Inducing good behavior

van der Veen, A.

Citation for published version (APA):
INDUCING GOOD BEHAVIOR

Ailko van der Veen
INDUCING GOOD BEHAVIOR

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het college voor promoties
ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op dinsdag 28 februari 2012, te 14:00 uur

door

Ailko van der Veen

geboren te Sleen
Promotiecommissie

Promotor: Prof. dr. T.J.S. Offerman

Co-promotor: Dr. A.M. Onderstal
Contents

1. Introduction 1

   2.1. Introduction 7
   2.2. Experimental Design and Procedures 10
   2.3. Results 14
      2.3.1. How to Subsidize Contributions to Public Goods 14
      2.3.2. Control Treatment 19
      2.3.3. Toward an Explanation of the Boiling Frog Effect 22
   2.4. Conclusion 25

3. Inducing Good Behavior: Bonuses versus Fines in Inspection Games 27
   3.1. Introduction 27
   3.2. Inspection Games 29
   3.3. Experimental Design and Procedures 30
   3.4. Results 32
      3.4.1. Inspecting and Shirking Probabilities 32
      3.4.2. Earnings 33
      3.4.3. Explaining Observed Behavior 34
   3.5. Conclusion 37

4. How to Prevent Workers from Shirking: the Use and Effectiveness of Rewards and Punishments in the Inspection Game 41
   4.1. Introduction 41
   4.2. Inspection Game and Theoretical Benchmark 43
   4.3. Experimental Design and Procedures 46
   4.4. Results 48
      4.4.1. Overview 48
      4.4.2. Dynamics and Explanation 53
   4.5. Conclusion 59
5. Keeping out Trojan Horses: Auctions and Bankruptcy in the Laboratory

5.1. Introduction

5.2. Experimental Design and Procedures

5.3. Hypotheses

5.4. Results

5.4.1. Comparisons between Auctions

5.4.2. Individual Behavior

5.5. Explanation of the Main Results

5.5.1. Risk Aversion

5.5.2. Asymmetric Equilibria

5.5.3. Cursed Bidders

5.6. Conclusion

Bibliography

A. Literature on the Boiling Frog Story

B. Instructions "How to Subsidize Contributions to Public Goods"

C. Instructions "Inducing Good Behavior"

D. How to Derive the Equilibrium Predictions of IBE and QRE with Loss Aversion in the Context of the Canonical Inspection Game

E. Instructions "Keeping out Trojan Horses"

F. Proofs of Propositions "Keeping out Trojan Horses"

G. Inleiding
List of Tables

2.1. Main Features of the Treatments .................................................. 11
2.2. Responses to the Subsidy ......................................................... 16
2.3. Estimates of the Main Treatment (hurdle model) ......................... 18
2.4. Estimates of the Dual Task Effect - Control Treatment - (hurdle model) ......................... 22
2.5. Dual Task Procedures and Frequency of Changes ................................. 22
2.6. Beliefs and Contributions in Treatments with Maximum Subsidy 0.75 .................................. 24

3.1. Choice Proportions, Average by Treatment .................................. 32
3.2. Earnings in Part Two, Average by Treatment ................................ 34
3.3. Predicted Choice Probabilities ................................................... 35

4.1. Experimental Design ........................................................................ 48
4.2. Actions in Stage 1 ........................................................................... 51
4.3. Actions in Stage 2 ........................................................................... 52
4.4. Assignment of Tokens .................................................................... 53
4.5. Efficiency and Earnings .................................................................. 54
4.6. Played Combinations and Transitions ............................................. 55
4.8. Employers’ Strategies and Earnings ............................................... 57
4.9. Questionnaire ................................................................................ 58

5.1. Summary of Treatments .................................................................. 64
5.2. Comparisons between Auctions and Liability Regimes .................... 69
5.3. Estimated Bidding Functions [5.12]–[5.14] ....................................... 70
List of Figures

2.1. Development of Subsidy over Time .......................... 13
2.2. Handout for Treatment Pred-75 ............................. 15
2.3. Average Contributions over Time in Main Treatments .......... 17
2.4. Interaction Individual and Group Task ........................ 20
2.5. Controlling for the Dual Task Procedure in Gradual ............ 21

3.1. Inspection Games ............................................. 29
3.2. Parameterization of the Inspection Games Used in the Experiment .. 31
3.3. Proportions of Shirking (left panel) and Inspecting (right panel) across Treatments .. 33
3.4. Changes in Shirking (left) and Inspecting (right) after Introduction of Bonuses and Fines 36

4.1. Inspection Game ............................................ 44
4.2. Inspection Game and the Possibility to Reward and Punish .......... 45
4.3. Equilibria in the Repeated Game (continuation probability 0.8) .......... 47
4.4. Timeseries Inspect and Shirking ................................ 50

5.1. Average Winning Bid and Fraction of Winners Making a Loss ........ 68
5.2. Theoretical and Estimated Bid Function for FP for the Case of Limited Liability ...... 71

D.1. Canonical Inspection Game, Transformed Game and Impulse Matrix .......... 102
Preface

In 2002 I decided to study Economics. Although starting with an interest in macro-economics, I had to do Industrial Organization in the second year, and Jeroen Hinloopen let us students play market games every week and the winner got chocolates. I was completely sold and during the rest of my studies I followed every course in game theory and Industrial Organization I could lay my hands on. When I finished in 2006, Theo Offerman and Sander Ondestal gave me the opportunity to do a PhD at CREED. They taught me the noble art of Experimental Economics. When I applied for the job and told them that I wanted to write a book, they explained that nobody would read a book written by a student. Their advice was to try to write articles for Journals, preferably in the top 5.

I would like to thank all my colleagues at CREED for their inspiration and their help. For Theo and Sander are not the only persons that I have to be grateful to. When I left CREED I was in a better physical shape than before, thanks to the two times a week indoor soccer with some of the most fanatic players I ever met. For a scientific institute there were a lot of sports: Göniil organized a dance party and a frisbee competition, and once a year I would visit Marcelo’s master dance class. The corridor of CREED was an inspiring musical environment, except when Matthijs was in the States. The discussions at CREED did a lot for my mental fitness as well. Aljaz, Martin, Michal, and Klaus knew an amazing lot about experimental economics. The daily lunches with the other PhD students including Roel, Thomas, Adrian, Jona, Pedro, Matthias, and Boris always led to sparkling conversations, in a haze of burned toast. Nadege knew a lot about the brain and Julian made impressive graphs. Joep could tell everything about making wine and liquor and gave me the possibility to teach microeconomics at the Beta Gamma faculty, which was fun to do. Other highlights were the trips to New York organized by Theo. I remember standing in a freezing, pitch black night on Brooklyn Bridge with a full moon over Manhattan, while Ben explained that we were looking at gratte-ciels all around. In a previous year we visited the Bourgeois Pig at Manhattan with Adam and Eve.

During most of the time at Creed, I shared my office (the official CREED library) with Audrey who, when not working at home, was a pleasure to talk to and drink tea with. For the last months Yang took over and I could explain to her the importance of the library and she will share these secrets with Anita. In addition I would like to thank the heads of CREED, first Frans van Winden a very inspiring man and later Arthur Schram, a man with fine attacking and defending skills, both built CREED from the very start.

Next to Sander and Theo, both Jos Theelen an excellent programmer who programmed most
of my experiments and bravely held his PSV ground in an AJAX environment, and Karin Breen who could make sense of the university and banking bureaucracy, were important for my Thesis. Doing research in Nottingham together with Martin Sefton and Daniele Nosenzo, while having one leg on a chair and being pampered by the NHS, also brings back fond memories.

Of course, it takes a lot more people than those at CREED to finish a PhD. Friends to go to the theater with, friends to play volleyball with, friends to go to the movie with, friends to lunch with, friends to talk with and even friends to play still more soccer and chess with. My family and family in law were very supportive, and I hope I was not too much of a social failure. I thank from the bottom of my heart, all those people that helped me. Especially I want to thank Gisela, who not only inspired me to change directions, but was a great help during my studies and my PhD phase in every (im)possible way. Finally, I hope you will enjoy the read as much as I did the research and proof them wrong.
1. Introduction

Using the experimental method, we analyze mechanisms to induce “good” behavior in four cases. Good is seen from the perspective of a superior in a hierarchical relation, where command and control is not an economically feasible option. The four cases are:

1. A government wants to induce good behavior with the help of subsidies. We analyze whether it is more effective to introduce the subsidy gradually or introduce it in one big step.

2. A government wants to induce good behavior with the help of automatic bonuses and automatic fines. We analyze which of the two instruments is the more effective one.

3. An employer wants to induce good behavior from a worker by rewarding desired behavior and punishing undesired behavior. This time the application of instruments is not automatic, but a discretionary power of the employer. Again, we analyze which of the instruments is more effective.

4. A government wants to induce good behavior from limitedly liable bidders in an auction. The government doesn’t want the bidders to overbid, in a situation where post auction bankruptcy is undesirable. We compare the English auction to the first-price sealed-bid auction, with respect to the likelihood of post auction bankruptcy.

We use laboratory experiments, while we also could have used mechanism design to construct a theoretically optimal mechanism. However, most models used in mechanism design, assume agents to be rational, selfish, and making decisions not affected by emotions. Experimental evidence, both from the lab and from the field, shows that said assumptions often do not hold. As we do not have at our disposal a unifying theory of human behavior, we rely on laboratory experiments to study the four cases. In each case, we will confront subjects with two commonly used mechanisms, and study which of the two performs best.

In Chapter 2 we investigate how to introduce subsidies aimed at steering behavior. In 2009, the Japanese government introduced a 10% subsidy on solar power panels. As the subsidy turned out to be less effective than planned, it is expected to be raised in the future. In the same year, the Chinese government announced a 50% subsidy on these panels, the highest

\[1\] For a discussion of mechanism design, see Myerson (1981).

\[2\] For an overview of this problem, see e.g., Froeb (2002).
such subsidy in the world (Ideas 2009). As subsidies are important instruments for governments, we test whether an introduction in one step or a gradual introduction is more effective.

In our experiment we use a public good game, where participants decide every round how much to contribute. The total contribution is raised by some fraction (20% in our case), free of cost for the participants. The pot is equally split and paid out to all participants, independent of their contributions. These rules make contributing 0 the dominant strategy for each participant. We augment this game with a subsidy. The subsidy we use is a reduction in the cost of contributing. If the subsidy is 0.45, contributing 10 to the common pot cost the participant only $(1 - 0.45) \times 10 = 5.5$.

We compare two treatments, the quick treatment and the gradual treatment. In both treatments the subsidy begins at a level of 0 and after a certain amount of rounds the subsidy starts increasing. In the quick treatment the subsidy switches to the target-level in one step. In the gradual treatment the subsidy is slowly raised each round until the target-level is reached. When it is reached, the subsidy stays at the target-level until the end of the experiment.

From experiments without subsidies, we already know that we can expect some participants to contribute. Furthermore, the subsidy makes contributions more effective and we know for example from Isaac and Walker (1988) and Isaac, Walker, and Williams (1994) that participants tend to contribute more when contributions are more effective. In the literature, two explanations are offered. One is the existence of material altruists, who care about the payoff of other people, and give more because their contribution is made more effective (Goeree, Holt, and Laury 2002). The other one is the existence of conditional cooperators in public good games (Offerman, Sonnemans, and Schram 1996; Fischbacher, Gächter, and Fehr 2001; Brandts and Schram 2001). Conditional cooperators choose their contribution conditional on their expectations concerning what others are going to contribute. If contributing is made more effective, they could become more optimistic about others contributing and therefore contribute more themselves.

In contrast to this literature we do not focus on the reason why people react to a subsidy, but focus on how they react to the implementation of the subsidy, either quick or gradual. Interestingly, the idea of conditional cooperators could still play a role. If conditional cooperators expect the other players to react more to an introduction in one step and less to an introduction in small steps, they could also be inclined to react stronger to an introduction in one step. Another option is that it is not so much expectations that drive the results, but anchoring (Tversky and Kahneman 1974). The initial subsidy serves as a reference point: participants only change their behavior if there is a noticeable change in the subsidy.

The experiment shows a difference in the change of contributions between the two treatments, but only if the target-level is high enough. We compare target-levels of 0.45 and 0.75. In treatments where the target-level is 0.45 subjects do not respond differently to a quick or gradual increase of the subsidy: contributions to the public good are hardly raised during the experiment.

To check whether a possible treatment effect could be explained by distraction (as faced in real live), we ran treatments with and treatments without a second task to be performed by the participants simultaneously with the public good task. This addition of an extra task does not produce a difference in contribution.
anyway. When the target-level is 0.75, again subjects hardly respond to a gradual increase of the subsidy, but when the subsidy is introduced in one step they very significantly raise their contribution to the public good. From the experiment we can conclude that to influence behavior, it is better to introduce a substantial subsidy at once than in small steps.

While in Chapter 2 we focus on authorities using subsidies to influence behavior, in Chapter 3 the focus is on authorities using either punishment or reward to encourage good behavior. In 2009, the Dutch tax authorities increased the fine for not reporting savings from 100% to 300% and announced further increases (Tweede Kamer 2009). In 2003, the South Korean tax authorities started rewarding taxpayers having high compliance levels (NTS 2004). Punishment of bad behavior and reward of good behavior are instruments often used by authorities. In an experiment, we test which instrument works better.

We investigate the question with the help of an inspection game, with two players, one called the inspector and the other one called the inspectee. In each round, both inspector and inspectee independently and simultaneously make a decision. The inspector decides whether to do a costly inspection and the inspectee decides whether to work, which is costly for the inspectee. The inspector has to pay the inspectee a wage (higher than the cost of working), except when the inspectee decided not to work and the inspector decided to inspect. When the inspectee works, the payoff of the inspector is enlarged more than an inspection costs.

To the baseline game we add either an automatic fine or an automatic bonus, but only if the inspector chose inspection. Fines are paid by the inspectee and received by the inspector; bonuses are paid by the inspector and received by the inspectee. After each round of the game, players are randomly rematched in new pairs of one inspector and one inspectee, but during the whole experiment a participant only plays one of the two roles.

We observe that the inspectee performs better under automatic fines than under automatic bonuses. This result is in line with predictions of the mixed strategy Nash equilibrium where players make their decisions dependent on the payoffs of the other player. If an inspectee knows that an automatic fine is introduced that adds to the payoff of the inspector, the inspectee will expect the inspector to inspect more often in order to collect the fine. To avoid the fine the inspectee will decide to work more often and this is what we see happen. However, this can not be the whole story. In line with the previous reasoning adding automatic bonuses should lead to less work, and this is not what we observe. There is only an insignificant difference in the decision to work for both treatments. These results can be fairly well explained by recent behavioral models based on respectively impulse balance equilibrium (Selten and Chmura 2008) and quantal response equilibrium (McKelvey and Palfrey 1995). We can conclude that automatic fines work better than automatic bonuses, but in contrast to the standard game theoretical prediction, automatic bonuses are not detrimental to the decision to work.
In Chapter 4 we focus again on punishment and reward, but this time in the context of employers and workers in a fairly standard labor relationship. With this context in mind we changed the set-up of the experiment on various points, although the basis is still the inspection game.

In contrast to the previous experiment, both punishing and rewarding are now at the discretion of the inspector (from now on called employer) and costly for the employer, while just as in the previous experiment, punishing reduces the payoff of the inspectee (from now on called worker) and rewarding increases the payoff of the worker. In each treatment, we use a cost/effect ratio of either 1:1 or a ratio of 1.3. A cost/effect ratio of $1:x$ means that a punishment [reward] that costs the employer 1, costs [contributes to] the worker $x$. Another difference is that employers and workers stay matched in the same pair for all rounds during the experiment. Finally, if the employer decides to inspect, an extra stage is added, in which the employer can choose either to punish the worker, reward the worker, or do nothing.

The literature gives us some indications for what we could expect to happen, but the literature is not conclusive. In the psychological literature, Skinner (1965) concludes from experiments on animals that unlike rewarding, punishing has no lasting effect. Furthermore, psychologists find that supervisors rewarding good behavior perform better in inducing hard work than supervisors punishing bad behavior (Sims, 1980; Podsakoff, Bommer, Podsakoff, and MacKenzie, 2006; George, 1995). However, this research is based on questionnaires which makes identifying cause and effect difficult.

In experimental economics, studies have investigated the strength of negative and positive reciprocity (Abbink, Irlenbusch, and Renner, 2000; Brandts and Sola, 2001; Charness and Rabin, 2002; Offerman, 2002; Brandts and Charness, 2004; Falk, Fehr, and Fischbacher, 2003; Charness, 2004; Al-Ubaydli and Lee, 2009). These studies found no or weak evidence for positive reciprocity, which would undermine the idea that workers would react positively to rewards. Although they found stronger evidence for negative reciprocity, it is difficult to draw conclusions from this evidence. On the one hand workers are perhaps more eager to avoid punishment, but on the other hand we could perhaps expect a negative spiral of punishment, less working, more punishment etc.

In our experiment, we see in general stronger results for treatments with a cost/effect ratio of 1:3 compared to those with a cost/effect ratio of 1:1, and we will further focus on the treatments with a cost/effect ratio of 1:3. We compare single instrument treatments in which employers have just one instrument, either punishment or reward, and the baseline treatment where they have no instrument at all. For the single instrument treatments, we find that the workers work more compared to the baseline treatment. Moreover, it does not matter whether the instrument is punishment or reward. With respect to inspection, we see less costly inspections in the punishment-only treatment compared to the baseline and the reward-only treatments. Therefore, with just one instrument available, the punishment-only treatment increases the payoff of the employer most.
We could expect that the payoff of employers in a treatment with both instruments would be at least as high as it is in a treatment with only punishment: employers could just ignore the reward instrument. This, however, is not the case. In the two instrument treatment, reward is used more often than punishment and in a questionnaire, participants in both roles (employer and worker) state that rewarding good behavior is more appropriate than punishing bad behavior. In the two instruments treatment, workers work as much as in the single instrument treatments, but what makes the punishment-only treatment more profitable for the employer than other treatments, is the fact that the employer needs fewer inspections. We conclude that only adding the possibility to punish to the baseline is most profitable for the employer, but when the possibility to reward is also added, the positive effect seems to decrease.

In Chapter 5 we deal with the question how an auctioneer could prevent winners in an auction from going bankrupt afterwards. The context is one in which winners have to file for bankruptcy, if it turns out that the value of the object is less than the price paid for it. Bankruptcy may be very undesirable in the case of license auctions where a government sells the right to exploit radio frequencies and where bankruptcy of the operator would interrupt communication via those frequencies or decrease competition. Another situation where post auction bankruptcy may be undesirable is when a government selects a (critical) supplier, using a procurement auction.

The problem of post auction bankruptcy is widespread in practice. An extreme example is the 1996 C-Block auction by the Federal Communications Commission in the US: all major bidders (winning bids $10.2 billion in total) went bankrupt (Zheng 2001). Governments have used various methods to overcome the bankruptcy risk. The literature mentions for example: surety bonds, a kind of third party guarantee (Calveras, Ganaza, and Hauk 2004), multi-sourcing, where bidders can only win part of the contract (Engel and Wambach 2006) and finally the average bid auction, where the winner is the one with a bid closest to the average (Decarolis 2010). We analyze whether a simple choice of auction type could mediate the problem. In a laboratory experiment, we compare the English auction and the first-price sealed-bid auction, two auction types that are used frequently to sell licenses and to procure goods and services.

Our experimental design is a straightforward implementation of the problem. Half of the participants take part in English auctions and the other half in first-price sealed-bid auctions. For each auction, the common value of the object is the sum of three numbers (signals) randomly generated. Each auction has three participants and each participant receives one of the signals, but is not informed about the value of the other signals. For each of the treatments, in half of the auctions if participants make a loss, they go bankrupt, and only incur a minimal cost. In the other half of the auctions participants have to cover their full losses.

\[4\] In the English auction, the auctioneer increases a counter indicating the price of an object. Each bidder can step out of the auction by stopping the counter. The other bidders are informed about the price where this bidder steps out, and the counter restarts from that point. The last bidder who remains in the auction wins the object and pays the price where the penultimate bidder steps out.

\[5\] In the first-price sealed-bid auction, all bidders simultaneously submit a bid. The highest bidder wins and pays a price equal to her own bid.
The literature gives us some intuition about what to expect. Klempner (2002) states for example that bidders that can go bankrupt, will bid more aggressive as the downside risk is capped by the bankruptcy option. However, the literature is inconclusive with respect to the question which auction-type will perform better. In case of auctions with a common value studied here, we can expect higher winning bids and therefore more bankruptcy in the English auction than in the first-price auction (Milgrom and Weber 1982). However, in English auctions bidders know when the other bidders step out of the bidding and they could use this information to make a more informed guess about the true value of the object and therefore overcome the bankruptcy risk. We find that when bankruptcy is a possibility, in auctions of both types more bidders make losses than in the unlimited liability case. This increase is not significantly different between both types of auction formats. The result contradicts the predictions of a Nash equilibrium analysis. Eyster and Rabin’s (2005) “cursed equilibrium” model explains our findings quite well. We conclude that a choice of either the English or the First-Price auction does not overcome the bankruptcy problem and that the cursed equilibrium model helps to explain this.
2. How to Subsidize Contributions to Public Goods: Does the Frog Jump out of the Boiling Water?\(^1\)

2.1. Introduction

Governments around the world subsidize contributions to public goods. In some cases, the subsidy is abruptly introduced in one step. For instance, the European Commission abolished in one time the 66.1% import duty on energy saving compact fluorescent lamps from China in October 2008.\(^2\) Similarly, in March 2009, the Chinese government announced the most aggressive subsidy on solar panels in the world. By providing a subsidy of 20 yuan per watt, the Chinese will essentially cover half the cost of entire installations at today’s solar panel prices. In other cases, the subsidy is introduced gradually in many small steps. As an example, in the Netherlands the duty on petrol was enhanced in numerous tiny amounts from 46.1% in 1993 to 69.7% in 2008. By increasing the duty on petrol, the Dutch effectively subsidize people who opt for public transport. In January 2009, Japan launched a rather modest subsidy on solar panels that corresponds to about 10 percent of the costs. The subsidy turned out to be less effective than planned, and it is expected that Japan will raise the subsidy in the future.

In this chapter, we investigate how subsidies of contributions to public goods should be introduced. In a series of experiments, we compare the effectiveness of an instantaneous rise in the subsidy to a slow rise of the subsidy to the same ultimate level. Doing so, we test a conjecture formulated by Al Gore in the 2006 movie *An inconvenient truth*. Gore claims that humans have a tendency to ignore changes in the environment when these changes occur at a very slow pace. Therefore, there is a danger that humans fail to respond while the climate deteriorates by the very gradual process of global warming. Gore draws an analogy between the boiling frog story and the inertia of humans: “If a frog jumps into a pot of boiling water, it jumps right out again,

---

\(^1\)This chapter is based on the identically titled paper joint with Theo Oerman and benefited from helpful comments of Rachel Croson, Tore Ellingsen, Guillaume Frechette, Andreas Leibbrandt, Charlie Plott, Andrew Schotter, Arthur Schram and Joep Sonnemans. We are grateful to CREED programmer Jos Theelen for programming the experiment.

\(^2\)The European Commission decided to impose the duty in 2001 after the European Lighting Companies Federation, a trade group for European producers, complained that China was flooding the market with cheap bulbs. The anti-dumping tariff was a huge setback for Chinese producers, for whom the exports to the European Union formed a substantial share of their market.
because it senses the danger. But the very same frog, if it jumps into a pot of lukewarm water that is slowly brought to a boil, will just sit there and it won’t move.” He concludes: “Our collective nervous system is like that frog’s nervous system. . . . If it seems gradual, . . . we are capable of just sitting there and not reacting.” Gore eloquently formulates a concern that is bothering many people from time to time. For instance, in a recent contribution, Krugman provides the same conjecture about how humans will fail to respond to “the creeping threat” of climate change. Gore and Krugman actually formulate two conjectures, one about frogs and one about humans. Although the boiling frog story is currently challenged, actual investigations on frogs published in the 19th century claim support for it (see Appendix A). The goal of our study is to investigate whether humans fail to react when slow changes in the environment increase the importance of contributions to the public good, as suggested by Gore and Krugman.

In the real world, contributing to a public good is one of many decisions that people continuously make. For instance, when we are cold in winter we may at any moment decide to put on an extra sweater or to set the thermostat a few degrees higher. At the same time, other activities continuously compete for our attention. To mimic this situation in the laboratory, we provide our subjects with a dual-task procedure. Our subjects continuously and simultaneously earn money with an individual task (their daily activities) and with their contributions to a public good. They can switch from the one task to the other task whenever they wish. While they are playing the game, we increase the subsidy to the contributions of the public good. The most important treatment variable is whether this increase occurs instantaneously or gradually.

In our experiments, we make use of a linear public good game where selfish subjects have a dominant strategy to completely free ride in the stage game for any level of the subsidy that we employed. Although the game was repeated for an unknown number of seconds, selfish subjects could not support cooperation in equilibrium because subjects did not receive information about others’ contributions during the public good game. Therefore, from a strategic point of view the game is essentially a one-shot game.

Nevertheless, there is a vast literature on public good games that furnished our conjecture that we would observe positive contributions when contributions were subsidized. One of the stylized facts in experiments on linear public good games is that subjects respond to how productive a contribution to the public good is. Isaac and Walker were among the first ones to find a positive effect of an increase in the Marginal Per Capita Return (MPCR), the marginal benefit that each player earns from the contribution of an extra dollar to the public good, on subjects’ contributions to the public good. In essence, a subsidy on subjects’ contributions to public goods corresponds to an increase in the MPCR. Therefore, it makes sense to expect a positive effect of a subsidy on subjects’ contributions.

3We chose to address the question in a public good game. Another possibility would have been to make use of a strategically equivalent public bad game. Then the question would be how subjects respond to different ways of taxing undesired taking from a common pool. Andreoni started a literature comparing subjects’ behavior in public good and public bad games. In many cases, subjects behave somewhat more cooperatively in the public good frame, but the evidence is not completely concurrent. Duwenberg, Gächter, and Hennig-Schmidt discuss the literature.
There are two possible causes behind subjects’ responsiveness to the MPCR. One possibility is that subjects do not only care about their own payoff but also about the material payoff of other subjects. Material altruists are more inclined to contribute with a higher MPCR because it makes their contribution more effective (Goeree, Holt, and Laury 2002). The other possibility is that a higher MPCR boosts contributions because it changes the beliefs that subjects have about the extent to which others cooperate. The recent literature on public good games has identified the presence of a substantial number of conditional cooperators (Offerman, Sommemaans, and Schram 1996; Fischbacher, Gächter, and Fehr 2001; Brandts and Schram 2001). If a larger MPCR makes the conditional cooperators more optimistic that others will contribute, they will be more inclined to contribute.

In this chapter, the main focus is not on why people respond to the MPCR/subsidy but on whether subjects respond differently when the MPCR/subsidy is changed gradually or instantaneously. The two questions may be related though. If subjects are cool and calculating material altruists, they will solely respond to the level of the subsidy. In this case we would not expect that humans fall prey to the boiling frog phenomenon. Conditional cooperators may believe that others will only fail to respond to a change in the subsidy if it is introduced in tiny steps. With such beliefs, conditional cooperators may only respond to the subsidy when it is introduced in one big step. A boiling frog phenomenon for humans in public good games may thus be driven by conditional cooperators who expect that others are sensitive to the way that the subsidy is introduced.

There is, however, also a possibility that a boiling frog effect in public good games is not driven by expectations but by anchoring (Tversky and Kahneman 1974). The initially chosen contribution level may serve as an anchor that prevents people from adapting their behavior unless a dramatic change in the subsidy occurs. Many studies have shown that people do not move sufficiently in the right direction away from their reference point or anchor. For instance, Northcraft and Neale (1987) find that respondents often quote a too high selling price for a house if they are given a reference point that is higher than the actual selling price and vice versa. Anchoring also explains why people often choose the firm’s default in the 401(k) savings plan (Madrian and Shea 2001). In a recent study, Schram and Sommemaans (2011) investigate how people choose their health insurance in a changing decision environment with a large set of alternatives that differ on a variety of dimensions. In a 2x2x2 design, Schram and Sommemaans vary the number of alternatives, switching costs, and the speed at which health deteriorates. With respect to the latter treatment variable, the authors find that if health deteriorates only gradually, individuals tend to stick to their chosen policy too long.

In a first series of experiments, we raised the subsidy level from 0% to 45%. Here, we do not observe significant differences between the treatment where the subsidy is introduced in one big step and the treatment where it is introduced in many small steps. With a maximum of 45%, the subsidy only marginally increases contributions in either case, though. Therefore, we decided to run an additional series of experiments where we raised the subsidy to 75%. Here,
there is a substantial effect of the subsidy when it is introduced instantaneously while there is at best a modest effect when it is introduced gradually. The difference in the fractions of people responding positively to the subsidy equals 27 percentage points. This difference is significant and persistent. Given that subject respond positively to the subsidy, they enhance their contributions to the same extent in both treatments.

Subjects may fail to respond to a gradual increase in the subsidy because they are distracted by a dual task. We investigated this possibility in a control treatment where subjects were not distracted by the individual task while the subsidy was gradually raised to 75%. If we look at the average contribution levels, subjects respond similarly to the subsidy in the single-task treatment as they do in the dual-task treatment. There is, however, a difference in how often subjects change their decisions. When they are not distracted by the dual task, subjects change their contribution level substantially more often.

An analysis of the beliefs reported by a group of subjects who did not contribute to the public good themselves discredits the explanation that the effect is driven by the beliefs of conditional cooperators. Instead, subjects simply seem to ignore changes in the environment if they are very small in size.

The remainder of this chapter is organized as follows. In Section 2.2 we describe our experimental design. Section 2.3 provides the results and Section 2.4 concludes. Appendix A reviews the existing evidence on the boiling frog story and Appendix B the instructions of the experiment.

### 2.2. Experimental Design and Procedures

The computerized experiment started with on-screen instructions (see Appendix B). After reading the instructions and answering some control questions, subjects received a summary of the instructions on paper. With their decisions, subjects earned points that were exchanged at the end of the experiment at a rate of 1 euro for 1800 points. Table 2.1 on the facing page summarizes the details of the 6 treatments. In total, 259 subjects participated who earned on average 23.1 euros (s.d. 9.4) in about 1 hour and 45 minutes. Each subject participated in one treatment only.

Subjects participated in a public good game that we adapted in different ways. After the public good game was finished, subjects received additional instructions and we obtained measures on their beliefs and social preferences. The dual task procedure formed the core of most of our treatments. We first discuss the main features of this procedure. Subjects performed a group task and an individual task at the same time. Subjects earned money with both tasks and could switch between the two tasks whenever they wanted. Subjects were informed that the earnings

---

4 Subjects who are distracted by a dual task sometimes behave differently. Darley and Batson 1973 find that students who are in a hurry to give a talk on the parable of the Good Samaritan were more likely to pass without stopping to help a shabbily dressed person in need than those who were not in a hurry. Mann and Ward 2004 report that dieters who have to remember a 9-digit number drink more from a high-calorie milkshake than dieters who are told to remember a 1-digit number (see also Ward and Mann 2000).
for the one task were independent of the earnings for the other task. To prevent an artificial endgame effect, we informed subjects that the two tasks would last between 25 and 40 minutes. It actually ended after exactly 28 minutes.

In the individual task, subjects earned money by keeping a randomly moving red dot inside a box. Subjects could move the box by pressing on one of the four buttons (up, down, left, right). At the end of each second the computer determined whether the dot was inside the box or not. The subject earned 15 points when the dot was inside the box and 0 points otherwise. Subjects could keep track of the total earnings for the individual task during the experiment.

For the group task, subjects were randomly assigned to a group of 6 people. They were not rematched during the experiment. In every second, subjects received an endowment of 10 points and determined how much of this endowment to contribute to the public good. Each point contributed to the public good was multiplied by 1.2 and then equally divided between the 6 group-members. So each group-member received 0.2 from each point contributed to the public good. At the start, each subject decided how much to contribute by setting the level of a slider equal to a number in the range from 0 to 10. In every subsequent second each subject had the possibility to change the contribution by moving the slider. If the subject refrained from changing the contribution, this person’s contribution automatically equaled the contribution in the previous period.

Subjects’ contributions were subsidized at a varying rate. If the subsidy equaled $s_t$ ($0 \leq s_t < 0.8$) in second $t$, subject $i$ actually paid a cost of $(1 - s_t) g_{i,t}$ for a contribution $g_{i,t}$ ($0 \leq g_{i,t} \leq 10$). Thus, in second $t$ subject $i$ earned the amount:

$$\pi_{i,t} (g_{i,t}) = 10 - (1 - s_i) g_{i,t} + 0.2 \sum_{j=1}^{6} g_{j,t}$$

Subjects knew that the subsidy would start at 0 and that it might change during the experiment but that it would never exceed 0.8, so making a donation would never become a dominant strategy. Above the slider, subjects observed the subsidy of that second. When the subsidy

---

**Table 2.1.: Main Features of the Treatments**

<table>
<thead>
<tr>
<th>Treatment</th>
<th>max subsidy</th>
<th>increase subsidy</th>
<th>dual task?</th>
<th>group-size</th>
<th>#subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>gradual-45</td>
<td>0.45</td>
<td>gradual</td>
<td>yes</td>
<td>6</td>
<td>48</td>
</tr>
<tr>
<td>quick-45</td>
<td>0.45</td>
<td>quick (start)</td>
<td>yes</td>
<td>6</td>
<td>54</td>
</tr>
<tr>
<td>gradual-75</td>
<td>0.45</td>
<td>gradual</td>
<td>yes</td>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td>quick-75</td>
<td>0.75</td>
<td>quick (start)</td>
<td>yes</td>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td>gradual-75-single</td>
<td>0.75</td>
<td>gradual</td>
<td>no</td>
<td>6</td>
<td>36</td>
</tr>
<tr>
<td>predict-75</td>
<td>predicted contribution levels gradual-75 and quick-75</td>
<td>yes</td>
<td>6</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>total</td>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td>259</td>
</tr>
</tbody>
</table>

---

5The box game was developed by John Krantz, see [http://psych.hanover.edu/JavaTest/CLE/Cognition/Cognition/dualtask_instructions.html](http://psych.hanover.edu/JavaTest/CLE/Cognition/Cognition/dualtask_instructions.html)
changed, the background of the subsidy-number turned red for a second. This way, subjects noted the change even when they were focused on the individual task. All subjects faced the same subsidy and they were explicitly informed that the change of the subsidy was outside of their control.

Subjects were NOT informed about the contributions made by the other group members while they participated in the public good game. They were also not informed about their earnings for the group-task until the end. This feature of the design was motivated by the observation that most consumers in the real world receive little or no information about other people's private energy consumption. A convenient consequence of this feature is that contribution decisions are independent across subjects.

We now turn to the differences between the treatments. The main treatments, gradual-45, quick-45, gradual-75 and quick-75, allow us to determine which way of changing the subsidy is most effective. In all these treatments, the subsidy remained at 0 during the first 4 minutes. Then in the quick-treatments, the subsidy jumped in one second from 0 to the maximum of 0.45 in quick-45 and from 0 to the maximum of 0.75 in quick-75. In the gradual treatments, the subsidy was raised with 0.001 per 2.2 seconds until it reached 0.45 in gradual-45, while it was raised with 0.001 per 1.3 seconds until it reached 0.75 in gradual-75, so that in either case the maximum was attained after 20 minutes and 40 seconds. In the remainder the subsidy stayed at the maximum until the end. Figure 2.1 on the next page displays the development of the subsidy across treatments.

To investigate the potential effect of the dual task procedure, we included treatment gradual-75-single where subjects only performed a single task. Like in the main treatments, subjects earned money from the group task and the individual task. However, subjects only had to decide themselves how much to contribute to the public good in the individual task, as the computer replicated for them the movements of the red dot as presented to and the choices made by one of the subjects in the individual task in a previous dual task treatment. Subjects could observe the choices that were made for the individual task by their counterpart in a previous experiment, but they could only affect their own earnings by their contribution decisions. This way, subjects could concentrate on the contribution task while their income was enhanced at the same pace as in the dual task experiment. A comparison of gradual-75-single and gradual-75 reveals the effect of the dual-task procedure.

After the public good game was finished, we obtained some measures that shed light upon the contribution decisions. We obtained a measure on subjects’ social preferences by eliciting their value orientations. Here, subjects received two amounts, a first one determined by their own choice and a second one determined by another subject’s choice. Subjects chose to allocate $I$ points to one self and $O$ points to a randomly chosen other person subject to the constraint $I^2 + O^2 = 4000^2$. In the experiment, subject used the mouse to select a point on a circle where
Finally, we collected some background information about our subjects.

In addition, we elicited the beliefs that subjects had about the contribution levels at the start and at the end in other quick and gradual groups. As expected, we found a positive correlation between beliefs about others’ contributions and own contributions. These data do not yet allow us to assess the role of conditional cooperators, because it is not clear whether the causal relation runs from beliefs to behavior or in the opposite direction.

To unravel the potential role of beliefs in the boiling frog phenomenon, we ran an additional treatment pred-75 where subjects neither played the public good game nor the box game. Instead, their task was to predict how much subjects had contributed in gradual-75 and quick-75 at specific moments. Subjects first received the instructions provided to the subjects in quick-75 and gradual-75 and then they received a handout that explained the development of the subsidy across time in the gradual mode and in the quick mode (see Figure 2.2 on page 15). We elicited subjects’ subjective beliefs about how much subjects contributed on average in previous sessions.

This circle test has been used for the first time by Sonnemans, Dijk, and Winden (2006).
We did this for the following three statements that refer to particular moments shown in the handout:

1. Your probability judgment for the statement: at the START, the average contribution was in the interval [0..2]; [2..4]; [4..6]; [6..8]; [8..10].

2. Your probability judgment for the statement: at the END, the average contribution in the GRADUAL groups (see hand-out) was in the interval [0..2]; [2..4]; [4..6]; [6..8]; [8..10].

3. Your probability judgment for the statement: at the END, the average contribution in the QUICK groups (see hand-out) was in the interval [0..2]; [2..4]; [4..6]; [6..8]; [8..10].

After providing the 5 probabilities connected to one statement, subjects were provided with a graphical presentation of the implied probability density, and they were allowed to make changes to their reported probabilities before they proceeded to the next statement. For half the subjects questions 2 and 3 were posed in the opposite order. Subjects were rewarded for reporting their beliefs seriously. In total, subjects reported 15 probabilities (3 statements × 5 intervals). At the end of the experiment, one of these 15 probabilities was drawn at random and every subject received a payment generated by the quadratic scoring rule. To correct the reported beliefs for risk attitudes, we employed the correction procedure described in Offerman, Sonnemans, van de Kuilen, and Wakker (2009). Basically, that procedure filters out the risk component in subjects’ reported beliefs. This is done by asking subjects to make probability judgments for an additional series of questions with given objective probabilities. These judgments are then used to map the originally reported probabilities into risk-corrected probabilities.

2.3. Results

We present the results in three parts. In Section 2.3.1 we look at how responsive our subjects are to the subsidy and we investigate whether subjects react stronger when the subsidy is quickly increased than when it is gradually enhanced. There we deal with our main treatments gradual-45, quick-45, gradual-75 and quick-75. In Section 2.3.2 we discuss the results of the control treatment that allows us to investigate whether the results are sensitive to the introduction of the dual task. In Section 2.3.3 we provide the evidence obtained in treatment pred-75 and we unravel the role that beliefs play in explaining the boiling frog phenomenon.

2.3.1. How to Subsidize Contributions to Public Goods

We chose to start with a low MPCR of 0.2 to allow for a positive effect of a subsidy on the contributions in all experiments. In gradual-45 and quickly-45, we increased the subsidy to a maximum of 0.45. This corresponds to an almost doubling of the MPCR from 0.2 to 0.2 / (1 - 0.45) = 0.364.

In their treatments with an MPCR of 0.3, Isaac and Walker (1988) and Isaac, Walker, and
Williams (1994) find a contribution level of roughly 35%-40% when their data of group sizes 4, 10 and 40 are pooled.

Table 2.2 on the next page shows how subjects responded to the increase of the subsidy. For each subject, we calculated the average contribution in the 50 seconds prior to the start of the rise of the subsidy and the average contribution in the 50 seconds after the subsidy reached its maximal level in a treatment. The columns “Pre” and “Post” report these statistics averaged across subjects. In the treatments with a maximum subsidy of 0.45, we observe a modest increase in the contribution level which reaches a weakly significant level in quick-45 but not in gradual-

Figure 2.2.: Handout for Treatment Pred-75

![Diagram of subsidy levels and time periods for Quick and Gradual conditions]
Table 2.2: Responses to the Subsidy

<table>
<thead>
<tr>
<th>Max Subsidy</th>
<th>Gradual</th>
<th>Quick</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre</td>
<td>Post</td>
</tr>
<tr>
<td>0.45</td>
<td>N: 48</td>
<td>(SD: 2.16)</td>
</tr>
<tr>
<td></td>
<td>(2.94)</td>
<td>(2.85)</td>
</tr>
<tr>
<td>0.75</td>
<td>N: 36</td>
<td>(SD: 1.85)</td>
</tr>
<tr>
<td></td>
<td>(2.86)</td>
<td>(3.75)</td>
</tr>
</tbody>
</table>

Notes: Table is based on data from gradual-45, quick-45, gradual-75, and quick-75; Pre [Post] gives the average contribution in the 50 seconds before the start [after the end] of the rise in subsidy; WMP: Wilcoxon Matched-Pairs Signed-Ranks Test; standard deviations between brackets.

45 Because our subjects responded less to the subsidy than we had expected we decided to run treatments where the subsidy increased to a maximum of 0.75. The table shows that in gradual-75 the increase in contributions is again modest and only weakly significant (at best). The increase in contributions in quick-75 is substantial and significant though.

Figure 2.3 on the facing page displays the average contributions across time in the four main treatments. The figure shows that there is a substantial and lasting effect of the subsidy in quick-75. In the other treatments there is only a modest effect of the introduction of the subsidy. A first glance at the data suggests that the boiling frog phenomenon only appears when the subsidy is increased instantaneously to a sufficiently high level.

To make the first impression from the figure statistically precise and to control for subjects’ background, we ran a regression that employed a “hurdle specification” (Papke and Wooldridge, 1996; McDowell, 2003). In our data, a fraction of the subjects responds positively to the subsidy. Given that subjects react to the subsidy, they do so at different absolute levels. A natural interpretation of such data is that the subjects first decide whether or not to respond to the subsidy. Only in case that they do respond to the subsidy, they decide on how much to increase their contribution. So the second decision is only made if the hurdle of the first decision is passed. Hurdle models are common in medical applications, where the factors that affect a patient’s decision to see a doctor may be different from the factors that affect the doctor’s and patient’s decision on how much to spend on medical care. As far as we know, Botelho, Harrison, Pinto, and Rutström (2009) were the first ones to apply hurdle models to public goods.

In all our treatments, subjects experienced the absence of the subsidy until the 240th second and the maximal subsidy after the 1240th second. Thus, all treatments are comparable before the 240th second and after the 1240th second. For each subject, we constructed 8 “periods” of 50 seconds after the 1240th second. For each of these 8 periods we computed the average contribution level, and from these levels we subtracted the subject’s average contribution level in 7 After running the treatments with a maximum subsidy of 0.45, we discovered that the modest response of our subjects to the subsidy is actually in line with the responses of subjects in Goeree, Holt, and Laury (2002) who also report substantial contributions for higher MPCR levels only.

8 In the paper of Botelho, Harrison, Pinto, and Rutström (2009), the factors that affect a subject’s decision to contribute or not are viewed as separate one the ones that affect a subject’s decision how much to contribute.
Figure 2.3.: Average Contributions over Time in Main Treatments

Notes: for each second, the average of contributions in the interval \([\text{second } - 25, \text{second } + 25]\) is displayed.

the 50 seconds just prior to the 240th second. This way we use normalized contributions that are corrected for individual differences in initial contributions. Because our data form a panel we use a clustering specification that takes into account the dependence of the data within subjects and the independence of the data across subjects. We estimate the fraction that positively responds to the subsidy separately from the increase in the contribution conditional on a positive response on the subsidy. [McDowell (2003)] shows that this approach provides the same consistent and efficient estimates as the procedure where the overall hurdle model is estimated in one time.

As explanatory variables we include dummies for the treatments that reveal the treatment effects relative to the omitted treatment gradual-45 as well as dummies for some background variables and dummies for the periods. Table 2.3 on the next page reports the results. The first column presents the estimates of the marginal effects of the explanatory variables on the probability that the subjects respond positively to the subsidy as calculated in a probit-regression. The second column reports the estimates of the marginal effects of the variables on the increase in contribution conditional on a positive response to the subsidy as calculated in an OLS-regression. The third column displays the estimates of the total marginal effects of the variables on the (un-
Table 2.3: Estimates of the Main Treatment (hurdle model)

<table>
<thead>
<tr>
<th>X</th>
<th>marginal effect (s.e.)</th>
<th>p</th>
<th>marginal effect (s.e.)</th>
<th>p</th>
<th>marginal effect (s.e.)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>quick-45</td>
<td>0.03 (0.09)</td>
<td>0.73</td>
<td>0.23 (0.53)</td>
<td>0.67</td>
<td>0.03 (0.40)</td>
<td>0.95</td>
</tr>
<tr>
<td>gradual-75</td>
<td>0.01 (0.10)</td>
<td>0.94</td>
<td>2.33 (0.76)</td>
<td>0.00</td>
<td>0.40 (0.65)</td>
<td>0.54</td>
</tr>
<tr>
<td>quick-75</td>
<td>0.28 (0.10)</td>
<td>0.01</td>
<td>2.67 (0.60)</td>
<td>0.00</td>
<td>2.43 (0.66)</td>
<td>0.00</td>
</tr>
<tr>
<td>Female</td>
<td>0.05 (0.07)</td>
<td>0.48</td>
<td>-1.65 (0.55)</td>
<td>0.00</td>
<td>-0.29 (0.44)</td>
<td>0.51</td>
</tr>
<tr>
<td>Cooperator</td>
<td>0.21 (0.08)</td>
<td>0.61</td>
<td>1.14 (0.50)</td>
<td>0.02</td>
<td>1.28 (0.54)</td>
<td>0.02</td>
</tr>
<tr>
<td>Economics</td>
<td>-0.07 (0.07)</td>
<td>0.31</td>
<td>0.23 (0.45)</td>
<td>0.61</td>
<td>-0.02 (0.47)</td>
<td>0.97</td>
</tr>
<tr>
<td>period-2</td>
<td>-0.02 (0.02)</td>
<td>0.36</td>
<td>0.32 (0.22)</td>
<td>0.15</td>
<td>0.05 (0.00)</td>
<td>0.62</td>
</tr>
<tr>
<td>period-3</td>
<td>0.01 (0.02)</td>
<td>0.57</td>
<td>0.32 (0.24)</td>
<td>0.18</td>
<td>0.18 (0.12)</td>
<td>0.11</td>
</tr>
<tr>
<td>period-4</td>
<td>-0.01 (0.02)</td>
<td>0.59</td>
<td>0.17 (0.24)</td>
<td>0.05</td>
<td>0.11 (0.13)</td>
<td>0.36</td>
</tr>
<tr>
<td>period-5</td>
<td>-0.02 (0.03)</td>
<td>0.34</td>
<td>0.24 (0.27)</td>
<td>0.36</td>
<td>-0.03 (0.14)</td>
<td>0.81</td>
</tr>
<tr>
<td>period-6</td>
<td>-0.05 (0.03)</td>
<td>0.09</td>
<td>0.11 (0.29)</td>
<td>0.17</td>
<td>-0.10 (0.15)</td>
<td>0.48</td>
</tr>
<tr>
<td>period-7</td>
<td>-0.04 (0.03)</td>
<td>0.21</td>
<td>0.39 (0.26)</td>
<td>0.14</td>
<td>-0.07 (0.14)</td>
<td>0.62</td>
</tr>
<tr>
<td>period-8</td>
<td>-0.04 (0.03)</td>
<td>0.15</td>
<td>0.34 (0.28)</td>
<td>0.24</td>
<td>-0.11 (0.15)</td>
<td>0.44</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Wald-tests</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>quick-45</td>
<td>0.03 (0.09)</td>
<td>0.73</td>
<td>0.23 (0.53)</td>
<td>0.67</td>
<td>0.03 (0.40)</td>
<td>0.95</td>
</tr>
<tr>
<td>quick-75–grad-75</td>
<td>0.27 (0.11)</td>
<td>0.02</td>
<td>0.34 (0.83)</td>
<td>0.68</td>
<td>2.03 (0.83)</td>
<td>0.01</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.07</td>
<td></td>
<td>0.26</td>
<td></td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>1392</td>
<td>524</td>
<td>1392</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: period-2-8 indicates second - eighth period blocks of 50 seconds after the 1240th second; for each subject, the average contribution in the 50 seconds before the subsidy starts changing is subtracted from each average contribution level in periods-2-8; regression based on gradual-45, gradual-75, quick-45 and quick-75; Column \( \Pr\{Y > 0\} \) shows the fraction of observations passing the hurdle \( Y > 0 \); \( Y|\{Y > 0\} \) displays the marginal effects given that the hurdle is passed; Column \( Y \) reports the total marginal effect; the omitted treatment is gradual-45; Female = 1 if subject is female, female = 0 if subject is male; 61% of the subjects were male; Cooperator = 1 if subject is altruistic or cooperative, coop = 0 if subject is individualistic or competitive; Economics = 1 if subject studies economics, econ = 0 if subject studies something else or does not study; the \( R^2 \) for the column \( \Pr\{Y > 0\} \) is a Pseudo \( R^2 \).

conditional) increase in contributions in an OLS-regression. For the behavioral reasons discussed above, we think that the results in the third column are based on an “incorrect” specification. We include them because they provide a summary of the overall marginal effect of the variables on the increase in contribution.

The treatment effects are listed in the bottom rows of the table (below “Wald tests”). The results are in line with the pattern emerging from the figures. The effect of the subsidy is in the expected direction for quick-45 and gradual-45, but rather small and far from significant. There, a quick increase in the subsidy neither affects the probability of reacting to the subsidy nor the level of the increase given that subjects reacted to the subsidy. The result is very different for the comparison of quick-75 and gradual-75. With a maximum subsidy of 0.75, the contributions are more than doubled when the subsidy is introduced instantaneously while there is only a modest effect when it changes gradually. The difference between the treatments is substantial and significant. We find that the fraction of subjects who respond positively to the increase in the subsidy is significantly larger in quick-75 than in gradual-75. The difference is 27 percentage
points. Interestingly, given that subjects do respond positively on an increase in the subsidy, there is no difference in how much they increase their contribution. Thus, the treatment effect is completely due to the enhanced probability of responding to the subsidy in quick-75.

The estimation results control for period and background effects. Females are as likely as men to react to the subsidy, but their conditional increase in contribution is smaller. Subjects who are identified as cooperator by the independent measurement of their value orientation are more likely to respond to the subsidy than those identified as individualists, and given that they do respond, they increase their contribution to a larger extent. The reported results are robust to excluding subjects' value orientation. When we run the regression without the dummy for cooperator, we get approximately the same results.

Economics students react slightly less to the subsidy but given that they do, they increase their contributions by a slightly larger amount. In total, the effect is small and not significant. The estimates of the coefficients for the period dummies are small and insignificant, in accordance with the fact that contribution levels were roughly stable after the 1240th second.

One possibility is that the difference in behavior between quick-75 and gradual-75 is completely determined by the switching costs between the two tasks. Switching costs between the two tasks may limit the number of times that subjects change the contribution level in the public good game. As a result, subjects may be further away from their subjectively optimal contribution level in gradual-75 where many changes are needed to accommodate the slowly changing subsidy. Figure 2.4 on the following page displays the decrease in hits around the time that a subject changed the contribution. In a time window of 20 seconds, subjects lose on average 36 points or 2 euro cents. Thus, the material switching costs seem to be rather limited. Still, subjects may behave differently when they are not distracted by the dual task. This is the topic of the next section.

2.3.2. Control Treatment

In this section, we deal with the sensitivity of the results with respect to the dual task procedure. This procedure may prevent subjects in gradual-75 to choose the subjectively optimal contribution level that they would have chosen when only faced with the public good task. To investigate this possibility, we ran treatment gradual-75-single, where subjects could concentrate on the public good task while they automatically received the same earnings for the individual task as one of the subjects in the dual-task treatments. Figure 2.5 on page 21 shows the average contribution levels over time in gradual-75-single together with the contributions in gradual-75. In gradual-75-single average contributions are slightly higher than in gradual-75 throughout the experiment. This is not surprising given that initial contributions are accidentally slightly higher (in the first 50 seconds, the difference in contribution levels is not significant, Mann-Whitney test, \( p = 0.28 \)). More importantly, the pattern in how people change their contributions when the subsidy is introduced is remarkably similar. In both treatments, the subsidy has only a modest effect on the long run contribution levels.
Figure 2.4: Interaction Individual and Group Task

Notes: this graph indicates the average number of hits in the individual task for each second in the period of 60 seconds before and 60 seconds after a second in which the slider indicating the contribution in the groups task moved; movements of the slider in successive seconds are taken as one; the graph is based on gradual-75 and quick-75.

We assessed the statistical importance of the dual task procedure in a hurdle regression similar to the one reported in Table 2.3 on page 18. In Table 2.4 on page 22, the dummy for treatment gradual-75 measures the treatment effect of the dual task compared to the omitted treatment gradual-75-single. There is neither a significant difference in the probability that subjects respond to the subsidy nor a significant difference in the extent to which subjects increase their
Figure 2.5.: Controlling for the Dual Task Procedure in Gradual

![Graph showing contribution levels over time for different conditions.](image)

Notes: for each second, the average of contributions in the interval \([\text{second} - 25, \text{second} + 25]\) is displayed.

contribution given that they do.\(^9\) Again, the regression results appear to be robust to excluding the dummy variable that independently measures whether a subject is cooperative.

The average contribution levels in Figure 2.5 mask some interesting patterns at the micro-level. Table 2.5 on the next page shows some statistics on the fractions of people that change their contribution at least once during the experiment and on how often these people change their decisions. In the single task experiment, the fraction of people changing their decisions exceeds the one in the dual-task experiment. The most remarkable difference is in how often subjects change their decisions (given that they do this at least once).

In the world outside the laboratory people are involved in multiple tasks all the time. The results of our experiment suggest that people change their decisions much more often when they face a single task. The reassuring news for previous experiments on public good games is that

---

\(^9\) In addition to the control treatment reported in this chapter, we ran a control to investigate whether the results in quick-45 are affected by the timing of the subsidy. We included treatment quick-45-end that was the same as quick-45, except that the change in subsidy occurred after 20 minutes and 40 seconds instead of after 4 minutes. We did not find any difference in how subjects responded to the subsidy in quick-45 and quick-45-end. We also ran controls for the dual task procedure in quick-45 and gradual-45, and also here we did not identify an effect of the dual task on subjects’ responses to the subsidy.
Table 2.4.: Estimates of the Dual Task Effect - Control Treatment - (hurdle model)

\[ Y = \text{change in contribution} \]

<table>
<thead>
<tr>
<th>X</th>
<th>marginal effect (s.e.)</th>
<th>p</th>
<th>marginal effect (s.e.)</th>
<th>p</th>
<th>marginal effect (s.e.)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>gradual-75\text{Single}</td>
<td>-0.01 (0.10)</td>
<td>0.93</td>
<td>0.04 (1.12)</td>
<td>0.97</td>
<td>0.24 (0.82)</td>
<td>0.77</td>
</tr>
<tr>
<td>Female</td>
<td>0.17 (0.10)</td>
<td>0.09</td>
<td>-1.26 (1.02)</td>
<td>0.22</td>
<td>0.62 (0.83)</td>
<td>0.46</td>
</tr>
<tr>
<td>Cooperate</td>
<td>0.34 (0.11)</td>
<td>0.00</td>
<td>0.00 (1.08)</td>
<td>1.00</td>
<td>1.59 (0.93)</td>
<td>0.09</td>
</tr>
<tr>
<td>Economics</td>
<td>0.13 (0.10)</td>
<td>0.20</td>
<td>1.58 (0.79)</td>
<td>0.05</td>
<td>0.62 (0.84)</td>
<td>0.46</td>
</tr>
<tr>
<td>period-2</td>
<td>-0.01 (0.03)</td>
<td>0.56</td>
<td>0.44 (0.38)</td>
<td>0.25</td>
<td>0.10 (0.16)</td>
<td>0.56</td>
</tr>
<tr>
<td>period-3</td>
<td>-0.03 (0.04)</td>
<td>0.40</td>
<td>0.74 (0.62)</td>
<td>0.23</td>
<td>0.16 (0.20)</td>
<td>0.50</td>
</tr>
<tr>
<td>period-4</td>
<td>-0.06 (0.04)</td>
<td>0.18</td>
<td>0.61 (0.66)</td>
<td>0.35</td>
<td>-0.01 (0.34)</td>
<td>0.97</td>
</tr>
<tr>
<td>period-5</td>
<td>-0.06 (0.05)</td>
<td>0.22</td>
<td>0.30 (0.67)</td>
<td>0.56</td>
<td>-0.20 (0.33)</td>
<td>0.54</td>
</tr>
<tr>
<td>period-6</td>
<td>-0.06 (0.06)</td>
<td>0.06</td>
<td>0.42 (0.74)</td>
<td>0.58</td>
<td>-0.35 (0.35)</td>
<td>0.31</td>
</tr>
<tr>
<td>period-7</td>
<td>-0.12 (0.05)</td>
<td>0.02</td>
<td>0.00 (0.69)</td>
<td>0.39</td>
<td>-0.45 (0.34)</td>
<td>0.18</td>
</tr>
<tr>
<td>period-8</td>
<td>-0.10 (0.05)</td>
<td>0.05</td>
<td>0.15 (0.63)</td>
<td>0.81</td>
<td>-0.55 (0.33)</td>
<td>0.33</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.12</td>
<td></td>
<td>0.14</td>
<td></td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>( N )</td>
<td>576</td>
<td></td>
<td>200</td>
<td></td>
<td>576</td>
<td></td>
</tr>
</tbody>
</table>

Notes: period-2-8 indicates second to eighth period blocks of 50 seconds after the 1240th second; for each subject, the average contribution in the 50 seconds before the subsidy starts changing is subtracted from each average contribution level in periods-2-8; regression based on gradual-75\text{Single} and gradual-75\text{Dual}; Column \( Pr(Y > 0) \) shows the fraction of observations passing the hurdle \( Y > 0 \); \( Y| (Y > 0) \) displays the marginal effects given that the hurdle is passed; Column \( Y \) reports the total marginal effect; the omitted treatment is gradual-75\text{Single}; Female = 1 if subject is female, female = 0 if subject is male; Cooperator = 1 if subject is altruistic or cooperative, coop = 0 if subject is individualistic or competitive; Economics = 1 if subject studies economics, econ = 0 if subject studies something else or does not study; the \( R^2 \) for the column \( Pr(Y > 0) \) is a Pseudo \( R^2 \).

average contribution levels do not seem to be affected by artificially limiting people to a single task.

2.3.3. Toward an Explanation of the Boiling Frog Effect

In the introduction we offered two possible explanations of a boiling frog effect in public good games. One possibility is that some subjects are conditional cooperators who want to match the expected contribution provided by the others. If conditional cooperators expect that others will not respond to a gradual increase but will react to an instantaneous increase in the subsidy, they will match their expectations and a boiling frog effect is born. The other possibility is that

Table 2.5.: Dual Task Procedures and Frequency of Changes

<table>
<thead>
<tr>
<th></th>
<th>Single</th>
<th>Dual</th>
<th>Single vs Dual</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N )</td>
<td>36</td>
<td>36</td>
<td>72</td>
</tr>
<tr>
<td>fraction subjects changing</td>
<td>0.89</td>
<td>0.61</td>
<td>( \chi^2; p = 0.01 )</td>
</tr>
<tr>
<td>numbers of changes per subject</td>
<td>72</td>
<td>11</td>
<td>MW; ( p = 0.00 )</td>
</tr>
</tbody>
</table>

Notes: a "subject" is recorded to be changing when there is at least one second, not being the first second, in which the contribution is different from that in a previous second; number of changes per subject is calculated on the basis of the persons who change; table based on gradual-75\text{Single} and gradual-75\text{Dual}; \( \chi^2 \) provides the result of a Chi-Square Test for \( r \times c \) Tables and MW presents the result of a Mann-Whitney rank test.
subjects start with a subjectively optimal initial contribution level when the subsidy is 0. When the subsidy is introduced, they only change their previously optimal decision if the change in subsidy in two subsequent seconds is sufficiently large. Such a myopic decision-making process may be the driving force behind a boiling frog phenomenon in public good games. Notice that the two explanations differ in the role assigned to subjects’ beliefs.

In the treatments where the subsidy was raised to a level of 0.45, we asked subjects to report their beliefs about how much other subjects contributed at particular moments in the experiment (before we communicated the results of the actual contribution levels). Like Croson (2007) and Dufwenberg, Gächter, and Hennig-Schmidt (2008), we find a positive relationship between beliefs about other’s contributions and own contributions. The Spearman-rank correlation between subject’s beliefs and the own behavior is substantial (0.31 at the start, 0.45 at the end in quick and 0.44 at the end in gradual) and significant ($p = 0.00$ in all three cases). This evidence is consistent with the explanation based on conditional cooperators. The evidence is far from conclusive, though, because the direction of the causality between beliefs and behavior remains unclear. We cannot exclude that subjects behave as they do because they myopically fail to respond to small changes in the environment, and, when asked about their beliefs of others’ contributions, simply project their own behavior on others.

To shed light upon the causality between beliefs and contributions, we ran treatment pred-75 where subjects played the role of predictor only. In pred-75, subjects were provided with the instructions received by subjects in quick-75 and gradual-75. In addition, these subjects were informed about the development of the subsidy in quick-75 and gradual-75. As shown in Figure 2.2 on page 15 they were then asked to predict the average contribution level in quick-75 and gradual-75 for three occasions: (i) at the 240th second, just before the subsidy started rising in either treatment; (ii) at the 1240th second in gradual-75, just after the subsidy stopped rising in gradual-75 (iii) at the 1240th second in quick-75. Notice that the predictors’ beliefs are not biased by their choices, because predictors never decided how much to contribute.

Table 2.6 on the next page presents the beliefs of the predictors together with the choices of the subjects in quick-75 and gradual-75. The upper-panel of the table shows that the predictors expect a substantial and significant effect of the subsidy in gradual-75 as well as in quick-75. This is only partly in agreement with the data, because the subsidy had a substantial and significant effect on contribution level in quick-75, but not in gradual-75. The lower-panel of the table presents statistics about how much the beliefs and the contribution levels changed as a result of the subsidy. Predictors expect a slightly larger effect of the subsidy in quick-75 than in gradual-

---

In this analysis, we excluded subjects when in the correction procedure the correlation between the subjects’ reported beliefs for the objective probabilities and the objective probabilities was lower than 0.35, when they had reported a probability of 50% for each of the 15 beliefs question or when they reported 50% for at least 9 of the 10 lottery questions.

A comparison of subjects’ beliefs and actual behavior of the other subjects reveals that subjects were on average too optimistic about the contributions of the others. The same bias in beliefs is reported in Offerman, Sönnerman, and Schram (1996) and Palley and Rossenthal (1991).

The procedure to investigate the causal direction between beliefs and contributions was developed by Dawes, McFavish, and Shaklee (1977).
The difference is weakly significant at $p = 0.07$. So predictors expect a weak boiling frog effect but the actual data reveal a strong effect. Predictors are better able to predict the effect of the subsidy in quick-75 than in gradual-75. In quick-75, predictors anticipate on average a smaller effect of the subsidy than actually exists, but the difference is far from significant. In gradual-75, predictors overestimate the effect of the subsidy substantially and significantly.

| Table 2.6.: Beliefs and Contributions in Treatments with Maximum Subsidy 0.75 |
|---------------------------------|-----------------|------------------|-------------------|-----------------|-----------------|-----------------|
|                                | start (I)       | quick (II)       | gradual (III)     | I vs II          | I vs III        | N               |
| beliefs                        |                 |                  |                   |                 |                 |                 |
| predictors                     | 3.76 (1.56)     | 5.85 (1.40)      | 3.53 (1.53)       | WC: $p = 0.00$  | WC: $p = 0.00$  | 42              |
| contributions                  |                 |                  |                   |                 |                 |                 |
| gradual-75                     | 1.85 (2.86)     | –                | 2.58 (3.75)       | WC: $p = 0.18$  |                 | 36              |
| beliefs                        | 1.46 (2.01)     | 4.32 (3.98)      | –                | WC: $p = 0.00$  | WC: $p = 0.00$  | 36              |
| Increase quick ($\Delta Q$)   |                 |                  |                   |                 |                 |                 |
| beliefs                        | 2.09 (1.87)     | 1.77 (1.93)      | –                | MW: $p = 0.07$  |                 |                 |
| contributions                  |                 |                  |                   |                 |                 |                 |
| playes (C)                     | 2.86 (3.56)     | 0.74 (3.25)      | –                | MW: $p = 0.03$  |                 |                 |
| B vs C                         | MW: $p = 0.66$  | MW: $p = 0.01$   |                  |                 |                 |                 |

Notes: Table is based on subjects in treatments quick-75, gradual-75 and pred-75; standard errors in parentheses; 7 from 49 subjects in pred-75 were excluded because of the criterion mentioned in footnote [10 on the preceding page]. Columns I, II and III report the expectation of the reported probability distributions (for details, see the end of Section 2.2); WC provides the result of a Wilcoxon rank test and MW presents the result of a Mann-Whitney rank test.

The evidence makes it less likely that the explanation based on conditional cooperators drives the boiling frog result. Subjects whose beliefs are not biased by their choices expect a substantial effect of the subsidy in gradual-75. If the explanation of conditional cooperators would drive the boiling frog phenomenon, we should have observed a substantial effect of the subsidy on contributions in gradual-75, which we did not. The results do not discredit the explanation based on anchoring. When subjects are actually absorbed in the game, they fail to respond to minor changes in the environment. Predictors who look at this process from a distance fail to appreciate this effect, and instead tend to think that people will respond in the same rational way as when the subsidy is introduced instantaneously.\textsuperscript{13}

\textsuperscript{13}This result is in line with some recent findings on the distinction between decision utility and experienced utility and findings on focusing illusion that are summarized by Kahneman and Thaler (2006). When making a decision, people often fail to accurately predict the utility that they will experience, or they mispredict how they will respond to changes in the environment. For instance, respondents think that people living in California are happier than people living in areas with a lesser climate such as the East or the Midwest, while this is actually not true (Schkade and Kahneman, 1998). Current assistant professors tend to overpredict the life satisfaction of obtaining a tenured position compared to being denied one (Gilbert, Pinel, Wilson, Blumberg, and Wheatley, 1998).
2.4. Conclusion

In this chapter, we investigated how humans react to an instantaneous versus a very gradual introduction of a subsidy to contribute to a public good. When the subsidy was raised to an intermediate level, we did not find support for the boiling frog story. This is not surprising, however, because even when the subsidy was introduced instantaneously, the effect of the subsidy on the contribution level was modest at best. When the subsidy was raised to a substantial level, a clear boiling frog effect emerged. Subjects hardly responded to the subsidy when it was introduced gradually while they reacted strongly when it was introduced in one shot. In particular, by introducing the subsidy in one time the fraction of subjects responding to the subsidy increased by 27%. Given that subjects did respond to the subsidy, there was no difference in the extent to which they increased their contribution between the two ways of introducing the subsidy.

Subjects who did not play the public good game but who were asked to report their beliefs about what contributors would do, predicted the effect of the subsidy more or less correctly when it was introduced at once. In contrast to what would be expected if the phenomenon were mediated by the beliefs of conditional cooperators, predictors failed to predict that the subsidy would not have an effect on the contributions when the subsidy was introduced gradually. The evidence does not discredit the explanation that the boiling frog phenomenon is caused by anchoring. In accordance with Al Gore’s and Paul Krugman’s conjecture, people simply fail to respond to tiny changes in the environment.
3. Inducing Good Behavior: Bonuses versus Fines in Inspection Games

3.1. Introduction

There are many situations where authorities have preferences over individuals’ choices. A tax authority wants taxpayers to truthfully report income, an employer wants an employee to work hard, a regulator wants a factory to comply with pollution regulations, police want motorists to observe speed limits, etc. A fundamental problem for authorities is how to induce compliance with desired behavior when individuals have incentives to deviate from such behavior. A standard approach is to monitor a proportion of individuals and penalize those caught misbehaving. To further encourage compliance, the authority may consider rewarding an individual who was inspected and found complying. For example, in 2003 the National Tax Service (NTS) of Korea introduced a system of bonuses for taxpayers found to have high compliance levels: bonuses included benefits such as providing a three-year exemption from tax audit and preferential treatment from financial institutions, e.g. reduced interest rates on loans (NTS, 2004, p. 31). Alternatively, the authority may consider increasing the sanctions on individuals who, upon inspection, are found not complying. For example, the Dutch government decided to increase the fine for undeclared savings from 100% to 300% in May 2009 (Tweede Kamer, 2009). In this chapter we study which of these two mechanisms is most successful in promoting good behavior.

The essence of such situations is captured by the ‘inspection game’, which we describe in Section 3.2. In this game an authority chooses to inspect or not, and an individual chooses to comply or not, and the unique Nash equilibrium is in mixed strategies, with positive probabilities of inspection and non-compliance. Perhaps unsurprisingly, fines for non-compliant behavior increase the equilibrium probability of compliance. On the other hand, and perhaps paradoxically, bonuses for compliant behavior reduce the equilibrium probability of compliant behavior. Thus, according to standard game theoretical reasoning, fines, and not bonuses, should be used to encourage compliance in such settings. Previous experiments have revealed limited success of the Nash equilibrium for predicting behavior in games where the equilibrium is in mixed strategies (Ochs, 1995; Potters and Winden, 1996; Goeree and Holt, 2001; Goeree, Holt, and Palfrey, 2003). One

\[1\] This chapter is based on the identically titled paper joint with Daniele Nosenzo, Theo Offerman, and Martin Sefton and benefited from helpful comments of Daniel Seidmann, participants at the 2010 ESA Conference in Copenhagen, the 2010 CREED-OxDEEx-CBESS Meeting in Amsterdam, and seminar audiences in Amsterdam. We are grateful to CREED programmer Jos Theelen for programming the experiment.
of the reasons why the Nash equilibrium does not provide an accurate description of behavior in these types of games is that it fails to capture 'own-payoff effects': players do change their behavior in response to changes in their own payoff, whereas the mixed strategy Nash equilibrium predicts that they will not. In the case of the inspection game, the own-payoff effect of introducing fines reinforces the theoretically expected effect: fines make non-compliance less attractive to the individual, and so the own-payoff effect points toward more compliance. However, the own-payoff effect of introducing bonuses for compliant behavior reduces the probability of non-compliance. Thus, Nash equilibrium and own-payoff effects point in different directions in this case, and so it is unclear whether the theoretical prediction that fines outperform bonuses in encouraging compliance will be supported in practice. We describe our experiment for comparing the effectiveness of bonuses and fines in Section 3.3. Our inspection game is framed as an employer-worker scenario where an employer can either inspect or not and a worker can either supply high or low effort. We designed three experimental treatments, each consisting of two parts. The first part was identical across treatments: subjects played a control version of the inspection game where the employer pays the worker a flat wage, unless she is inspected and found supplying low effort in which case the wage is not paid. In the second part of the BONUS treatment, subjects played a version of the game where the employer paid an additional bonus to the worker when the employer inspected and the worker supplied high effort. In the second part of the FINE treatment, subjects played a version of the game where the worker paid a fine to the employer if the employer inspected and the worker supplied low effort. Finally, in the second part of the CONTROL treatment, subjects continued playing the same game as in the first part. This design allows us to examine whether bonuses or fines are more effective in encouraging working/discouraging shirking. In addition, we are able to compare the efficiency properties of rewarding versus punishing mechanisms. We report our results in Section 3.4. We find that fines are more effective than bonuses in encouraging working and in raising combined earnings. This is in line with standard game theoretic predictions. However, the prediction that bonuses discourage working receives little support: although subjects shirk slightly more in the BONUS treatment than CONTROL the difference is small and not statistically significant. Moreover, the prediction that introducing bonuses will reduce combined earnings is not supported: the losses to employers are almost exactly offset by gains to workers. In general, standard comparative static predictions work well when own-payoff effects point in the same direction, but not otherwise. We show that observed deviations from Nash equilibrium predictions can be explained quite well by behavioral theories that incorporate loss aversion and can accommodate own payoff effects: Impulse Balance Equilibrium (Selten and Chmura 2008) and an augmented version of Quantal Response Equilibrium (McKelvey and Palfrey 1995). In Section 3.5 we discuss these results in relation to the existing literature and conclude.
### 3.2. Inspection Games

We study a simple simultaneous move inspection game. An employer can either inspect (I) or not inspect (N), and a worker can supply either high (H) or low (L) effort. The employer incurs a cost of $h$ from inspecting, and high effort results in the worker incurring a cost of $c$ and the employer receiving revenue of $v$. The employer pays the worker a wage of $w$, unless the worker supplies low effort and the employer inspects. The resulting payoffs are shown in the leftmost panel of Figure 3.1. We assume that all variables are positive and $v > c$, $w > h$, $w > c$. Note that joint payoffs are maximized when the worker supplies high effort and the employer does not inspect. Following Fudenberg and Tirole (1992, p. 17), we refer to this as the canonical version of the game. For a review of the theory of inspection games see Avenhaus, Von Stengel, and Zamir (2002).

The canonical game has a unique Nash equilibrium where the employer inspects with probability $p_c = c/w$ and the worker chooses low effort ("shirks") with probability $q_c = h/w$. In this equilibrium the employer’s expected payoff is $\pi_c^{employer} = v - w - hv/w$, the worker’s expected payoff is $\pi_c^{worker} = w - c$, and joint expected payoffs are $\pi_c = v - c - hv/w$. We now compare two possibilities for encouraging high effort relative to the canonical version of the game: imposing an additional fine on workers caught supplying low effort, versus paying a bonus to workers who are inspected and found supplying high effort. Suppose an additional fine $f$ is imposed on a worker caught shirking, resulting in the payoff matrix shown in the middle panel of Figure 3.1. Note that the fine is a transfer between the worker and the employer. Now the unique Nash equilibrium has the employer inspect with probability $p_f = c/(w + f)$ and the worker shirk with probability $q_f = h/(w + f)$. Thus, according to Nash equilibrium, fines discourage both inspections and shirking. In Nash equilibrium expected payoffs are $\pi_f^{employer} = v - w - hv/(w + f)$, and $\pi_f^{worker} = w - c$, and so the employer benefits from the introduction of fines, while the worker’s expected payoff is independent of fines. According to Nash equilibrium, fines enhance efficiency because joint expected payoffs are reduced by low effort and/or inspection, and both of these are discouraged by a fine on workers caught shirking. Next, we examine the case where the employer pays a bonus $b$ to a worker who is inspected and found to have chosen high effort. The payoff matrix for this game is shown in the rightmost panel of Figure 3.1. Now in equilibrium the employer inspects...
with probability \( p_b = c/(w + b) \) and the worker shirks with probability \( q_b = (h + b)/(w + b) \). According to Nash equilibrium bonuses reduce the probability of inspection and increase the probability of shirking. The workers equilibrium expected payoff is \( \pi^\text{worker}_b = w - c + cb/(w + b) \), increasing in \( b \), while the employer's is \( \pi^\text{employer}_b = v - w - v(h + b)/(w + b) \), decreasing in \( b \). Overall, bonuses reduce joint expected payoffs because the beneficial effect of less frequent inspection is outweighed by the detrimental effect of increased shirking. As is well known, comparative static predictions based on mixed strategy Nash equilibrium can often be counter-intuitive. This is because a player's equilibrium probability must keep her opponent indifferent among actions, and so a player's own decision probabilities are determined by the opponent payoffs and not by own payoffs. Consider, for example, how the introduction of a bonus affects own-payoffs from the perspective of the worker. Introducing the bonus has no effect on the expected payoff from shirking, but increases the expected payoff from choosing high effort (for a given inspection probability). Based on this own-payoff effect, one might expect the worker to shirk less frequently following the introduction of bonuses. However, the Nash equilibrium prediction goes in the opposite direction: bonuses lead to an increase in the equilibrium shirking probability. Previous experimental work (e.g., Ochs 1995; Goeree and Holt 2001; Goeree, Holt, and Palfrey 2003) shows that counterintuitive Nash equilibrium predictions are often rejected by the data: changing a player's own payoff does have an impact on that player's decision probabilities. Goeree and Holt (2001) observe own-payoff effects in one-shot games; Ochs (1995) and Goeree, Holt, and Palfrey (2003) observe own-payoff effects even after players have had ample opportunities to learn. Note that own-payoff effects may either reinforce or counteract equilibrium forces. Introducing fines into the inspection game generates an own-payoff effect that pulls workers' behavior in the same direction as Nash equilibrium predictions: introducing fines does not change the expected payoff from choosing high effort but does reduce the expected payoff from shirking. Thus the own-payoff effect discourages shirking, and this is consistent with the Nash equilibrium comparative static prediction. Similarly, own-payoff effects reinforce Nash equilibrium predictions about inspection probabilities in the inspection game with bonuses, but counteract Nash equilibrium predictions in inspection games with fines. In summary, given the evidence on the importance of own-payoff effects in previous experiments, it is not clear that experimental evidence will support the standard game theoretical analysis outlined above. In particular, the own-payoff effects arising when bonuses are paid to workers who are inspected and found supplying high effort may make them a more effective tool for encouraging effort than suggested by standard theory.

3.3. Experimental Design and Procedures

The experiment consisted of fifteen sessions at the University of Nottingham. Ten subjects participated in each session. Subjects were recruited from a campus-wide distribution list and
Figure 3.2.: Parameterization of the Inspection Games Used in the Experiment

<table>
<thead>
<tr>
<th>Canonical Game</th>
<th>Game with Fines</th>
<th>Game with Bonuses</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>H</strong></td>
<td><strong>L</strong></td>
<td><strong>I</strong></td>
</tr>
<tr>
<td>52</td>
<td>12</td>
<td>60</td>
</tr>
<tr>
<td>25</td>
<td>40</td>
<td>25</td>
</tr>
</tbody>
</table>

Notes: Employer is the ROW player, Worker is the COLUMN player. Within each cell, the Employer’s payoff is shown at the top and the Worker’s payoff at the bottom.

No subject participated in more than one session. No communication between subjects was permitted throughout a session. At the beginning of a session subjects were randomly assigned to computer terminals and were informed that the experimental session would consist of two parts, during each of which they could earn ‘points’. Subjects were also told that their cash earnings for the session would be based on all points accumulated in both parts of the experiment. Instructions for Part One were then distributed and read aloud. At the end of these subjects had to answer a series of questions to test their comprehension of the instructions. A monitor checked the answers and dealt with any questions in private. We did not continue with the experiment until all subjects had correctly answered all the questions. Part One then consisted of 40 rounds. At the beginning of the first round subjects learned their role: five subjects were assigned the role of ‘Employer’ and five the role of ‘Worker’. Subjects kept these roles for the entire session (i.e. for both Part One and Part Two). Across rounds subjects were randomly matched in pairs consisting of one Employer and one Worker, and in each round each pair played the canonical inspection game shown in the leftmost panel of Figure 3.2. At the end of each round subjects were informed of their own and their opponents’ choices and point earnings. Subjects were also shown their accumulated point earnings and a table with the distribution of choices across all subjects in the session for the previous twenty rounds.

At the end of Part One subjects were given instructions for Part Two, which were then read aloud. These explained that the second part consisted of another 80 rounds, again with pairings randomly determined at the beginning of each round. In our five CONTROL sessions these rounds used the same earnings table as in Part One. In our five FINE sessions the earnings table was as in Part One except that the worker would pay a fine of 20 points to the employer if the worker chose low effort and the employer chose to inspect. Thus in Part Two of the

---

2Subjects were recruited through the online recruitment system ORSEE [Greiner, 2004]. Instructions are available in Appendix C.

3Point earnings were derived from the game described in the previous section (see Figure 1) with $v = 60$, $c = 15$, $h = 8$, $w = 20$, and with 20 points added to all outcomes to ensure that subjects could not make losses in any of the games used in the experiment. These parameters were chosen so that Nash equilibrium probabilities are not too close to 0, 0.5 or 1 (all probabilities lie in the intervals [0.2, 0.4] or [0.6, 0.8]). We also sought separation between games with and without bonuses or fines so that, where a change in behavior is predicted by standard theory, the predicted change in probabilities across games is at least 20 percentage points.
Table 3.1.: Choice Proportions, Average by Treatment

<table>
<thead>
<tr>
<th></th>
<th>Part One</th>
<th></th>
<th>Part Two</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CONTROL</td>
<td>FINE</td>
<td>BONUS</td>
<td>CONTROL</td>
</tr>
<tr>
<td>Proportion of Shirking</td>
<td>0.29</td>
<td>0.52</td>
<td>0.45</td>
<td>0.44</td>
</tr>
<tr>
<td>Nash</td>
<td>0.40</td>
<td>0.40</td>
<td>0.40</td>
<td>0.40</td>
</tr>
<tr>
<td>Proportion of Inspecting</td>
<td>0.80</td>
<td>0.77</td>
<td>0.78</td>
<td>0.81</td>
</tr>
<tr>
<td>Nash</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Notes: table shows the proportion of shirking/inspecting decisions in the last 20 rounds of each Part of the experiment.

experiment subjects in the FINE sessions played the inspection game shown in the middle panel of Figure 3.2 on the preceding page. In our five BONUS sessions the earnings table was as in Part One except that the employer would pay a bonus of 20 points to the worker if the worker chose high effort and the employer chose to inspect (rightmost panel of Figure 3.2). At the end of Part Two subjects were paid in cash according to their accumulated point earnings from all rounds using an exchange rate of £0.004 per point. Sessions took about 40 minutes on average and earnings ranged between £10.2 and £23.1, averaging £14.9 (approximately US$24 at the time of the experiment).

3.4. Results

3.4.1. Inspecting and Shirking Probabilities

Figure 3.3 on the next page displays the smoothed proportions of inspecting and shirking decisions across all the rounds of the experiment. For some cases there is a clear change in behavior in round 41, following the transition from Part One to Part Two and the introduction of fines or bonuses, but otherwise the observed proportions appear quite stable across rounds. Table 3.1 reports the proportions of shirking and inspecting over the last 20 rounds of each Part of the experiment. The Nash equilibrium predictions for choice probabilities are also reported for comparison. The first 40 rounds of the experiment (Part One) are common to the three treatments, and we do not find any significant differences in the proportions of shirking or inspecting across treatments (Kruskal-Wallis test p-values are 0.37 for shirk and 0.78 for inspect). Averaged across all sessions the observed proportion of shirking decisions is 45% and the observed proportion of inspecting decisions is 78%; both statistics compare favorably with predictions made by Nash equilibrium (40% and 75%, respectively).  

4Our non-parametric analysis is based on two-tailed tests applied to 5 independent observations per treatment. We consider data from each session as one independent observation. Tests are applied to averages based on the last 20 rounds of each Part of the experiment. The data analysis does not lead to different results if we focus on all rounds.

5Treating data from each session as an independent observation and using a one-sample sign test, we cannot reject the hypothesis that in Part One the proportions of shirking and inspecting across the 15 sessions are equal to Nash equilibrium predictions (p = 1.00 for shirking and p = 0.18 for inspecting).
In Part Two of the experiment the proportions of shirking and inspecting diverge significantly across treatments (Kruskal-Wallis test: \( p = 0.02 \) for shirk, and \( p = 0.01 \) for inspect). Clearly, the changes in payoff matrices introduced in Part Two of the different treatments caused subjects to adjust their behavior. For pair-wise statistical comparisons between treatments we use Mann-Whitney rank-sum tests. As predicted, we find less shirking in FINE (23%) than in CONTROL (44%), and the difference is statistically significant \( (p = 0.02) \). Although Nash equilibrium predicts workers will shirk considerably more in BONUS than in CONTROL (70% vs. 40%), shirking in BONUS is only slightly higher than in CONTROL (50% vs. 44%), and the difference is not statistically significant \( (p = 0.55) \). As for inspection probabilities, these are significantly lower in FINE than CONTROL \( (p = 0.01) \) and BONUS than CONTROL \( (p = 0.01) \). We also note, however, that the inspection probability in FINE is considerably higher than predicted (62% vs. 37.5%), while the proportion of inspections in BONUS is closer to the theoretical level (45% vs. 37.5%). In fact, whereas Nash equilibrium predicts that introducing bonuses and fines have the same effect on inspection probabilities, we find a statistically significant difference in the proportions of inspections between FINE and BONUS \( (p = 0.01) \).

### 3.4.2. Earnings

Table 3.2 reports average earnings per game across treatments in the last 20 rounds of Part Two of the experiment. Nash equilibrium predictions are also reported for comparison.

In principle, joint earnings can range from 32 points (when the employer inspects and the worker shirks) to 85 (when the employer does not inspect and the worker works). Theory predicts

---

6According to one-sample sign tests, the proportion of shirking is significantly different from the equilibrium prediction in Part Two of BONUS \( (p = 0.06) \), but not in FINE \( (p = 0.37) \) or CONTROL \( (p = 1.00) \). The proportion of inspecting in Part Two of the experiment differs significantly from the Nash prediction in FINE and BONUS \( (p = 0.06 \text{ both cases}) \), but not in CONTROL \( (p = 0.37) \). These p-values are each based on five independent sessions so insignificant results should be treated with caution.
that joint earnings are equal to 61 points in the game used in CONTROL. In the experiment, earnings in our CONTROL sessions are close to this, averaging 58.7 points across the last 20 rounds of Part Two. Theory also predicts that fines are beneficial and bonuses are detrimental for efficiency. Using Mann-Whitney rank-sum tests, we find that, consistent with these predictions, joint earnings in FINE are higher than in CONTROL, and the difference in the distributions is statistically significant ($p = 0.01$). On the contrary, we find no evidence that bonuses hamper efficiency: in fact, introducing bonuses slightly increases on average joint earnings relative to CONTROL, although the effect is not statistically significant ($p = 0.85$). A second aspect of our data is worth discussing: while according to Nash equilibrium the introduction of fines is Pareto improving, as it is predicted to leave the workers’ earnings unchanged relative to CONTROL and to increase the employer’s payoff, we find that fines are in fact detrimental for workers. In FINE, workers earn about 1.5 points per game less than in CONTROL ($p = 0.06$). Fines are instead beneficial for the employer as predicted ($p = 0.01$). Thus, the introduction of fines has distributive consequences that are not fully accounted for by standard theory: employers are better off when fines are introduced, but this occurs at the expenses of workers who are worse off relative to CONTROL, although the latter effect is small in magnitude and only weakly statistically significant. The introduction of bonuses has instead the predicted distributive consequences: it significantly increases the worker’s payoff and decreases the employer’s payoff ($p = 0.01$ and $p = 0.02$ respectively).

### 3.4.3. Explaining Observed Behavior

Whereas Nash equilibrium predictions seem to capture well the comparative static effects of fines on shirking behavior and bonuses on inspecting behavior, they do not capture observed effects of fines on inspections or bonuses on effort. It is notable that the instances where Nash predictions fail are those where own-payoff effects, as discussed in Section 3.2 on page 29, work in the opposite direction to equilibrium effects. Table 3.3 on the facing page contains predicted choice probabilities made by two alternative concepts: Quantal Response Equilibrium (QRE)

<table>
<thead>
<tr>
<th></th>
<th>CONTROL</th>
<th>FINE</th>
<th>BONUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joint Earnings</td>
<td>61.0</td>
<td>73.0</td>
<td>50.5</td>
</tr>
<tr>
<td>Nash</td>
<td>24.2 (1.08)</td>
<td>22.5 (1.38)</td>
<td>32.7 (1.01)</td>
</tr>
<tr>
<td>Worker Earnings</td>
<td>34.5 (5.11)</td>
<td>47.1 (1.35)</td>
<td>26.1 (2.30)</td>
</tr>
<tr>
<td>Nash</td>
<td>36.0</td>
<td>48.0</td>
<td>18.0</td>
</tr>
</tbody>
</table>
| Notes: table shows average point earnings per game (last 20 rounds only). Standard deviations based on session averages in parentheses.
Table 3.3: Predicted Choice Probabilities

<table>
<thead>
<tr>
<th></th>
<th>Probability of Shirking</th>
<th>Probability of Inspecting</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CONTROL</td>
<td>FINE</td>
</tr>
<tr>
<td>Results</td>
<td>0.44</td>
<td>0.23</td>
</tr>
<tr>
<td>Nash</td>
<td>0.40</td>
<td>0.20</td>
</tr>
<tr>
<td>QRE ($\lambda = 0.989$)</td>
<td>0.46</td>
<td>0.19</td>
</tr>
<tr>
<td>IBE</td>
<td>0.41</td>
<td>0.16</td>
</tr>
<tr>
<td>Nash with loss-aversion</td>
<td>0.25</td>
<td>0.11</td>
</tr>
<tr>
<td>QRE with loss-aversion ($\lambda = 0.289$)</td>
<td>0.42</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Notes: Results shows the proportion of shirking/inspecting decisions in the last 20 rounds of the second part; the other rows give the predictions according to the different equilibrium concepts.

and Impulse Balance Equilibrium (IBE). The predictions are for our Part Two data. In QRE players' choices are stochastic. Better responses (i.e., yielding a higher expected payoff) are predicted to be played more frequently than worse responses, but not with 100% certainty. The degree of precision $\lambda$ with which players choose their responses determines the extent to which QRE predictions deviate from Nash equilibrium predictions. When $\lambda = 0$ players choose actions equi-probably and in the limit as $\lambda$ approaches $\infty$ players always choose their best-response. Part One data is used to estimate the QRE precision parameter $\lambda$ in our experimental setting. For the estimated value of $\lambda$ QRE predictions are generally close to Nash equilibrium predictions.

IBE is based on the idea that players look at forgone payoffs when they adjust their decision probabilities: choosing an option that yields a lower payoff than the alternative option generates an 'impulse' in the direction of the non-chosen option. Impulses generated by forgone payoffs that represent a 'loss' relative to a player's security payoff level (her pure strategy maximin value) weigh twice as much as forgone 'gains'. In equilibrium, players choose the decision probabilities such that the impulses of forgone payoffs are equal across options. IBE predictions differ markedly from Nash equilibrium when own payoff and Nash equilibrium effects are in conflict: the IBE predicted probability of shirking in BONUS is 43% (versus the 70% Nash prediction) and the predicted probability of inspecting in FINE is 61% (versus 37.5%). The fact that Nash equilibrium and QRE are not augmented by loss-aversion while IBE is has generated a recent debate about whether the incorporation of loss-aversion is what drives the observed differences in performance across these equilibrium concepts (see Selten and Chmura 2008; Brunner, Camerer, and Goeree 2011; Selten, Chmura, and Goerg 2011). To examine this possibility, Table 3.3 also reports predictions made by Nash equilibrium and QRE when these concepts are augmented with loss-aversion. Incorporating loss-aversion into the concepts generally improves the performance.

7 Appendix D contains details on the procedures used to derive the equilibrium predictions for IBE and QRE.
8 As in Selten and Chmura 2008 and Brunner, Camerer, and Goeree 2011, we calculate the best fitting overall estimate for $\lambda$ in our data by minimizing the sum of mean squared distances of the predicted QRE probabilities from the observed session-averaged choice probabilities in the experiment. This yields an estimated $\lambda$ of 0.989.
9 As in Selten and Chmura 2008 we incorporate loss-aversion by transforming payoffs above the security level as follows. If $x$ is the payoff and $m$ is the security level, any payoff $x > m$ is transformed into $x' = m + (x - m)/2$. This estimated value of $\lambda$ was obtained using data from Part One as this allows us to make out-of-sample predictions for behavior in the games used in Part Two of the experiment.
Figure 3.4.: Changes in Shirk (left) and Inspect (right) after Introduction of Bonuses and Fines.

*Notes: in each round, the average is displayed of the proportions of (max) 5 previous rounds, the current round and (max) 5 future rounds.*

of QRE, but not the performance of Nash equilibrium. Overall, the comparative static effects observed in our experiment are generally better captured by IBE and QRE with loss-aversion than by Nash equilibrium analysis or by QRE without loss-aversion. This is summarized in Figure 3.4. The Figure shows how the introduction of bonuses and fines affect the probability of shirking and inspecting relative to CONTROL according to the three solution concepts, as well as in the data for the last 20 rounds of Part Two.

When Nash equilibrium effects and own-payoff effects work in the same direction (i.e. for the impact of fines on shirking and the impact of bonuses on inspections) there is little to choose among the various solution concepts. When Nash equilibrium effects and own-payoff effects work in opposite directions (i.e. for the impact of fines on inspecting and the impact of bonuses on shirking), Nash equilibrium (with or without loss-aversion) is outperformed by the alternative concepts. Among these, IBE and QRE augmented by loss-aversion perform better than QRE without loss-aversion. Nash equilibrium predicts that bonuses increase shirking by 30% relative to CONTROL, whereas shirking only increases by about 6% in our data. This observed effect compares quite favorably with the comparative static predictions made by IBE (a predicted 2% increase in shirking) and QRE augmented by loss-aversion (a predicted 4% increase), but not with the comparative static predictions made by QRE without loss-aversion (a predicted 22% increase). Similarly, Nash equilibrium predicts that fines reduce inspection rate by about 37% relative to CONTROL, whereas inspection rates actually fall by about 19%. QRE without loss-aversion predicts a decrease in inspecting by 35%, whereas the predicted magnitude of the decrease is smaller in IBE and QRE with loss-aversion (about 20% or less).

---

The exact procedure is discussed in Appendix D.
3.5. Conclusion

We compare the effectiveness of bonuses and fines as instruments for encouraging compliance in inspection games. In our setting the incentive for a worker to work is given by the monitoring activity of an employer and the costs/benefits incurred by the worker when she is inspected and found to have worked or shirked. The unique Nash equilibrium of the game is in mixed strategies with positive probabilities of inspection and shirking. We find that bonuses targeted at those inspected and found working are not effective in encouraging working: in fact, subjects in our experiment shirk slightly more often when bonuses are present, although the effect is not statistically significant. On the other hand, we find that introducing harsher fines for shirkers is an effective tool for encouraging working. The question of whether rewards or punishments are a better tool for inducing socially desirable behavior has been addressed in previous experimental work. Most of the literature has used two-stage games where in the second stage, after having observed choices made in the first stage, players can incur costs to punish or reward other players. Players are not predicted to use costly rewards or punishments if they are solely concerned about own earnings, but they might if they have preferences for reciprocity. In fact, a large experimental literature documents the willingness of some people to eschew private interests and react positively toward those that treat them well (positive reciprocity) or negatively toward those that treat them poorly (negative reciprocity). In particular, early studies of games that allow for both positive and negative reciprocity found that the latter has a particularly strong impact (Abbink, Irlenbusch, and Renner 2000; Offerman 2002; Charness and Rabin 2002). These findings are echoed in Andreoni, Harbaugh, and Vesterlund (2003) who investigate the effects of rewards and punishments in a proposer-responder game where the proposer chooses an amount to transfer to the responder and the responder can then either punish or reward the proposer. They find that proposers’ transfers are particularly sensitive to the threat of punishment, although rewards have also positive effects. Similarly, Sefton, Shupp, and Walker (2007) examine the effect of rewards and punishments on contributions in a repeated public good game and find that punishments help subjects to sustain higher cooperation levels compared to a control game with no reward/punishment opportunities, whereas the possibility of rewards has only a transient effect.\footnote{More recent research has shown that the effectiveness of rewards and punishments in settings such as this depends on the rewarding/punishing technology. Sutter, Häsler, and Kocher (2010) find that when the benefit/cost of receiving reward/punishment is three times larger than the cost of delivering it (i.e. with a 3:1 technology), both mechanisms are effective in encouraging contributions. Similarly, Rand, Dreber, Ellingsen, Fudenberg, and Nowak (2009) find that rewards are as effective as punishments in sustaining cooperation in a repeated public good game experiment with unknown time horizon and with a 3:1 reward/punishment technology. Güler, Irlenbusch, and Rockenbach (2006) study a public good game where the rewarding mechanism displays a 1:1 technology and a punishment mechanism displays a 3:1 technology. They find that only the latter have an impact on contributions. Güler, Irlenbusch, and Rockenbach (2009) use a public goods game where one group member (the leader) can reward or punish the other contributors. Although both rewarding and punishment mechanisms display a 3:1 technology, they find that contributions are higher when punishments are used.}

10
Moreover, we study bonuses and fines that are pure transfers from one party to another, and so have no direct efficiency implications. Thus, bonuses or fines can only enhance performance to the extent that they succeed in inducing behavior that is more aligned with the group interest. Finally, unlike previous research on the effect of rewards/punishments in social dilemmas, in our game standard theory predicts that bonuses and fines will affect performance. As far as we are aware there have only been two experimental studies of inspection games. Dorris and Glimcher [2004] observe the behavior of human and monkey subjects in inspection games with different parameterizations of the inspection cost. In some experiments they had humans playing against humans, whereas in others they had humans or monkeys in the role of Worker playing against a computer in the role of Inspector. They find that (human and monkey) Workers' behavior is close to Nash equilibrium predictions only for high inspection costs. Dorris and Glimcher [2004] do not study the impact of bonus or fines in their setup. Rauhut [2009] studies the impact of the severity of the punishment in an inspection game. His set up differs from ours in that the punishment hurts the inspectee but does not affect the payoff of the inspector in any way. A consequence is that an increase in the punishment decreases the probability of inspection but leaves the probability of shirking unaffected in the Nash equilibrium. Nevertheless, he finds that inspectees shirk less often when the punishment is increased, in agreement with the own-payoff effect. Our study differs from his also in that we study reward as well as punishment. As far as we are aware ours is the first study to compare positive and negative incentives in inspection games. Our study also contributes to a recent literature evaluating different solution concepts for predicting behavior in games with mixed strategy equilibria (e.g., Selten and Chmura, 2008; Brunner, Camerer, and Goeree, 2011; Selten, Chmura, and Goerg, 2011). Standard game theoretical analysis applied to the game used in our experiment yields the perhaps paradoxical result that introducing bonuses increases considerably the probability that the employee will shirk. While in our experiment we do observe a slight increase in shirking in the presence of bonuses, this effect is much smaller than predicted by Nash equilibrium and is not statistically significant. This is more in line with the predictions made by alternative concepts such as Impulse Balance Equilibrium and Quantal Response Equilibrium (although, for our data, the latter concept performs better than Nash equilibrium only if it incorporates loss aversion). More generally, our results show that when Nash equilibrium and alternative predictions diverge we find more support for the latter than for the former. In this study we have focused on the case where rewards and punishments are simple transfers between the interacting parties (e.g., monetary fines for misconduct or bonuses for good conduct). This seems to be a useful starting point as the connections between incentives, behavior, and earnings depend on how their contributions compare with others. Dickinson [2001] assigns rewards/punishment points to the highest/lowest contributor in the group, and Parkin, Fehr, Gächter, and Winter-Ebmer [2000] assigns rewards/punishments to those who contribute more/less than average.

11 There have been public good game experiments where rewards/punishments are automatically assigned to players depending on how their contributions compare with others. Dickinson [2001] assigns rewards/punishment points to the highest/lowest contributor in the group, and Parkin, Fehr, Gächter, and Winter-Ebmer [2000] assigns rewards/punishments to those who contribute more/less than average.

12 See also Glimcher, Dorris, and Bayer [2005].

13 In fact, Rauhut studies a game where two inspectors interact with two inspectees who are involved in a prisoners' dilemma. Under some assumptions, this expanded game has the same characteristics as an inspection game.
are straightforward to interpret: bonuses and fines have no direct efficiency consequences unless they induce a change in behavior. We find that fines, but not bonuses, enhance efficiency. An interesting extension would be one where the costs and benefits of rewarding/being rewarded are asymmetric (e.g., when bonuses consist of medals and prizes, that may have more value for the person receiving them than for the person awarding them). If the bonus remains equally costly to the inspector while it becomes more beneficial to the inspectee, our results suggest that the inspectee will shirk less often because of the enhanced own-payoff effect of working. Thus, in such a setup bonuses may have a positive effect on inspectees’ good behavior. Also, in this study we examine the performance of exogenously imposed mechanisms. In our experiment, workers chose whether to work or shirk and employers chose whether to inspect or not inspect. Fines and bonuses were then triggered automatically in response to the actions chosen by the players. Another interesting avenue for further research would be to explore the endogenous choice of punishing and rewarding mechanisms.
4. How to Prevent Workers from Shirking: the Use and Effectiveness of Rewards and Punishments in the Inspection Game

4.1. Introduction

In the labor market, employers usually want workers to perform in a way that, left to themselves, they would not do. In many situations, workers will only deliver the desired performance level if there is a serious possibility that their work is inspected by the employer. Monitoring a worker is costly to the employer, though, and the employer would prefer not to do so if he were sufficiently sure that the worker would work hard. The essence of the interaction in such situations is described in the inspection game. In this game, the employer chooses to inspect or not, and the worker chooses to provide low or high effort. In every situation one of the players prefers to have chosen a different action. Basically, the inspection game is an asymmetric matching pennies game and the unique equilibrium is in mixed strategies.

To further encourage good behavior, after inspection the employer may consider punishing a worker who was found providing low effort or rewarding a worker who was found providing high effort. In this chapter, we investigate experimentally whether employers use rewards or punishments to incentivize their workers, and we compare the effectiveness of the two possibilities. Whether rewards for good behavior or punishments for bad behavior are more effective in preventing shirking is still an open question. Folk wisdom suggests that rewards may be more effective. As Benjamin Franklin (1744), one of America’s founding fathers, put it: “... a spoonful of honey will catch more flies than (a) gallon of vinegar”. This folk wisdom is backed up by a strand of literature in psychology started by Skinner (1965). From his studies on animals, he concluded that rewards dominate punishments as punishments lose their effectiveness in the long term. In agreement with this conclusion, psychologists have reported that supervisors rewarding good behavior are more successful in encouraging subordinates to work hard than supervisors punishing bad behavior (Sims 1980; Podsakoff, Bommer, Podsakoff, and MacKenzie 2006; George 1995).

\[\text{This chapter is based on the identically titled paper joint with Daniele Nosenzo, Theo Offerman, and Martin Sefton. We are grateful to CREED programmer Jos Theelen for programming the experiment.}\]
Typically, these studies draw their conclusions on the basis of questionnaires for employers and employees. This complicates the interpretation of the results because it is a priori not clear that rewards and punishments cause worker’s behavior or vice versa.

Controlled laboratory experiments investigating the strength of positive and negative reciprocity have been run, but not in the context of the inspection game. Previous studies consistently found relatively strong evidence for negative reciprocity and weak (or no) evidence for positive reciprocity (Abbink, Irlenbusch, and Renner 2000; Brandts and Sola 2001; Charness and Rabin 2002; Offerman 2002; Brandts and Charness 2004; Falk, Fehr, and Fischbacher 2003; Charness 2004; Al-Ubaydi and Lee 2009). The weak evidence for positive reciprocity casts doubt on the effectiveness of rewards in employer/worker relations. Ex ante it is hard to say what should be inferred from the stronger evidence for negative reciprocity for the case of the inspection game. On the one hand, employers using punishments may trigger a negative spiral of ongoing shirking and punishments, so that punishments may even have a counterproductive effect. On the other hand, workers may fear the possibility of punishment and work hard simply to avoid them. This would happen if the findings in the ultimatum game generalize to the inspection game. In the ultimatum game, proposers tend to behave well and propose fair offers to avoid the rejection (punishment) by responders (for a meta-study of ultimatum game experiments, see Oosterbeek, Sloof, and van de Kuilen 2004). So evidence collected in controlled laboratory experiments in different environments is also rather inconclusive.

We collect controlled evidence on the use and effectiveness of rewards and punishments in the inspection game in a $(1 + 3 \times 2)$ design. In all treatments, pairs are formed that consist of a worker and an employer interacting repeatedly for an indeterminate length of time. In the baseline treatment, subjects do not have the possibility to reward or punish, and they only interact through the inspection game. In the other treatments, two treatment variables are introduced. The first one is the tool to incentivize workers, which takes the form of (i) reward only, (ii) punish only, or (iii) reward and punish. The second treatment variable concerns the effectiveness of the tool itself, which is either low or high. With the low ratio, each reward or

---

2Our study also contributes to investigations of rewards and punishments in other applications. Andreoni, Harbaugh, and Vesterlund (2003) study the effects of rewards and punishments in a bargaining game where the proposer chooses an amount to transfer to the responder and the responder can then either punish or reward the proposer. They find that proposers' transfers are particularly responsive to the threat of punishment, although rewards have a positive effect. Sefton, Shupp, and Walker (2007) examine the effect of rewards and punishments on contributions in a repeated public good game and find that punishments help sustaining higher cooperation levels in comparison to a baseline without reward/punishment opportunities, whereas the possibility of rewards has only a transient effect.

3In other settings, the effectiveness of rewards and punishments appears to depend on the rewarding/punishing technology. Satter, Hailer, and Kocher (2010) obtain the result that when the benefit/cost of receiving reward/punishment is three times the cost of delivering it (i.e. with a 3:1 technology), both mechanisms are effective in encouraging contributions. Likewise, Rand, Dreber, Ellingsen, Fudenberg, and Nowak (2009) find that rewards are equally effective as punishments in sustaining cooperation in a repeated public good game with unknown time horizon and with a 3:1 reward/punishment technology. Güererk, Irlenbusch, and Rockenbach (2006) study a public good game with a 1:1 rewarding mechanism and a 3:1 punishment mechanism technology and find that only the latter affect contributions. Güererk, Irlenbusch, and Rockenbach (2006) study a public good game where one group member (the leader) can reward or punish the other contributors. Although both rewarding and punishment mechanisms employ a 3:1 technology, they find that punishments are more effective.
punishment point assigned by the employer yields or costs the worker one point and with the high ratio, each assigned reward or punishment point yields or costs the worker three points.

We obtain the following results. Like in public good games, the possibility to reward and/or punish has rather small effects on the interaction between employers and workers with the low ratio. With the high ratio, the following pattern emerges in our data. When employers can either only punish or only reward, workers shirk substantially less often than in the baseline game. The reduction in shirking behavior is approximately equally large with the two tools. With punishments, it is achieved with fewer inspections than with rewards. Therefore, employers are better off with punishments than with rewards. However, when employers have the possibility to use the two tools simultaneously, subjects still tend to employ the reward tool more often. This surprising result can be explained in the following way. When employers can use both tools simultaneously, punishments seem to be relatively less effective than in the case where only punishments are allowed, while rewards do not lose their effectiveness. Results from a questionnaire suggest that our subjects find rewards the more appropriate tool to incentivize workers. Thus, when both tools are available, employers can no longer hide behind the excuse that punishments provided the only way to get the workers to work hard. So there may be two factors contributing to the effect. On the one hand, workers seem to resist punishments when both rewards and punishments are possible, and on the other hand, employers prefer to make use of rewards instead of punishments. As a result, employers do not prefer the use of punishments when both tools are allowed.

This chapter is organized in the following way. Section 4.2 describes the game and provides the standard theoretical benchmark based on selfish rational players. Section 4.3 presents the experimental design. Section 4.4 presents the experimental results and Section 4.5 concludes.

4.2. Inspection Game and Theoretical Benchmark

The inspection game involves two players and simultaneous moves. The employer chooses between inspect and not inspect, and the worker shirks or works. In the standard version of the game (see, e.g., Fudenberg and Tirole 1992, p. 17), the employer incurs a cost of $h$ from inspecting. If the worker provides high effort, the worker incurs a cost of $c$ and the employer receives a revenue of $v$. If the employer does not inspect, the worker always receives a wage of $w$. If the employer inspects, the worker receives nothing when she shirks and she receives the wage when she works. The resulting payoffs are shown in the left panel of Figure 4.1 on the next page. We assume that all variables are positive and $v > c$, $w > h$, $w > c$. Note that joint payoffs are maximized when the worker supplies high effort and the employer does not inspect. The right panel presents the payoffs that we used in the experiment.\footnote{This means that in the experiment, we used the parameters $v = 40$, $w = 20$, $c = 15$ and $h = 15$. We added 15 to each of the worker’s potential payoffs and 25 to each of the employer’s possible payoffs because we wanted to prevent negative outcomes (which are problematic to implement in an experiment) and because we wanted the expected earnings in equilibrium not to differ too much between the two types of players.}
Let $p$ denote the probability of inspection and $q$ denote the probability of shirking. In the unique Nash equilibrium, the probabilities $p$ and $q$ are determined endogenously and must leave the players indifferent between actions. Thus, in equilibrium the employer inspects with probability $p = c/w$ and the worker chooses to shirk with probability $q = h/w$. The employer receives an expected payoff of $\pi_{\text{employee}} = v - w - hv/w$, the worker receives an expected payoff of $\pi_{\text{worker}} = w - c$, and joint payoffs are $\pi_c = v - w - hv/w$. In the version of the game used in the experiment, the employer inspects with probability $p = 3/4$ and the worker shirks with probability $q = 3/4$, and the employer’s expected payoff equals 15 while the worker’s expected payoff equals 20. The inspection game is the stage game in the baseline treatment.

In the games where we allow for punishments and rewards, the stage game of the baseline treatment is augmented in the following way. If the employer inspects, he observes the worker’s choice to shirk or work, and then chooses between ‘No action’, ‘Punish’ and ‘Reward’. If he chooses No action, then the payoffs are simply determined by the payoffs of the Inspection game. If he chooses Reward, he must assign the reward level $k$ from the set $0, 1, 2, 3, 4, 5$ and the employer’s payoff from the inspection game is diminished by $k$ while the worker’s payoff is increased by $\alpha k$. If he chooses Punish, he sets the punishment level $l$ from the same set $0, 1, 2, 3, 4, 5$ and the employer’s payoff from the inspection game is diminished by $l$ while the worker’s payoff is decreased by $\alpha l$. With the low ratio $\alpha = 1$ and with the high ratio $\alpha = 3$. Figure 4.2 on the facing page presents the augmented game graphically. In the games where we allow for reward only, the punishment option is chopped off from the game in Figure 4.2 and in the games where we allow for punishment only, the reward option is eliminated.

The subgame perfect equilibrium outcome of the augmented game is identified by backward induction. After inspection, a selfish and rational employer will either choose No action or choose free punishment ($k = 0$) or free reward ($l = 0$). This behavior is anticipated by the worker and the employer, and as a result, play in the phase preceding the final phase remains unaffected. Thus, in the subgame perfect equilibrium outcome subjects mix between their actions Inspect and Not inspect and actions Work and Shirk in precisely the same way as in the baseline treatment, i.e., $p = 3/4$ and $q = 3/4$.

---

Notes: Employer is the ROW player, Worker is the COLUMN player. Within each cell, the Employer’s payoff is shown at the top and the Worker’s payoff at the bottom.
In the actual labor market, as well as in our experiment, employers and workers are engaged in a repeated interaction. Here, we consider the case where in each stage the game described above is played and where players' earnings are simply the sum of the earnings in all stage games. After each stage game, there will be a new stage game with independent probability $\delta$ and this process continues until it is terminated by chance. In such a setup, it is well-known that a continuum of outcomes can be supported in equilibrium when the continuation probability is sufficiently large. In particular, the cooperative outcome (Not inspect, Work) can be supported in equilibrium by threatening to set the other player on her minimax payoff if she ever deviates from the equilibrium path.

Instead of pursuing a full analysis of the repeated game (which is impossible because the number of possibilities explodes), we provide an intuitive argument for why it is easier to support cooperation in the versions of the game where punishments are allowed. In Figure 4.3 on page 47, we display in gray the pairs of $(p, q)$ that correspond to equilibria where the players play according to a ‘normal stationary stage game strategy’ in each stage game, unless one of them deviates, in which case the deviating player is set on her minimax payoff forever. We assume that in the normal stationary stage game strategy, subjects mix with constant probabilities $(p, q)$, and after inspection employers punish a worker maximally if they find the worker shirking and if

This can only be accomplished if (i) he never inspects or (ii) he inspects with positive probability and the worker always works. In (i), the worker will want to shirk with $q = 1$, in which case the employer’s strategy ceases to be a best response. In (ii), the employer prefers to deviate and never inspect. Likewise, it is easy to see that the employer cannot employ positive rewards $k > 0$ in any Nash equilibrium.
they are allowed to punish, and employers reward a worker maximally if they find the worker working and if they are allowed to reward. In the games that allow the employer to punish a deviating worker, cooperation can effectively be pursued. The expected future losses due to the unforgiving punishment outweigh the temptation to shirk. Without the possibility of punishment, full cooperation cannot be sustained in equilibrium. The promise that good behavior is rewarded may seduce the worker to work hard for a while, but if the employer never inspects and the reward therefore never materializes, the worker will be tempted to shirk. So from this perspective, games in which the employer can punish workers who are found shirking are expected to be more successful in generating actual cooperation.

4.3. Experimental Design and Procedures

The computerized experiment was carried out at the University of Nottingham. Subjects were recruited from a campus-wide distribution list. In total, 250 subjects participated in 21 sessions. Each session contained either five or six pairs of participants. Each subject participated in one session only. During a session no communication between subjects was allowed. Of each of the seven treatments, we carried out three sessions.

At the end of the session, subjects were paid in cash according to their accumulated point earnings from all rounds using an exchange rate of £0.007 per point. Sessions took about 40 minutes on average and earnings ranged between £5.6 and £23.0, averaging £12.1 (approximately US$19.1 at the time of the experiment). Sessions started with a random assignment of subjects to computer terminals. Subjects received the instructions on paper, so that they could read along while an experimenter read the instructions out loud. The instructions concluded with a series of questions testing subjects' understanding of the instructions. Answers were checked by the experimenters, who dealt privately with any remaining questions.

At the start of the experiment, subjects were assigned to pairs and roles. Within each pair, one subject received the role of 'Employer' and the other the role of 'Worker'. Subjects knew that they would stay in the same role and in the same pair during the whole experiment. They were informed that each session consisted of at least 70 rounds, from round 70 on each round could be the last one with probability 1/5. For comparability we kept the (computerized) random stopping draws constant across treatments: each treatment consisted therefore of three sessions with 71, 73 and 83 rounds, respectively.

In each treatment, a round started with a stage where at the same time the worker chose between 'high' (shirk) and 'low' (work) and the employer between 'inspect' and 'not inspect' which led to the payoffs presented in the right panel of Figure 4.1 on page 44. In the Baseline treatment, these were the only choices made in the round and subjects were immediately informed about the choices and payoff consequences for each one of them. At any time, subjects were informed of all choices and earnings of the own pair in previous rounds.

The other 6 treatments varied from the Baseline treatment in the tool that employers received
Figure 4.3: Equilibria in the Repeated Game (continuation probability 0.8)

Notes: the pairs \((p, q)\) in gray present the pairs that can be supported in this particular class of equilibria, while the pairs \((p, q)\) in black cannot be supported in this class. In the ‘normal phase’, subjects mix with constant probabilities \((p, q)\) in every stage game, and after inspection employers punish a worker maximally if they find the worker shirking and if they are allowed to punish, and employers reward a worker maximally if they find the worker working and if they are allowed to reward. The punish/reward games are based on the low ratio (1:1 technology). If a player deviates from the normal phase, she is set on her minimax payoff forever by the other player. In the Punish and Reward&Punish games, the minimax payoff of the worker decreases by 5 (because of the availability of a punishment of 5). In the games that allow punishments and rewards, the players may ignore the reward/punishment possibility, in which case the analysis coincides with the one for the baseline game. In this way, these graphs present additional equilibria offered by the relevant tool. We assume that a deviation of \((p, q)\) is always immediately noticed, even with interior values of \(p\) and \(q\). In reality, the normal phase should be carried out in “cycles” and players can only start punishing deviating players after a deviation from a cycle is observed. Therefore, in a “more realistic analysis”, the area of equilibrium pairs would diminish in each game, but the main qualitative features of the graphs would be preserved. For the more effective 1:3 technology, the pictures look very similar.
Table 4.1: Experimental Design

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Reward</th>
<th>Punishment</th>
<th>Technology</th>
<th>Number of pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>BL</td>
<td>no</td>
<td>no</td>
<td>-</td>
<td>17</td>
</tr>
<tr>
<td>R1:1</td>
<td>yes</td>
<td>no</td>
<td>1:1</td>
<td>18</td>
</tr>
<tr>
<td>P1:1</td>
<td>no</td>
<td>yes</td>
<td>1:1</td>
<td>18</td>
</tr>
<tr>
<td>R&amp;P1:1</td>
<td>yes</td>
<td>yes</td>
<td>1:1</td>
<td>18</td>
</tr>
<tr>
<td>R1:3</td>
<td>yes</td>
<td>no</td>
<td>1:3</td>
<td>18</td>
</tr>
<tr>
<td>P1:3</td>
<td>no</td>
<td>yes</td>
<td>1:3</td>
<td>18</td>
</tr>
<tr>
<td>R&amp;P1:3</td>
<td>yes</td>
<td>yes</td>
<td>1:3</td>
<td>18</td>
</tr>
</tbody>
</table>

To incentivize workers (Reward, Punish or Reward & Punish) and the effectiveness of the tool (Low or High). In each of these other treatments, the round was extended with an extra stage if the employer had chosen to inspect. In the extra stage, only the employer had to make a choice after receiving information of the worker’s choice between shirk and work. In the ‘R1:1’ and ‘R1:3’ treatments, the employer chose between ‘no action’ and ‘reward’, in the ‘P1:1’ and ‘P1:3’ treatments, between ‘no action’ and ‘punish’ and in the ‘R&P1:1’ and ‘R&P1:3’ treatments between ‘no action’, ‘reward’, and ‘punish’. If reward [punish] was chosen in the second stage, the employer chose the number of reward [punishment] tokens, a number from the set 0, 1, 2, 3, 4, 5. The employer paid a cost of 1 point per token. In the ‘1:1’ treatments the effectiveness ratio of the reward/punishment technology was low, meaning that each token increased (in case of reward) or decreased (in case of punishment) the payoff of the worker by one point. In the ‘1:3’ treatments, we employed a more effective 1:3 reward/punishment technology, in which case the worker’s payoff increased or decreased by three points for each token. Finally, both players in the pair were informed of the results in the pair (all choices and payoffs). Table 4.1 summarizes the experimental design.

4.4. Results

We present the experimental results in two parts. In Section 4.4.1, we present an overview of the aggregate results. This part provides the main answers to our research questions. In Section 4.4.2, we delve deeper into the data. There, we present the dynamics in the data and we provide an explanation of the main findings.

4.4.1. Overview

Figure 4.4 on page 50 displays how the inspect decisions of the employers and the shirk decisions of the workers developed over time. The two upper panels compare the Baseline treatment with the treatments with the low ratio. In all these treatments, there is a moderate upward trend in the frequency of inspection. In the second half of the experiment, the inspection probabilities are quite close to the stage game Nash benchmark of 75%. With the low ratio, inspection
probabilities do not differ much between treatments, although employers inspect to a somewhat lesser extent in P1:1 than in R1:1, R&P1:1 and BL. In contrast, the frequencies of shirking remain pretty constant across time in the low ratio treatments, at a substantially lower level than the stage game Nash benchmark. The treatments that allow for rewards and or punishments trigger somewhat less shirking than the Baseline treatment, but differences are modest.

The two lower panels provide the picture for the treatments with the high ratio. Here, the differences with the Baseline treatment are more pronounced. In R1:3 and Baseline, inspection frequencies are similar at the start and eventually grow to approximately the same level in the final rounds. In contrast, the inspection levels in R&P1:3 and P1:3 stay approximately constant, at lower levels than in the other two treatments. The right lower panel shows that subjects shirk substantially less in the treatments with the possibility of rewards and/or punishments than in the Baseline treatment. There are hardly any differences in the three treatments where employers have the possibility to incentivize workers through rewards and/or punishments. Thus, the decrease in inspection level in R&P1:3 and the even bigger decrease in inspection level in P1:3 do not come at the cost of higher shirking.

Because we are mainly interested in the comparison of the treatments after subjects have become familiar with the experiment, we focus on the second part of the experiment in the remainder of this chapter (unless we explicitly mention otherwise). Table 4.2 on page 51 presents the raw averages of inspections and shirking together with test results of hypotheses comparing the levels across treatments. Throughout this chapter, we employ a prudent test procedure with independent average statistics per pair of subjects. So each pair of subjects yields one data-point. We report the results of two-sided non-parametric ranksum tests.

When the punishment/reward technology is relatively ineffective (1:1), the modest differences between the treatments appear not to be significant, with as only exception the comparison of the inspection level between P1:1 and Baseline, which is weakly significant at \(p=0.10\). P1:1 is the only 1:1 treatment where the inspection level is (weakly) significantly less than the stage game Nash benchmark of 75% \((p = 0.06)\). In the 1:1 treatments as well as the Baseline treatment, the shirking levels are significantly below the stage game Nash benchmark of 75%.

With a highly effective punishment/reward technology (1:3), the picture for inspections is qualitatively similar, but some differences are statistically more pronounced. The comparisons of the inspection levels remain insignificant with two exceptions: in P1:3 lower inspection levels are observed than in R1:3 \((p = 0.06)\) and P1:3 is the only 1:3 treatment where the inspection level is significantly below the stage game Nash benchmark. In contrast, the shirking levels in R1:3, P1:3 and R&P1:3 are all substantially and significantly below the Baseline treatment. With regard to shirking, the 1:3 treatments are statistically indistinguishable from each other.

In the comparison of the 1:1 treatments and the 1:3 treatments, the differences in shirking are significant in the Reward treatments \((p=0.06)\) and the Punish treatments \((p=0.01)\).
Table 4.3 on page 52 shows how often employers chose no action, reward and punish after they inspected the worker and observed her decision to work or shirk. In total, employers rewarded workers more often than that they punished them. In R&P 1:1, after inspection employers rewarded workers in 53% of the cases and punished them in only 7% of the cases. In R&P1:3, rewards were assigned in 47% of the cases and punishments in 20% of the cases. Further insight is obtained if these numbers are broken down for whether the worker behaved well or shirked. Unsurprisingly, after the employer observed the worker shirking, he hardly rewarded her and after he observed the worker working he hardly punished her. In R1:1, the employer rewards working in 55% of the cases and in P1:1 the employer punishes shirking in 51% of the cases. Likewise, in R1:3 the employer rewards working in 64% of the cases and in P1:3 the employer punishes shirking in 52% of the cases. So conditional on the tool being appropriate for the action taken, it is used with an approximately equal frequency. In R&P1:1 a remarkable shift in the relative frequencies is observed: here, working is rewarded in 76% of the cases while shirking is only punished in 22% of the cases. So with the low ratio, employers favor rewards over punishments when either tool is allowed. A similar shift is not observed in R&P1:3, though. There, working
Table 4.2.: Actions in Stage 1

<table>
<thead>
<tr>
<th>Treatment</th>
<th>N</th>
<th>Mean</th>
<th>R1:1</th>
<th>P1:1</th>
<th>R&amp;P1:1</th>
<th>=75%</th>
<th>Mean</th>
<th>R1:1</th>
<th>P1:1</th>
<th>R&amp;P1:1</th>
<th>=75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>BL</td>
<td>17</td>
<td>74%</td>
<td>0.72</td>
<td>0.10</td>
<td>0.87</td>
<td>52%</td>
<td>47%</td>
<td>0.51</td>
<td>0.48</td>
<td>0.15</td>
<td>0.00</td>
</tr>
<tr>
<td>R1:1</td>
<td>18</td>
<td>75%</td>
<td>0.14</td>
<td>0.90</td>
<td>0.81</td>
<td>40%</td>
<td>0.96</td>
<td>0.35</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1:1</td>
<td>18</td>
<td>65%</td>
<td>0.34</td>
<td>0.06</td>
<td>0.96</td>
<td>42%</td>
<td>0.23</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R&amp;P1:1</td>
<td>18</td>
<td>72%</td>
<td>0.88</td>
<td>47%</td>
<td>0.52</td>
<td>33%</td>
<td>47%</td>
<td>0.51</td>
<td>0.48</td>
<td>0.15</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Treatment</th>
<th>N</th>
<th>Mean</th>
<th>R1:3</th>
<th>P1:3</th>
<th>R&amp;P1:3</th>
<th>=75%</th>
<th>Mean</th>
<th>R1:3</th>
<th>P1:3</th>
<th>R&amp;P1:3</th>
<th>=75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>BL</td>
<td>17</td>
<td>74%</td>
<td>0.54</td>
<td>0.21</td>
<td>0.30</td>
<td>52%</td>
<td>47%</td>
<td>0.01</td>
<td>0.02</td>
<td>0.03</td>
<td>0.00</td>
</tr>
<tr>
<td>R1:3</td>
<td>18</td>
<td>79%</td>
<td>0.06</td>
<td>0.12</td>
<td>0.42</td>
<td>27%</td>
<td>0.81</td>
<td>0.62</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1:3</td>
<td>18</td>
<td>57%</td>
<td>0.48</td>
<td>0.05</td>
<td>0.96</td>
<td>27%</td>
<td>0.70</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R&amp;P1:3</td>
<td>18</td>
<td>67%</td>
<td>0.20</td>
<td>33%</td>
<td>0.52</td>
<td>33%</td>
<td>0.70</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R1:1 vs R1:3 p=0.37 p=0.06
P1:1 vs P1:3 p=0.54 p=0.01
R&P1:1 vs R&P1:3 p=0.26 p=0.79

Notes: in the columns mean the average of the means of all pairs is displayed; the p-values are the results of the rank-sum tests between treatments within technologies; =75% gives the result of comparing inspect and shirk with the one shot mixed Nash equilibrium benchmark (75%, 75%); bottom 3 rows present the outcomes of ranksum tests between technologies within treatments. Rounds 36-70 only.

is rewarded in 61% of the cases while shirking is punished in 62% of the cases.

In the Baseline treatment, we observe an approximately equal number of inspect/work outcomes as inspect/shirk outcomes. In contrast, Table 4.3 on the following page shows that when employers chose to inspect, they encountered working much more often than shirking in the treatments where punishments and/or rewards are allowed. Thus, even though conditional on the appropriate action employers used each tool about equally frequently, we observe much more reward decisions than punishment decisions because inspect/work occurred substantially more often than inspect/shirk.

Table 4.4 on page 53 provides an overview of the number of tokens assigned by the employer, conditional on choosing a reward or a punishment. The Table shows that in all treatments the expected punishment of shirking behavior is approximately equally large, in the range of 3.34 to 3.90. In contrast, there is more variation in the extent to which employers reward working. In the 1:1 treatments, the expected rewards of working behavior (4.15 in R1:1 and 4.15 in R&P1:1) are higher than the expected punishments of shirking behavior, while in the 1:3 treatments the expected rewards of working behavior (3.21 in R1:3 and 2.74 in R&P1:1) are lower than the expected punishments of shirking behavior. Thus, the level of the reward depends on the technology, and subjects reward less when the ratio is high. Possibly this result is due to inequality aversion considerations.

Furthermore, in the 1:1 treatments the mode of the distribution is to assign 5 tokens in all cases. That is, given than an employer chose to reward or punish, he tended to assign the maximum number of tokens. Again, the picture looks differently for rewards in the 1:3 treatments; there the
mode of the distribution shifts to cheaper rewards of 2 or 3 tokens. It is also worth mentioning that employers sometimes used free punishments of 0 tokens if the worker shirked, while they almost never used free rewards of 0 points to reward if the worker worked. Possibly, employers regard a punishment of 0 tokens as a useful warning while they fear that a free reward backfires.

Table 4.5 on page 54 presents the efficiency levels of the firms on the left hand side and employer’s and worker’s total earnings on the right hand side. We define efficiency as the sum of the worker’s and employer’s earnings in stage 1. Arguably, this is the statistic that would be most interesting to the owners of the firm because it deals with the primary money streams in the firm (in actual firms rewards and punishments are not necessarily expressed in monetary terms).

When the technology is relatively ineffective (1:1), efficiency is only marginally and usually insignificantly enhanced by the possibility to reward and/or punish. Treatment P1:1 provides the exception, where the efficiency level is weakly significantly increased compared to the BL treatment. This is due to the fact that the same level of shirking is accomplished with fewer inspections in P1:1. Interestingly, in the 1:1 treatments the employer does not benefit from the possibility to reward and/or punish, while the worker is better off when rewards are allowed (both in R1:1 and R&P1:1, workers earn significantly more than in BL).

The picture is different in the 1:3 treatments where rewards and punishments are more effective. There, the efficiency levels are significantly enhanced when rewards and/or punishments are
### Table 4.4: Assignment of Tokens

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Actions</th>
<th>Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>stage II</td>
<td>stage I</td>
</tr>
<tr>
<td>R1:1 reward</td>
<td>Work</td>
<td>157</td>
</tr>
<tr>
<td></td>
<td>Shirk</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>162</td>
</tr>
<tr>
<td>P1:1 punish</td>
<td>Work</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Shirk</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>88</td>
</tr>
<tr>
<td>R&amp;P1:1 reward</td>
<td>Work</td>
<td>238</td>
</tr>
<tr>
<td></td>
<td>Shirk</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>242</td>
</tr>
<tr>
<td></td>
<td>punish</td>
<td>Work</td>
</tr>
<tr>
<td></td>
<td>Shirk</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>31</td>
</tr>
<tr>
<td>R1:3 reward</td>
<td>Work</td>
<td>229</td>
</tr>
<tr>
<td></td>
<td>Shirk</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>245</td>
</tr>
<tr>
<td>P1:3 punish</td>
<td>Work</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Shirk</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>69</td>
</tr>
<tr>
<td>R&amp;P1:3 reward</td>
<td>Work</td>
<td>188</td>
</tr>
<tr>
<td></td>
<td>Shirk</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>199</td>
</tr>
<tr>
<td></td>
<td>punish</td>
<td>Work</td>
</tr>
<tr>
<td></td>
<td>Shirk</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>84</td>
</tr>
</tbody>
</table>

Notes: conditional on a reward or punishment decision, the average relative frequency of the number of tokens assigned in a treatment for the worker’s decision is listed. The expected value is calculated as the sum of the products of the tokens and the relative frequencies; rounds 36-70 only.

allowed and employers are better off compared to the BL treatment. Remarkably, although the employers are the ones who decide whether they want to punish or reward, and therefore could ignore the possibility to reward if both tools are allowed, employers earned less in P1:3 than in R&P1:3. The difference is (weakly) significant at $p = 0.09$. In Section 4.4.2, we come back to this surprising result. The workers also benefit significantly from employers’ ability to incentivize them, except in the treatment P1:3 where only punishments are allowed, in which case they earned approximately the same as in the BL.

### 4.4.2. Dynamics and Explanation

The previous section dealt with the aggregate static outcomes of the experiment. In this section, we present the behavioral dynamics and we provide an explanation of the main results. Table 4.5 on page 56 presents how often combinations of employer and worker decisions occurred in the different treatments. In addition, it displays transitions by listing the frequencies of outcomes in a new round conditional on the outcomes in the previous round.

In the columns ‘freq’, the relative frequencies of employer/worker decisions are listed. In BL, the most common combinations are inspect/work and inspect/shirk, which occur approximately...
Table 4.5: Efficiency and Earnings

<table>
<thead>
<tr>
<th>Treatment</th>
<th>N</th>
<th>Mean (s.d.)</th>
<th>p-values</th>
<th>Mean (s.d.)</th>
<th>p-values</th>
<th>Mean (s.d.)</th>
<th>p-values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>R1:1</td>
<td>P1:1</td>
<td>R&amp;P1:1</td>
<td>R1:1</td>
<td>P1:1</td>
<td>R&amp;P1:1</td>
</tr>
<tr>
<td>BL</td>
<td>17</td>
<td>42.19 (9.05)</td>
<td>0.31</td>
<td>0.09</td>
<td>0.15</td>
<td>22.64 (7.96)</td>
<td>0.70</td>
</tr>
<tr>
<td>R1:1</td>
<td>18</td>
<td>43.77 (5.20)</td>
<td>0.54</td>
<td>0.28</td>
<td>0.72</td>
<td>22.54 (4.46)</td>
<td>0.79</td>
</tr>
<tr>
<td>P1:1</td>
<td>18</td>
<td>44.91 (5.11)</td>
<td>0.65</td>
<td>0.89</td>
<td>0.09</td>
<td>23.36 (5.16)</td>
<td>0.95</td>
</tr>
<tr>
<td>R&amp;P1:1</td>
<td>18</td>
<td>46.01 (7.64)</td>
<td>0.05</td>
<td>0.01</td>
<td>0.01</td>
<td>23.90 (6.97)</td>
<td>0.04</td>
</tr>
<tr>
<td>R1:1 vs R1:3</td>
<td>p = 0.10</td>
<td>p = 0.05</td>
<td>p = 0.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1:1 vs P1:3</td>
<td>p = 0.04</td>
<td>p = 0.02</td>
<td>p = 0.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R&amp;P1:1 vs R&amp;P1:3</td>
<td>p = 0.53</td>
<td>p = 0.23</td>
<td>p = 0.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: the column efficiency concerns the sum of the earnings of the employer and the worker in the first stage (excluding rewards and punishments). The column Employer (Worker) concerns the total earnings of the employer (worker) in both stages. The p-values list the results of rank-sum tests. Bottom 3 rows present results of rank-sum tests between technologies. Table is based on rounds 36-70.

Equally often. In all other treatments, the outcome inspect/work is more often observed than any of the other outcomes. A striking result is that the cooperative outcome (not inspect/work) occurs rather infrequently, usually in less than 20% of the cases, with as main exception treatment P1:3. There, with an effective punishment tool, employers are able to get the workers to work without inspecting that often. This feature of the data is in line with the game theoretic intuition provided in Section 4.2 suggesting that the cooperative outcome was most easily pursued when punishments were allowed. It is remarkable that the relative frequency of the cooperative outcome again falls when the possibility to reward is added in R&P1:3.

In the BL treatment the outcomes not inspect/work, inspect/work and inspect/shirk were often repeated in the next round, while the outcome not inspect/shirk was much less stable. In fact, after not inspect/shirk almost anything could happen with about equal probability.

In the reward treatments R1:1 and R1:3, very different dynamics are observed. Here, the outcome inspect/work attracts most of the outcomes, especially when the effective technology is employed in R1:3. The exception is when the bad outcome is reached where the employer
Table 4.6: Played Combinations and Transitions

<table>
<thead>
<tr>
<th>Treatment</th>
<th>t=t freq.</th>
<th>ni/w</th>
<th>ni/s</th>
<th>in/w</th>
<th>in/s</th>
<th>t=t+1 freq.</th>
<th>ni/w</th>
<th>ni/s</th>
<th>in/w</th>
<th>in/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>BL</td>
<td>16%</td>
<td>47%</td>
<td>19%</td>
<td>16%</td>
<td>17%</td>
<td>16%</td>
<td>47%</td>
<td>19%</td>
<td>16%</td>
<td>17%</td>
</tr>
<tr>
<td></td>
<td>9%</td>
<td>22%</td>
<td>24%</td>
<td>22%</td>
<td>33%</td>
<td>9%</td>
<td>22%</td>
<td>24%</td>
<td>22%</td>
<td>33%</td>
</tr>
<tr>
<td></td>
<td>37%</td>
<td>14%</td>
<td>4%</td>
<td>60%</td>
<td>23%</td>
<td>37%</td>
<td>14%</td>
<td>4%</td>
<td>60%</td>
<td>23%</td>
</tr>
<tr>
<td>R1:1</td>
<td>14%</td>
<td>20%</td>
<td>36%</td>
<td>30%</td>
<td>17%</td>
<td>14%</td>
<td>20%</td>
<td>36%</td>
<td>30%</td>
<td>17%</td>
</tr>
<tr>
<td></td>
<td>11%</td>
<td>27%</td>
<td>16%</td>
<td>25%</td>
<td>31%</td>
<td>11%</td>
<td>27%</td>
<td>16%</td>
<td>25%</td>
<td>31%</td>
</tr>
<tr>
<td></td>
<td>45%</td>
<td>15%</td>
<td>6%</td>
<td>63%</td>
<td>15%</td>
<td>45%</td>
<td>15%</td>
<td>6%</td>
<td>63%</td>
<td>15%</td>
</tr>
<tr>
<td>P1:1</td>
<td>20%</td>
<td>29%</td>
<td>26%</td>
<td>16%</td>
<td>22%</td>
<td>20%</td>
<td>29%</td>
<td>26%</td>
<td>16%</td>
<td>22%</td>
</tr>
<tr>
<td></td>
<td>16%</td>
<td>27%</td>
<td>28%</td>
<td>25%</td>
<td>19%</td>
<td>16%</td>
<td>27%</td>
<td>28%</td>
<td>25%</td>
<td>19%</td>
</tr>
<tr>
<td></td>
<td>39%</td>
<td>21%</td>
<td>9%</td>
<td>53%</td>
<td>18%</td>
<td>39%</td>
<td>21%</td>
<td>9%</td>
<td>53%</td>
<td>18%</td>
</tr>
<tr>
<td>R&amp;P1:1</td>
<td>18%</td>
<td>46%</td>
<td>12%</td>
<td>28%</td>
<td>14%</td>
<td>18%</td>
<td>46%</td>
<td>12%</td>
<td>28%</td>
<td>14%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>28%</td>
<td>25%</td>
<td>26%</td>
<td>27%</td>
<td>10%</td>
<td>28%</td>
<td>25%</td>
<td>26%</td>
<td>27%</td>
</tr>
<tr>
<td></td>
<td>50%</td>
<td>12%</td>
<td>5%</td>
<td>71%</td>
<td>11%</td>
<td>50%</td>
<td>12%</td>
<td>5%</td>
<td>71%</td>
<td>11%</td>
</tr>
<tr>
<td></td>
<td>23%</td>
<td>6%</td>
<td>11%</td>
<td>30%</td>
<td>54%</td>
<td>23%</td>
<td>6%</td>
<td>11%</td>
<td>30%</td>
<td>54%</td>
</tr>
</tbody>
</table>

Notes: freq. gives the frequencies of all combinations employer/worker decisions in rounds 36-70; t=t presents the frequency in the current round and t=t+1 presents the outcomes in the subsequent round conditional on the combination of the current round; ni=not inspect, in=inspect, w=work, s=shirk.

inspects and the worker shirks, in which case subjects often stubbornly repeat their previous choices.

In the Punish treatment P1:3, the efficient outcome not inspect/work is repeated in a clear majority of the cases where it occurs. Likewise, inspect/work and inspect/shirk are also often repeated, both in P1:1 and P1:3. In contrast, in P1:3, the outcome not inspect/shirk is almost always abandoned, most often in favor of the outcome where the worker gives in (not inspect/work). In this treatment, the fear of punishment seems to loom large. In the reward and punish treatment R&P1:3, the dynamics are similar as in the reward treatment to the extent that the combination of inspect and work absorbs many previous outcomes. In R&P1:3 the outcome of inspect and work is repeated even more often once it is reached, but here it does not absorb behavior from the other cells. Here, the outcomes not inspect/work and inspect/shirk tend to be repeated, while after no inspect/shirk any outcome may occur.

A striking feature shared by all treatments is that both theemployer and the worker tended to stubbornly repeat their choices when the bad outcome was reached where the employer inspects and the worker shirks. Table 4.7 on the following page zooms in on the question how likely such ‘battles of the will’ were, how long they lasted and how they tended to be resolved. In the 1:1 treatments, runs occurred approximately equally frequently in R1:1 and P1:1 as in BL, but they occurred to a lesser extent in R&P1:1. In the treatments where punishments and/or rewards were possible, the average lengths of these runs were smaller than in the baseline treatment. In contrast, in all effective technology 1:3 treatments, runs occurred much less frequently that in the baseline treatment, and if they occurred, they lasted shorter, except for R1:3. In all cases, it was the worker who was more likely to give in after a battle of the wills by changing her behavior.

55
Table 4.7.: Battle of the Wills: Who Gives in?

<table>
<thead>
<tr>
<th>Treatment</th>
<th>#runs</th>
<th>length mean (sd)</th>
<th>work. empl.</th>
<th>both</th>
<th>Treatment</th>
<th>#runs</th>
<th>length mean (sd)</th>
<th>work. empl.</th>
<th>both</th>
</tr>
</thead>
<tbody>
<tr>
<td>BL</td>
<td>18</td>
<td>4.83 (2.75)</td>
<td>67%</td>
<td>22%</td>
<td>11%</td>
<td>18</td>
<td>67%</td>
<td>22%</td>
<td>11%</td>
</tr>
<tr>
<td>R1:1</td>
<td>19</td>
<td>3.84 (0.76)</td>
<td>42%</td>
<td>42%</td>
<td>16%</td>
<td>10</td>
<td>5.60 (3.86)</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>P1:1</td>
<td>15</td>
<td>4.27 (2.12)</td>
<td>93%</td>
<td>7%</td>
<td>0%</td>
<td>11</td>
<td>3.64 (1.80)</td>
<td>64%</td>
<td>27%</td>
</tr>
<tr>
<td>R&amp;P1:1</td>
<td>10</td>
<td>4.20 (2.49)</td>
<td>60%</td>
<td>30%</td>
<td>10%</td>
<td>9</td>
<td>3.67 (0.71)</td>
<td>78%</td>
<td>22%</td>
</tr>
</tbody>
</table>

Notes: a run is a series of consecutive rounds where the worker shirks and the employer inspects; runs shorter than 3 are discarded; we only consider runs that had their first round and their last round between 36 and 69.

In Section 4.4.1, we reported the remarkable result that even though employers made more money when they used punishments to incentivize workers in P1:3 than when they used rewards to encourage workers in R1:3, they did not shift toward using punishments when both tools were allowed in R&P1:3. Ideally, to investigate the success of rewarding versus punishing, one would like to classify employers as ‘punishers’, ‘rewarders’, ‘punishers and rewarders’ and ‘no-punishers and no-rewarders’ and the workers as ‘shirkers’ or ‘workers’ on the basis of an external measure. Then we could compare the occurrence of either type of employers across treatments, and we could compare their performance when matched with shirkers, and when matched with workers.

We do not have such independent measures in our experiment, and therefore use behavior in the first 10 rounds as a proxy for the measure, and we use the rounds 11-70 to determine the success of various strategies. Table 4.8 on the next page presents employers’ earnings as a function of their own type and the type of worker they were matched with.

For completeness, the Table presents the results for the 1:1 treatments as well as the 1:3 treatments. Here, we focus on the 1:3 treatments because in these treatments we observed real differences between the treatments. In the treatment where employers are restricted to using rewards R1:3, employers classified as rewarder make clearly more money when they are matched with a worker who is not a shirker than employers who do not make use of the possibility to reward. If rewarders are matched with shirkers they make approximately the same amount as money as employers who do not use the reward tool. In the treatment where employers can make use of punishments but not rewards P1:3, when matched with a shirker employers make substantially more money when they are punishers than when they are not. In contrast, when matched with workers who work, the punishment strategy is counter productive and punishers earn less than the employers who refrain from punishing. Remarkably, when matched with workers who work, employers who refrain from punishing in P1:3 earn substantially more than employers who refrain from rewarding in R1:3. Possibly, the latent threat of (not used) punishments encouraged workers to behave well in P1:3.
Table 4.8: Employers’ Strategies and Earnings

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Worker</th>
<th>punisher</th>
<th>no punisher/rewarder</th>
<th>punisher/rewarder</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>mean(s)</td>
<td>N</td>
<td>mean(s)</td>
</tr>
<tr>
<td>R1:1</td>
<td>Worker</td>
<td>3</td>
<td>20.21(1.57)</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Shirk er</td>
<td>6</td>
<td>21.27(3.01)</td>
<td>3</td>
</tr>
<tr>
<td>P1:1</td>
<td>Worker</td>
<td>3</td>
<td>25.34(4.16)</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Shirk er</td>
<td>5</td>
<td>20.39(1.70)</td>
<td>4</td>
</tr>
<tr>
<td>R&amp;P1:1</td>
<td>Worker</td>
<td>1</td>
<td>13.25(2.89)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Shirk er</td>
<td>2</td>
<td>30.00(2.07)</td>
<td>3</td>
</tr>
<tr>
<td>R1:3</td>
<td>Worker</td>
<td>2</td>
<td>24.57(1.08)</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Shirk er</td>
<td>4</td>
<td>22.96(2.58)</td>
<td>5</td>
</tr>
<tr>
<td>P1:3</td>
<td>Worker</td>
<td>3</td>
<td>25.48(3.79)</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Shirk er</td>
<td>5</td>
<td>27.76(4.42)</td>
<td>4</td>
</tr>
<tr>
<td>R&amp;P1:3</td>
<td>Worker</td>
<td>3</td>
<td>22.09(2.94)</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Shirk er</td>
<td>2</td>
<td>23.53(4.78)</td>
<td>2</td>
</tr>
</tbody>
</table>

Notes: workers and employers are classified on the basis of their behavior in the first 10 rounds; employers’ average earnings are based on rounds 11-70 (stage 1 and 2 earnings added); workers are classified on the basis of how often they shirked in the first 10 round, the 9 workers shirking fewest are classified as “workers”, the other 9 as “shirkers”; employers are classified on the basis of the average assigned reward tokens (x1) and the average punish tokens (x2) over the first 10 rounds: if max (x1, x2) < 0.5 then the employer is classified as “no punisher/no rewireder”, if max (x1, x2) ≥ 0.5 and |x1 - x2| < 0.25 then the employer is classified as “punisher/rewarder”, if max (x1, x2) ≥ 0.5 and x1 - x2 ≥ 0.25 then the employer is classified as “rewarder”, if max (x1, x2) ≥ 0.5 and x2 - x1 ≥ 0.25 then the employer is classified as “punisher”. When both tools become available in R&P1:3, the picture becomes different. Unlike in P1:3, employers who are matched with shirkers earn hardly more when they act as punisher than when they refrain from punishing and rewarding. So punishing loses much of its bite when both tools are available. In contrast, employers who are matched with workers who work earn much more when they pursue a rewarding strategy than when they refrain from using, and the difference is bigger than in R1:3. So rewarding workers who behave well seems to become more remunerative when both tools are allowed. Another striking feature is that employers who are matched with well-behaving workers and who refrain from punishing and rewarding in R&P1:3 earn much less than employers who are matched with well-behaving workers and who refrain from punishing in P1&3. This suggests that the unused threat of punishing loses much of its force when employers can use rewards as well as punishments.
Table 4.9.: Questionnaire

<table>
<thead>
<tr>
<th>Treatment Type</th>
<th>q1 reward (sd)</th>
<th>q2 punishment (sd)</th>
<th>q3 reward (sd)</th>
<th>q4 punishment (sd)</th>
<th>q5 reward (sd)</th>
<th>q6 punishment (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1:1 employer</td>
<td>4.08 (2.23)</td>
<td>5.42 (2.15)</td>
<td>5.83 (1.47)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>worker</td>
<td>3.92 (2.11)</td>
<td>3.58 (2.19)</td>
<td>6.67 (0.65)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>employer vs worker MW</td>
<td>p = 0.84</td>
<td>p = 0.04</td>
<td>p = 0.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1:1 employer</td>
<td>2.50 (1.68)</td>
<td>4.42 (2.68)</td>
<td>4.08 (1.88)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>worker</td>
<td>2.50 (2.11)</td>
<td>4.00 (2.59)</td>
<td>4.92 (1.41)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>employer vs worker MW</td>
<td>p = 0.69</td>
<td>p = 0.70</td>
<td>p = 0.29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1:1 vs P1:1 MW</td>
<td>p = 0.07</td>
<td>p = 0.30</td>
<td>p = 0.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R&amp;P1:1 employer</td>
<td>4.92 (1.98)</td>
<td>2.92 (1.88)</td>
<td>5.42 (2.19)</td>
<td>4.58 (2.50)</td>
<td>5.67 (1.78)</td>
<td>4.67 (2.27)</td>
</tr>
<tr>
<td>worker</td>
<td>3.67 (2.39)</td>
<td>3.00 (2.91)</td>
<td>4.42 (2.39)</td>
<td>3.92 (2.93)</td>
<td>6.75 (0.45)</td>
<td>5.08 (1.93)</td>
</tr>
<tr>
<td>employer vs worker MW</td>
<td>p = 0.18</td>
<td>p = 0.93</td>
<td>p = 0.71</td>
<td>p = 0.58</td>
<td>p = 0.05</td>
<td>p = 0.70</td>
</tr>
</tbody>
</table>

Wilcoxon

<table>
<thead>
<tr>
<th>q1 vs q2</th>
<th>q3 vs q4</th>
<th>q5 vs q6</th>
</tr>
</thead>
<tbody>
<tr>
<td>R&amp;P1:1 employer</td>
<td>p = 0.02</td>
<td>p = 0.26</td>
</tr>
<tr>
<td>worker</td>
<td>p = 0.09</td>
<td>p = 0.51</td>
</tr>
</tbody>
</table>

Notes: the questionnaire was filled out by the subjects of the last 6 sessions equally divided over R1:1; P1:1 and R&P1:1; MW=Mann-Whitney test; 7[1] = completely disagree; q1="After inspection, I enjoyed rewarding the worker if he or she provided high effort; I think the employer enjoyed rewarding me after inspecting if I provided high"; q2="After inspection, I enjoyed punishing the worker if he or she provided low effort; I think the employer enjoyed punishing me after inspecting if I provided low"; q3="I assigned reward points to reinforce the worker’s behavior; I think the employer assigned reward points to reinforce my behavior"; q4="I assigned punishment points to change the worker’s behavior; I think the employer assigned punishment points to change my behavior"; q5="It is appropriate to reward a worker who provides high effort"; q6="It is appropriate to punish a worker who provides low effort".

The success of the different strategies lines up with their actual use. In P1:3 where punishments were effective, 56% (5 out of 9) of the employers who were matched with a shirker pursued a punishing strategy. In R&P1:3, the percentage of employers exclusively relying on punishments decreased to 22% (2 out of 9).

In the final 6 sessions, we administered a questionnaire to further explore the reasons for an asymmetry between rewards and punishments. In the questionnaire, we asked employers as well as workers whether they felt that the employer enjoyed punishing/rewarding, whether the employer’s aim was to influence the worker’s behavior and to what extent the uses of punishments and rewards were appropriate. Table 4.9 presents the results. Employers and workers tend
to agree that employers enjoy rewarding good behavior, while they do not enjoy punishing bad behavior. Employers as well as workers think that rewards and punishments are used to influence the worker’s behavior. Interestingly, the employers agree more with these statements than workers do, although the differences are usually not significant. Most informative are the answers regarding the appropriateness of the uses of rewards and punishments. Both employers and workers agree very much with the statement that it is appropriate to reward a well-behaving worker, while they agree substantially and significantly less with the statement that punishments are appropriate when the worker shirks. The difference in feelings about the appropriateness of the two tools may explain why many employers primarily chose to reward and why punishments lost part of their effectiveness when both tools were available.

4.5. Conclusion

Employers who want to stimulate workers to work hard may consider using rewards and punishments to achieve their goal. The use and effectiveness of rewards and punishments by employers is often hotly debated. Many people have strong opinions on how workers should be encouraged. It is surprising that this important discussion has not yet been backed up by controlled laboratory evidence. In this chapter, we have contributed to filling this gap.

We have obtained the following results. When rewards and punishments are relatively ineffective, as in our 1:1 treatments, rewards and punishments have only modest effects that are often not significant. Instead, when we introduced effective rewards and punishments in our 1:3 treatments, we observed substantial and significant effects. In the treatments where employers could use only punishments or only rewards, as well as in the treatment where both tools were allowed, we observed a common substantial decrease in the rate of shirking compared to the baseline treatment. In the treatment where employers were restricted to punishments, this was accomplished with much fewer costly inspections than when employers were restricted to rewards. As a result, employers earned more when they could only use punishments than when they could only use rewards. A remarkable result was that when employers could use both rewards and punishments, they did not shift in the direction of using punishments. To the contrary, employers continued to reward more often than punish when both tools were allowed.

A closer analysis reveals that the punishment strategy loses much of its force when both rewards and punishments are allowed. Pursuing a punishment strategy is more remunerative when employers cannot reward than when they can. In addition, employers as well as workers report that they feel that rewarding a well-behaving worker is more appropriate than punishing a shirker. The bottom line is that when employers can use rewards and punishments, our results suggest that they will primarily incentivize their workers through rewards, and for good reasons because the effectiveness of punishments may be eroded when rewards are possible. From the firm’s perspective, shirking behavior is most efficiently reduced when the manager does not have the possibility to reward good behavior of the workers. So if the government (or the owners of
the firm) limits the extent to which bonuses can be given, superior results for the firm may be obtained.
5. Keeping out Trojan Horses: Auctions and Bankruptcy in the Laboratory

5.1. Introduction

Confronted with a large wooden horse outside their gate, the Trojans discussed how to deal with it. Some, like the soothsayer Cassandra, advised destruction. Her father, King Priam, decided otherwise, which had the well-known dire consequences for Troy. Nowadays, governments may be confronted with a similar situation when auctioning the right to market a good: The bids may look very attractive at the onset, but the auction can turn into a nightmare if the winner goes bankrupt.

Indeed, a license auction or a procurement procedure can hardly be considered a success if the winning bidder defaults on its obligations. If the winner of a license auction files for bankruptcy, the market power of the remaining competitors will increase, potentially at the cost of consumers. This situation may last for several years if the licenses are tied up in bankruptcy litigation. If the winner of a procurement procedure goes bankrupt, the delivery of goods and services may be considerably delayed and the procuring organization may have to buy those for a higher price from a different supplier.

The problem of defaulting bidders is not only of academic interest. In the 1996 C-block auction by the Federal Communications Commission (FCC) in the US, all major bidders went bankrupt. While in total these bidders bid $10.2 billion almost nothing was paid [Zheng, 2001]. Additionally, in the construction industry in the US between 1990 and 1997, 80,000 contractors filed for bankruptcy. The liabilities for public and private clients are estimated to lie above $21 billion [Calveras, Guinza, and Hauk, 2004].

Firms on the edge of bankruptcy may have an incentive to bid aggressively, because they bid for “options on prizes” rather than on “prizes”. If the object turns out to be more valuable than expected, they make a nice profit. However, if it leads to losses, the firms will default, which they probably would have done even if they had not participated in the auction [Klemperer, 2002; Board, 2007]. Therefore, they have an advantage over financially healthy firms because

---

[1] This chapter is based on the identically titled paper joint with Sander Onderstal and benefited from helpful comments of Susan Athey, Gary Charness, Marcus Cole, Simon Gächter, Charley Holt, Audrey Hu, Thomas Kittsteiner, Dan Levin, Theo Offerman, Marion Ott, Sarah Parlane, Tim Salmon, and participants at conference and seminar presentations at the University of Amsterdam, the University of Nottingham, NAKE 2010, M-BEES 2010, CEDEX 2010, ESA 2010, and EARIE 2010.
the latter have to take the downward risks of the project into account and are willing to bid less aggressively than underfinanced firms (Zheng [2001] Klempner [2002]).

In this chapter, we examine how an auctioneer can mitigate the likelihood of bidders going bankrupt. In particular, we answer the following question using a laboratory experiment: How do first-price auctions (like the first-price sealed-bid auction) and second-price auctions (like the English auction) perform in terms of the likelihood of bankruptcy? This question is particularly interesting because procurement auctions are usually first-price auctions while license auctions typically tend to be of the second-price type. If one of the two auction types tends to be less sensitive to ex post bankruptcy, the auctioneer may have a reason to switch to the other auction type.

The literature only partially answers our research question. In theory, in settings with (stochastic) private values, the probability of bankruptcy in second-price auctions is higher than in first-price auctions (Parlane [2003] Engel and Wambach [2006] Board [2007]). The intuition is the following. Bidders like taking risks if they are limitedly liable because they are not hurt as much by the downside risk as bidders with sufficient resources. Because the dispersion of the equilibrium price in second-price auctions is larger than in first-price auctions, bidders are willing to bid higher in second-price auctions. As a consequence, it is more likely that bankruptcies arise in second-price auctions than in first-price auctions.

Common value auctions with limitedly liable bidders have hardly been studied theoretically. For settings with unlimited liability, it is well known that in common value auctions, second-price auctions result in higher equilibrium prices than first-price auctions (Milgrom and Weber [1982]). Therefore, second-price auctions may be more sensitive to bankruptcy. However, in some settings such as ours, bidders can take into account information contained in others’ bids in second-price auctions but not in first-price auctions. So, if this information relates to the value of the object, bidders may bid cautiously in case of “bad news” resulting in a low probability of bankruptcy. Therefore, second-price auctions may perform better than first-price auctions in terms of bankruptcy.

Our study relates to the experimental literature on common value auctions and the winner’s curse. Levin, Kagel, and Richard (1996) find that the first-price sealed-bid auction (FP) and the English auction (EN) do not differ systematically in terms of average revenue unless the uncertainty about the common value is relatively small. Although their experimental design was not aimed at studying limited liability, it has some features of it. Subjects interacted in a

---

2 In practice, there are several mechanisms other than (standard) auctions that may perform well in terms of preventing bankrupt bidders, including the use of surety bonds (Calveras, Gamaon, and Hank [2004], multisourcing (Engel and Wambach [2006]), and the “average bid auction” (Decarolis [2010]). Burguet, Gamaon, and Hank [2009] study expected cost minimizing procurement auctions for settings with limitedly liable contractors.


series of auctions. Profits were added to and losses were subtracted from their starting capital. When their cash balance was exhausted, they were declared bankrupt and they had to leave the experiment. It turned out that some students indeed went bankrupt.

Roelofs (2002) and Saral (2009) study the effect of limited liability on bidding behavior in the laboratory. Roelofs observes that in the first-price sealed-bid auction, bidders increase their bid if default is possible compared to a situation where it is not. Saral analyzes bidding in second-price auctions under unlimited liability and two types of limited liability: market-based limited liability (inter-bidder resale following the auction) and statutory limited liability (a bidder pays a penalty if she makes a loss). She finds that bids are lower under unlimited liability than under market-based limited liability and statutory limited liability with a low default penalty. In the case of a high default penalty, the average bid does not differ between statutory limited liability and unlimited liability. Neither Roelofs nor Saral study the relative performance of standard auctions, which is the target of our study.

We examine bidding under limited liability in FP and EN. We do so in a laboratory experiment in an independent private signals common-value setting. In Sections 5.2 and 5.3 we present our experimental design and hypotheses. Our model is a three-bidder wallet game (Klemperer 1998). Subjects are limitedly liable in the same way as in Saral’s (2009) statutory limited liability regime. In our design, subjects always go bankrupt if they win the auction for a price exceeding the object’s value. In the case of bankruptcy, subjects do not leave the experiment, but they incur some bankruptcy costs which they have to cover from their starting capital. This set-up makes it relatively easy to derive the Nash equilibria and construct hypotheses on the basis of those. We show that EN has a symmetric equilibrium in which none of the bidders goes bankrupt. The equilibrium of FP is analytically not solvable, but we numerically derive that bidders bid more aggressively than in EN resulting in a strictly positive probability of bankruptcy.

Section 5.4 contains our experimental results. We observe that in both auctions, subjects bid more aggressively and, in turn, go bankrupt more often than predicted by theory. Moreover, bidders do not bid more aggressively and do not go bankrupt more frequently in FP than in EN. These results remain valid when comparing the experimental outcomes with the outcomes in settings in which subjects had to cover their losses.

In Section 5.5 we check whether our data are consistent with risk aversion, asymmetric equilibria, and Eyster and Rabin’s (2005) χ-cursedness. We argue that χ-cursedness gives a robust explanation of where our experimental observations differ from our initial theoretical results, in contrast to risk aversion and asymmetric equilibria. Section 5.6 concludes.
Table 5.1.: Summary of Treatments

<table>
<thead>
<tr>
<th>Auction</th>
<th>Order of Liability Regimes</th>
<th># Sessions</th>
<th># matching groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>EN</td>
<td>ULUL</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>LULU</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>FP</td>
<td>ULUL</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>LULU</td>
<td>2</td>
<td>6</td>
</tr>
</tbody>
</table>

Notes: U [L] stands for unlimited liability [limited liability]

5.2. Experimental Design and Procedures

We ran our experiment at the Center for Research in Experimental Economics and political Decision making (CREED) at the University of Amsterdam. From the student population, 144 undergraduates were publicly recruited and split into 4 groups of 36 students, one group for each treatment. Each session consisted of 4 parts of 12 rounds. Subjects read the computerized instructions at the start of each part. Test questions were included in the instructions of parts 1 and 2 to check the subjects’ understanding of the instructions. As parts 3 and 4 were equal to parts 1 and 2 respectively, we did not ask test questions for those parts.

Each session took about 2 hours and participants earned on average €19.28 (with a minimum of €7.24 and a maximum of €33.14). Earnings were denoted in experimental “francs”, having an exchange rate of 100 francs for €3.50. The experiment and the instructions were programed within the AJAX framework in JavaScript and PHP Script.

Two treatments consisted of English auctions and two consisted of first-price sealed-bid auctions. All sessions alternated with 2 parts in which participants were limitedly liable and 2 parts where they were unlimitedly liable. We included rounds with unlimited liability so that we could identify the effect of limiting liability on bidding behavior. Subjects were given a starting capital of 50 [150] francs before the beginning of each part in the case of [un]limited liability. To control for order effects, we ran the parts in half of the treatments in an ULUL sequence (unlimited, limited, unlimited, and limited) and the other half in a LULU sequence. The first two parts of every session were meant to give the participants the opportunity to gain experience. For the duration of each session, the group of participants was randomly split into fixed matching groups of 6, out of which for all rounds, 2 bidding groups of 3 bidders each were randomly chosen by the software. Table 5.1 gives an overview of the four treatments.

The subjects interacted in the three-bidder wallet game [Klemperer 1998]. Before the auction, the three bidders \(i \in \{1, 2, 3\}\) were each presented with a private signal \(\theta_i\), randomly and independently drawn from a uniform distribution on \([0, 100]\). We kept draws constant across treatments for the sake of comparability of the results. The value of the object was the sum of the three private signals:

\[
v = \theta_1 + \theta_2 + \theta_3.
\]

\[\text{For the instructions, see Appendix E} \]

64
In FP, subjects independently entered a bid between 0 and 300. The highest bidder won and paid a price equal to his own bid. EN consisted of two phases. In phase 1, the price started at zero and was increased by one every 1/6th of a second. The first phase ended as soon as a subject quit the auction by pressing a “stop” button. Before the start of the second phase, the other participants were informed that one of the bidders stepped out and the level of her bid. After 5 seconds, the price was increased again until one of the two remaining bidders dropped out. The remaining bidder won the object for the price at which the second-highest bidder quit. To mirror the maximum price of 300 in FP, we let all bidders automatically step out at a price of 300 if they had not quit beforehand. In both auctions, ties were resolved randomly. Between rounds, subjects were informed about the true value of the object, the winning bid, but not about the signals of others.

The payoffs for each round were as follows. In the limited liability regime, bidder $i$’s utility is given by

$$U^\ell_i(v, p, w) = \begin{cases} 
  v - p & \text{if } w = i \text{ and } v \geq p \\
  -c & \text{if } w = i \text{ and } v < p \\
  0 & \text{if } w \neq i
\end{cases}$$

(5.2)

where $w \in \{1, 2, 3\}$ denotes the winner of the auction, $p$ the price the winner pays, and $c > 0$ bankruptcy costs. In the experiment, $c = 4$. Note that the 50 francs endowment at the start of each part of 12 rounds ensured that subjects always obtained positive earnings. This model captures a situation where the winning bidder goes bankrupt if she makes a loss, in which case she incurs some (fixed) bankruptcy costs instead of the loss. Notice that these costs can be higher than the loss. For example, if the price exceeds the value by 3, the incurred loss equals 4 instead of 3.

In the unlimited liability regime, payoffs are

$$U^\infty_i(v, p, w, s) = \begin{cases} 
  \max(v - p, -s) & \text{if } w = i \\
  0 & \text{if } w \neq i
\end{cases}$$

(5.3)

where $s$ denotes the total score of the participant $i$ before the start of that round, i.e., the payoffs in this part up to the current round including the initial endowment in this part. Therefore, under the unlimited liability regime the total score of a participant could also never become negative. By choosing the 150 francs endowment, we feel that we found a good balance between mimicking a setting with truly unlimited liability (which requires an extremely high starting capital) and giving subjects sufficient incentives to earn money on top of the endowment (which favors a low starting capital)\footnote{Bankruptcy costs may refer to the bidder losing her job, reputation damage, legal costs, and so forth.}

\footnote{In parts 3 and 4, 3 out of the 144 participants did not have to cover all losses in at least one round because the accumulated losses would otherwise exceed their endowment. Of these participants, one took part in FP and two in EN. The fact that subjects did not have to cover losses above their endowment may have induced them to bid more aggressively relative to a setting with truly unlimited liability. Note that this is unfavorable to our hypothesis that bidders bid at least as aggressively under limited liability as under unlimited liability.}
5.3. Hypotheses

The equilibrium strategies for risk-neutral bidders can be straightforwardly derived from the literature. The symmetric Bayesian Nash equilibrium of EN with unlimited liability is given by

\[ B_1^E(\theta) = 3\theta; \quad B_2^E(\theta, \tilde{B}_1^E) = 2\theta + \frac{\tilde{B}_1^E}{3} \]  

(5.4)

where \( B_1^E \) is the price at which a bidder steps out of the auction in phase \( \varphi = 1, 2 \) of the auction and \( \tilde{B}_1^E \) is the price at which the lowest bidder leaves the auction. It is readily verified that the winning bidder will always make a positive profit in equilibrium so that the equilibrium under unlimited liability is also an equilibrium in the case of limited liability. Let \( \theta^{(k)} \) denote the \( k \)th highest value from \( \{\theta_1, \theta_2, \theta_3\} \), \( k = 1, 2, 3 \). In equilibrium, the expected winning bid equals

\[ R_\infty^E = R_\infty^F = \mathbb{E}\left\{ B_2^E(\theta^{(2)}), B_1^E(\theta^{(3)}) \right\} = 125 \]  

(5.5)

where \( R_\infty^E \) \([R_\infty^F]\) is the expected winning bid of EN with unlimited \([limited]\) liability.

The unique equilibrium of FP with unlimited liability is given by

\[ B_F(\theta) = \frac{5}{3}\theta \]  

(5.6)

If bidders are unlimitedly liable, the expected winning bid in FP equals

\[ R_\infty^F = \mathbb{E}\left\{ B_F(\theta^{(1)}) \right\} = 125. \]  

(5.7)

Therefore, the expected winning bid in FP and EN is the same, which is not surprising in view of Myerson’s (1981) revenue equivalence theorem.

In FP, the winner makes a loss with some probability because

\[ v - B_F(\theta^{(1)}) = \frac{2}{3}\theta^{(1)} + \theta^{(2)} + \theta^{(3)} < 0 \]  

(5.8)

for low values of \( \theta^{(2)} \) and \( \theta^{(3)} \). More specifically,

\[ \Pr\{v-B_F(\theta^{(1)}) < 0|\theta^{(1)} = \theta\} = \Pr\{\theta^{(2)} + \theta^{(3)} < \frac{2}{3}\theta^{(1)}|\theta^{(1)} = \theta\} = \Pr\{\theta_1 + \theta_2 < \frac{2}{3}\theta|\theta_1 < \theta_2\} = \frac{2}{9}. \]  

(5.9)

So, the probability that the winner makes a loss is independent of the winner’s signal, which makes sense because the signals for the second- and third-highest bidder are uniformly distributed between 0 and the highest signal. With respect to equilibrium bidding in FP in the case of limited liability, the wallet game is a special case of Milgrom and Weber’s (1982) affiliated signals model. Milgrom and Weber derive symmetric equilibria for the English auction and the first-price sealed-bid auction with unlimited liability. These equilibria are presented here. Equilibrium uniqueness follows from a standard argument (see e.g., Balow, Huang, and Klemperer 1999).

66
liability, we derive the following result.\textsuperscript{10}

**Proposition 5.1.** FP has a symmetric Bayesian Nash equilibrium which follows from the following differential equation:

\[
b_F' (\theta) = \frac{10\theta^2 - 4\theta b_F(\theta)}{\theta^2 + 2\theta b_F(\theta) - (b_F(\theta))^2 + 2c(b_F(\theta) - \theta)}
\]  

with boundary condition \( b_F(0) = 0 \).

Because the differential equation is not solvable analytically, we rely on the fourth order Runge-Kutta method to approximate a solution using signals starting at zero with increments of 0.01.\textsuperscript{11}

We find that if \( c = 4 \), expected winning bid in FP is approximately

\[
R_F^e \approx 137.
\]  

The probability that the winner makes a loss and goes bankrupt is around 34%. So, in the case of limited liability, both the expected winning bid and the probability of bankruptcy is higher in FP than in EN.

Comparing settings with limited and unlimited liability, we observe that the expected winning bid remains the same in EN, while it increases in FP. Moreover, according to theory, bidders never make losses in EN regardless of their liability. This is in contrast to FP, in which bidders make losses in both liability settings. In particular, winners are expected to go negative more often under limited liability than under unlimited liability. These results allow us to construct the following hypotheses related to our main research questions:

**Hypothesis 1** In the case of limited liability, the average winning bid in FP is higher than in EN. In FP, bidders incur losses more often than in EN.

**Hypothesis 2** For EN, limitation of liability increases neither the average winning bid nor the probability of overbidding.

**Hypothesis 3** For FP, limitation of liability increases both the average winning bid and the probability of overbidding.

### 5.4. Results

We present the results of our experiment in two sections. First, we deal with differences in winning bids and the presence of winners with negative payoffs between auctions. Second, we explore individual bidding behavior including learning and order effects.

\textsuperscript{10}We relegate proofs of propositions to Appendix F.

\textsuperscript{11}It is readily verified that if \( c = 0 \), the equilibrium bidding function is \( b_F(\theta) = 2\theta \). In this equilibrium, the probability that the winning bidder goes bankrupt is equal to 50% and expected winning bid equals 150.
5.4.1. Comparisons between Auctions

In this section, we focus on the aggregate results from parts 3 and 4, i.e., we only consider experienced bidders. The left panel of Figure 5.1 indicates that the average winning bid is higher under limited liability than under unlimited liability for both FP and EN. While this was expected for FP, our analysis predicted no difference for EN. Moreover, in the limited liability regime, the average winning bid in EN is higher than in FP, although the difference between auctions is smaller than the difference between liability regimes. This observation is also in contrast with our theoretical predictions that bidders bid more aggressively in FP than in EN in the case of limited liability.

When we aggregate the fraction of winners having negative payoffs (right panel, Figure 5.1), the above pattern is confirmed: There is a (slightly) higher frequency of negative payoff in EN than in FP and substantially more bankruptcies in the case of limited liability than losses in the case of unlimited liability. Furthermore, Figure 5.1 indicates a much higher than expected number of winners scoring a negative payoff\(^{12}\).

Table 5.2 compares the auction types with respect to the winning bid, the fraction of winners with a negative payoff, and the losses made for both liability regimes. The statistical tests are based on aggregate data per matching group. To make the losses made comparable for limited and unlimited liability regimes, we present for both the difference between the value of the object and the price of the object, ignoring the protection that limitation of liability would offer to bidders making a loss. We do not find support for the hypothesis that bidders protected by limited liability bid more aggressively in FP than in EN. On the contrary, EN generates significantly higher winning bids than FP and also the number of winners going bankrupt is higher, albeit not significantly so. Moreover, using a difference-in-difference approach, all differences are not significant. With respect to losses made, we cannot reject the hypothesis that these are the

\(^{12}\) On the basis of the drawn signals, we predict 0\% for the EN treatments and 8.3\% and 20.8\% for unlimited and limited liability respectively in the FP treatments. The realized fractions are clearly higher.
### Table 5.2: Comparisons between Auctions and Liability Regimes

<table>
<thead>
<tr>
<th>Variable</th>
<th>Liability</th>
<th>FP Nash (s.d.)</th>
<th>EN Nash (s.d.)</th>
<th>FP vs EN p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winning bid</td>
<td>Unlimited</td>
<td>120.8 (6.5)</td>
<td>130.0 (9.7)</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>Limited</td>
<td>132.4 (12.7)</td>
<td>130.0 (9.0)</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Diff-in-diff</td>
<td>11.6 (12.6)</td>
<td>20.6 (7.6)</td>
<td>0.25</td>
</tr>
<tr>
<td>%Losing</td>
<td>Unlimited</td>
<td>8.3% (8.3%)</td>
<td>0% (12.7%)</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>Limited</td>
<td>20.8% (9.4%)</td>
<td>0% (10.0%)</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>Diff-in-diff</td>
<td>12.5% (10.6%)</td>
<td>0% (13.8%)</td>
<td>0.33</td>
</tr>
<tr>
<td>Losses made</td>
<td>Unlimited</td>
<td>10.8 (7.4)</td>
<td>0% (8.5)</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>Limited</td>
<td>19.4 (11.2)</td>
<td>0% (7.5)</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>Diff-in-diff</td>
<td>8.6 (11.6)</td>
<td>0% (9.5)</td>
<td>0.95</td>
</tr>
</tbody>
</table>

*Notes: The Nash predictions here are based on the signals actually drawn for the participants, the unit of observation is the average per matching group, %Losing refers to the fraction of winners with negative payoffs, Losses Made are the average losses when the winner has a negative payoff, Diff-in-diff is the outcome of the difference for the auction type between the limited and unlimited regime, and s.d. stands for standard deviation. The p-values emerge from the Mann-Whitney test.*

same for the two types of auction, both on the level of the liability regimes and with respect to the difference between regimes. Finally, looking between liability regimes, for both auctions, we find a significantly higher winning bid and fraction of winners making a loss under the limited liability regime than under the unlimited liability regime.

### 5.4.2. Individual Behavior

In this section, we study subjects’ individual bidding behavior, which serves as a stepping stone to our analysis in Section 5.5 in which we try to unravel why observed behavior differs from the theoretical predictions. The importance of a close look at individual behavior is indicated by the simple fact that on average only in between 60% and 70% of the cases, does the bidder with the highest signal win\(^{13}\), which is highly contrasting to our theoretical prediction that in equilibrium, all participants bid according to the same bid function that is monotonically increasing in their signal.

To examine bidding behavior in greater detail, we estimated a random effects model with a

\(^{13}\) To be more specific, in the case of unlimited liability, 70% [62%] of the winners in FP and 64% [63%] of the winners in EN has the highest signal.
clustering specification to get robust p-values. We estimated three bidding functions: $B_{ijt}^F$ for bidders in FP, $B_{ijt}^E_1$ and $B_{ijt}^E_2$ for the first [second] bidder to step out in EN, where $ijt$ indicates bidder $i$ in matching group $j$ in round $t$:

$$B_{ijt}^F = \beta_F^i + \beta_F^\theta \theta_{ijt} + \beta_F^L L_{ijt} + \beta_F^\theta L \theta_{ijt} + \beta_F^\theta X \theta_{ijt} + \alpha_F^i + \epsilon_{ijt},$$  

(5.12)

$$B_{ijt}^E_1 = \beta_{E_1}^i + \beta_{E_1}^\theta \theta_{ijt} + \beta_{E_1}^L L_{ijt} + \beta_{E_1}^\theta L \theta_{ijt} + \beta_{E_1}^\theta X \theta_{ijt} + \alpha_{E_1}^i + \epsilon_{ijt},$$  

(5.13)

$$B_{ijt}^E_2 = \beta_{E_2}^i + \beta_{E_2}^\theta \theta_{ijt} + \beta_{E_2}^L L_{ijt} + \beta_{E_2}^\theta L \theta_{ijt} + \beta_{E_2}^\theta X \theta_{ijt} + \alpha_{E_2}^i + \epsilon_{ijt},$$  

(5.14)

where $L$ is a dummy that equals 1 if and only if liability is limited, $Lulu$ is a dummy which is equal to 1 if and only if subjects play the LULU sequence, $X$ is a dummy referring to a subject’s experience (1 for parts 3 and 4), and $\tilde{B}_{E_1}$ denotes the price at which the first bidder stepped out in EN. The $\beta$’s are the parameters of the model.

### Table 5.3: Estimated Bidding Functions (5.12)-(5.14)

<table>
<thead>
<tr>
<th></th>
<th>FP</th>
<th>EN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bid</td>
<td>Lowest bid</td>
</tr>
<tr>
<td></td>
<td>Coef (s.e.)</td>
<td>Coef (s.e.)</td>
</tr>
<tr>
<td>Constant</td>
<td>58.76 (4.33)**</td>
<td>73.80 (4.69)**</td>
</tr>
<tr>
<td>Signal ($\theta$)</td>
<td>0.95 (0.06)**</td>
<td>0.76 (0.13)**</td>
</tr>
<tr>
<td>Lowest bid ($\tilde{B}_{E_1}$)</td>
<td>16.83 (4.51)**</td>
<td>12.73 (6.37)*</td>
</tr>
<tr>
<td>Signal $\times$ Limited liability ($\theta L$)</td>
<td>-0.06 (0.08)</td>
<td>-0.00 (0.18)</td>
</tr>
<tr>
<td>LULU</td>
<td>-4.28 (6.27)</td>
<td>8.78 (7.99)</td>
</tr>
<tr>
<td>Signal $\times$ LULU ($\theta Lulu$)</td>
<td>0.07 (0.74)</td>
<td>0.05 (0.15)</td>
</tr>
<tr>
<td>Experienced ($X$)</td>
<td>-1.23 (3.62)</td>
<td>0.23 (2.92)</td>
</tr>
<tr>
<td>Signal $\times$ Experienced ($\theta X$)</td>
<td>0.11 (0.05)*</td>
<td>0.43 (0.09)**</td>
</tr>
</tbody>
</table>

Notes: ** [*] indicates statistical significance at the 1% [5%] level, and s.e. stands for (robust) standard error.

Table 5.3 contains statistical significance at the 1% [5%] level, and s.e. stands for (robust) standard error.

---

Those differences are statistically significant according to Wald tests.
bid, the opposite holds true: a lower constant and a lower slope for EN than for FP. The reason can be seen in the regression for the highest bid where participants react strongly to the level at which the first bidder stepped out. Bidding turns out to be quite aggressive in phase 1 of the auction, while in phase 2, bidders step out relatively quickly. Subjects behave as though they can always safely step out of the auction in the second phase of EN. Still, bidders use the information contained in the behavior of the first bidder in that the earlier another bidder steps out in the first phase, the earlier they quit in the second phase.

In the regression, we added the last four variables in Table 5.3 on the preceding page to control for order effects and learning. This turned out not to change the significance and direction of the other coefficients. We do not observe order effects, but there seems to be some learning. In FP, bidders adapt their bidding behavior, albeit in the wrong direction: In parts 3 and 4, they bid more aggressively than in the first two parts, overbidding even more relative to the Nash equilibrium. For EN, we observe experienced bidders letting their bids depend more on their signal than inexperienced ones. However, given that the expected second-highest signal equals 50, the net effect of experience on the average winning bid is minimal.

5.5. Explanation of the Main Results

In this section, we attempt to explain the differences between our data and the theoretical predictions. In particular, in both auctions and under both liability regimes, bidders tend to overbid relative to the Nash equilibrium. Moreover, we reject the hypothesis that in the case of limited liability, bidding is more aggressive in FP than in EN. We explore risk aversion, asymmetric equilibria, and χ-cursedness as potential explanations.
5.5.1. Risk Aversion

To which extent is our data consistent with equilibrium bidding for risk-averse bidders? Suppose that all three bidders have the same common utility function $u$, where $u$ is differentiable, strictly increasing, and strictly concave, with $u(0) = 0$. In EN, equilibrium bidding is not affected by bidders’ risk attitudes: In both phases of the auction, bidders drop out at the price at which their payoff would be zero if the remaining competitor(s) dropped out at that price. In FP, the effect of risk aversion is not clear a priori. In the standard symmetric independent private values model, risk-averse bidders bid more aggressively than risk-neutral ones (Maskin and Riley, 1984). However, in the case of a common value, from a bidder’s viewpoint, the object’s value is stochastic because she does not know the signals of the other bidders. This tends to drive down bids. Holt and Sherman (2000) show that these two effects exactly cancel in a two-bidder wallet game. In equilibrium, risk-averse bidders bid as if they were risk-neutral. In the case of three bidders, intuitively, the second effect dominates the first: More competition drives up the price so that a risk-averse bidder has lower incentives to further increase her bid while she is more inclined to shade the risk-neutral equilibrium bid because she has less information about the common value. The following proposition confirms this intuition.

**Proposition 5.2.** In the case of unlimited liability, for risk-averse bidders, the symmetric Bayesian Nash equilibrium of FP has the property that

$$B_{F}^{r} (\theta) < \frac{5}{3} \theta = B_{F}(\theta).$$

(5.15)

All in all, risk aversion does not seem to be the (sole) reason why subjects tend to overbid in either auction.

5.5.2. Asymmetric Equilibria

Alternatively, subjects may have played different equilibria than the above symmetric equilibria. However, for FP this cannot be the case as the symmetric equilibrium is the unique equilibrium. In contrast, EN has a continuum of asymmetric equilibria as the following proposition by Engelmann and Wolfstetter (2009) shows.

**Proposition 5.3.** In the case of unlimited liability, EN has the following equilibria:

$$B_{E,i}^{1} (\theta) = \gamma_{i} \theta; B_{E,i}^{2} (\theta, \tilde{B}_{E,k}^{1}, k) = \delta_{i} \theta + \frac{\tilde{B}_{E,k}^{1}}{\gamma_{k}},$$

(5.16)

where $B_{E,i}^{k} (\theta) / [B_{E,i}^{k} (\theta, \tilde{B}_{E,k}^{1}, k)]$ denotes the price at which bidder steps out when no one bidder $k \in \{1, 2, 3\} \backslash \{i\}$ has stepped out [at price $\tilde{B}_{E}^{1}$], $i = 1, 2, 3$, and

$$\gamma_{i}, \delta_{i} > 0, \quad \gamma_{1} \gamma_{2} > \gamma_{1} + \gamma_{2}; \quad \gamma_{3} = \frac{\gamma_{1} \gamma_{2}}{\gamma_{1} \gamma_{2} - \gamma_{1} - \gamma_{2}}.$$  

(5.17)
\[ \delta_m = \frac{\delta_n}{1 - \delta_n}, \{m, n\} = \{1, 2, 3\} \backslash \{k\}. \quad (5.18) \]

**Corollary 5.1.** The expected winning bid in the symmetric equilibrium (Equilibrium bid English unlimited liability) of EN is at least as high as in any of the equilibria in Proposition Asymmetric EN.

The asymmetric equilibria of EN share two properties that are inconsistent with our data. First, the equilibrium price is always below the value of the object so that bidders never make a loss. This implies that the above strategies are also an equilibrium for a setting with limited liability. In other words, asymmetric equilibria cannot explain why bidders bid more aggressively in the case of limited liability compared to the case of unlimited liability. Second, the expected winning bid in the asymmetric equilibria is always lower than in the symmetric one. This is clearly inconsistent with our observation in the experiment, that the average winning bid is much higher than in the symmetric equilibrium.

Also the explanation that subjects miscoordinate on an asymmetric equilibrium does not seem appealing. Clearly, an asymmetric equilibrium requires bidders to coordinate as to who bids aggressively and who does not. However, we did not find evidence that bidders adapted their strategies over time in the direction of an asymmetric equilibrium. Moreover, even in the case of miscoordination, the first-phase bidding functions should have a zero constant, which we clearly rejected when estimating bidding functions in Section 5.4.

We conclude that our data cannot be (solely) explained by bidders playing asymmetric equilibria.

### 5.5.3. Cursed Bidders

Finally, subjects may have behaved as “cursed” bidders in line with Eyster and Rabin’s (2005) \( \chi \)-cursed equilibrium. We start by deriving the \( \chi \)-cursed equilibrium for the two auctions if bidders are unlimitedly liable.

**Proposition 5.4.** The symmetric \( \chi \)-cursed equilibrium of EN with unlimited liability is given by

\[
B^1_E(\theta) = 100 \chi + (3 - 2 \chi) \theta; \quad B^2_E(\theta, \bar{B}^1_E) = \left(2 \theta + \frac{\bar{B}^1_E - 100 \chi}{3 - 2 \chi}\right) (1 - \chi) + (\theta + 100) \chi. \quad (5.19)
\]

**Proposition 5.5.** The symmetric \( \chi \)-cursed equilibrium of FP with unlimited liability is given by

\[
B^E_F(\theta) = 100 \chi + \left(\frac{5}{3} - \chi\right) \theta. \quad (5.20)
\]

The following corollary shows that the expected winning bid for the seller is the same for both auctions, given that all bidders possess the same level of \( \chi \)-cursedness.
Corollary 5.2. In the case of unlimited liability, if bidders play the symmetric $\chi$-cursed equilibrium, FP and EN generate the same expected winning bid, which equals

$$R_F^{\infty, \chi} = R_E^{\infty, \chi} = 125 + 25\chi. \quad (5.21)$$

The estimated coefficients for the bidding function for FP in Table 5.3 on page 70 indicate that on aggregate, bidding strategies correspond to an average $\chi$-cursedness level of about 0.65. For EN, the estimated bidding functions are less appropriate to estimate the average $\chi$ because we only observe the lowest two bids. The average winning bid for EN produces a better approximation for the average $\chi$ because the bid in the middle determines the winning bid. Using this, the average $\chi$ is about 0.87. Eyster and Rabin (2005) find that the average $\chi$-cursedness level for experienced subjects in Avery and Kagel’s (1997) experiment on the two-bidder wallet game equals 0.64. Our estimates seem reasonably close to that. Moreover, subjects may differ in the level of $\chi$-cursedness, which could explain the observation that it is not always the bidder with the highest signal who wins. The difference in estimated average $\chi$-cursedness level between EN and FP may be explained by “auction fever”. To some extent, cursed bidders compete as if bidding in a setting with uncertain private values. In a lab experiment, Elrhart, Ott, and Abele (2008) show that in an environment with uncertain private values, bidders tend to be affected by auction fever in that they bid higher in ascending auctions than in strategically equivalent sealed-bid auctions.

For the limited liability setting, our data reject the theoretical prediction that FP yields more aggressive bidding and more bankruptcies than EN. Cursedness could offer an explanation here as well. Fully cursed bidders (for whom $\chi = 1$) experience the auction as a pure private value auction because they do not take into account that the fact of winning impacts the expected value for the object. As is well known for (stochastic) private value auctions, in the case of limited liability, the expected winning bid is higher and the winner is more likely to go bankrupt in EN than in FP (Parlane, 2003; Engel and Wambach, 2006; Board, 2007). This result also holds true in our setting as the propositions below show. Define

$$\bar{U}(p, \theta_1) \equiv \mathbb{E}_{\theta_2, \theta_3} \{\max(0, v - p)\} - c\mathbb{P}\{v < p\} \quad (5.22)$$

as the perceived expected utility of a 1-cursed bidder with signal $\theta_1$ when winning at price $p$.

Proposition 5.6. In the case of limited liability, in the symmetric $\chi$-cursed equilibrium of EN, a bidder with signal $\theta$ steps out at $b_E^{\chi-1}(\theta)$ which is implicitly defined by

$$\bar{U}(b_E^{\chi-1}(\theta), \theta) = 0. \quad (5.23)$$

To solve for the bidding function, assume that $b_E^{\chi-1}(\theta_1) > 100 + \theta_1$ for all $\theta_1 \in [0, 100]$. Bidder 1 solves

$$\frac{1}{6,000,000} (200 - p + \theta_1)^3 - \frac{c}{10,000} \left[10,000 - \frac{1}{2} (200 - p + \theta_1)^2\right] = 0. \quad (5.24)$$
The first term on the left-hand side refers to the situation in which bidder 1 does not go bankrupt. The resulting bidding function is approximately

\[ b^{\chi=1}_E(\theta) \approx \theta + 200 - \sqrt{60,000c} + \sqrt{\frac{c^2}{60,000}} + c \approx 141.9 + \theta. \] (5.25)

Indeed, \( b^{\chi=1}_E(\theta_1) > 100 + \theta_1 \), like we assumed. The corresponding expected winning bid equals

\[ R^{\ell,\chi=1}_E \approx 191.9. \] (5.26)

Proposition 5.7. In the case of limited liability, the symmetric 1-cursed equilibrium of FP follows from the following differential equation:

\[ b^{\chi=1}_F(\theta) = -2 \frac{\bar{U}(b^{\chi=1}_E(\theta), \theta)}{\bar{U}_1(b^{\chi=1}_E(\theta), \theta)} \] (5.27)

with boundary condition

\[ \bar{U}(b^{\chi=1}_E(0), 0) = 0. \] (5.28)

Numerically, we derive that the expected winning bid equals approximately

\[ R^{\ell,\chi=1}_F \approx 188.1, \] (5.29)

which is less than in EN. Indeed, the ranking between auctions in terms of expected winning bid reverses for fully cursed bidders compared to a setting with fully rational bidders, like we observe in our data. The following corollary formalizes this result.

Corollary 5.3. In the case of limited liability, in the symmetric 1-cursed equilibrium, the average winning bid in EN is higher than in FP.

Note that for FP, the observed average winning bid is roughly in the middle between the theoretical predictions for fully rational and fully cursed bidders. For EN, the observed winning bid is closer to the prediction for fully cursed bidders than the one for rational bidders. This observation is in line with the higher estimated \( \chi \)-cursedness level in the case of unlimited liability for EN than for FP, which may be explained by auction fever as in Elberth, Ott, and Abele (2008).

To summarize, \( \chi \)-cursedness explains our experimental observations quite well, at least on the aggregate level.\[^{15}\]

\[^{15}\] Obviously, it could be the case behavior is explained by a mixture of \( \chi \)-cursedness, risk aversion, and asymmetric equilibria.
5.6. Conclusion

In a laboratory experiment, we have studied which standard auction is least conducive to bankruptcy. More precisely, we have analyzed the first-price sealed-bid auction and the English auction in a common value context. Our data strongly reject our theoretical prediction that the English auction leads to less aggressive bids and fewer bankruptcies than the first-price sealed-bid auction. In particular, we observe no statistical difference between the two auctions in terms of bankruptcy. Our results suggest that for license auctions and procurement procedures, it will not be helpful for governments to run a second-price auction instead of a first-price auction (or the other way around) if they wish to mitigate the likelihood of bidders going bankrupt.
Bibliography


A. Literature on the Boiling Frog Story

Currently, the correctness of the boiling frog story is questioned (Gibbons, 2002). On the basis of their work with other animals, contemporary zoologists think that frogs will try to escape irrespective of whether the heating occurs instantaneously or gradually. There is, however, some tension between the current view and the 19th century investigations where frogs were actually heated in experiments.

Goltz (1869, p. 127-130) describes an experiment with two frogs, one decapitated frog and one normal frog. Goltz immersed the frogs in water leaving out only a small part of the frog. He raised the temperature of the water in about ten minutes from 17.5° C to 56° C. From a temperature of 25° C the healthy frog tried to escape from the water and died a terrible death at 42° C because the experimental setup did not allow the frog to get away. The decapitated frog scarcely moved until the temperature reached 56° C when it made some spastic movements.

Notice that Goltz heated the frogs rather quickly. In fact, his aim was not to test the boiling frog phenomenon. Instead, he wanted to find the location of the frog’s soul. Because he believed that it was seated in the brain, he wanted to find differences between how a brainless frog and a healthy frog reacted to being boiled and therefore he chose to heat the frogs quickly.

Heinzmann (1872) reports on experiments with in total 27 frogs. He set out to work with decapitated and brain damaged frogs. After his first trials where he heated the frogs locally with a leg in the water, he moved to a setup where the frog was seated on a cork floating in a cylinder of water. He heated the frogs in about 90 minutes from a temperature of about 21° C to about 37.5° C. (So he stopped short of literally frying the frogs because some pre-trials had convinced him that from 37.5° C the frogs became paralyzed until death entered). After thus fine-tuning the experiment, he continued to work with normal undamaged frogs. In his 12th trial he managed for the first time to heat a healthy frog from 23° C to 39° C without any movement of the frog, even though the frog could jump away from the setup at any moment if it wanted to. Two of the next three trials were successful repetitions of the 12th trial. Then Heinzmann set out to reach the opposite goal, that is, to gradually freeze frogs without a movement, and again, after some initial trials where he used damaged frogs he managed to accomplish his goal.

1In personal communication, Dr. Victor Hutchinson, a Research Professor Emeritus from the University of Oklahoma’s Department of Zoology, whose research interests include physiological ecology of thermal relations of amphibians and reptiles, formulated the current skepticism as follows: “It [the boiling frog story] makes a nice story, but it really is a myth. In fact, most animals, vertebrate and invertebrate (all we have tested) exposed to increasing heat respond similarly — they attempt to escape noxious conditions (chemicals, etc.) that could lead to their death. This is an expected survival response as logic might indicate.”

2Goltz mentions a third decapitated frog that he does not boil and that serves as a control frog for the decapitated frog that is boiled.
with healthy frogs.

Unaware of the study by Heinzmann, Foster (1873) confirmed Goltz’s finding that uninjured frogs become violent in their attempts to escape when the temperature is heated above 30° C. Foster carried out trials where he heated the water slowly and trials where he heated the water quickly. Unfortunately, he does not describe how fast he heated the water. The paper of Foster is mainly dedicated to explaining why Goltz’s decapitated frog did not respond to the heating, a finding that Foster found puzzling.

Hall and Motora (1887) mention that Fratscher (1875) successfully repeated Heinzmann’s results. Fratscher even succeeded in inducing rigor mortis in normal frogs by immersing only a small part of frog in the fluid. Sedgwick (1883) at Johns Hopkins is the person with an overview of the entire literature on the heating of frogs up until 1882. His intuition was that the variance in the speed of the heating explains the difference in Goltz’s and Foster’s results and Heinzmann’s and Fratscher’s results. In agreement with his intuition, he reports that he was able to replicate all previous results by varying the speed of the heating process. At the end of the 19th century, the consensus is that it is possible to boil frogs without movement if it is done sufficiently slowly.

A related question is whether rapid heating induces frogs to try to escape at lower temperatures than slow heating. Foster mentions this possibility, but says that he did not pay attention to this issue while he did his experiments. Arguably, this “lite version” of the boiling frog story is the more relevant one for Al Gore’s analogy. As far as we know, the lite version of the story has not been tested with frogs, but there are some physiological studies with humans showing that the smallest perceptible change in weight of an object placed on the fingertip varies with the speed of the change in weight. (Hall and Motora 1887; Schripture 1897).
B. Instructions “How to Subsidize Contributions to Public Goods”

This appendix contains the instructions that were presented on-screen to the participants. For the six treatments we only needed three different sets of instructions. In the treatments gradual-45, quick-45, gradual-75, and quick 75 subjects received exactly the same instructions. These treatments only differed in the way the subsidy was changed during the experiment.

Instructions treatments: gradual-45, quick-45, gradual-75, and quick 75

Instructions

Welcome to this experiment. Please read the following instructions with care. If something is not clear, raise your hand and we will help you. After everyone has finished reading the instructions and before the experiment starts, you will receive a handout with a summary of the instructions. You can use this handout throughout the experiment.

You will be asked to make a number of decisions. The experiment consists of two parts. Below this section, you will find the instructions for the first part. After part 1 has been completed you will receive instructions for the second part. Your decisions and the decisions of other participants will determine how much money you earn.

During the experiment, your earnings will be denoted in points. Your earnings in the experiment will be equal to the sum of your earnings in part 1 and in part 2. At the end of the experiment, your earnings (in points) will be converted into money. For each 18 points you earn, you receive 1 eurocent. Hence, 1800 points are equal to 1 euro. Your earnings will be privately paid to you in cash.

Part 1

In part 1, you will earn money with two different tasks. One task is an individual task and the other is a group task. You will perform both of them at the same time. The individual task will be on the left side of your screen and the group task on the right side. You will earn points
for both tasks simultaneously. Your earnings for the individual task do not depend on your actions in the group task and your earnings for the group task do not depend on your actions in the individual task. Part 1 will last between 25 and 45 minutes. Both tasks will stop at the same time. The computer will inform you when part 1 is finished. Although both tasks run at the same time, it is up to you to decide how much time you want to spend on each task. You can switch between the tasks whenever you want.

**Individual task**

In the individual task, you will earn points by keeping a randomly moving red dot inside a box. In the big window on the left side of the screen you will see a red dot making random movements. The dot starts inside the box, and your task is to keep it inside that box by moving the box. You can move the box by pressing (with your mouse) on one of the four arrow buttons above the white field. The box will move in the same direction as the direction of the arrow (up, down, left, right).

At the end of every second the computer determines whether the dot is inside or outside the box. If it is inside the box you will receive 15 points, if it is outside you will receive 0 points for that second. You start with zero points and your earnings for this task equal the sum of earnings in all seconds. While you perform the individual task, your total earnings for this task will be listed in the upper left part of the screen.

**Group task**

You are randomly assigned to a group of 6 participants (including yourself). Throughout this task, you will remain in this group of 6 persons. For the group task each participant will decide about how much to contribute to the group. Your earnings for this task depend on your own decisions as well as on the decisions of the other participants in your group.

For each second, the computer calculates how many points you get for that second and these points are added to the total for the group task. Your earnings for every second depend on the endowment you get every second, your contribution to the group in that second, the contributions that the others in your group make in that second and the level of the subsidy in that second.

Each group-member receives an endowment of 10 points in every second. In the beginning, each group-member decides how much to contribute to the group (a contribution equals at least 0 points and at most 10 points). In each subsequent second, each group-member may change the own contribution. If a group-member does not change the contribution, this person’s contribution equals the contribution that he or she made in the previous second.
Contributing to the group has two effects on your payoff: a benefit effect and a cost effect. We will first deal with the benefit effect. Your contribution benefits yourself and the other members of your group in the following way. Every second, the computer adds up all contributions made in your group and multiplies the sum with 1.2. The resulting number of points is equally divided between the 6 group-members. This means that in each second you will receive 0.2 point for each point contributed to the group.

Now we deal with the cost effect of your contribution. Contributing points to the group is costly for you. Your contribution will be subsidized though, which means that part of the money that you spend on contributing is returned to you. The higher the subsidy, the less you actually pay for your contribution. In this sense, the subsidy determines how costly your contribution is. The subsidy denotes the part of your contribution that you do not have to pay. For instance, if the subsidy is 0.000, each point that you will contribute to the group will cost you 1 point. If the subsidy is 0.250, each point that you will contribute to the group will cost you 0.750 point, if the subsidy is 0.500, each point that you will contribute to the group will cost you 0.500 point, etc.

The subsidy may change during the experiment. It is at least 0.000 and at most 0.800. Whether it changes or not is outside of your control. All participants in the group face the same subsidy. All participants will be clearly informed when and how the subsidy changes. AT THE START OF THE EXPERIMENT, THE SUBSIDY EQUALS 0.000.

Summarizing, in each second:
(i) costs of contributing = own contribution*(1-subsidy)
(ii) earnings group task = 10 - costs of contributing + 0.2*sum contributions

During part 1 group-members will NOT be informed about the contributions of the others in the group. There will also be no information about the earnings for the group task. This information will only be revealed at the end of part 2.

Making your decisions in part 1

Below you see a picture of the screen that will be used in part 1 to enter your decisions. On the left part of the screen you find the window used for the individual task. During the experiment the red dot will move randomly and your goal is to move the white box such that the red dot stays in the box. You move the white box by pressing the arrows above the window. On the right part of the screen you find the window used for the group task. In the gray area you see a slider. With that slider you will indicate how much you want to contribute to the group task. You can change your contribution by changing the position of that slider.
Above the slider you see the subsidy for that second. Each time the subsidy changes the background of the subsidy number turns red for a second.

On the next screen you will be requested to answer some control questions. Please answer these questions now.

**Instructions treatment: gradual-75-single**

**Instructions**

Welcome to this experiment. Please read the following instructions with care. If something is not clear, raise your hand and we will help you. After everyone has finished reading the instructions and before the experiment starts, you will receive a handout with a summary of the instructions. You can use this handout throughout the experiment.

You will be asked to make a number of decisions. The experiment consists of two parts. Below this section, you will find the instructions for the first part. After part 1 has been completed you will receive instructions for the second part. Your decisions and the decisions of other participants will determine how much money you earn.
During the experiment, your earnings will be denoted in points. Your earnings in the experiment will be equal to the sum of your earnings in part 1 and in part 2. At the end of the experiment, your earnings (in points) will be converted into money. For each 18 points you earn, you receive 1 eurocent. Hence, 1800 points are equal to 1 euro. Your earnings will be privately paid to you in cash.

Part 1

In part 1, you will earn money with two different tasks. One task is an individual task and the other is a group task. What is special about the individual task is that the computer forces you to make the same choices as a participant of a previous experiment. The individual task will be on the left side of your screen and the group task on the right side. You will earn points for both tasks simultaneously. Your actions only affect your earnings for the group task. Your earnings for the group task do not depend on the actions of the previous participant for the individual task. Part 1 will last between 25 and 45 minutes. Both tasks will stop at the same time. The computer will inform you when part 1 is finished.

Individual task

In the individual task, the previous participant earned points by keeping a randomly moving red dot inside a box. In the big window on the left side of the screen you will see a red dot making random movements. The dot starts inside the box, like it did for the previous participant. The previous participant's task was to keep it inside that box by moving the box. He or she could move the box by pressing (with the mouse) on one of the four arrow buttons above the white field. The box will move in the same direction as the direction of the arrow (up, down, left, right) pushed by the previous participant. You cannot influence this process.

At the end of every second the computer determines whether the dot is inside or outside the box. If it is inside the box you will receive 15 points (like the previous participant did), if it is outside you will receive 0 points for that second (again, like the previous participant did). You start with zero points and your earnings for this task equal the sum of earnings in all seconds. Your total earnings for this task will be listed in the upper left part of the screen.

Group task

You are randomly assigned to a group of 6 participants (including yourself). Throughout this task, you will remain in this group of 6 persons. For the group task each participant will decide about how much to contribute to the group. Your earnings for this task depend on your own
decisions as well as on the decisions of the other participants in your group.

For each second, the computer calculates how many points you get for that second and these points are added to the total for the group task. Your earnings for every second depend on the endowment you get every second, your contribution to the group in that second, the contributions that the others in your group make in that second and the level of the subsidy in that second.

Each group-member receives an endowment of 10 points in every second. In the beginning, each group-member decides how much to contribute to the group (a contribution equals at least 0 points and at most 10 points). In each subsequent second, each group-member may change the own contribution. If a group-member does not change the contribution, this person’s contribution equals the contribution that he or she made in the previous second.

Contributing to the group has two effects on your payoff: a benefit effect and a cost effect. We will first deal with the benefit effect. Your contribution benefits yourself and the other members of your group in the following way. Every second, the computer adds up all contributions made in your group and multiplies the sum with 1.2. The resulting number of points is equally divided between the 6 group-members. This means that in each second you will receive 0.2 point for each point contributed to the group.

Now we deal with the cost effect of your contribution. Contributing points to the group is costly for you. Your contribution will be subsidized though, which means that part of the money that you spend on contributing is returned to you. The higher the subsidy, the less you actually pay for your contribution. In this sense, the subsidy determines how costly your contribution is. The subsidy denotes the part of your contribution that you do not have to pay. For instance, if the subsidy is 0.000, each point that you will contribute to the group will cost you 1 point. If the subsidy is 0.250, each point that you will contribute to the group will cost you 0.750 point, if the subsidy is 0.500, each point that you will contribute to the group will cost you 0.500 point, etc.

The subsidy may change during the experiment. It is at least 0.000 and at most 0.800. Whether it changes or not is outside of your control. All participants in the group face the same subsidy. All participants will be clearly informed when and how the subsidy changes. AT THE START OF THE EXPERIMENT, THE SUBSIDY EQUALS 0.000.

Summarizing, in each second:

(i) costs of contributing = own contribution*(1-subsidy)

(ii) earnings group task = 10 – costs of contributing + 0.2*sum contributions

During part 1 group-members will NOT be informed about the contributions of the others in the
group. There will also be no information about the earnings for the group task. This information will only be revealed at the end of part 2.

Making your decisions in part 1

Below you see a picture of the screen that will be used in part 1 to enter your decisions for the group task. On the left part of the screen you find the window used for the individual task. During the experiment the red dot will move randomly and you will observe how the previous participant moved the white box. On the right part of the screen you find the window used for the group task. In the gray area you see a slider. With that slider you will indicate how much you want to contribute to the group task. You can change your contribution by changing the position of that slider.

Above the slider you see the subsidy for that second. Each time the subsidy changes the background of the subsidy number turns red for a second.

On the next screen you will be requested to answer some control questions. Please answer these questions now.
Instructions treatments: predict-75

Introduction

Welcome to this session. In this session we will ask you to state your beliefs about what has happened in a previous experiment. You will not carry out that experiment yourself, but we will ask you to state your beliefs about what participants did in that experiment. The closer your beliefs are to how the previous participants actually behaved, the more you will earn.

After you have finished stating your beliefs, we will ask you to make two other types of decisions that allow you to make additional money. You will earn points during this session. At the end of the session your points for all three types of decisions will be added up and combined with a starting capital of 8000 points. The resulting total number of points will be exchanged into euros at a rate of 1000 points is 1 euro. Only at the end of the session you will be informed how much you earned with each type of decisions. You will receive the instructions for a next part only when a previous part is finished.

On your table you will find a hardcopy of the instructions given to the participants in that previous experiment. During your session you are allowed to keep them. We want you to study these instructions now.

Now you have studied the instructions of the previous experiment, we will explain what you will be asked to do. You will be asked several times to state your probability judgment about certain statements. For each of three statements there will be five sub-questions. After you have finished all sub-questions, one of the fifteen sub-questions is chosen at random by the computer and your answer on that sub-question determines the points you earnings for this part.

During the group task in the previous experiment, contributing became less costly over time as a result of an increase of the subsidy. Participants of the previous experiment participated in one of the "gradual" groups or one of the "quick" groups. We will ask you some questions about how the participants of the two type of groups behaved.

1. In the "gradual" groups the subsidy started increasing after exactly 4 minutes. During 16 minutes and 40 seconds it was raised gradually until it reached 0.75 after exactly 20 minutes and 40 seconds. Then it stayed at 0.75 until the end of the task after exactly 28 minutes.

2. In the "quick" groups the subsidy was increased in one time from 0 to 0.75 after exactly 4 minutes and it stayed at 0.75 until the end of the task after exactly 28 minutes.

As you could see in the instructions of the previous experiment, participants only knew that
their experiment would start with a subsidy of 0 and that this subsidy could change during the experiment.

We will present you with 3 statements and ask you 5 sub-questions per statement. The statements refer to the handout with the figures that show how the subsidy changed in the “gradual” group and the “quick” group.

The statements and sub-questions we will ask you are:

1. For all participants, the subsidy started at the same level (see both figures of the handout). What is your probability judgment (in %) that at the START the average contribution of all participants was in the interval …

2. For the participants in the GRADUAL groups, the subsidy changed as indicated in in the lower figure of the handout. What is your probability judgment (in %) that at the END the average contribution of the participants in the GRADUAL groups was in the interval …

3. For the participants in the QUICK groups, the subsidy changed as indicated in the upper figure of the handout. What is your probability judgment (in %) that at the END the average contribution of the participants in the QUICK groups was in the interval …

These statements are presented one after another. For each statement you have to give your probability judgment for five intervals. Each time an interval is shown you choose a percentage.

The table that is handed out to you shows how much you earn for a particular probability judgment for an interval. The table contains three columns. The first column shows the percentage of your probability judgment, the second column displays your earnings if the real average contribution level (“the true value”) is in the interval and the third column shows what you get if it is not in the interval. You find your earnings by looking in the row that corresponds to your probability judgment and the column that corresponds to the real average contribution (second column if the average contribution is inside the interval, third column if it is outside the interval).

You will make your decision on the computer in the following way. After you have typed in your probability judgment, the computer will open the same table as the one handed out on paper. The row that corresponds with your chosen probability judgment is preselected. You can pick a different row in the table if you prefer to change your probability judgment. You can do this by selecting the up or down arrow, or by clicking the mouse in the menu and scroll to another probability judgment. Next, when you click on <confirm> your choice is final and you continue with the next statement. When you are finished, press <confirm> with your mouse.

After you pressed <confirm> with your mouse, you will be asked your probability judgment.
for the next interval. When you have provided judgments for all five intervals / sub-questions, you will continue to the next statement.

After you have completed the three statements with the fifteen sub-questions the computer randomly draws one of the fifteen sub-questions. Your answer together with the actual average contribution for the relevant sub-question determines your earnings for reporting your probability judgment.

On the next screen you will be requested to answer some control questions. Please answer these questions now.
C. Instructions “Inducing Good Behavior”

Introduction
This is an experiment about decision-making. In the room, there are ten people who are participating in this experiment. You must not communicate with any other participant in any way during the experiment. At the end of the experiment you will be paid in private and in cash. The amount of money you earn will depend on the decisions that you and the other participants make. The experiment consists of two parts, each part consisting of a number of rounds. In each round you can earn points. At the end of the experiment you will be paid according to the sum of your total point earnings from all rounds in both parts at a rate of 0.4 pence per point. You will receive the instructions for the second part after the first part is finished.

Part One
At the beginning of Part One five of the participants will get the role of "employers" and five will get the role of "workers". You will find out whether you are an employer or worker when the decision-making part of the experiment begins. If you are an employer you will remain an employer throughout the first part, and if you are a worker you will remain a worker throughout the first part.

Part One will consist of 40 rounds. In each round the employers will be paired with the workers. Thus, if you are an employer you will be paired with one of the workers, and if you are a worker you will be paired with one of the employers. The people you are paired with will change randomly from round to round.

At the beginning of a round all participants will make their decisions. Employers must choose either INSPECT or NOT INSPECT. Workers must choose either HIGH effort or LOW effort. At the end of the round, after everyone has made their decision, the computer will inform you of the choices made by you and the person you were paired with and your point earnings for the round.

The number of points you earn in a round will depend on the decisions made by you and the
person you are paired with in that round, as described in the tables below:

<table>
<thead>
<tr>
<th>Employer’s point earnings</th>
<th>Worker’s point earnings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HIGH</td>
</tr>
<tr>
<td>INSPECT</td>
<td>52</td>
</tr>
<tr>
<td>NOT INSPECT</td>
<td>60</td>
</tr>
</tbody>
</table>

For example, if the employer chooses NOT INSPECT and the worker chooses LOW the employer earns 0 points and the worker earns 40 points.

In addition, on your screen you will see your accumulated point earnings so far, and a table summarizing the decisions made by all participants in previous rounds. The table will be like the one shown below (although the data in the table has been chosen for illustrative purposes only; in the experiment the data will correspond to the actual decisions made by participants).

<table>
<thead>
<tr>
<th>Results of last 20 rounds</th>
<th>HIGH</th>
<th>LOW</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>INSPECT</td>
<td>10%</td>
<td>20%</td>
<td>30%</td>
</tr>
<tr>
<td>NOT INSPECT</td>
<td>30%</td>
<td>40%</td>
<td>70%</td>
</tr>
<tr>
<td>Total</td>
<td>40%</td>
<td>60%</td>
<td>100%</td>
</tr>
</tbody>
</table>

For example, the table tells you that the combination (INSPECT, HIGH) occurred in 10% of the cases, that the employers chose INSPECT in 30% of the cases, and the workers chose HIGH in 40% of the cases. The table is based on the results of the most recent 20 rounds only.

To make sure everyone understands the instructions so far, please complete the questions about Part One below. In a couple of minutes someone will come to your desk to check the answers.

1. Will you be matched with the same person from round to round? ——

2. How many points will you earn in a round if you are an employer, choose NOT INSPECT, and the worker you are matched with chooses HIGH? ——

3. How many points will you earn in a round if you are a worker, choose HIGH, and the employer you are matched with chooses NOT INSPECT? ——

4. Is the following statement true: the screen summarizing the history so far always contains information on all previous rounds ——

5. Is the following statement true: the screen summarizing the history so far contains information on the choices of all 10 participants in the room ——

**Part Two**

In Part Two you will keep the same role as you had in Part One. Again, you will be matched
with a different person in the other role in each round. Part Two will consist of an additional 80 rounds, starting with round 41 and ending after round 120. Your decisions together with the decisions of the people that you will be matched with will determine your earnings that will be added to your total earnings in points from Part One. At the beginning of a round, employers must again choose either INSPECT or NOT INSPECT, while workers must choose either HIGH effort or LOW effort. At the end of the round, the computer will inform you of the outcome of the round for you and the person you are paired with.

[CONTROL: The point earnings that the employer and worker receive in each of the four cases (INSPECT, HIGH); (INSPECT, LOW); (NOT INSPECT, HIGH); (NOT INSPECT, LOW) will remain exactly the same as in Part One, as shown below.

<table>
<thead>
<tr>
<th>Employer’s point earnings</th>
<th>Worker’s point earnings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HIGH</td>
</tr>
<tr>
<td>INSPECT</td>
<td>52</td>
</tr>
<tr>
<td>NOT INSPECT</td>
<td>60</td>
</tr>
</tbody>
</table>

[FINE: The only difference between Part One and Two will be that the worker will pay a fine of 20 points to the employer when the worker was inspected and chose low effort. So after INSPECT and LOW the employer’s point earnings increase by 20 points and the worker’s point earnings decrease by 20 points, as shown in the tables below:

<table>
<thead>
<tr>
<th>Employer’s point earnings</th>
<th>Worker’s point earnings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HIGH</td>
</tr>
<tr>
<td>INSPECT</td>
<td>52</td>
</tr>
<tr>
<td>NOT INSPECT</td>
<td>60</td>
</tr>
</tbody>
</table>

Thus, if the employer chooses INSPECT and the worker chooses LOW the employer earns 32 points and the worker earns 0 points. In all other cases the payoffs remain the same as in Part One.]

[BONUS: The only difference between Part One and Two will be that the employer will give a reward of 20 points to the worker when he or she inspected the worker and found out that the worker chose high effort. So after INSPECT and HIGH the employer’s point earnings decrease by 20 points and the worker’s point earnings increase by 20 points, as shown in the new earnings tables below:

<table>
<thead>
<tr>
<th>Employer’s point earnings</th>
<th>Worker’s point earnings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HIGH</td>
</tr>
<tr>
<td>INSPECT</td>
<td>52</td>
</tr>
<tr>
<td>NOT INSPECT</td>
<td>60</td>
</tr>
</tbody>
</table>
Thus, if the employer chooses INSPECT and the worker chooses HIGH the employer earns 32 points and the worker earns 45 points. In all other cases the payoffs remain the same as in Part One.

As before, your screen will display your accumulated point earnings (including your earnings from Part One). You will also see a table summarizing the decisions made by all participants in previous rounds. At the start of period 41, this table will be empty. The table will again list the results of the most recent 20 rounds after round 41.

**Ending the session**

At the end of round 120 your total points from all rounds will be converted to cash at a rate of 0.4 pence per point and you will be paid this amount in private and in cash. Now please begin making your Part Two decisions.

<table>
<thead>
<tr>
<th></th>
<th>Employer’s point earnings</th>
<th>Worker’s point earnings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HIGH</td>
<td>LOW</td>
</tr>
<tr>
<td>INSPECT</td>
<td>32</td>
<td>12</td>
</tr>
<tr>
<td>NOT INSPECT</td>
<td>60</td>
<td>0</td>
</tr>
</tbody>
</table>
D. How to Derive the Equilibrium Predictions of IBE and QRE with Loss Aversion in the Context of the Canonical Inspection Game

In this appendix, we explain the procedure to derive the equilibrium predictions of IBE and QRE with loss aversion in the context of the canonical inspection game. Selten and Chmura (2008) provide a more general discussion for IBE and Brunner, Camerer, and Goeree (2011) for QRE.

In IBE, players judge the payoffs according to how they relate to their security level. A player’s security level \( s \) is determined by the player’s pure maximin payoff, the maximum of the minimum payoffs corresponding to the player’s actions. The left panel of Figure D.1 presents the canonical inspection game, in which the inspector can secure a payoff of 12 and the worker a payoff of 25. The payoff matrix is then transformed to account for loss aversion in the following way. From each payoff exceeding a player’s security level half the difference between the payoff and the security level is subtracted (the other payoffs remain unchanged). Or, each payoff \( x \) is replaced by \( x - \max\left\{ \frac{1}{2}(x - s), 0 \right\} \). As a consequence, losses compared to the reference point weigh twice the amount that gains weigh. The middle panel of Figure D.1 presents the Transformed inspection game. From the Transformed game, the Impulse matrix is derived with the following procedure. Each set of two payoffs of a player corresponding to the same action of the other player is transformed such that the highest payoff becomes 0 and the lowest becomes the difference between the highest and the lowest. The resulting numbers represent the impulses to choose the other action given the action chosen by the other player. The impulse matrix is presented in the right panel of Figure D.1.

In the IBE, a player’s expected impulse from one action to the other equals the expected impulse from the other action to the one action. Let \( p \) represent the probability that the employer chooses I, and \( q \) the probability that the worker chooses L, then \( p \) and \( q \) follow from the solution of the impulse balance equations:

\[
4p(1 - q) = 12(1 - p)q \\
7.5(1 - p)(1 - q) = 5pq
\]
In QRE, players maximize expected utility taking the actual response function of the other player into account, but make mistakes. Let $E_{\text{player}}[a]$ represent a player’s expected utility from choosing action $a$, then:

$$p = \frac{e^{\lambda E_{\text{employer}}[I]}}{e^{\lambda E_{\text{employer}}[I]} + e^{\lambda E_{\text{employer}}[N]}}$$

$$q = \frac{e^{\lambda E_{\text{worker}}[L]}}{e^{\lambda E_{\text{worker}}[L]} + e^{\lambda E_{\text{worker}}[H]}}$$

where $\lambda$ represents the player’s rationality parameter that is estimated from the data. For QRE with loss aversion, the payoffs of the Transformed inspection game are used. In this case, $p$ and $q$ follow from the solution of:

$$p = \frac{e^{\lambda[32(1-q)+12q]}}{e^{\lambda[32(1-q)+12q]} + e^{\lambda[36(1-q)]}}$$

$$q = \frac{e^{\lambda[25]}}{e^{\lambda[25]} + e^{\lambda[20]+32.5(1-p)}}$$

The QRE prediction for the game without loss aversion is similarly found using the ordinary payoffs listed in the left panel of Figure D.1.
E. Instructions “Keeping out Trojan Horses”

Before each part of experiment computerized instructions were shown, that could be viewed by participants in their own pace. The instructions differ in length depending on the part of the experiment they precede. At the start of the experiment, there was a relatively large text with all the details of the auction and the payoffs. Before each of the following parts the instructions give only the changes with respect to the previous block.

They differ also in content depending on the type of auction and the sequence of limited and unlimited liability (LULU vs ULUL). While most of the explanations is the same for all types of auctions and liability, there are some differences explaining the various types of auctions and liabilities. We will give you the instructions for every part for the LULU design and indicate where they differ and indicate for which variation(s) it is applicable.

Instructions for part 1

Introduction
You are about to participate in an economic experiment. The instructions are simple. If you follow them carefully, you can make a substantial amount of money. Your earnings will be paid to you in euros at the end of the experiment. This will be done in private, one participant at a time.

Earnings in the experiment will be denoted by “francs”. At the end of the experiment, francs will be exchanged for euros. The exchange rate will be 3.5 eurocent per franc, or 3.5 euro for each 100 francs.

At the top of your screen, you will see the button “ready”. Please, click this when you have completely finished the instructions.

Auctions
In today’s experiment, you will participate in auctions. In these auctions you may try to obtain a fictitious good. In the remainder of these instructions we will explain the way in which the auction is organized and the rules you must follow.

Rounds
Today's experiment consists of 48 rounds. In each round, a fictitious good will be auctioned. The 48 rounds are split into 4 blocks of 12. We will now explain the instructions for the first 12 rounds. Instructions for later rounds will appear after round 12. In every round, you will be a member of a group. This group consists of you and two other people. It is unknown to you and to other participants who is in which group. In addition, we will make new groups in every round. Thus, the members of your group will change from round to round.

**The Value of the Auctioned Good**

The value of the fictitious good will be the same for all three bidders in your group. More precisely, the fictitious good is a bundle of three objects. The total value of the good equals the total value of the three objects:

\[
\text{(Value of the good)} = \text{(Value of object 1)} + \text{(Value of object 2)} + \text{(Value of object 3)}
\]

Before you participate in the auction in any round, you will be informed about the value of one of the three objects. We will call this information your “signal”. This signal can be any number (randomly determined by the computer) between 0 and 100 francs. Similarly, each of the two other participants in your group will be informed about the value of one of the other objects. So, the total value of the good is equal to the sum of the signals of the three bidders in your group.

Note the following about the signals:

1. The signal for each bidder is determined independently of the signals of the other two bidders;
2. A signal can be any number between 0 and 100;
3. Any signal between 0 and 100 is equally likely.

For example, if your signal equals 50, and the signals of the other two bidders in your group are 25 and 75 respectively, the value of the fictitious good will be:

\[
\text{(Value of the good)} = 50 + 25 + 75 = 150
\]

Note that the value of the fictitious good will always lie between 0 and 300.

**The Auction**

[For EN:] In the auction, the computer will gradually raise the price from 0 to 300. At each price, you and the other members of your group can indicate to step out of the auction.

When the first bidder steps out of the auction, the auction will stop for a few seconds. The other two bidders will be informed at which price the first bidder stepped out. The auction ends
when the second bidder steps out of the auction. The remaining bidder gets the good: he or she will obtain the value of the three objects. This bidder pays the price at which the second bidder stepped out of the auction.

If two or three participants step out of the auction at the same price, the computer will randomly determine which one will actually step out. The other(s) will remain in the auction. If two or three bidders remain in the auction up to a price of 300, the computer will randomly determine who wins the object. This bidder has to pay 300 francs.

[For FP:] In the auction, you and the other members of your group will submit a bid. This must be a number between 0 and 300. The bidder submitting the highest bid gets the good. He or she will obtain the value of the three objects for a price equal to his or her bid. If two or three participants submit the same bid, the computer will randomly determine which one will win. The winner pays his or her own bid.

Earnings
[For LULU:] If the winner in a certain round pays less than the value of the good, his or her earnings in that round will be:

\[(\text{Earnings}) = (\text{Value of the good}) - (\text{Price})\]

In contrast, the price paid by the winner may turn out to be higher than the good’s value. If this is the case, then the winner does not have to cover the loss in the auction. However, the bidder will face a cost of 4 francs, which will be subtracted from his or her earnings so far. Note that the winner will pay 4 francs even if his or her loss is only 1, 2, or 3 francs.

If not winning, a bidder’s earnings will be zero.

[For ULUL:] The winner’s earnings in a round will be:

\[(\text{Earnings}) = (\text{Value of the good}) - (\text{Price})\]

Note that the price paid by the winner may turn out to be higher than the object’s value. If this is the case, then the winner makes a loss, which will be subtracted from his or her total earnings in this part so far.

Starting Capital
[For LULU:] At the beginning of part 1, each participant will obtain a starting capital of 50 francs. This starting capital may be used to cover potential losses made in part 1. So, your total
earnings in this part will be the starting capital of 50 plus earnings in the auctions minus the cost in case of a loss in the auction.

[For ULUL:] At the beginning of part 1, each participant will obtain a starting capital of 150 francs. This starting capital may be used to cover potential losses made in part 1. So, your total earnings in this part will be the starting capital of 150 plus earnings in the auctions. You cannot earn less than zero in this part. If your total earnings end up below zero after a certain round, you will start at zero in the next round.

Instructions for part 2

We will now start the second part of the experiment. Part 2 will be almost the same as part 1. The same fictitious good will be sold in the same auction. Again, the good consists of three objects, and each bidder will obtain a signal equal to the value of one of the objects. The exchange rate remains 3.5 euro cent per franc, or 3.5 euro for each 100 francs. Part 2 will also consist of 12 rounds.

[For LULU:] The only difference is that in part 2, the winner of the good has to cover the loss if the price turns out to be higher than the value of the good. Therefore, the winner’s earnings in a round will be as follows.

[For ULUL:] The main difference is that in part 2, the winner of the good does not have to cover the loss if the price of the good turns out to be higher than its value. The winner’s earnings in a round will be as follows.

Earnings

[For LULU:] The winner’s earnings in a round will be:

\[(\text{Earnings}) = (\text{Value for the good}) - (\text{Price})\]

Note that the price paid by the winner may turn out to be higher than the object’s value. If this is the case, then the winner makes a loss, which will be subtracted from his or her total earnings in this part so far.

[For ULUL:] If the winner in a certain round pays less than the value of the good, his or her earnings in that round will be:

\[(\text{Earnings}) = (\text{Value of the good}) - (\text{Price})\]
In contrast, the price paid by the winner may turn out to be higher than the good's value. If this is the case, then the winner does not have to cover the loss in the auction. However, the bidder will face a cost of 4 francs, which will be subtracted from his or her earnings so far. Note that the winner will pay 4 francs even if his or her loss is only 1, 2, or 3 francs.

If not winning, a bidder's earnings will be zero.

Starting Capital

[For LULU:] At the beginning of part 2, each participant will obtain a starting capital of 150 francs. This starting capital may be used to cover potential losses made in part 2. So, your total earnings in this part will be the starting capital of 150 plus earnings in the auctions. You cannot earn less than zero in this part. If your total earnings end up below zero after a certain round, you will start at zero in the next round. So, you will not lose part of your earnings in part 1 if your starting capital of 150 francs turns out not to be sufficient to cover losses in part 2.

[For ULUL:] At the beginning of part 2, each participant will obtain a starting capital of 50 francs. This starting capital may be used to cover potential losses made in part 2. So, your total earnings in this part will be the starting capital of 50 plus earnings in the auctions minus the cost in case of a loss in the auction.

Instructions for part 3 for the LULU treatment

We will now start the third part of the experiment. Part 3 will be exactly the same as part 1. So, the same fictitious good will be sold in the same auction. Again, the good consists of three objects, and each bidder will obtain a signal equal to the value of one of the objects. The exchange rate remains 3.5 eurocent per franc, or 3.5 euro for each 100 francs. Part 3 will also consist of 12 rounds.

Recall that the only difference between part 3 and part 2 is that in part 3, the winner of the good does not have to cover the loss if the price of the good turns out to be higher than its value. Therefore, the winner's earnings in a round will be as follows.

Earnings

If the winner in a certain round pays less than the value of the good, his or her earnings in that round will be:

\[ \text{(Earnings)} = \text{(Value of the good)} - \text{(Price)} \]

In contrast, the price paid by the winner may turn out to be higher than the good's value. If this is the case, then the winner does not have to cover the loss in the auction. However, the

\(^1\)The instructions for the ULUL treatment are very similar, with parts 3 and 4 swapped.
bidders will face a cost of 4 francs, which will be subtracted from his or her earnings so far. Note that the winner will pay 4 francs even if his or her loss is only 1, 2, or 3 francs.

If not winning, a bidder's earnings will be zero.

**Starting Capital**

At the beginning of part 3, each participant will obtain a starting capital of 50 francs. This starting capital may be used to cover potential losses made in part 1. So, your total earnings in this part will be the starting capital of 50 plus earnings in the auctions minus the cost in case of a loss in the auction.

**Instructions for part 4 for the LULU treatment**

We will now start the fourth and last part of the experiment. Part 4 will be exactly the same as part 2. The same fictitious good will be sold in the same auction. Again, the good consists of three objects, and each bidder will obtain a signal equal to the value of one of the objects. The exchange rate remains 3.5 eurocent per franc, or 3.5 euro for each 100 francs. Part 4 will also consist of 12 rounds.

Recall that the only difference between part 3 and part 4 is that in part 4, the winner of the good has to cover the loss if the price turns out to be higher than the value of the good. Therefore, the winner's earnings in a round will be as follows.

**Earnings**

The winner's earnings in a round will be:

\[
(\text{Earnings}) = (\text{Value for the good}) - (\text{Price})
\]

Note that the price paid by the winner may turn out to be higher than the object's value. If this is the case, then the winner makes a loss, which will be subtracted from his or her total earnings in this part so far.

**Starting Capital**

As in part 2, at the beginning of part 4, each participant will obtain a starting capital of 150 francs. This starting capital may be used to cover potential losses made in part 4. So, your total earnings in this part will be the starting capital of 150 plus earnings in the auctions.

You cannot earn less than zero in this part. If your total earnings end up below zero after a certain round, you will start at zero in the next round. So, you will not lose part of your
earnings in parts 1, 2 and 3 if your starting capital of 150 francs turns out not to be sufficient to cover losses in part 4.
F. Proofs of Propositions “Keeping out Trojan Horses”

Proof of Proposition 5.1. Let $\bar{u}(θ, \tilde{θ})$ be the utility of bidder 1 with type $θ$ who bids as if having type $\tilde{θ}$ “close” to $θ$ while the other two bidders bid according to the same strictly increasing bidding function $B$ with $B(θ) < 2θ$. Then,

$$
\bar{u}(θ, \tilde{θ}) = \frac{1}{60,000} \left[ θ + 2\tilde{θ} - B(\tilde{θ}) \right]^3 - \frac{1}{30,000} \left[ θ + \tilde{θ} - B(\tilde{θ}) \right]^3 - \frac{1}{20,000} c \left[ B(\tilde{θ}) - θ \right]^2.
$$

The first [second] term on the right-hand side in the first line refers to situations in which bidder 1 does not go [goes] bankrupt. The first-order condition of the equilibrium is given by

$$
\frac{\partial \bar{u}(θ, \tilde{θ})}{\partial θ} \bigg|_{\tilde{θ}=θ} = \frac{1}{10,000} \left[ 3θ - B(θ) \right]^2 \left[ 2 - B'(θ) \right] - \left[ 2θ - B(θ) \right]^2 \left[ 1 - B'(θ) \right] - cB'(θ) \left[ B(θ) - θ \right] = 0
$$

from which differential equation (5.10) follows.

Proof of Proposition 5.2. Let $B$ be the equilibrium bid function. According to the ranking lemma (see e.g., Milgrom 2004), the proposition holds true if $B(0) = 0$ and if $B(θ) = \frac{5}{3}θ$ implies that $B'(θ) < \frac{5}{3}$. It is standard that $B(0) = 0$ must hold in a symmetric equilibrium. Moreover, suppose that bidders 2 and 3 bid according to $B$ and that bidder 1 with signal $θ$ bids as if having signal $\tilde{θ}$. Bidder 1’s utility equals

$$
\bar{u}(θ, \tilde{θ}) = \int_0^\tilde{θ} \int_0^\tilde{θ} u(θ + θ_2 + θ_3 - B(\tilde{θ})) d\frac{θ_2}{100} d\frac{θ_3}{100}.
$$
The first-order condition of the equilibrium implies that if $B(\theta) = \frac{5}{3} \theta$,

$$0 = 10,000 + \hat{u}_2(\theta, \theta)$$

$$= 2 \int_0^\theta u(2\theta + \theta_2 - B(\theta))d\theta_2 - \int_0^\theta \int_0^\theta u'(\theta + \theta_2 + \theta_3 - B(\theta))d\theta_2d\theta_3$$

$$= 2 \int_0^\theta u \left( \frac{1}{3} \theta + \theta_2 \right) d\theta_2 - \int_0^\theta \int_0^\theta \left[ u \left( \frac{1}{3} \theta + \theta_2 \right) - u \left( \theta_2 - \frac{2}{3} \theta \right) \right] d\theta_2 \Rightarrow$$

$$B'(\theta) = \frac{2 \int_0^\theta u \left( \frac{1}{3} \theta + \theta_2 \right) d\theta_2}{\int_0^\theta [u \left( \frac{1}{3} \theta + \theta_2 \right) - u \left( \theta_2 - \frac{2}{3} \theta \right)] d\theta_2} < \frac{5}{3}.$$  

The third equality follows by direct integration and by substituting $B(\theta) = \frac{5}{3} \theta$. The inequality follows because the strict concavity of implies that

$$\int_0^\theta \left[ u \left( \frac{1}{3} \theta + \theta_2 \right) + 5u \left( \theta_2 - \frac{2}{3} \theta \right) \right] d\theta_2 < u'(0) \int_0^\theta \left[ \left( \frac{1}{3} \theta + \theta_2 \right) + 5 \left( \theta_2 - \frac{2}{3} \theta \right) \right] d\theta_2 = 0.$$

\[\square\]

**Proof of Corollary 5.1** The expected winning bid equals

$$\mathbb{E} \left\{ \min \left( \frac{\delta_n \theta_n}{1 - \delta_n}, \delta_n \theta_n \right) + \theta_k \right\} \leq \mathbb{E} \left\{ \delta_n \theta_n + \theta_k \right\} \leq \mathbb{E} \left\{ \theta_n + \theta_k \right\} \leq \mathbb{E} \left\{ \theta^{(1)} + \theta^{(2)} \right\} = 125 = R^\infty_E,$$

from which the result immediately follows.  

\[\square\]

**Proof of Proposition 5.4** Suppose both opponents of bidder 1 bid according to (5.19). Bidder 1 wishes to step out of the auction at a price equal to her (perceived) expected value. If both of her opponents step out at the same price $p$, bidder 1 knows that both have signal

$$\theta = \frac{p - 100\chi}{3 - 2\chi}.$$  

She steps out at price $p$ equal to her perceived expected value, i.e.,

$$v = \theta_1 + 2(1 - \chi)\theta + 100\chi = \theta_1 + 2(1 - \chi)\frac{p - 100\chi}{3 - 2\chi} + 100\chi = p.$$  

It is readily verified that $B^{1,\chi}_E$ in (5.19) is a solution. Similarly, $B^{2,\chi}_E$ follows by taking into account that bidder 1 updates her beliefs about the signal of the lowest bidder with probability $1 - \chi$.  

\[\square\]

**Proof of Proposition 5.5** Let $\tilde{u}(\theta, \bar{\theta})$ be the perceived utility of bidder 1 with type $\theta$ who bids
as if having type $\tilde{\theta}$ while the other two bidders bid according to the same strictly increasing bidding function $B$. Then,

$$\tilde{u}(\theta, \tilde{\theta}) = \tilde{\theta}^2 \left[ (1 - \chi) \left( \theta + \tilde{\theta} \right) + \chi (\theta + 100) - B(\tilde{\theta}) \right].$$

The first-order condition of the equilibrium is given by

$$\frac{\partial \tilde{u}(\theta, \tilde{\theta})}{\partial \tilde{\theta}} \bigg|_{\tilde{\theta} = \theta} = 2\theta \left[ 2(1 - \chi) + \chi (\theta + 100) - B(\theta) \right] + \theta^2 \left[ (1 - \chi) - B'(\theta) \right] = 0.$$

It is readily verified that (5.20) is a solution.

---

**Proof of Proposition 5.6.** Bidder 1 steps out at price $p$ equal to her perceived expected value of winning given that her two opponents bid according to equilibrium. Because bidder 1 is fully cursed, she assumes that the other two bidders’ signals are uniformly distributed on $[0, 100]$ regardless of her winning the auction and regardless of the price at which an opponent steps out. Therefore, she indeed steps out at a price $p$ which solves $\tilde{U}(p, \theta) = 0$.

---

**Proof of Proposition 5.7.** Let $\tilde{u}(\theta, \tilde{\theta})$ be the utility of bidder 1 with type $\theta$ who bids as if having type $\tilde{\theta}$ while the other two bidders bid according to the same strictly increasing bidding function $B$. Then

$$\tilde{u}(\theta, \tilde{\theta}) = G(\tilde{\theta}) \tilde{U}(B(\tilde{\theta}), \theta)$$

where

$$G(\theta) \equiv \frac{\theta^2}{10,000}$$

is the distribution function of the higher of two draws from $U[0,100]$. Equation (5.27) follows immediately from the first-order condition of the equilibrium:

$$\left. \frac{\partial \tilde{u}(\theta, \tilde{\theta})}{\partial \tilde{\theta}} \right|_{\tilde{\theta} = \theta} = G'(\theta) \tilde{U}(B(\theta), \theta) + G(\theta) \tilde{U}_1(B(\theta), \theta) B'(\theta) = 0.$$

---

**Proof of Corollary 5.3.** (The proof proceeds along the same lines as Maskin and Riley’s (1984) proof of their Theorem 4.) Conditional on a bidder with type $\theta$ winning, the expected winning
bid in EN is given by

\[ R_E(\theta) = \int_0^\theta \frac{b_E^{\chi=1}(t)}{G'(\theta)} dG(t) \]

where \( G \) is the distribution function of the higher of two draws from \( U[0,100] \). Consequently,

\[ R'_E(\theta) = \left[ b_E^{\chi=1}(\theta) - R_E(\theta) \right] \frac{G'(\theta)}{G(\theta)} \]

The winning bid in FP equals \( R_F(\theta) = b_F^{\chi=1}(\theta) \). Therefore,

\[ R'_F(\theta) = b_F^{\chi=1}(\theta) = \frac{\tilde{U}(b_F^{\chi=1}(\theta), \theta)}{U_1(b_F^{\chi=1}(\theta), \theta)} G'(\theta). \]

Because \( b_E(0) = b_F(0) \), it follows that \( R_E(0) = R_F(0) \). According to the ranking lemma (see e.g., Milgrom [2004]), the proposition follows if \( R_E(\theta) = R_F(\theta) \Rightarrow R'_E(\theta) > R'_F(\theta) \), which is equivalent to

\[ b_E^{\chi=1}(\theta) - b_F^{\chi=1}(\theta) > \frac{\tilde{U}(b_F^{\chi=1}(\theta), \theta)}{U_1(b_F^{\chi=1}(\theta), \theta)}. \]

Consider the left- and right-hand sides as functions of \( b_F \). For \( b_F = b_E \), both sides vanish. The derivative of the right-hand side is equal to \(-1 + \frac{\tilde{U}}{U_1(b_F^{\chi=1}(\theta), \theta)} < -1 \) whereas the derivative of the left-hand side equals -1. Therefore, because \( b_F^{\chi=1}(\theta) < b_E^{\chi=1}(\theta) \), we conclude that the inequality is satisfied.

\( \square \)
G. Inleiding

In dit proefschrift onderzoeken we vier vragen, die betrekking hebben op het gedrag van één of meerdere ondergeschikten in een hiërarchische relatie, waarbij een superieur bepaald gedrag (hier ‘goed’ genoemd) economisch gezien prefereert boven ander gedrag. In al deze gevallen, kan middels controle en sturing de ondergeschikte(n) misschien wel tot het gewenste gedrag worden bewogen, maar deze weg is voor de superieur (te) kostbaar. Het gaat om de volgende vragen:

1. Een overheid wil door het invoeren van een subsidie bepaald gedrag bevorderen. De vraag die we ons stellen, is of het effectievere is om een dergelijke subsidie in één stap in te voeren dan wel geleidelijk in kleine stapjes.

2. Een overheid wil via beloningen en/of via boetes gewenst gedrag bevorderen. Hierbij gaat het om beloningen en boetes die automatisch volgen op gewenst respectievelijk ongewenst gedrag. De vraag is welke van de twee instrumenten effectievere is.

3. Een werkgever wil via belonen en/of straffen bepaald gedrag van een werknemer bevorderen. In dit geval gaat het om instrumenten die de werkgever naar eigen inzicht kan hanteren. De vraag is ook hier welk instrument is effectievere.

4. Een overheid die gebruik maakt van veilingen, bijvoorbeeld voor de verkoop van frequentie licenties of de inkoop van goederen, wil niet dat de hoogste bieder na afloop van de veiling failliet gaat. De vraag is of het risico voor dit type faillissement beperkt kan worden door te kiezen voor een bepaald type veiling. We vergelijken twee veelvoorkomende veilingtypen, de Engelse veiling en de eerste-prijs gesloten-bod veiling.

Voor ons onderzoek gebruiken we laboratorium experimenten, terwijl we via het zogenaamde mechanism design ook hadden kunnen proberen de optimale instrumenten te ontwerpen. De meeste modellen die gebruikt worden in deze benadering gaan echter uit van rationele, zelfzuchtige en/of emotieloze mensen. Experimenten, uitgevoerd zowel in het laboratorium als in het veld, laten zien dat dergelijke vooronderstellingen meestal niet opgaan. Omdat we geen allesomvattende theorie van het menselijk gedrag tot onze beschikking hebben, gebruiken we laboratorium experimenten om bovengenoemde vragen te onderzoeken. De vraag is steeds welke van de twee in de praktijk vaak gebruikte instrumenten het beste werkt.

1Voor een bespreking van mechanism design, zie Myerson (1981).
2Voor een overzicht, zie bijvoorbeeld, Tirok (2002).
In hoofdstuk 2 onderzoeken we op welke manier subsidies het best kunnen worden ingevoerd als de subsidievestrekker bepaald gedrag wil stimuleren. In 2009 introduceerde de Japanse overheid een subsidie van 10% op zonnepanelen. Omdat die subsidie minder effect bleek te hebben dan gepland, wordt verwacht dat deze subsidie in de toekomst verhoogd zal worden [Leader 2009]. In datzelfde jaar, kondigde de Chinese overheid aan een 50% subsidie aan op zonnepanelen, de hoogste subsidie in zijn soort ter wereld [Ideas 2009]. Subsidieverstrekking is een belangrijk instrument van overheden en we testen of een invoering in één stap effectiever is dan een invoering in kleine stappen.

In ons experiment maken we gebruik van een zogenaamd publiek goed spel. In een publiek goed spel beslissen de deelnemers elke ronde hoeveel ze volledig anoniem bijdragen aan een algemene pot. Elke bijdrage aan de pot wordt vervolgens kosteloos verhoogd met een bepaald percentage (20% in ons geval). De totale inzet wordt vervolgens gelijk verdeeld over alle deelnemers, ongeacht of en hoeveel een deelnemer heeft bijgedragen. Deze regels zorgen ervoor dat het voor elke deelnemer afzonderlijk financieel gezien altijd voordeeliger is om niets bij te dragen. Aan de basisopzet voegen we een subsidie toe, die de kosten van een bijdrage verlaagt. Als een deelnemer 10 bijdraagt, terwijl de subsidie .45 is dan kost de bijdrage de deelnemer $(1 - 0.45) \times 10 = 5.5$.

De deelnemers van het experiment worden in twee groepen ingedeeld. De ene groep volgt de snelle treatment en de andere groep de langzame treatment. Beide treatments starten met een subsidie van 0.00 en na 4 minuten wordt de subsidie verhoogd. In de snelle treatment gaat de subsidie in één keer naar het beoogde niveau en in de langzame treatment stapje voor stapje. Voor beide treatments geldt dat als het beoogde niveau is bereikt, dit gehandhaafd blijft tot het einde van het experiment, 28 minuten na de start.3


3Om te onderzoeken of een evenwichtige verschil, net als in het gewone leven, zou kunnen worden toegeschreven aan het feit dat mensen voortdurend afgeleid worden door andere zaken die aandacht vragen, maakten we treatments met en treatments zonder een extra spel dat de aandacht kan afleiden. Dit 'afleidende spel' kon door de deelnemers tegelijk met het publieke spel gespeeld worden en voor beide spelen kon geld verdiend worden. Het al of niet toevoegen van het afleidende spel blijkt echter geen significante invloed te hebben op de hoogte van de bijdragen.
zullen bijdragen. Het zal in dit geval afhangen van de verwachtingen die de voorwaardelijke coöperatoren hebben met betrekking tot de bijdragen van de andere deelnemers, wat de voorwaardelijke coöperatoren zelf zullen bijdragen. Het zou kunnen zijn dat als bijdragen effectiever worden, zij optimistischer worden over de hoogte van de bijdragen van de andere deelnemers en daarom zelf meer gaan bijdragen.

Terwijl deze literatuur zich richt op de vraag waarom mensen reageren op subsidies, ligt bij ons de focus op de reactie op twee verschillende manieren waarop subsidies worden geïmplementeerd, snel of langzaam. Interessant is dat het concept van voorwaardelijke coöperatoren ook hier een rol zou kunnen spelen. Indien voorwaardelijke coöperatoren verwachten dat de andere deelnemers sterker reageren op een sneller dan op een langzame invoering, zal dit voor hen een reden kunnen zijn, om zelf ook meer bij te dragen. Een andere mogelijke oorzaak voor een dergelijk effect zou het zogenoemde anchoring effect kunnen zijn (Tversky and Kahneman, 1974). De begin subsidie dient als een referentie punt: deelnemers zullen hun gedrag alleen veranderen als er een waarneembare verandering in het subsidienniveau optreedt.

De uitkomst van het experiment is dat er een verschil is in de wijze, waarop in beide treatments de bijdragen aan het publieke goed veranderen, maar dat dit alleen optreedt als de subsidie hoog genoeg is. Als de subsidie .45 is, is het verschil tussen langzame en snelle invoering niet significant. In beide gevallen is er sowieso geen significant verschil tussen de bijdrage voor en na de invoering van de subsidie. Als de subsidie daarentegen .75 is, dan zien we nog steeds dat er voor en na de langzame invoering van de subsidie geen significant verschil is, maar tussen voor en na een snelle invoering is het verschil uitermate significant. Uit het experiment zouden we dan ook kunnen concluderen dat een relatief hoge subsidie beter in één stap kan worden ingevoerd.

Terwijl het in hoofdstuk 2 gaat om overheden die gedrag willen sturen via subsidies, gaat het in hoofdstuk 3 om gedragsbeïnvloeding via straffen en belonen. In 2000 verhoogde de Nederlandse overheid de boete voor het niet aan de belasting opgeven van spaargeld van 10% tot 25% van het verzorgeng bedrag en verdere verhogingen zijn reeds aangekondigd (Tweede Kamer, 2000). In 2003 begon de Zuid-Koreaanse overheid met het belonen van belastingbetalers met een goede staat van dienst (NTS, 2004). Het bestraffen van ongewenst gedrag en het belonen van gewenst gedrag zijn twee instrumenten die vaak gebruikt worden door autoriteiten.

We onderzoeken welk instrument beter werkt met behulp van een inspectie spel. In elke ronde van dit spel, nemen een inspecteur en een geïnspecteerde tegelijkertijd en onafhankelijk van elkaar een besluit. De inspecteur beslist of hij een voor zichzelf kostbare inspectie van het werk van de geïnspecteerde uit gaat voeren en de geïnspecteerde besluit al dan niet te gaan werken. De inspecteur moet de geïnspecteerde een loon uit betalen, dat hoger ligt dan de kosten van het werk voor de geïnspecteerde, tenzij de inspecteur heeft besloten te inspecteren en de geïnspecteerde heeft besloten om niet te werken. Het loon is hoger dan de kosten van de inspectie.
Aan dit inspectie spel voegen we een automatische boete toe in het geval de de inspecteur inspecteert en de geïnspecteerde niet werkt en een automatische beloning indien de inspecteur inspecteert en de geïnspecteerde werkt. Boetes gaan ten koste van de geïnspecteerde en komen ten goede van de inspecteur, beloningen gaan ten koste van de inspecteur en komen ten goede van de geïnspecteerde. Voor elke ronde worden inspecteurs willekeurig gekoppeld aan geïnspecteerden, al is het wel zo dat iedere deelnemer steeds dezelfde rol speelt gedurende het hele experiment.

We zien dat de geïnspecteerde vaker besluit te werken onder een regime van automatische boetes dan onder een regime van automatische beloningen. Dit resultaat komt overeen met de voorspellingen van een standaard speltheoretische benadering uitgaande van een gemengd NASH evenwicht, waar de spelers hun beslissingen laten afhangen van de beloningstructuur voor de andere speler. Indien een geïnspecteerde weet dat er een automatische boete is ingevoerd, die bijdraagt aan de verdiensten van de inspecteur, dan zal de geïnspecteerde verwachten dat de inspecteur vaker zal inspecteren om zo de boete te kunnen incasseren. Om die boete te vermindjden zal de geïnspecteerde vaker gaan werken. Dit gemengd NASH evenwicht kan echter niet het hele verhaal zijn. In dezelfde lijn geredeneerd zou de toevoeging van een automatische beloning moeten leiden tot het minder vaak werken door de geïnspecteerde en dat zien we in het experiment niet gebeuren. Er wordt slechts insignifiant minder gewerkt in beide treatments. Dit tegenstrijdige resultaat blijkt beter verklaard te kunnen worden door recente gedragsmodellen die uitgaan van een impulse balance evenwicht (Selten and Chmura, 2008) of een quantal response evenwicht (McKelvey and Palfrey, 1995). Samengeteld, automatisch straffen werkt beter dan automatisch belonen, maar in tegenstelling tot de voorspellingen uit het standaard speltheoretische model is het niet zo dat automatische beloningen leidt tot minder vaak werken door de geïnspecteerde.

In hoofdstuk 4 richten we ons opnieuw op straf versus beloning, maar deze keer in de context van werkgevers en werknemers in een standaard arbeidsverhouding. De set-up van het experiment op verschillende punten aangepast, hoewel de basis van het experiment het inspectie spel blijft.

In tegenstelling tot het vorige experiment, staat het al of niet belonen dan wel straffen nu helemaal ter discretie van de inspecteur (die we vanaf hier de werkgever noemen). Beide instrumenten zowel belonen als straffen zijn nu kostbaar voor de werkgever, terwijl net als in het vorige experiment straffen de geïnspecteerde (vanaf nu de werknemer genoemd) punten kosteren en belonen de geïnspecteerde punten opleveren. In elk van de treatments hanteren we een kost/gevolg verhouding die of 1:1 of 1:3 is. Een kost/gevolg verhouding van 1 : x betekent dat een straf [beloning] die de werkgever 1 punt kosteren, de werknemer x kost [oplevert]. Een ander verschil is dat in dit experiment dezelfde werkgever en dezelfde werknemer gedurende het hele experiment in alle ronden aan elkaar gekoppeld zijn. Tenslotte, als de werkgever besluit om te straffen, voegen we een extra onderdeel aan de ronde toe, waarin de de werkgever kan besluiten om te straffen, te belonen of om niets te doen.
Weliswaar verschafft de literatuur enige aanknopingspunten om de uitzonderlijke van het experiment te voorspellen, maar de literatuur is niet eenduidig. In de psychologische literatuur, concludeert Skinner (1965) aan de hand van experimenten met dieren dat in tegenstelling tot belonen, straffen geen blijvend effect heeft. Verder hebben psychologen gevonden dat opzichten die goed gedrag belonen er beter in slagen om ongedwongen te laten werken dan opzichten die slecht ongewenst gedrag bestraffen (Sims, 1980; Podsakoff, Bommer, Podsakoff, and MacKenzie, 2006; George, 1995). Het probleem is echter dat het laatste onderzoek is gebaseerd op vragenlijsten en het dus niet goed mogelijk is om vast te stellen wat oorzaak en wat gevolg is.

In de experimentele economie, is onderzoek gedaan naar de kracht van negatieve en van positieve wederkerigheid (Abbink, Irlenbusch, and Renner, 2000; Brandts and Sola, 2001; Charness and Rabin, 2002; Ofran, 2002; Brandts and Charness, 2004; Falk, Fehr, and Fischbacher, 2003; Charness, 2004; Al-Ubaydli and Lee, 2009). Er blijkt maar weinig bewijs te zijn voor positieve wederkerigheid en dit ondermijnt het idee dat werknemers reageren op beloningen. Het bewijs voor negatieve wederkerigheid is sterker, maar daaruit is het moeilijker een eenduidige conclusie te trekken. Aan de ene kant zou negatieve wederkerigheid werknemers kunnen stimuleren om straffen te vermijden, maar aan de andere kant zou deze negatieve wederkerigheid ook kunnen leiden tot een negatieve spiraal van straffen, minder werken en weer terug naar meer straffen.

In ons experiment zien we duidelijker resultaten voor de treatments met een kost/gevolg verhouding van 1:3 vergeleken met treatments met een kost/gevolg verhouding van 1:1. We zullen ons verder focussen op de treatments met een kost/gevolg verhouding van 1:3. We vergelijken treatments waar de werkgever enkel over het instrument belonen beschikt en die waarbij de werkgever alleen over het instrument straffen beschikt met het basis treatment zonder instrumenten. We zien dat vergeleken met het basis treatment in de beide treatments met precies één instrument werknemers vaker werken. Verder zien we dat dit verschil even groot is en het niet uitmaakt of dat ene instrument belonen dan wel straffen is. Als we kijken naar het aantal inspecties dan zien we een significant lager aantal inspecties in treatments met alleen straffen dan in treatments met alleen belonen of zonder instrumenten. Dit maakt voor de werkgever de situatie waarbij deze alleen beschikt over de mogelijkheid om te straffen financieel gezien het meest aantrekkelijk.

Omdat werkgevers het extra instrument belonen, zouden moeten kunnen negeren, verwachten we dat zij het in een treatment met beide instrumenten (belonen en straffen) net zo goed zouden moeten doen als in een treatment waar ze alleen kunnen straffen. Dat blijkt echter niet het geval, als de werkgevers ook de beschikking krijgen over het beloningsinstrument gebruiken ze dat veel vaker dan het instrument straffen. Aan het einde van het treatment met beide instrumenten experiment kreeg een deel van de deelnemers een vragenlijst. Op de vraag of het gepaster zou zijn goed gedrag te belonen dan wel ongewenst gedrag te bestraffen, gaven zowel deelnemers die in de rol van ondernemer speelden als ook deelnemers die in de rol van werknemer speelden gemiddeld aan dat het belonen van goed gedrag gepaster is. Wat we zien is dat als de werkgevers beschikken over belonen en straffen, werknemers evenveel werken als in een treatment waarin
alleen gestraft kan worden. Wat voor de werkgever de ‘alleen straffen’ treatment winstgevender maakt, is dat er minder inspecties nodig zijn. We kunnen dus concluderen dat voor werkgevers het toevoegen van enkel straffen aan de standaard opzet het meest winstgevend is, maar dat het effect minder wordt als ook de mogelijkheid om te belonen wordt toegevoegd.

In hoofdstuk 5 onderzoeken we de vraag hoe een overheid die een veiling organiseert kan voorkomen dat die veiling wordt gewonnen door een bidder die vervolgens failliet gaat. De context van deze veiling is er een, waarbij winnaars failliet gaan als achteraf blijkt dat de waarde van het geveilde goed lager is dan de prijs die ze ervoor betaald hebben en het faillissement schadelijk is voor de organiserende partij. We kunnen hierbij denken aan radiofrequenties die geveild worden en waarbij het faillissement van de winnaar een onderbreking van de communicatie via die frequenties inhoudt. Een ander voorbeeld waarbij een faillissement achteraf schadelijk is voor de organisator is als er een veiling is georganiseerd om de inkoop van (essentiële) goederen te regelen.

Het probleem van het faillissement achteraf is wijdverspreid. Een extreem voorbeeld is de veiling van de zogenoemde C-Blok frequenties in 1996 door de Federal Communications Committee in de VS: alle belangrijke grote winnaars, die samen $10.2 miljard hadden betaald, gingen failliet (Zheng 2001). Overigens hebben overheden verschillende methoden gebruikt om het risico op dit type faillissement te voorkomen. In de literatuur worden bijvoorbeeld surety bonds genoemd, een soort garantiestellingen door een derde partij (Calveras, Ganuza, and Hank 2004), daarnaast multi-sourcing, waarbij bidders slechts een deel van het contract kunnen verwerven (Engel and Wambach 2006) en veilingen die gewonnen worden door de bidder die het dichtst bij het gemiddelde bod zit (Decarolis 2010). Wij daarentegen onderzoeken of het uitmaakt welk van twee veel gebruikte veilingtypen, de Engelse veiling⁴ en de eerste-prijs gesloten-bod veiling⁵ wordt gekozen.

Het ontwerp van het experiment is direct afgeleid uit het probleem. De ene helft van de deelnemers neemt deel aan een set Engelse veilingen, de andere helft aan eerste-prijs gesloten-bod veilingen. In elke veiling zijn drie deelnemers, voor elk van de deelnemers wordt afzonderlijk een willekeurige getal getrokken. De waarde van het te veilen object is de som van de drie getrokken getallen. De winnaars van de veiling maken winst als de prijs die ze moeten betalen lager is dan de waarde van het object en maken een verlies als de prijs hoger is. In de helft van de veilingen waarin de deelnemers actief zijn, gaan ze failliet als ze een verlies lijden en wordt daardoor hun verlies beperkt tot een geringe waarde. In de andere helft gaan ze niet failliet als ze verlies maken en dragen dan het volledige verlies.

⁴In een Engelse veiling, verhoogt de veilingmeester telkens de prijs van het object. Iedere bidder kan op elk moment uit de veiling stappen. De overgebleven bidders krijgen te weten bij welke prijs er een bidder is uitgestapt en bij die prijs gaat de veiling verder. De bidder die het langst in de veiling blijft wint het object en betaalt de prijs waarbij de voorlaatste bidder is uitgestapt.

⁵In de eerste-prijs gesloten-bod veiling doen alle bidders tegelijkertijd en onafhankeelijk van elkaar een bod en de hoogste bidder wint.
De literatuur geeft ons aan wel enig inzicht in wat we kunnen verwachten. Klempner (2002) geeft bijvoorbeeld aan dat bieders die bankroet kunnen gaan, agressiever zullen bieden, omdat het mogelijke verlies is beperkt door de mogelijkheid failliet te gaan. Waar de literatuur echter geen uitsluiting over geeft, is in welke van de twee typen veilingen die we vergelijken dit verschijnsel het meest zal voorkomen. Bij veilingen zoals de onze waar de waarde van het object dezelfde waarde heeft heeft voor alle bieders, kunnen we volgens Milgrom and Weber (1982) in het algemeen verwachten dat in Engelse veilingen hoger geboden zal worden en dat in deze veilingen dus meer faillissementen zullen optreden. Echter in de door ons gebruikte opzet, weten de deelnemers wanneer andere deelnemers uit de veiling stappen en deze informatie kunnen ze gebruiken om de waarde van het object beter in te schatten. De resultaten van het experiment laten zien dat indien bieders failliet kunnen gaan, er in beide veilingen als verwacht agressiever wordt geboden en dat dit vaker leidt tot verliezen bij de winnaars. We zien echter geen significant verschil in het aantal faillissementen en de hoogte van de biedingen. Dit resultaat staat haaks op een voorspelling afgeleid uit een analyse van het NASH evenwicht. Als we in plaats van deze NASH analyse Eyster and Rabin’s (2005) ‘cursed equilibrium’ model gebruiken zien we dat we hiermee de uitkomsten van het experiment, beter kunnen verklaren. Onze conclusie is dan ook dat het simpelweg kiezen tussen de twee standaard veiling typen het probleem van faillissement na afloop van de veilingen niet oplost en dat het cursed equilibrium model ons helpt dit te verklaren.