1 and $50, $5 and $20, and $5 and $50, with p=0.012, p=0.012, p=0.019, and p=0.019, respectively; other pairwise comparisons are not significant (p>0.1). When we test for an ascending order in a Jonckheere-Terpstra test, we find a highly significant effect of stakes on the fraction of lies, with p<0.001.

In the robustness checks using a Probit regression (see Table 4.3), as can be seen from Model 2 (columns 2 and 2m), we find a significant incentives effect in the mind games. The marginal effects of *$5, $20 and $50* dummies are significant and increasing with the size of the incentive (column 2m). The regression model also shows lying in the mind game does not change if we introduce the second player. From the insignificant coefficients of *Second player* (see columns 2 and 2m) follows that introducing the second player, and therewith social preferences, does not change participants' lying behavior. Using pairwise non-parametric comparisons between the original mind game and the two-player mind game, we find between-treatments differences are not significant for $1, $20, and $50 (p>0.1, Fisher test), but the difference is significant for $5, with p=0.043. Based on the insignificant *Second player* coefficient in the regressions and most of the non-parametric test results being insignificant, we pool the data from both mind games for our tests, as well as report the non-pooled data tests.

Figure 4.2 presents the pooled results for the cheating games and mind games. Figure 2 shows a clear large incentive effect on lying in the mind games but not in the cheating games. The non-parametric tests results from the pooled data, which are the same as in the separate data analyses, support this effect. In particular, a highly significant positive incentives effect on lying exists in the mind games (p<0.001), whereas no such effect exists in the cheating games (p=0.719, Jonckheere Terpstra).

Figure 4.2: Lying in the cheating and mind-games groups



Notes: The lines represent levels of cheating (above the expected 16.7%) in the cheating and mind games by incentives.

The above analysis reveals that our data do not reject Hypothesis 3. In particular, we find the following:

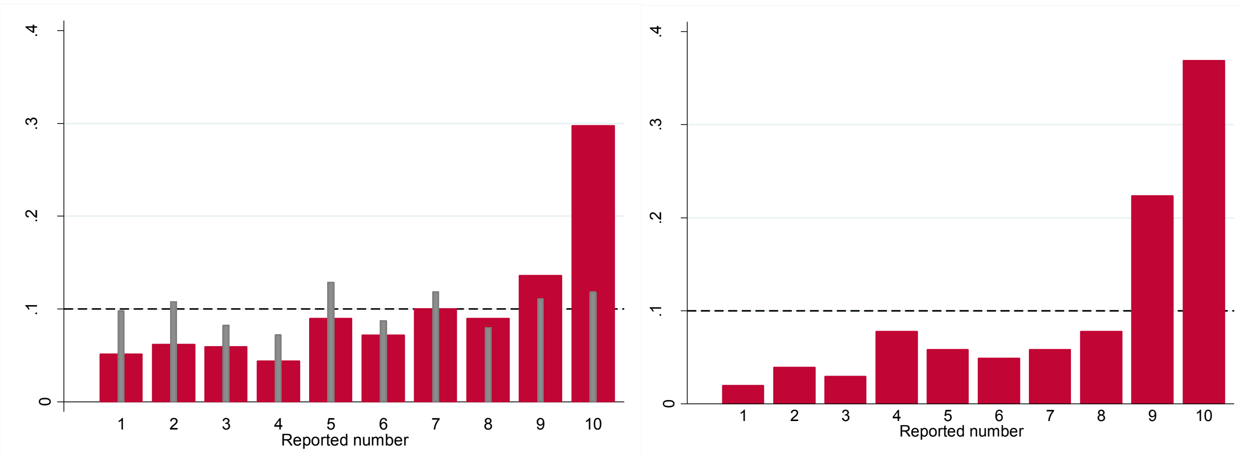
Result 3. Increasing incentives in the mind game increases lying.

4.5.2 The functional form of lying cost

We now turn to the results of the "Cost Structure" experiment. Figure 4.3 illustrates the results from the observed game. The figure shows the distributions of actual and reported numbers on the aggregate level and is the first indicator that the reported numbers are higher than the actual numbers in the observed game.

Because, differently from the previous literature, we observe the actual outcomes in this game, we can measure lying on the *individual level.* To do so, we compare the distribution of actual outcomes with the distribution of reported numbers in a Wilcoxon Matched-Pairs Signed-Ranks test. We find the reported numbers are significantly higher than the actual outcomes, with p<0.001. Overall, 26% (103 out of 390) of participants report incorrect numbers. They lie by over-reporting the numbers; none of the participants underreport the actual outcome.

Figure 4.3: Distribution of reported numbers in the observed and non-observed games

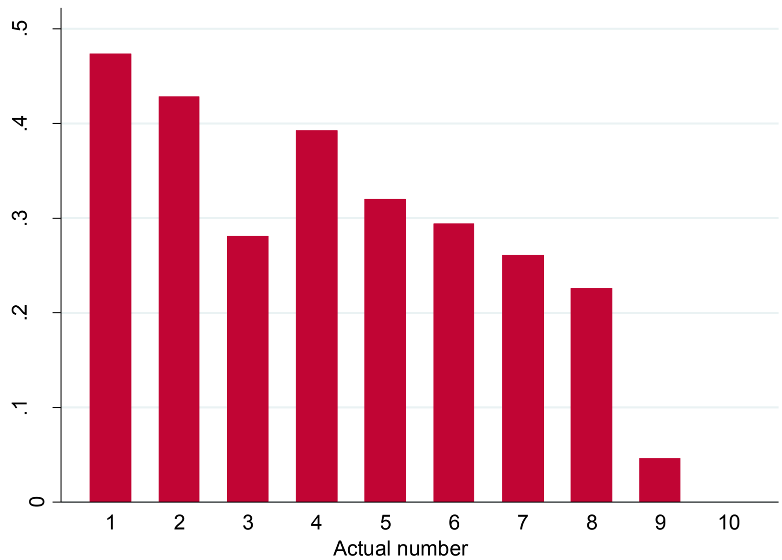


(a) (b)

Note: Figure 4.3a represents the distribution of reported and actual numbers in the observed game, and Figure 4.3b represents the distribution of reported numbers in the non-observed game. The thick dark bars show the reported numbers, whereas the thin light gray bars show the actual numbers. The dashed lines show the theoretical distribution of the outcomes; note that in previous experiments (including those in the first set of treatments above and the non-observed game), the theoretical prediction was used for the testing of hypotheses.

From Figure 4.4, we can see the lower the actual outcome is, the more likely participants are to lie. For example, only 5% of participants who observed a 9 overreport their number, whereas 47% of the participants who observed a 1 did. A Jonckheere Terpstra test indicates lying is significantly descending when the actual outcome increases (p<0.001). A Spearman correlation supports the result that lying is higher for low outcomes, with a Spearman's rho of -0.318 and p<0.001.

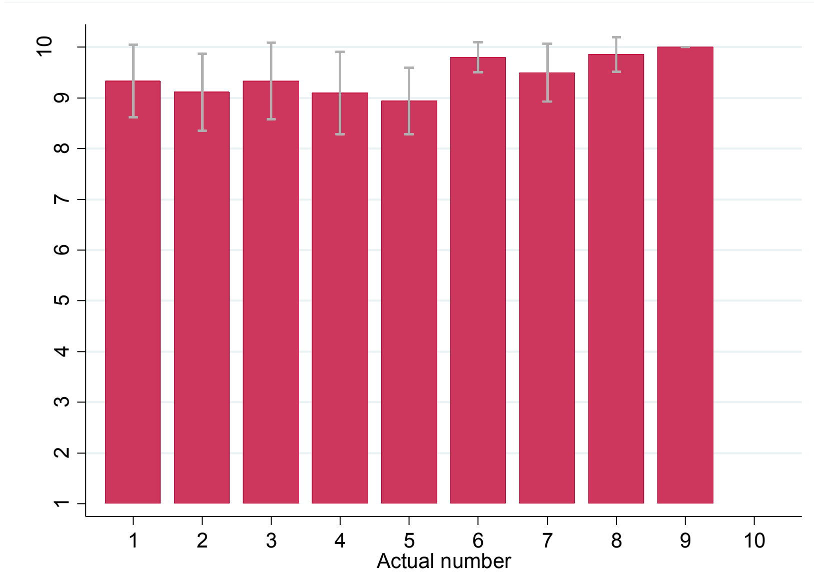
Figure 4.4: Fraction of lying conditional on the actual outcome



Note that in the observed game, the incentive to lie is between 0 and 9 Euros (reporting 10 when seeing 1). Thus, the situation is comparable to the low-stakes domain in the cheating game above, where the concern about being exposed is relatively low. The data are consistent with participants comparing the benefit of the lie with the cost of the lie. The higher the difference between the actual outcome and 10, the higher fraction of participants lie, showing incentives matter in the observed game.

Next, we test the extent to which participants lie. Figure 4.5 displays the average numbers reported for participants who lie, as a function of the actual outcome. We find that for each actual number, participants report high outcomes, with the average varying between 8.94 and 10 and the overall average of 9.32. Importantly, and in contrast to the convex-cost-of-lying hypothesis, the number that those who lie reported does not depend on the actual outcome (Spearman's rho=0.052, p>0.1). That is, participants do not lie by a bit and increase their actual outcome by an increment: 68% of the participants who lie report the maximal possible outcome of 10. Figures 4.6 and 4.7 present the distribution of reported numbers.

Figure 4.5: Average numbers reported by participants who lie



Note: Error bars are standard errors of the mean.

Figure 4.6: Distribution of numbers reported by participants who lie

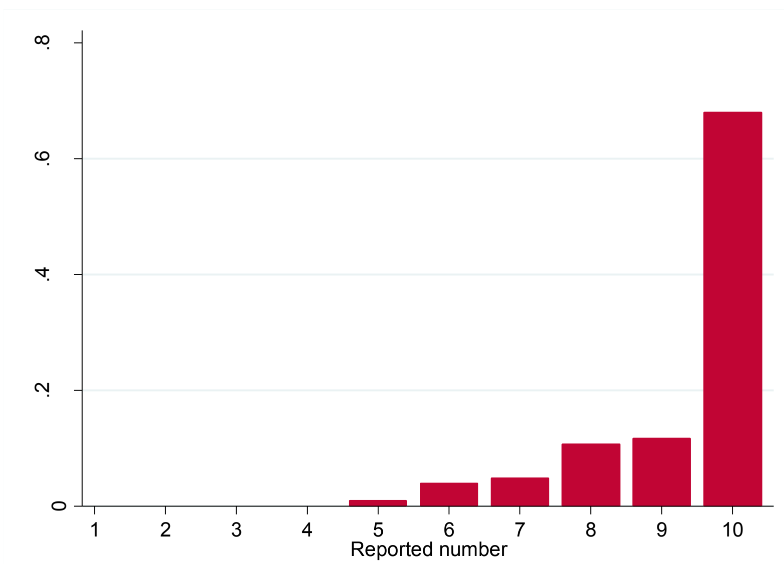
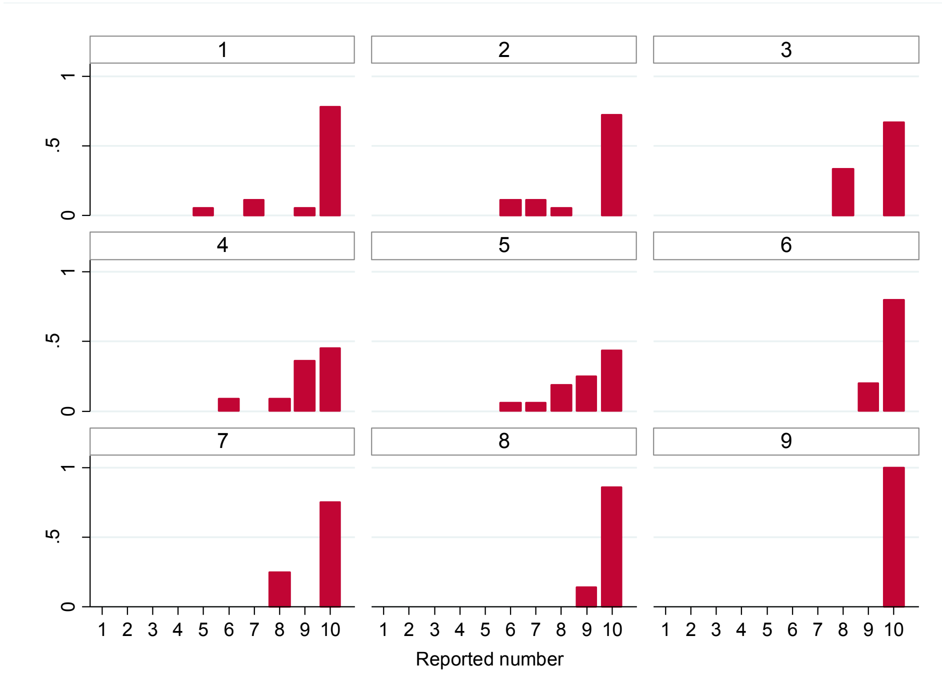


Figure 4.7: Distribution of numbers reported by participants who lie, for each actual outcome



Notes: Figure 4.7 represents the distribution of reported numbers conditional on lying for each actual outcome separately. The number on the top of the cell stands for the actual outcome, and the numbers on the x-axis, for the reported numbers.

The above findings show the hypothesis that participants lie by a little bit, which is argued widely in the literature, is rejected in our experiment. The evidence from our experiment is in line with the fixed cost of lying—most of the participants who lie do so to the fullest extent. Therefore, the level of lying is not correlated with the actual outcome. The above analysis reveals our data do not reject Hypothesis 4a but do reject Hypothesis 4b. In particular, we find the following:

Result 4. A convex cost of lying is rejected; the results are consistent with a fixed cost of lying.

With respect to the gender effects, we find men report significantly higher numbers than women, with an average reported number of 6.71 for women and 7.40 for men (p=0.037, MWU). When comparing the actual outcomes with the reported numbers, we find 23% of the numbers that women report are lies, whereas the fraction of lies for men is 31%. The difference is marginally significant in a Fisher test, with p=0.085. The result that men lie more than women in the observed game is in line with the result found in the cheating game.

4.5.3 Trying to appear honest

We now compare the behavior of participants in the observed game with that of participants in the treatment in which we cannot observe the actual outcome. In the non-observed game, participants who lie may choose to hide behind high but not maximal numbers. A reason to do so is to signal to the experimenter that they do not lie. Figure 4.3b presents the distribution of reported outcomes in the non-observed game. As the figure shows, participants report higher numbers than in the theoretical distribution; a Kolmogorov Smirnov test confirms participants lie significantly in the non-observed game ($p<0.001$). In addition, overall lying is higher in the non-observed game than in the observed game ($p=0.016$, MWU).

The figure also shows that in the non-observed game, a substantial fraction of participants report a 9, which we do not find in the observed game. Whereas only 14 % report a 9 in the observed game and do not deviate significantly from the actual fraction who observed a 9 ($p=0.106$, binomial test), 22 % report a 9 in the non-observed game, which is significantly more than the theoretical prediction ($p<0.001$). Moreover, the difference in reporting a 9 is significant between the games, with $p=0.033$ (Fisher test). That is, some participants report a high but not the maximal outcome, which confirms the signaling-to-others hypothesis—some participants try to appear not to be liars when lying. As a result, we observe more partial lying in the non-observed game, in which such signaling is possible. The above analysis reveals our data do not reject Hypothesis 5. In particular, we find the following:

Result 5. A higher fraction of participants lies partially in the non-observed game than in the observed game.

Finally, in line with the results from the cheating game and the observed game, men lie significantly more than women in the non-observed game. The average number that women report is 7.26; for men, it is 8.41 (p=0.046, MWU).

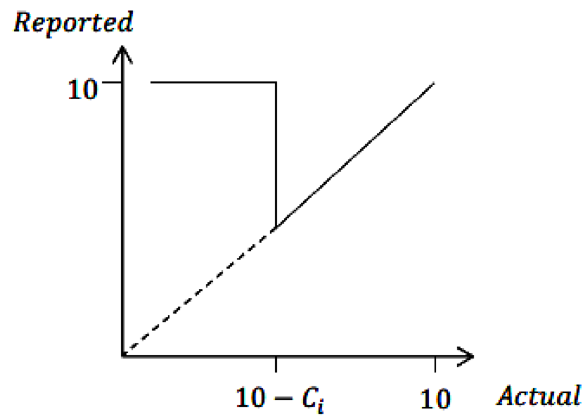
4.6 Conclusion

Although the economics literature has extensively studied lying, many questions regarding individual lying behavior remain unanswered, in particular the structure of the intrinsic cost of lying. For example, the evidence reported in the literature shows people do not cheat more when the incentives to do so are increased. This evidence suggest that people have either zero lying costs and will always lie when doing so benefits them, or those with infinite costs who will never lie. When we asked people in the audience during a seminar about their type, only one out of 130 people did not self-identify as a third type: someone who has a finite lying cost and would lie if the benefit outweighs the cost.

In this paper, we show that when we modify the standard cheating game to reduce the concern about being exposed as a liar, participants do react to the incentives to lie. This observation indicates that the finding in the literature regarding an intrinsic lying cost picked up a different element in the utility of the participant: incentives did not affect lying, because increasing the incentives to lie also increased the participant's concern about being exposed, and the effect was not due to the structure of the intrinsic cost of lying. Whereas our results show that incentives increase cheating in the mind game, they also show this effect is limited, because at least half of our participants chose not to lie. The large fraction of people who did not lie could be caused by our incentives not being large enough, by an additional concern that we did not identify, or because many people are just not willing to lie for money.

After showing that incentives affect lying, we turn to studying the structure of this intrinsic lying cost. The evidence in the "Cost Structure" experiment indicates again a significant incentives effect. The evidence also supports the hypothesis regarding a fixed lying cost, rejecting the hypothesis regarding a convex cost. We find that if people lie, they do so to the full extent, whereas partial lying is infrequent and significant only when one can signal to others by reporting a non-maximal outcome. Figure 4.8 offers a simple demonstration of the lying decision with a fixed lying cost.

Figure 4.8: Demonstration of choices under a fixed lying cost



As Figure 4.8 illustrates, the decision maker is doing a simple cost-benefit calculation. When the benefit of the lie is larger than the cost of lying, she will lie to the full extent; otherwise, she will tell the truth. In our experiment, whenever the difference (10 minus the actual number) is larger than the intrinsic cost of lying C_i, the participant will lie and say 10. Otherwise, she will say the actual number.

Together, the results show that ethical behavior associated with the intrinsic cost of lying follows standard economic reasoning in our experiments. When the benefit from lying increases without increasing the concern of being exposed, more people choose to lie, and once a person chooses to violate the norm of not lying, she does so to the full extent.

These findings have important implications for constructing policies to reduce cheating. When the interaction is closer to the cheating game one, in which peoples' concern regarding being exposed increases with incentives, it might be better to make the incentives associated with the decision to cheat high. On the other hand, when the interaction is closer to the mind game one, in which people are less concerned about the chance of being exposed, it might make sense to reduce the incentives to cheat. Consider, for example, tests in schools. The policy maker can decide whether to give the students one large exam on the topic, or few smaller ones. A large exam will increase the incentive to cheat but might also decrease the chance of cheating if it sufficiently increases the students' concern about being detected. But if the concern of being detected is not increased enough, while the incentives to cheat are increased, then the students might be more likely to cheat. Hence, if the perceived chance of being exposed is high enough and increases with incentives, our results suggest that the policy maker will reduce cheating by administer one very important exam. However, if the perceived chance of being exposed is not high, then the policy maker will be better off designing a few less important tests instead of one.

Another policy implication that could be tested in future research relates to detecting cheating. The result showing that our participants are best characterized by fixed cost of lying implies that it is more likely to find cheaters close to the maximum level they can report. When the actions themselves are verifiable by the audit technology, as in our observed game, most of the cheating occurs at the maximum. When the audit technology is less precise, our results (as well as the findings in the literature) suggest that much of the cheating is done at levels that are slightly below the maximum. These findings can help in detecting cheating when the detection technology is costly. Of course, one will have to take cheaters' strategic consideration into account in such games.

4.7 Appendix: Experimental Instructions

The instructions for the cheating and mind games include only the treatments with a $1 incentive. The instructions for the observed and non-observed games are translated from German.

Cheating game

Thank you for taking part in our experiment. This experiment is part of a study we are conducting, and your help is highly appreciated.

We have given you a die, and ask you to roll it in private after you read the instructions.

We will then ask you to write at the bottom of this form the number that you rolled. Your payment will be $0 if you write “1”, “2”, “3”, “4,” or “6,” and it will be $1 if you write “5.” That is, if you write “5,” you will be paid $1; if you write a different number, you will not be paid. The payoffs are as follows:

Number thrown	1	2	3	4	5	6
Payoff	$0	$0	$0	$0	$1	$0

After you have rolled the die, please write down the number you rolled. Then come to the other room of the lab to fill out a post-questionnaire and receive your payment. Please bring this form with you.

Do you have any questions?

Please roll the die in private now and report the number that came up: _____

Thank you for participating!

Cheating game with a real-effort stage

Part 1

Thank you for taking part in our experiment. This experiment is part of a study we are conducting, and your help is highly appreciated. This experiment consists of two parts. After you have finished the first part, you will receive the instructions for the second part.

Part 1: We gave you a packet with 200 copies of questionnaire forms. Participants in a previous experiment filled out these questionnaires. Your task is to copy the data into the excel sheet on your computer screen.

First, please write down your participant number on the upper left side. Also, please write down what time you started working on the task and what time you finish. We ask you to work on the task for 15 minutes.

We ask you to fill in six columns: “participant number,” “gender,” “ethnicity,” “age,” “field of study,” and “year.” If you cannot read a questionnaire form, please just continue to the next one.

At the end of the 15 minutes, please raise your hand and a research assistant will come over to you to save the excel file and give you the instructions for the second part of the experiment.

Do you have any questions?

Thank you!

Part 2

These are the instructions for the second part of the experiment.

We have given you a die, and ask you to roll it in private after you read the instructions.

We will then ask you to write at the bottom of this form the number that you rolled. Your payment will be $0 if you write “1”, “2”, “3”, “4,” or “6,” and it will be $1 if you will write “5.” That is, if you write “5,” you will be paid $1; if you write a different number, you will not be paid. The payoffs are as follows:

Number thrown	1	2	3	4	5	6
Payoff	$0	$0	$0	$0	$1	$0

After you have rolled the die, please write down the number you rolled. Then come to the other room of the lab to fill out a post-questionnaire and receive your payment. Please bring this form with you.

Do you have any questions?

Please roll the die in private now and report the number that came up: _____

Thank you for participating!

Mind game

Thank you for taking part in our experiment. This experiment is part of a study we are conducting, and your help is highly appreciated.

We have given you a die, and ask you to roll it in private after you read the instructions.

Now think of a number from 1 to 6 and remember it. Then, at the bottom of this form, please write the number that you rolled at the bottom of this form.

If the number you rolled is the number you thought of, please mark “Yes” at the bottom of the form, and your payment will be $1. If the number you rolled is not the number you thought of, please mark “No” at the bottom of the form, and your payment will be $0.

After rolling the die, writing down the number you rolled, and marking “Yes” or “No,” please come to the other room of the lab to fill out a post-questionnaire and receive your payment. Please bring this form with you.

Do you have any questions?

Please roll the die in private now and report the number that came up: _____

Please indicate whether the number you rolled is the number that you thought of:

□ Yes

□ No

Thank you for participating!

Two-player mind game

Thank you for taking part in our experiment. This experiment is part of a study we are conducting, and your help is highly appreciated. In the experiment, you will be matched with another participant.

We have given you a die, and ask you to roll it in private after you read the instructions.

We ask you now to think of a number from 1 to 6 and remember it. Then, at the bottom of this form, please write the number you rolled.

If the number you rolled is the number you thought of, please mark "Yes" at the bottom of the form. Your payment will be $1, and the person you are matched with will be paid $0. If the number you rolled is not the number you thought of, please mark "No" at the bottom of the form. Your payoff will be $0, and the person you are matched with will be paid $1.

After rolling the die, writing down the number you rolled, and marking "Yes" or "No," please come to the other room of the lab to fill out a post-questionnaire and receive your payment. Please bring this form with you.

Do you have any questions?

Please roll the die in private now and report the number that came up: _____

Please indicate whether the number you rolled is the number that you thought about:

□ Yes

□ No

Thank you for participating!

Observed game

/Screen 1/

Welcome to our experiment. Please read the instructions carefully. If you have a question, please raise your hand and we will come over to you. Please do not communicate with other participants during the experiment.

Every participant will receive 2.50 Euros for attending, which will be paid out independently of the decisions made in the experiment.

Furthermore, you will be able to earn additional money. At the end of the experiment, you will receive the income that you earned over the course of the experiment, plus the 2.50 Euros for attending, in cash.

Your decisions are private, and no other participant will know about them.

/Screen 2/

On the next screen, you will see 10 boxes with numbers hidden behind them. The numbers in the boxes are 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10, and they are placed in a random order. We will ask you to click on one box.

Once you click on the box, you will see a number that we ask you to remember and later report to us.

The number you report determines how much money you will be paid. You will be paid the equivalent in Euros to the number you report. In other words, if you write "1," you will receive 1€. If you write "2," you will receive 2€; if you write "3," you will receive 3€, and so on.

If you have any questions, please raise your hand and we will come to you!

/Screen 3/

Please click on one of the boxes: _____

Please remember the number you saw.

/Screen 4/

On a sheet of paper we will give you soon, please report the number that you saw. Please also write down your cubical number on that sheet.

We will collect the sheets of paper from every participant, as soon as everybody is done, and will prepare the payments.

/On paper/

Please report the number that you saw: _____

Your cubical number: ______

Non-observed game

/Screen 1/

Welcome to our experiment. Please read the instructions carefully. If you have a question, please raise your hand and we will come over to you. Please do not communicate with other participants during the experiment.

Every participant will receive 2.50 Euros for attending, which will be paid out independently of the decisions made in the experiment.

Furthermore, you will be able to earn additional money. At the end of the experiment, you will receive the income that you earned over the course of the experiment, plus the 2.50 Euros for attending, in cash.

Your decisions are private and no other participant will know about them.

/Screen 2/

We will give you an envelope with 10 pieces of paper in it. The numbers on the paper pieces are 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10, and they are placed in a random order. We will ask you to take out one piece of paper and later to put it back in the envelope.

Once you pull a piece of paper out of the envelope, you will see a number that we ask you to remember and later report to us.

The number you report determines how much money you will be paid. You will be paid the equivalent in Euros to the number you report. In other words, if you write "1," you will receive 1€. If you write "2," you will receive 2€; if you write "3," you will receive 3€, and so on.

If you have any questions, please raise your hand and we will come to you!

/Screen 3/

Please take one number out of the envelope.

Please remember the number you saw.

Please put the piece of paper back in the envelope.

/Screen 4/

On a sheet of paper we will give you soon, please report the number that you saw. Please also write down your cubical number on that sheet.

We will collect the sheets of paper from every participant, as soon as everybody is done, and will prepare the payments.

/On paper/

Please report the number that you saw: _____

Your cubical number: ______

References

Abeler, J., Becker, A., Falk, A. (2014). Representative evidence on lying costs. *Journal of Public Economic*s, 113, 96-104.

Allingham, M.G., Sandmo, A. (1972). Income Tax Evasion: A Theoretical Analysis. *Journal of Public Economics*, 1, 323-338.

Andreoni, J. (1989). Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence. *Journal of Political Economy*, 97(6), 1447-1458.

Andreoni, J. (1990). Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. *The Economic Journal*, 100, 464-477

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

Ariely, D., Bracha, A., Meier, S. (2009). Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially. *American Economic Review*, 99(1), 544-555.

Augustine (421). Enchiridion on faith, hope and love. Washington, DC: Regnery Publishing, Inc., 1996.

Baron, D. P. (2007). Corporate Social Responsibility and Social Entrepreneurship. *Journal of Economics & Management Strategy*, 16(3), 683-717.

Bartling, B., Engl, F., Weber, R.A. (2014). Does Willful Ignorance Deflect Punishment? - An Experimental Study. *European Economic Review,* 70, 512-524.

Bartling, B., Weber R.A., Yao, L. (2015). Do Markets Erode Social Responsibility? *Quarterly Journal of Economics*, 130(1), 219-66.

Battigalli, P., Dufwenberg, M. (2007). Guilt in Games. *American Economic Review*, 97(2): 170-176.

Battigalli, P., Charness, G., Dufwenberg, M. (2013). Deception: The role of guilt. *Journal of Economic Behavior & Organization*, 93(C), 227-232.

Becker, G.S. (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy*, 76(2), 169-217.

Berger, J., Harbring, C., Sliwka, D. (2013). Performance Appraisals and the Impact of Forced Distribution- An Experimental Investigation. *Management Science*, 59(1), 54-68.

Bénabou, R., Tirole, J. (2006). Incentives and Prosocial Behavior. *American Economic Review*, 96(5): 1652-1678.

Bénabou, R., Tirole, J. (2012). Laws and Norms. *IZA Discussion Paper 6290, Institute for the Study of Labor (IZA).*

Bicchieri, C. (2006). The Grammar of Society: the Nature and Dynamics of Social Norms, Cambridge University Press.

Blount, S. (1995). When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences. In: *Organizational Behavior and Human Decision Process*, 63, 131-144.

Bolton, G. E., Ockenfels A. (2000). ERC: A Theory of Equity, Reciprocity, and Competition. *American Economic Review*, 90(1), 166-193.

Brandts, J., Charness, G. (2011). The Strategy versus the Direct-response Method: A First Survey of Experimental Comparisons. *Experimental Economics*, 14, 375-398.

Cassar, L. (2015). Contract Design with Motivated Agents: Experimental Evidence. Working Paper.

Charness, G. (2004). Attribution and Reciprocity in an Experimental Labor Market. In: *Journal of Labor Economics*, 22, 665–688.

Charness, G., Cobo-Reyes, R., Sanchez, A. (2014). The effect of charitable giving on workers' performance: Experimental evidence. *ThE Papers 14/06, Department of Economic Theory and Economic History of the University of Granada.*

Charness, G., Gneezy, U. (2012). Strong Evidence for Gender Differences in Risk Taking. *Journal of Economic Behavior & Organization*, 83(1), 50-58.

Charness, G., Levine, D.I. (2007). Intention and Stochastic Outcomes: An Experimental Study. *The Economic Journal*, 117, 1051-1072.

Charness,G., Masclet, D., Villeval, M.C. (2014). The Dark Side of Competition for Status. *Management Science*, 60(1), 38-55.

Charness, G., Rabin, M. (2002). Understanding Social Preferences with Simple Tests. *Quarterly Journal of Economics*, 117(3), 817-869.

Cohn, A., Fehr, E., Maréchal, M. (2014). Business culture and dishonesty in the banking industry. *Nature*, 516, 86-89.

Conrads, J., Irlenbusch, B. (2013). Strategic ignorance in ultimatum bargaining. *Journal of Economic Behavior & Organization*, 92, 104-115.

Conrads, J., Lotz, S. (2015). The Effect of Communication Channels on Dishonest Behavior. *Journal of Behavioral and Experimental Economics*, 58, 88-93.

Cooper, D.J., Kagel, J. H. (forthcoming). Other-Regarding Preferences: A Selective Survey of Experimental Results. In: Handbook of Experimental Economics, Vol. 2, Kagel, J.H., Roth, A.E. (eds), Princeton University Press.

Crumpler, H., Grossman, P. (2008). An experimental test of warm glow giving. *Journal of Public Economics*, 92, 1011-1021.

Crawford, V., Sobel, J. (1982). Strategic information transmission. *Econometrica*, 50, 1431–1451.

Croson, R., Gneezy, U., 2009. Gender differences in preferences. *Journal of Economic Literature*, 47, 448-474.

Dana, J., Cain, D.M., Dawes, R.M. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and Human Decision Processes*, 100(2), 193-201.

Dana, J., Weber, R.A., Kuang, J.X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, 33, 67-80.

de Haan, T., Offerman, T., & Sloof, R. (2011). Noisy signaling: theory and experiment. *Games and Economic Behavior*, 73(2), 402-428.

Dreber, A., Johannesson, M. (2008). Gender differences in deception. *Economics Letters*, 99(1), 197-199.

Ekman, P. (1985). Telling Lies. Clues to Deceit in the Marketplace, Politics and Marriage. New York: Norton.

Ekman, P. (1988). Lying and nonverbal behavior: Theoretical issues and new findings. *Journal of Nonverbal Behavior*, 12(3), 163-175.

Elster, J. (1989). Social Norms and Economic Theory. *Journal of Economic Perspectives*, 3(4), 99-117.

Erat, S., Gneezy, U. (2012). White Lies. *Management Science*, 58(4), 723-733.

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

Falk, A., Kosfeld, M. (2006). The Hidden Costs of Control. *American Economic Review*, 96(5), 1611-1630.

Fehr, E., Gächter, S. (2000). Fairness and Retaliation: The Economics of Reciprocity. *Journal of Economic Perspectives*, 14(3), 159–181.

Fehr, E., Schmidt, K.M. (1999). A Theory of Fairness, Competition and Cooperation. *Quarterly Journal of Economics*, 114(3), 817-868.

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

Fischbacher, U., Föllmi-Heusi, F. (2014). Lies in Disguise – An Experimental Study on Cheating. *Journal of the European Economic Association*, 11(3), 525-547.

Fong, C.M., Oberholzer-Gee, F. (2011). Truth in giving: Experimental evidence on the welfare effects of informed giving to the poor. *Journal of Public Economics*, 95, 436-444.

Friedman, M. (1970). The Social Responsibility of Firms Is to Increase Its Profits. *New York Times*, September 13, 32-33.

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

Gneezy, U. (2005). Deception: The Role of Consequences. *American Economic Review*, 95(1), 384-394.

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

Gourieroux, C., Monfort, A. (1989). Statistique et Modèles Économétriques, Vol. 2. Paris: Economica.

Greiner, B. (2004). An online recruitment system for economic experiments. In: *Forschung und wissenschaftliches Rechnen* 2003. GWDG Bericht 63, Kremer, K., Macho, V. (eds), 79-93, Göttingen.

Greiner, B. (2015). Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE. *Journal of the Economic Science Association*, 1(1), 114-125.

Grossman, Z. (2014). Strategic Ignorance and the Robustness of Social Preferences. *Management Science*, 60(11), 2659-2665.

Grossman, Z., van der Weele, J. (2013). Self-image and Strategic Ignorance in Moral Dilemmas. *Working Paper, University of California at Santa Barbara.*

Holmstrom, B. (1979). Moral Hazard and Observability. *Bell Journal of Economics*, 10(1), 74-91.

Houser, D., Vetter, S., Winter, J. (2012). Fairness and cheating. *European Economic Review*, 56(8), 1645-1655.

Imas, A. (2014). Working for the "warm glow": On the benefits and limits of prosocial incentives. In: *Journal of Public Economics*, 114, 14-18.

Jiang, T. (2013). Cheating in the mind games: The subtlety of rules matters. *Journal of Economic Behavior & Organization,* 93, 328-336.

Kaiser, J. (2007). An exact and a Monte Carlo proposal to the Fisher-Pitman permutation tests for paired replicates and for independent samples. *The Stata Journal*, 7(3), 402-412.

Kant, I. (1787). On a Supposed Right to Lie from Benevolent Motives. In Lewis W. Beck, ed., The critique of practical reason and other writings in moral philosophy. Chicago: University of Chicago Press, 1949, 364-350.

Kartik, N. (2009). Strategic Communication with Lying Costs. *Review of Economic Studies,* 76, 1359-1395.

Koppel, H., Regner, T. (2014). Corporate Social Responsibility in the work place- Experimental evidence from a gift-exchange game. In: *Experimental Economics*, 17(3), 347-370.

Larson, T., Capra, M.C., (2009). Exploiting moral wiggle room: Illusory preference for fairness? A coment. *Judgement and Decision Making*, 4(6), 467-474.

Levine, D.I. (1998). Modeling Altruism and Spitefulness in Experiments. *Review of Economic Dynamics*, 1, 593- 622.

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

Margolis, J. D., Elfenbein, H.A., Walsh, J.P. (2007). Does it pay to be good? A meta-analysis and redirection of research on the relationship between corporate social and financial performance. *Working paper, Harvard Business School, Cambridge MA.*

Matthey, A., Regner, T. (2011). Do I Really Want to Know? A Cognitive Dissonance-Based Explanation of Other-Regarding Behavior. *Games*, 2, 114-135.

Mazar, N., Amir, O., Ariely, D. (2008). The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research*, 45(6), 633-644.

McGoey, L. (2012). The logic of strategic ignorance. *The British Journal of Sociology*, 63(3), 553-576.

Muehlheusser, G., Roider, A., Wallmeier, N. (2015). Gender Differences in Honesty: Groups versus Individuals. *Economics Letters*, 128, 25-29.

Offerman, T. (2002). Hurting hurts more than helping helps. *European Economic Review*, 46, 1423-1437.

Pigors, M., Rockenbach, B. (forthcoming). Consumer Social Responsibility. In: *Management Science*.

Rabin, M. (1993). Incorporating Fairness into Game Theory and Economics. *American Economic Review*, 83, 1281-1302.

Roth A.E. (1995). Introduction to Experimental Economics. In: Handbook of Experimental Economics, Kagel J.H., Roth A.E. (eds), Princeton University Press.

Selten, R. (1967). Die Strategiemethode zur Erforschung des eingeschränkt rationale Verhaltens im Rahmes eines Oligopolexperimentes. In: *Beiträge zur experimentellen Wirtschaftsforschung*, Sauermann, H. (ed), Tübingen: J.C.B. Mohr (Paul Siebeck).

Shalvi, S., De Dreu, C. K. W. (2014). Oxytocin promotes group serving dishonesty. *Proceedings of the National Academy of Science*, 111, 5503-5507.

Simon, W. H. (2005). Wrongs of Ignorance and Ambiguity: Lawyer Responsibility for Collective Misconduct. *Yale Journal of Regulation*, 22(1), 1-35.

Stanwick, P.A., Stanwick, S.D. (1998). The Relationship Between Corporate Social Performance and Organizational Size, Financial Performance, and Environmental Performance: An Empirical Examination. In: *Journal of Business Ethics*, 17 (2), 195-204.

Sutter, M. (2009). Deception through telling the truth? Experimental evidence from individuals and teams. *Economic Journal*, 119, 47-60.

Tonin, M., Vlassopoulus, M. (2010). Disentangling the sources of pro-socially motivated effort: A field experiment. In: *Journal of Public Economics*, 94, 1086-1092.

Tonin, M., Vlassopoulus, M. (2015). Corporate Philanthropy and Productivity: Evidence from an Online Real Effort Experiment. In: *Management Science.*

van der Weele , J. (2014). Inconvenient Truths: Determinants of Strategic Ignorance in Moral Dilemmas. *SSRN working paper.*

van der Weele, J.; Kulisa, J.; Kosfeld, M. and Friebel, G. (2014). Resisting Moral Wiggle Room: How Robust Is Reciprocal Behavior? *American Economic Journal: Microeconomics*, 6:3, 256-64.

Vrij, A. (2008). Detecting lies and deceit: Pitfalls and opportunities. Chichester: John Wiley.

Weisel, O., Shalvi, S. (2015). The collaborative roots of corruption. *Proceedings of the National Academy of Science*, 112, 10651-10656.

AGNE KAJACKAITE

CONTACT INFORMATION

University of Cologne
Albertus Magnus Platz
50937 Köln, Germany
Phone: +49 176 4184 7517
E-Mail: kajackaite@wiso.uni-koeln.de
Homepage: https://sites.google.com/site/agnekajackaite/

ACADEMIC POSITION

Since 10/2012	Ph.D. Candidate Department of Management, Economics and Social Sciences, University of Cologne; advisor: Dirk Sliwka Expected Completion Date: 05/2016

REFERENCES

Uri Gneezy
UCSD
ugneezy@ucsd.edu

Axel Ockenfels
University of Cologne
ockenfels@uni-koeln.de

Dirk Sliwka
University of Cologne
sliwka@wiso.uni-koeln.de

RESEARCH INTERESTS

Behavioral Economics, Experimental Economics, Unethical Behavior, Incentives in Organizations

EDUCATION

Since 10/2012	PhD in Economics, University of Cologne
10/2009-06/2012	Master of Science in Economics (with distinction), University of Cologne
09/2005-07/2009	Bachelor of Science in Economics, University of Vilnius

ACADEMIC VISITS

09/2014-03/2015	Rady School of Management, UC San Diego
11/2015	Rady School of Management, UC San Diego

PUBLICATIONS

Kajackaite, A. (2015). If I close my eyes, nobody will get hurt. The effect of ignorance on performance in a real effort experiment. *Journal of Economic Behavior & Organization,* 116, 518-524.

Kajackaite, A., Werner, P. (2015). The incentive effects of performance requirements – A real effort experiment. *Journal of Economic Psychology,* 49, 84-94.

WORKING PAPERS AND WORK IN PROGRESS

Lying Costs and Incentives (with Uri Gneezy), *under review*

Social Responsibility and Incentives in the Lab: Why Do Agents Exert More Effort when Principals Donate? (with Dirk Sliwka), *under review*

Risk-preference elicitation, choice structures, and gender effects (with Gary Charness, Catherine Eckel, and Uri Gneezy)

Delay of decisions (with Steffen Andersen, Uri Gneezy, and Julie Marx)

Can a social CEO be beneficial for the shareholders? (with Dirk Sliwka)

Cost of Lying: Size of Lie (with Uri Gneezy and Joel Sobel)

INVITED TALKS

02/2016	University of Southern California
01/2016	University of Amsterdam
01/2016	University of Essex
11/2015	Caltech
11/2015	Rady School of Management, UC San Diego
10/2015	Maastricht University
04/2015	CREED, University of Amsterdam
12/2014	UC Santa Barbara
11/2014	Rady School of Management, UC San Diego

TALKS AND CONFERENCES

09/2015	ESA European Meetings, Heidelberg
08/2015	PhD Workshop at NHH, Bergen
05/2015	Rady visitors conference on Incentives and Behavior Change, Modica
04/2014	IMEBESS conference in Nuffield College, Oxford
10/2013	AEW Workshop in LUISS, Rome

SUMMER AND SPRING SCHOOLS

03/2015	Spring School in Behavioral Economics, UCSD

TEACHING

Fall 2015: TA Strategic Human Resource Management (Master), University of Cologne

Spring 2015, 2014, and 2013: TA Economics of Incentives in Organizations (Master and PhD), University of Cologne

Fall 2013: TA Seminar on Corporate Development (Bachelor), University of Cologne

Since Summer 2013: Supervisor of various bachelor and master theses, mostly in Personnel Economics and Behavioral Economics, University of Cologne

AWARDS AND SCHOLARSHIPS

02/2016	Behavioral Economics Research Spearhead, University of Amsterdam; Funding for experimental research
11/2015	C-SEB Visitor Program Scholarship
05/2015	Behavioral Economics Research Spearhead, University of Amsterdam; Funding for experimental research
10/2012-09/2015	Full Scholarship for Ph.D. Studies granted by the Federal State NRW
11/2010	Dean's Award for outstanding academic achievement, University of Cologne
10/2009-04/2012	Mummert Scholarship for the Executives of Tomorrow from Central and South Eastern Europe: Full Scholarship for Master Studies (Robert Bosch Foundation)
09/2007-08/2008	DAAD ERP Scholarship for a year abroad

REFEREEING

Management Science, Journal of Economic Behavior & Organization, Journal of Economics & Management Strategy, Social Choice and Welfare, International Review of Economics

OTHER PROFESSIONAL EXPERIENCE

05/2011 – 08/2011	GASCADE Gastransport GmbH, Kassel, Germany Regulation management (Internship)
03/2008 – 04/2008	Deutsche Bundesbank, Munich, Germany Banking Supervision Department (Internship)

OTHER RELEVANT SKILLS

Languages: Lithuanian (native), German (fluent), English (fluent), Italian (basic), Russian (basic)
Computer: Microsoft Office, STATA, SPSS, zTree