Essays in experimental economics

van Veldhuizen, R.R.

Citation for published version (APA):
Experimental Economics uses experimental methods to study economic behavior. Economic experiments allow for causal relationships between two or more variables to be established in settings where this cannot be achieved with naturally-occurring data. This dissertation applies experimental methods to four economic questions. Chapter 2 examines whether nonrenewable resource markets are more likely to be in line with economic theory when resource stocks are low. Chapter 3 examines whether a wage increase in the public sector decreases corruption. Chapter 4 examines whether willpower is associated with framing effects in economic decisions. Finally, chapter 5 examines whether workers are influenced by the productivity of coworkers who can observe them.

Roel van Veldhuizen (1986) obtained a BA in Social Science from University College Utrecht (2007) and received his MPhil and MSc degrees in Economics from Erasmus University Rotterdam and the Tinbergen Institute in 2009. Roel began work on his PhD thesis at the University of Amsterdam in 2009, where he was affiliated with the Center for Research in Experimental Economics and Political Decision Making. Currently, he is a postdoctoral researcher at the Social Science Research Center Berlin (WZB).
ESSAYS IN EXPERIMENTAL ECONOMICS

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 donderdag 31 Januari 2013, te 10:00 uur

doort

Roel Rogier van Veldhuizen
geboren te Wageningen
PROMOTIECOMMISSIE

PROMOTOR:
Prof. Dr. J.H. Sonnemans

OVERIGE LEDEN:
Prof. Dr. B. Irlenbusch
Prof. Dr. H. Oosterbeek
Prof. Dr. J.J.M. Potters
Prof. Dr. A.J.H.C. Schram
Prof. Dr. A.R. Soetevent

Faculteit Economie en Bedrijfskunde
I would like to take this opportunity to thank the people who have made this thesis possible both inside and outside the University of Amsterdam.

I would like to start by thanking my supervisor Joep Sonnemans for his guidance and encouragement. Over the past three years, Joep provided me with a great deal of freedom to work on topics I wanted to pursue. The open door policy he subscribed to also meant that I could always walk into his office with comments and questions; this is a rare luxury for a PhD student for which I am very grateful. I would also like to thank him for helping me design, program and write up the experiments of chapters 2 and 5 as a co-author, as well as providing invaluable comments for chapters 3 and 4.

Others have also directly contributed to this thesis as co-authors: I am grateful to Thomas de Haan and Hessel Oosterbeek for their contributions to chapter 4 and chapter 5 respectively. I would also like to thank Bernd Irlenbusch, Hessel Oosterbeek, Jan Potters, Arthur Schram and Adriaan Soetevent for agreeing to be on my doctoral committee.

I consider myself lucky to have had many great colleagues at the University of Amsterdam, both at CREED and beyond. Yang, Thomas, Thomas, Te, Sander, Robert, Noémi, Nadège, Matze, Matthijs, Karin, Julián, Jos, Jona, Jeroen, Gönül, Frans, Boris, Ben, Anita, Aljaž, Ailko and Adrian: thanks for making the University of Amsterdam such a pleasant environment to work in. I would like to especially thank my roommates Pedro and Marcelo: doing research requires just the right combination of labor and dialogue, which our years of sharing an office provided. I would also like to express my gratitude to Theo, who helped me arrange a visit to the University of California, San Diego, from where I am now finishing this thesis. Finally, I would like to thank the University of Amsterdam for providing me with a first-hand experience of what it is like to work at a construction site during the final 1,5 years of writing.
Beyond the University of Amsterdam there was also the Tinbergen Institute, where I spent many days taking and teaching courses, being part of the Educational Board and the TI Magazine and organizing the PhD lunch seminar series. Thanks to Berber, Eva, Gijs, Guangyuan, Jordan, Marcin, Marcel, Stephen, Tiantian, Tim, the Poker gang and others for making TI life enjoyable especially during the first two years. Special thanks also to Nan, Petr, Xin, Te and Grega for experiencing the thrills of the job market with me.

Beyond academia there were those who provided welcome distractions to academic life. I would like to express my gratitude to Ruben and Frans for spending many evenings playing Game of Thrones, RoboRally and other games, to Jan for sharing a season ticket to a great football team for the last five years, to Dirk for having great potential (most of the time) and to Harmen for all of the above. I also managed to greatly improve my understanding of the beautiful game thanks to the skills of the BallaBalla team, which is bound to serve me well when playing football in Germany in the future.

Last but not least, I would like to thank my family, starting with my brothers Micha and Sofian, who share with me a bond that only siblings can have. And finally also my parents, without whom I would not have been around and who I have always been able to rely on for unquestioning support.

La Jolla, 28 November 2012
## Contents

1 Introduction ............................................. 1
   1.1 Overview ............................................ 2

2 Nonrenewable Resources, Strategic Behavior and the Hotelling Rule ............................................. 5
   2.1 Introduction ............................................ 5
   2.2 Literature Review ..................................... 12
   2.3 Theoretical Framework ................................ 15
   2.4 Hypotheses ............................................ 19
   2.5 Experimental Design ................................... 23
   2.6 Results ................................................. 27
      2.6.1 Prologue ........................................... 30
      2.6.2 Producer Focus in the Main Part .................. 31
      2.6.3 Comparing the Prologue and the Main Part ....... 33
      2.6.4 Market Outcomes in the Main Part ............... 35
   2.7 Discussion ............................................ 42
   Appendix 2.A Re-Examining Predictions ................. 44
   Appendix 2.B Experimental Instructions ................. 48
   Appendix 2.C Screenshot ................................... 58

3 The Influence of Wages on Public Officials’ Corruptibility: a Laboratory Investigation ................. 59
   3.1 Introduction ............................................ 60
   3.2 Background ............................................ 61
   3.3 The Bribery Model ..................................... 64
   3.4 Experimental Design ................................... 66
## CONTENTS

3.5 Hypotheses ................................................. 69
3.6 Results ...................................................... 70
3.6.1 Bribe Acceptance ........................................ 71
3.6.2 G and B Choices ........................................ 74
3.7 Robustness: Bribery without Monitoring .................. 74
3.8 Discussion .................................................. 79
Appendix 3.A Instructions ...................................... 81
Appendix 3.B Screenshot ....................................... 85

4 Willpower Depletion and Framing Effects .................. 87
4.1 Introduction ................................................ 87
4.2 Experimental Design ...................................... 90
4.2.1 Depletion Task .......................................... 90
4.2.2 Secondary Tasks ......................................... 91
4.2.3 Procedure ............................................... 96
4.3 Research Questions ........................................ 97
4.4 Results ..................................................... 100
4.4.1 Stroop Task .............................................. 100
4.4.2 Attraction Effect Task ................................. 102
4.4.3 Compromise Effect Task ............................... 103
4.4.4 Anchoring Effect Task ................................. 105
4.4.5 Prisoner’s Dilemma Task .............................. 106
4.4.6 Cognitive Task .......................................... 108
4.4.7 Depletion over Successive Tasks ...................... 108
4.4.8 Correlations between Tasks ............................ 109
4.5 Discussion .................................................. 111
Appendix 4.A Lotteries of the Compromise and Attraction Effect Tasks 113
Appendix 4.B Experimental Instructions ....................... 115
4.B.1 Initial Instructions ..................................... 115
4.B.2 Compromise/Attraction Effect Task .................... 116
4.B.3 Anchoring Effect Task .................................. 117
4.B.4 Prisoner’s Dilemma Task ............................... 117
4.B.5 Cognitive Task .......................................... 118
Appendix 4.C Screenshot of the Stroop Task .................... 120
CONTENTS

5 Peers at Work in the Lab ................................. 121
  5.1 Introduction ........................................ 121
  5.2 Overview of Mas and Moretti (2009) ................. 124
  5.3 Experimental Design ................................ 127
    5.3.1 Baseline Stage .................................. 127
    5.3.2 Production Stage ................................ 128
    5.3.3 End of the Experiment ......................... 130
  5.4 Hypotheses ........................................... 130
  5.5 Results ............................................... 132
    5.5.1 Peer Effects .................................... 132
    5.5.2 Observability .................................... 133
    5.5.3 Ability ........................................... 134
    5.5.4 Additional Effects ............................... 135
    5.5.5 Comparing Treatments ........................... 140
    5.5.6 Why no Peer Effects? ......................... 140
  5.6 Discussion ........................................... 144
Appendix 5.A Instructions ................................. 146
Appendix 5.B Screenshots ................................. 154

Bibliography ................................................. 156

Samenvatting (Summary in Dutch) ......................... 173
List of Tables

2.1 Experimental Time Line .................................................. 21
2.2 Unconstrained Oligopoly Production Function .................... 30
2.3 Main Part Production Function .......................................... 34
2.4 Correlations between the Prologue and the Main Part ............ 35
2.5 Main Part Market Quantity and Benchmarks ....................... 40
2.6 Main Part Production Function with Instrumental Variables ..... 45
2.7 Main Part Production Function in Rounds with and without Expectation Elicitation ........................................... 46
3.1 Motivations for Transfer Acceptance .................................. 73
3.2 Overview of Transfer Acceptance Rates .............................. 78
3.3 Importance of Charity in Transfer Acceptance ..................... 79
4.1 Payoffs in the Prisoner’s Dilemma Task ............................ 96
4.2 Stroop Task Statistics ...................................................... 102
4.3 Attraction Effect Task Statistics ....................................... 103
4.4 Compromise Effect Task Statistics .................................... 104
4.5 Compromise Effect Task Statistics by Riskiness ................... 104
4.6 Anchoring Effect Task Estimates ...................................... 106
4.7 Prisoner’s Dilemma Task Statistics ................................... 106
4.8 Prisoner’s Dilemma la Probit Estimates ............................. 108
4.9 Cognitive Task Statistics .................................................. 108
4.10 Stroop Task Earnings per Round ...................................... 109
4.11 Correlations between Tasks ............................................. 110
4.12 Lotteries of the Attraction Effect Task .............................. 113
4.13 Lotteries of the Compromise Effect Task ............................ 114
5.1 Main Results of Mas and Moretti (2009) ........................................ 126
5.2 Peer Effects Estimates .......................................................... 133
5.3 Peer Effects Estimates by Observability ..................................... 134
5.4 Peer Effects Estimates by Ability ............................................. 135
5.5 Peer Effects Estimates in Contemporaneous Production ............ 137
5.6 Peer Effects Panel Estimates .................................................... 138
5.7 Peer Effects Estimates by Self-Monitoring ................................. 139
5.8 Peer Effects Estimates by Treatment ...................................... 142
5.9 Recall of Production Stage ...................................................... 143
5.10 Recall of Permanent Productivity .......................................... 143
5.11 Variation in Production Speed ............................................... 144
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Crude Oil Prices (2008 Dollars)</td>
<td>7</td>
</tr>
<tr>
<td>2.2</td>
<td>Resource Prices (1949 Dollars)</td>
<td>8</td>
</tr>
<tr>
<td>2.3</td>
<td>Benchmark Prices and Quantities</td>
<td>18</td>
</tr>
<tr>
<td>2.4</td>
<td>Monopoly Results</td>
<td>28</td>
</tr>
<tr>
<td>2.5</td>
<td>Unconstrained Oligopoly Results</td>
<td>29</td>
</tr>
<tr>
<td>2.6</td>
<td>Treatment LOW Scarcity Rents</td>
<td>36</td>
</tr>
<tr>
<td>2.7</td>
<td>Treatment LOW Density Plots</td>
<td>37</td>
</tr>
<tr>
<td>2.8</td>
<td>Treatment HIGH Scarcity Rents</td>
<td>38</td>
</tr>
<tr>
<td>2.9</td>
<td>Treatment HIGH Density Plots</td>
<td>39</td>
</tr>
<tr>
<td>2.10</td>
<td>Decision Screen</td>
<td>58</td>
</tr>
<tr>
<td>3.1</td>
<td>The Experimental Game Tree</td>
<td>65</td>
</tr>
<tr>
<td>3.2</td>
<td>Incidence of Transfers and B Choices</td>
<td>71</td>
</tr>
<tr>
<td>3.3</td>
<td>Fraction of Accepted Transfers</td>
<td>72</td>
</tr>
<tr>
<td>3.4</td>
<td>Fraction of B Choices</td>
<td>75</td>
</tr>
<tr>
<td>3.5</td>
<td>Transfer Acceptance Rate with and without Monitoring</td>
<td>77</td>
</tr>
<tr>
<td>3.6</td>
<td>Decision Screen</td>
<td>85</td>
</tr>
<tr>
<td>4.1</td>
<td>The Distribution of Earnings in the Stroop Task (in Cents)</td>
<td>101</td>
</tr>
<tr>
<td>4.2</td>
<td>Depletion Effects over Successive Rounds</td>
<td>110</td>
</tr>
<tr>
<td>4.3</td>
<td>Example Used for the Compromise and Attraction Effect Tasks</td>
<td>116</td>
</tr>
<tr>
<td>4.4</td>
<td>Example Used for the Anchoring Effect Task</td>
<td>117</td>
</tr>
<tr>
<td>4.5</td>
<td>Example Used for the Prisoner’s Dilemma Task</td>
<td>118</td>
</tr>
<tr>
<td>4.6</td>
<td>Example Used for the Cognitive Task</td>
<td>119</td>
</tr>
<tr>
<td>4.7</td>
<td>Screenshot of the Stroop Task</td>
<td>120</td>
</tr>
<tr>
<td>5.1</td>
<td>Team Overview</td>
<td>129</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>--------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>5.2</td>
<td>Production Speeds</td>
<td>136</td>
</tr>
<tr>
<td>5.3</td>
<td>Production Speeds by Treatment</td>
<td>141</td>
</tr>
<tr>
<td>5.4</td>
<td>Screenshot of the Baseline Phase</td>
<td>154</td>
</tr>
<tr>
<td>5.5</td>
<td>Screenshot of the Production Phase</td>
<td>155</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

As the title of this thesis suggests, this thesis is not unified by one overarching topic. Rather, the common theme of all chapters is their use of laboratory experiments. Laboratory experiments have been used in the sciences for many centuries to establish causal relationships between two or more variables. In economics, they have been adopted more recently but serve a similar purpose. In particular, they make it possible to establish causality by actively intervening in a controlled economic environment that has been created by researchers to test a specific hypothesis. Laboratory experiments are most appropriate if real world data are not available and cannot be generated using field experiments or if testing a particular hypothesis requires a level of control and cleanliness that cannot be achieved outside a laboratory setting.

Laboratory experiments can serve multiple roles, which are reflected in the different chapters of this thesis. Perhaps the first application of experimental economics was to test economic theories and see under which circumstances they are more (or less) likely to hold. I take a similar approach in chapter 2 in which I present an experimental test of a well known theory from resource economics. Experiments have also been used to test proposed economic institutions or policies. Chapter 3 uses this approach to investigate if increasing the wages of public officials decreases their corruptibility. Chapter 4 goes back to the approach of chapter 2 and takes a theory about willpower from a different discipline (psychology) and tests if this theory also applies in economic situations. Chapter 5 represents a newer approach in the literature. Rather than testing theories, institutions or policies, this chapter tries to replicate the results of a well-known empirical paper
that has been hard to replicate in the field.

In the remainder of this introduction, I give a brief summary of subsequent chapters. Each of these chapters were originally separate articles written either individually (chapter 3) or with different sets of coauthors (other chapters). The order of presentation is not thematic but chronological; chapter 2 was written first. A summary in Dutch can be found towards the end of this thesis.

1.1 Overview

Chapter 2 describes an experiment on nonrenewable resource markets. One of the most important theoretical findings in nonrenewable resource economics is that prices (or more accurately ‘scarcity rents’) should increase over time at the rate of interest. However, this finding -the so-called Hotelling rule- is not observed in practice. In real terms, resource prices have actually remained remarkably constant over time. Although many explanations for this empirical finding have been proposed, none of them have proven fully satisfactory. This encouraged Joep Sonnemans and myself to think about whether a laboratory experiment could present an alternative way to address this finding.

In this chapter we propose that resource owners follow the Hotelling rule when they have a relatively small amount of resources remaining. On the other hand, when resources are plentiful, resource owners deviate from the Hotelling rule as they focus on more immediate concerns. In the experiment, participants play the role of resource owners who have to allocate their resources over several time periods. Some have a small remaining stock of resources, whereas others have a large stock of resources available. We find that the former group on average follows the Hotelling rule almost perfectly, whereas the latter group extracts too many resources, in line with our predictions. Thus, we conclude that Hotelling’s theory is more likely to hold up in situations where resource owners have relatively few resources remaining.

In chapter 3 I investigate if increasing public officials’ wages decreases their corruptibility. Theory (starting with Becker and Stigler 1974) suggests that increasing the wages of public officials should reduce their corruptibility. If this relationship holds, it provides governments with a policy instrument that falls directly under its control and would therefore be relatively easy to implement.
However, empirical evidence for this relationship has been hard to come by. Since corrupt activities are usually not incorporated into official accounts, direct measures of corruption are only rarely available. As a consequence, empirical studies have often been forced to rely on indirect measures of corruption. Additionally, even if more direct measures of corruption are available, establishing causality may still be problematic.

Laboratory experiments allow us to obtain a more direct measure of corruption as well as establish causality. This chapter presents the results of a laboratory experiment in which participants in the role of public officials either accept or reject a bribe and then decide between a neutral and a corrupt action. The corrupt action benefits the briber but poses a large negative externality on a charity. In the experiment, I exogenously vary public officials’ wages and find that increasing public officials’ wages reduces their corruptibility. In particular, it makes experienced public officials less likely to accept a bribe and reduces the number of corrupt actions as well.

Chapter 4 starts from a recent literature in social psychology that argues that exercising self-control requires willpower and that people’s willpower should be regarded as a limited resource that can be temporarily depleted. In this chapter, Thomas de Haan and I use a laboratory experiment to investigate if depleting people’s willpower increases their susceptibility to framing effects. Framing effects occur when people’s decisions are affected by seemingly irrelevant aspects of decision problems. A well known example is the anchoring effect, which shows that presenting people with a random or unrelated number changes their subsequent estimate of the true value of a statistic. Typically, framing effects can be mitigated but the mitigation process makes use of higher order cognitive processes, which require willpower.

In the experiment, we first deplete participants’ willpower and subsequently have them take part in a series of framing tasks, including a framed prisoner’s dilemma, an attraction effect task, a compromise effect task and an anchoring task. We expected willpower depleted participants to be more susceptible to framing effects. However, we found no evidence that framing effects are indeed more prevalent in willpower depleted participants than in controls. This suggests that further research should investigate under what conditions the effect of willpower depletion on framing is likely to appear.
Finally, in chapter 5 Hessel Oosterbeek, Joep Sonnemans and I attempt to replicate the results of Mas and Moretti (2009) in the laboratory. In what has become a very influential study, Mas and Moretti (2009) show that cashiers are influenced by the productivity of those coworkers who can observe them during the production process (which is a peer effect). Many subsequent studies have assumed that the results of this study generalize to other contexts. Whether this assumption is warranted should ultimately be an empirical question, however. Replications in other settings could inform us on whether their results can indeed be extrapolated to other contexts. However, in practice such replications are difficult to do because of the level of detail required in the data. Hence, we use a laboratory experiment to approximate the key characteristics of Mas and Moretti (2009). To the extent that the findings of Mas and Moretti (2009) reflect fundamental aspects of human behavior that apply outside the original context, they should be replicable in a laboratory environment as well.

We designed the experiment specifically to capture the fundamental characteristics of the production process of Mas and Moretti (2009). Participants in the role of workers are members of teams and are not financially dependent on the effort of other workers but instead face a higher workload if their team members exert lower effort. Workers perform a real effort task, we vary worker observability and we obtain a direct measure of (permanent) productivity. In line with Mas and Moretti (2009), we expected workers to work harder when partnered with more productive coworkers who can observe them during the production process. However, we find that this result does not generalize to the laboratory. We find no evidence of peer effects despite finding that workers are well aware of the production and productivities of their coworkers when this information is available and have the ability to both increase and decrease their effort levels. This suggests that the findings of Mas and Moretti (2009) may not be as general as has been assumed.
Chapter 2

Nonrenewable Resources, Strategic Behavior and the Hotelling Rule

In this chapter, I use the methods of experimental economics to investigate possible causes for the failure of the Hotelling rule for nonrenewable resources. The argument is that as long as resource stocks are large enough, producers may choose to (partially) ignore the dynamic component of their production decision, shifting production to the present and focusing more on strategic behavior. Experimentally varying stock size in a nonrenewable resource duopoly setting shows that producers with large stocks indeed pay significantly less attention to variables related to dynamic optimization, which leads to a failure of the Hotelling rule.

2.1 Introduction

Today those who plan for the future prosperity of their nation realize the extent to which other raw materials are essential to the general well-being, and for some of these we can see no adequate substitutes. Foremost among these most useful and least abundant (...) commodi-

---

1This chapter is based on Van Veldhuizen and Sonnemans (2012). We would like to thank Adrian de Groot Ruiz, Thomas de Haan, Jona Linde, Arthur Schram, Jeroen van de Ven and Ailko van der Veen. We are also grateful to seminar participants at the European Economic Review Talented Economists Clinic 2010, the University of Heidelberg, the University of Amsterdam, the Tinbergen Institute, the 2010 ESA World Meeting, IMEBE 2011 and M-BEES 2011 for their helpful comments. Financial support from the University of Amsterdam Research Priority Area in Behavioral Economics is gratefully acknowledged.
ties stands mineral oil. (...) [Even] the most optimistic American may well ask himself, Where will my children and children’s children get the oil? - George Otis Smith, National Geographic (1920)

From the 19th century American gold rushes to the 21st century quest for drilling rights on the North Pole, there has always been something special about nonrenewable resources. Nonrenewable resources share the characteristic that they cannot be replenished, meaning that persistent use will eventually lead to physical or economic depletion (i.e. such that the remaining stock will not be worth extracting anymore). The nonrenewable resource family contains a broad variety of materials from oil and natural gas to iron and phosphate. Many nonrenewable resources have been an important part of our daily lives for so long that even in 1920 the geologist George Otis Smith deemed them “essential” (quote above; Smith, 1920).

It is precisely because of the importance of these nonrenewable resources that politicians, geologists and companies alike have been concerned with resource depletion for a long time. Indeed, the words George Otis Smith wrote down some 90 years ago are surprisingly similar to comments made in recent years about coal depletion (e.g. Heinberg, 2007), phosphate depletion (e.g. D´ery and Anderson, 2007) or oil depletion (e.g. Deffeyes, 2005). These concerns are rooted in the fear that if we continue to remain dependent on nonrenewable resources, we run the risk of economic collapse once these resources are no longer available. This has led to calls for governments to actively intervene and aid in the development of renewable alternatives. Indeed, former president Bill Clinton remarked in 2006 that “we may not have as much oil as we think, so we need to get in gear [and reduce oil dependence]” Bulletin (2006), with then president George W. Bush going one step further by stating that the United States should “get off oil” (Mouawad, 2008).

Luckily, economic theory suggests that the situation may not be quite so bad. Hotelling (1931) showed that in a perfectly competitive industry, nonrenewable resource producers will deplete the resource at the socially optimal rate. Moreover, in the presence of market power (Solow, 1974) or the presence of a constant severance tax (Heaps, 1985), the market will actually extract at a lower rate. To the extent that these factors are important, we should therefore be worried about nonrenewable resources being exhausted too slowly. In theory, the ideological
successors of George Otis Smith can thus relax knowing that depletion -when it occurs- is likely to occur at the socially optimal time or later, provided that Hotelling’s framework holds. Yet how confident can we be that nonrenewable resource producers actually follow the Hotelling approach? Hotelling (1931) showed that in a perfectly competitive environment with zero marginal costs and constant demand (real) prices should grow at the rate of interest -a result which has become known as the Hotelling rule. More generally, prices may in fact grow at a larger or smaller rate depending on the assumptions, yet they should always grow in the long run. How well, then, does Hotelling’s framework fit the real world? Figure 2.1 gives a time series of crude oil prices since the 1860s. Clearly, there are occasional periods of increasing prices, yet real prices have overall remained around the same level despite an enormous increase in production (Hall and Hall 1984, Adelman 2002). Moreover, this pattern is by no means unique to oil prices; figure 2.2 shows that -like crude oil- copper, zinc and

\[ \text{In the short run they may temporarily decrease under some assumptions, for example if extraction costs are positive and decreasing over time. However, prolonged stretches of non-increasing resource prices are implausible; see the next section for more details.} \]
CHAPTER 2. NONRENEWABLE RESOURCES

Figure 2.2: Resource Prices (1949 Dollars)

Notes. This figure is reprinted from “Should We Worry About The Failure Of The Hotelling Rule” by T. Kronenberg, 2008, Journal of Economic Surveys, 22(4), 774-93.

Iron ore prices have also not increased. More formally, in reviews of the empirical literature [Krautkraemer (1998), Kronenberg (2008) and Livernois (2009)] argue that support for the Hotelling framework is very limited.

Yet if the Hotelling framework is normatively the best way to approach the nonrenewable resource problem, this raises the question of what reason nonrenewable resource owners have had for not adopting it. In this chapter, we argue that the failure of the Hotelling rule may be the result of the multifacetedness of the nonrenewable resource problem. In particular, we argue that the nonrenewable resource problem consists of many different aspects (e.g. exploration, strategic behavior, technological developments, dynamic optimization etc.) and that in practice, producers may not be willing or able to take every aspect fully into account. Moreover we argue that the degree to which a nonrenewable resource producer pays attention to a given aspect of the resource problem depends on whether it can be feasibly included in the optimization problem, whether the benefits of including it outweigh the costs and whether the aspect is salient to the producer.

Indeed, many nonrenewable resource owners may not have sufficient computational capacity to take every aspect into account for all future periods. In fact, even including more than one aspect into a single model has proven very difficult.\footnote{See e.g. Groot, Withagen, and De Zeeuw (2003) for a discussion of some of the difficulties associated with incorporating both dynamic optimization and strategic behavior into a single model.}
Moreover even if a nonrenewable resource producer did have the ability to include all aspects of the nonrenewable resource problem in its decision making process, it might not be beneficial for her to do so from a cost-benefit perspective. For example, making accurate predictions about market demand in 15 or 20 years is likely to be quite costly, whilst a transient change would have a negligible effect on present-day extraction rates. Moreover not all aspects of the nonrenewable resource problem may be equally salient to a producer. For example, the manager of a resource firm may not be directly concerned with long-run profits if she expects to retire long before the date of exhaustion has been reached.

Although in principle there are many possible aspects to consider, in this chapter we will focus on the two aspects that are in our opinion the most crucial parts of the nonrenewable resource problem. The first key aspect of the nonrenewable resource problem is that producers always have to take into account that their current extraction decision is going to affect future extraction possibilities. This is a necessary characteristic of the Hotelling framework and a necessary condition for the Hotelling rule to hold; we will refer to it as the dynamic optimization aspect. The less attention producers pay to the dynamic optimization aspect, the more they will overproduce relative to the optimal Hotelling rule.

The other key element of most real-life incarnations of the nonrenewable resource problem is that multiple producers are active on the market, leading to the possibility of strategic behavior with respect to other producers. Though strategic behavior may not be present on all nonrenewable resource markets, it is still an element of great economic interest, as evidenced by the large number of papers focusing on this topic (see Newbery, 1981; Lewis and Schmalensee, 1980; Groot, Withagen, and De Zeeuw, 2003; Loury, 1986; Smith, 2005; among others). The less attention producers pay to the strategic behavior aspect, the less they update their production decision on the basis of the production decision of other producers.

Moreover we argue that the degree to which nonrenewable resource producers pay attention to a given aspect of the resource problem depends on the size or longevity of their resource stock. In particular, the larger (smaller) the resource model. Pindyck (1981) and Farrow (1985) give evidence that resource firms may altogether ignore some aspects of the resource problem in practice. Cairns (1986) argues that mining firms may ignore the dynamic optimization aspect in the Nickel industry.
CHAPTER 2. NONRENEWABLE RESOURCES

stock is, the less (more) attention a producer will pay to dynamic optimization (strategic behavior). Indeed, for a producer with a large stock the date of exhaustion is still far in the future, which may make it computationally difficult to stick to a dynamically optimal time path for all periods, whereas the benefits of doing so may not outweigh the costs in any case. Moreover, neither she nor her stockholders or head of state may be particularly interested in getting a dynamically optimal production path; present profits may be much more salient. On the other hand, it will be relatively profitable to behave strategically with respect to other producers and perhaps even create a cooperative agreement. Similarly, the date of exhaustion for a small stock producer is more imminent, making it more beneficial, computationally easier and more salient to take exhaustion into account. Since most nonrenewable resource producers in practice still have a large remaining stock, we should thus expect them to focus more on strategic behavior than on dynamic optimization, leading to a failure of the Hotelling rule.

Ideally, it would be possible to investigate the relationship between stock size and the applicability of the Hotelling framework using field data. However, using field data to investigate this relationship might be problematic for several reasons. One problem is that field data may be biased (a well-known example is OPEC ‘proven reserve’ data). Field data may also be unavailable altogether (especially marginal cost data; [1998]) or may simply be very noisy (e.g. because of unobserved demand shifts, small changes in technology etc., see e.g. [1985]). Moreover, even if good data are available, it may be hard to compare large stock and small stock producers, since they are likely to differ on more than just the stock dimension. Also, any observed production differences may be the result of changes in factors outside the model of interest (such as government interventions, oil booms on the stock market, see e.g. [2009]), which might not be extractable from the data or otherwise may not be easily incorporated into a dynamic model. Moreover, output changes may be the result of revised expectations, which are also rarely available from field data.

---

5 In particular, since OPEC production quotas became based on proven reserves in the early 1980s, the official estimates of some OPEC states (including Saudi Arabia, UAE, Iran and Iraq) have shown suspiciously large upward jumps in reserve levels. For example, the UAE’s proven reserve increased by nearly 200% from 1985 to 1986 (BP, 2010). See Gerlagh and Lisk for a theoretical model that provides one explanation why it may be optimal to overstate reserves.

6 Also, small stock nonrenewable resource markets are quite hard to find, since in most cases resource pools are still projected to be sufficient for several decades.
2.1. INTRODUCTION

These data concerns can, however, be addressed using laboratory experiments. In a controlled laboratory environment, it is possible to exclude factors outside the model as well as possible biases or noise by keeping the environment fixed between sessions. Expectations can also be obtained, such that revised expectations can be taken into account and be disentangled from strategic concerns. Indeed, the field of experimental economics has a large tradition of experiments in oligopoly. To our knowledge, this is the first study to investigate producer behavior in a nonrenewable resource oligopoly in a laboratory experiment.

We run an experiment in which two producers with a limited stock of non-renewable resources are paired on a nonrenewable resource market. In this way the experimental setting allows for strategic behavior and dynamic optimization whilst abstracting away from other aspects. We experimentally vary stock size and find that producers with small stocks pay significantly more attention to variables related to dynamic optimization, although the evidence for strategic behavior is not so clear-cut. The change in focus on dynamic optimization is reflected by market outcomes: in the large stock treatment extraction rates are persistently above the Nash level, whereas in the small stock treatment they are never higher than the Nash level in any period. As a consequence, the Hotelling rule is almost perfectly observed in the small stock treatment, whereas in the large stock treatment it is persistently violated through overproduction.

In the next section, we will review Hotelling’s work as well as several previous attempts at explaining the failure of the Hotelling rule. In section 3 we formulate the model that forms the basis of the experiment, which brings us to the hypotheses for the experiment in section 4. In section 5, we then go over the design of the experiment before we show the results in section 6. Finally, section 7 provides a short discussion of the results.

\footnote{See for example \cite{Huck1999, Huck2000, Abbink2008, Abbink2009, Apesteguia2007, Apesteguia2010} or see \cite{Engel2010} for an overview of the literature. See also \cite{Chermak2002, Fischer2004, Sadrieh2003, Brown2009} for experiments on dynamic optimization tasks.}
2.2 Literature Review

The origins of the field of nonrenewable resource economics can be traced back to Harold Hotelling (1931). In the spirit of an earlier work by Gray (1914), Hotelling sets out the problem of a firm - in his case the owner of a mine - facing a limited stock of resources. Hotelling’s work is notable for its novelty and for its sheer scope: it addresses not just a then new economic problem but also discusses many relevant extensions, including uncertainty, the possibility of exploration and market power.

Hotelling starts his analysis by examining the problem of a resource-constrained firm in a fully competitive market. Firms in a competitive market face a trade-off between extracting their resource in the present and extracting it at some future date. For the market to be in equilibrium and to prevent arbitrage opportunities, firms have to be indifferent about when to extract their resource. Hotelling shows that in a competitive environment with zero marginal costs, the only way to keep resource owners indifferent between extracting in the present and extracting in the future is for resource prices to grow at the rate of interest. That way, extracting a marginal unit in the present means a loss of today’s price plus the interest over today’s price, and this is equal to the benefit of extracting a marginal unit in the future. This result has become known as the Hotelling rule.

The Hotelling rule in its original form is valid only in a competitive environment with zero marginal costs. However, it can be generalized to other environments as well. In a more general form, the Hotelling rule states that the scarcity rent should grow at the rate of interest. The scarcity rent represents the excess return that producers get to compensate them for exhausting their resource. The scarcity rent is thus equal to the difference between the equilibrium price on a nonrenewable resource market and the equilibrium price on the same market if the resource had been abundant. It is also sometimes referred to as the in situ value, (marginal) user cost or shadow price of the resource. Examples of generalized Hotelling rules are presented in studies which allow for exploration possibilities or technical innovation (Pindyck 1978, 1980; Arrow and Chang 1978), allow producers to have non-profit maximizing motives (Mead 1979) and allow the market to be

---

8 In this chapter, we shall use the terms firms, producers and resource owners interchangeably.
9 See Devarajan and Fisher 1981 for an early overview of the impact of Hotelling’s work on the field.
2.2. LITERATURE REVIEW


Many of these generalizations were created to provide an explanation for the failure of the original Hotelling (1931) rule. It is possible for a generalized Hotelling rule to imply non-increasing prices under certain conditions. Intuitively, in any Hotelling-type model prices are pushed upwards over time by increasing scarcity rents. For a model to be consistent with non-increasing prices, there thus needs to be an alternative force that provides enough downward pressure on prices to compensate the upward pressure created by the increasing scarcity rents. Previous work has suggested several mechanisms through which non-increasing prices can occur within a generalized Hotelling rule.

Firstly, including exploration possibilities can lead to a U-shaped price pattern if there are stock effects in the cost function (Pindyck, 1978). That is, newly found resource stocks may be cheaper to extract, which means that marginal cost decreases may more than match increasing scarcity concerns, leading to decreasing prices. Relatedly, technological developments can also lead to decreasing marginal costs and (non-increasing or) decreasing price patterns in the short to medium run (Slade, 1982). In both cases price decreases are the result of marginal costs decreases which more than match scarcity rent increases. However, since marginal costs are bounded from below, prices will eventually have to start rising. Thus, either exploration possibilities or technological developments can only explain non-increasing resource prices in the short run; in the long run they imply a U-shaped price pattern. However, there is little evidence for a long-run U-shaped price pattern for any nonrenewable resource.

It is also possible for non-increasing prices to occur for strategic reasons. For example, if price is taken as a signal of resource abundance, it may be beneficial for resource owners to keep prices artificially low to prevent a third party from developing a renewable alternative as in Gerlagh and Liski (2011). However, their model with discounting predicts increasing prices in the short run and falling prices in the long run, which seems hard to reconcile with current price data. Alternatively, non-increasing prices can also be caused by insecure property

\[10\]

A notable exception is formed by oil prices from 1870 to 1978. Indeed Slade (1982) finds a U-shaped time pattern for this time period. However, prices have since fallen back to World War II levels. Thus her results may no longer be applicable if price data are extended beyond the 1970s.
This applies for example to the early history of American oil drilling, when property rights applied to land parcels and not oil fields, meaning that there were often multiple pumpjacks extracting oil from the same field. More recently, it also applied to the Middle East oil fields of the 1960s and 1970s, when the big American oil firms correctly anticipated that their resources would be confiscated in the near future (Mead, 1979). Yet although property rights may explain non-increasing prices for some resources in some periods, they have been quite well defined for other resources and other time periods and there, too, prices have rarely consistently increased.

There are several more extensions of the Hotelling set-up which allow prices to be non-increasing, including capacity constraints and stochastic exploration (see Krautkraemer’s 1998, Gaudet’s 2007 or Livernois’ 2009 survey of the literature for more details). Each of these mechanisms could explain the empirically observed pattern of non-increasing resource prices in the short run. At the same time, scarcity rents should still be increasing even in the short run. However, studies examining (constructed estimates of) scarcity rents have also failed to consistently reveal increasing trends (Farrow, 1985; Halvorsen and Smith, 1991; Cairns and Davis, 1998). For example, Farrow (1985) gives a case where scarcity rents actually seem to decrease over time.

What all these extensions have in common is that they attempt to reverse the implications of the basic Hotelling rule (i.e. find a model that predicts decreasing prices instead of increasing prices) while keeping the main assumption—firms dynamically optimize profits over a long time horizon—intact. However, following Pindyck (1981) and Cairns (1986) in the next sections we argue that in fact the assumption that firms dynamically optimize profits over a long time horizon—though normatively appealing—may not be descriptively accurate. Indeed, as Adelman (2002) and Hamilton (2009) argue, another way to interpret historical data on oil prices is to say that “oil prices historically hav[e] been influenced little or none at all by the issue of exhaustability” (Hamilton, 2009). This line of

---

11 This led Smith (1920) to lament “the waste of capital and labor under conditions of competitive drilling”.
12 An important exception to this point are possible stock degradation effects. Indeed, if extraction costs increase sufficiently strongly as the resource stock gets depleted, it is possible for scarcity rents to decrease over time, whilst prices would then be increasing (Livernois and Martin, 2001). However, this pattern is inconsistent with the empirically observed pattern of non-increasing prices.
reasoning is consistent with non-increasing resource prices; it is also consistent
with anecdotal and empirical evidence suggesting that simple heuristics (mining
practice) may provide a better description of actual behavior (see Farrow, 1985;
Cairns, 1986). We will come back to this idea in section IV. First, however, we
will derive the Hotelling model that forms the benchmark for the remainder of
the chapter.

2.3 Theoretical Framework

We generalize the Hotelling set-up by allowing for market power in the Cournot
sense. This way, the model allows for both dynamic optimization and strategic
behavior. Other than allowing for market power, we stick to the original
Hotelling set-up as much as possible. Hence, we abstract away from possibilities
of exploration, capital investments et cetera.

Let there thus be \( N \) symmetric producers indexed \( i \) with a per-period profit
function \( \Pi(q_i^t, Q_t) \) that depends on the producer’s quantity of the resource sold
in period \( t \) \( (q_i^t) \) as well as the market quantity sold in period \( t \) \( (Q_t = \sum_{j=1}^{N} q_j^t) \).
Moreover, each producer \( i \) faces a resource constraint which limits total production
over all periods to be no larger than an initial private resource stock \( S_0^i \). There
is a common discount factor \( \delta \) which is equal to \( \frac{1}{1+r} \), where \( r > 0 \) is the market
interest rate.

A first thing to note about this setup is that we use a discrete time rather than
a continuous time framework. Although a continuous time framework is more
commonplace in the literature, a discrete time framework fits in better with the
experiment. To keep the experiment as simple as possible for participants, we also
adopt a linear demand framework with \( a \) the choke price and \( b \) the slope of the
demand function. We also assume that marginal costs are constant and (without
further loss of generality) equal to zero. We then get the following specification
for the profit function:

\[
\Pi(q_i^t, Q_t) = (a - bQ_t)q_i^t
\]

Producers maximize the sum of discounted profits subject to the resource
constraint. The solution to the producer problem depends on the assumptions
that the producer makes about the market quantity $Q_t$. Offerman, Potters, and Sonnemans (2002) mention three benchmarks, which differ only in the degree to which individual producers think they can influence the market quantity $Q_t$. For the Nash equilibrium benchmark, producers assume that they can only influence their own production strategies; they treat the production strategies of other producers as given. In the second benchmark -Collusion- producers maximize joint profits. Finally, for the Walras (or competitive) benchmark, producers (mistakenly) believe that no firm has the ability to influence the market price and hence the market quantity (i.e. producers assume that $Q_t \perp q^i_t \forall i$).

Of the three benchmarks, the Collusive and Walras benchmarks are essentially individual optimization problems, since in both cases producers assume that there are no other parties on the market that can influence their profits. Thus, both the Collusive problem and the Walras problem can be solved using calculus of variations. Letting $0 < T \leq \infty$ be the maximum number of periods, the Lagrangian becomes:

$$L = \sum_{t=0}^{T} \delta^t (a - bQ_t)q^i_t - \lambda^t (\sum_{t=0}^{T} q^i_t - S^i_0)$$

Here $Q_t = Nq^i_t$ for the Collusive benchmark and $Q_t = Nq^W_t$ for the Walras benchmark, where $q^W_t$ is the average quantity on the Walrasian market. Plugging these expressions for $Q_t$ into the Lagrangian, taking the derivative with respect to $q^i_t$ and $q^i_0$, and then by symmetry putting $q^i_t = q^W_t$ for the Walras benchmark yields the following expression:

$$q^i_t \geq q_U - \frac{q_U - q^i_0}{\delta^t} \text{ with } q^C_U = \frac{a}{2Nb} \text{ and } q^W_U = \frac{a}{Nb}$$

(2.1)

This is the Hotelling rule for Walrasian or Collusive symmetric oligopolies expressed in terms of quantities. Here, $q_U$ is the unconstrained or static benchmark quantity, which differs depending on the benchmark that is adopted; it is equal to the quantity that would be produced in the absence of resource scarcity, see below.
2.3. THEORETICAL FRAMEWORK

By summing over all firms, equation 2.1 can also be rewritten in terms of prices:

\[ p_i^t \leq p_U + \frac{p_0 - p_U}{\delta t} \quad \text{with} \quad p_U^C = \frac{a}{2} \quad \text{and} \quad p_U^W = 0 \] (2.2)

The two remaining steps are to use the resource constraint to find the optimal \( q_0 \) (or equivalently \( p_0 \)) and the optimal time of exhaustion \( t_* \). This procedure, though mathematically straightforward, is somewhat tedious and thus omitted. Turning our attention back to equation 2.2, the first term on the right is the unconstrained benchmark price. The difference between the actual market price and the unconstrained market price is made up by the second term on the right \((\frac{p_0-p_U}{\delta t})\) which is the scarcity rent of the resource. This term is positive and exponentially increasing; as a result prices will increase exponentially with respect to the unconstrained benchmark.

Solving for the Nash equilibrium requires the use of dynamic game theory (see Başar and Olsder, 1999). Salo and Tahvonen (2001) solve for the Nash equilibrium for a continuous and infinite time framework with a continuous action space. However, the setup we use in the experiment is simpler to analyze because it uses a finite time horizon, a discrete time framework and integer production quantities. This allows us to solve for the Nash equilibrium numerically using a recursive procedure. Figure 2.3 shows the benchmark price and quantity levels for one of the parameter combinations used in the experiment (for treatment HIGH). The figure shows that prices are increasing at the highest rate in the Walras benchmark and at the lowest rate in the Collusive benchmark. This implies that \( p_0 \) is lowest for the Walras benchmark and highest for the Collusive benchmark.

It is important to note that for figure 2.3 we assumed that producers stick to each benchmark perfectly. However, both in the experiment and in real life it is possible that producers make mistakes or switch between benchmarks after period 1. To allow for these possibilities, we also calculated the Nash, Collusive

\footnote{In the terminology of Başar and Olsder, 1999, we are solving for the feedback Nash equilibrium. Since the producer problem in the experiment is a ladder-nested multi-act feedback game, the numerical procedure we use for the feedback Nash equilibrium is the one described by Başar and Olsder, 1999, on page 119-121.}

\footnote{Since high prices and low production levels go together, collusion actually leads to slower extraction and greater conservation of the resource. This point was also noted by Solow, 1974, who argued that “if a conservationist is someone who would like to see resources conserved beyond the pace that competition would adopt, then the monopolist is the conservationists friend. No doubt they would both be surprised to know it.” (Solow, 1974, p. 5)}
Figure 2.3: Benchmark Prices and Quantities

Notes. This figure plots the symmetric benchmark market prices and quantities for treatment HIGH of the experiment for all periods.
and Walras strategies for every possible state of the market (i.e. every possible period/stock combination); these are the benchmarks we compare our results to in the results section.

Finally, in what follows we will sometimes refer to unconstrained or static benchmarks. The unconstrained benchmark quantities are the quantities that would be adopted in the absence of resource scarcity; the market quantities $Q_t$ are equal to $\frac{N}{N+1} \frac{a}{b}$, $\frac{1}{2} \frac{a}{b}$ and $\frac{a}{b}$ for Nash, Collusion and Walras respectively. From equation (2.1) it is easy to see that unconstrained benchmarks always encompass larger production levels (and thus lower prices) than their dynamic counterparts.

### 2.4 Hypotheses

In previous sections we saw that the Hotelling rule does not seem to describe the data very well. In this chapter, we argue that the failure of the Hotelling rule may be the result of the multifacetedness of the Hotelling framework. Indeed the nonrenewable resource problem is a mixture of several aspects, including for example exploration, strategic behavior, technological developments, dynamic optimization etc. Moreover, we argue that resource owners do not always take every aspect fully into account in their decision making process. In this section, we will examine this line of reasoning in greater detail and relate it to the hypotheses that are tested in the experiment.

One reason why producers do not take all aspects fully into account simultaneously is that doing so may be computationally impossible. Indeed, despite the enormous financial capabilities of some nonrenewable resource producers, even a very rich nonrenewable resource owner may not have a large enough computational capacity to take every aspect into account for all future periods. Pindyck (1981) argues that having a limited computational capacity may induce producers to (partially) ignore the dynamic consequences of their extraction decision. In general, including more than two aspects into a Hotelling framework may make the optimization problem intractable. Thus, resource owners have to make choices on what aspects of the decision problem they are going to focus their computational resources.

Secondly, even if a producer did have the ability to include every aspect of the nonrenewable resource problem in its decision making process, it might not be
beneficial for her to do so from a cost-benefit perspective. For example, computing the dynamically optimal time path would require a knowledge of future demand elasticities. However, making accurate predictions about future demand elasticities is likely to be quite costly. An accurate prediction would for example require incorporating the expected availability of a backstop technology in the future, the expected sensitivity of consumers to environmental issues, expected population growth etc. At the same time, even a sizable change in the expected demand elasticity in 15 or 20 years may not affect the optimal current extraction rate very much. Indeed, for a resource stock that will not be depleted for many decades, even ignoring the resource constraint completely may not lead to a very different time path in the short run, since initial production levels may already be quite close to the unconstrained level. Thus, if fully incorporating certain aspects of the resource problem is costly and the benefits of doing so are small at least in the short run, then incorporating these aspects into the decision process may not be worth the costs.

Thirdly, not all aspects of the nonrenewable resource problem may be equally salient to a producer. For example, the manager of an oil firm with a substantial remaining resource pool - particularly one who expects to retire or move jobs in the not too distant future - may not be very mindful of finding a dynamically optimal production strategy. Indeed, doing so will be beneficial for the company in the long run only and may even reduce profits in the short run. As Cairns (1986) and Slade (1988) have argued, this means that in practice the concerns of mining firms are likely to be dominated by price volatility, capital accumulation or cost control rather than dynamic optimization. Similarly, resource state companies might be under pressure from their government to acquire immediate income to finance public spending (see e.g. Ezzati, 1976; Teece, 1982; Griffin, 1985). Since the goal of increasing short term revenue is likely to be in direct conflict with maximizing long term profits, this will also lead the producer off the dynamically optimal path.

Thus, the degree to which a nonrenewable resource producer pays attention

\footnote{An interesting example is provided by the island republic of Nauru. The surface of the island of Nauru consisted almost entirely of phosphate, a nonrenewable resource used to produce fertilizer. In the late 60s and early 70s, phosphate production had made Nauru so rich that as a country it had the highest GDP per capita in the world. However, its government used the proceeds to finance lavish public spending which eventually led to bankruptcy when the phosphate income started to fall in the 1990s (see e.g. Cox and Kennedy, 2005).}
2.4. HYPOTHESES

Table 2.1: Experimental Time Line

<table>
<thead>
<tr>
<th>Prologue 1: Nonrenewable Resource Monopoly</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction &amp; Check up Questions</td>
</tr>
<tr>
<td>2. Practice</td>
</tr>
<tr>
<td>3. 1 Incentivized Round</td>
</tr>
<tr>
<td>(35 min)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prologue 2: Unconstrained Oligopoly</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction &amp; Check up Questions</td>
</tr>
<tr>
<td>2. 3 Incentivized Rounds</td>
</tr>
<tr>
<td>(30 min)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Main: Nonrenewable Resource Oligopoly</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction &amp; Check up Questions</td>
</tr>
<tr>
<td>2. 10 Incentivized Rounds</td>
</tr>
<tr>
<td>(65 min)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROLOGUE 1</td>
</tr>
<tr>
<td>Stock</td>
</tr>
<tr>
<td>a</td>
</tr>
<tr>
<td>b</td>
</tr>
<tr>
<td>Interest Rate</td>
</tr>
<tr>
<td>Periods</td>
</tr>
<tr>
<td>Producers</td>
</tr>
</tbody>
</table>

Parameters to a given aspect of the resource problem depends on whether it can be feasibly included in the optimization problem, whether the benefits of including it outweigh the costs and whether the aspect is salient to the producer. In this chapter, we focus on what we have previously argued are two of the most important aspects of the nonrenewable resource problem: dynamic optimization and strategic behavior. Here, the dynamic optimization aspect refers to the ability to allocate resource production over time in a dynamically efficient way, whereas the strategic behavior aspect refers to the ability to base a production strategy on the expected production strategy of other producers. Moreover, we propose that the degree to which producers focus on either dynamic optimization or strategic behavior is a function of the size of the remaining private resource stock.

We experimentally induced variation in stock size by running two treatments

---

16 Although the rest of the chapter will focus on these two aspects, we do believe that the subsequent analysis also extends to other aspects.
of a non-renewable resource duopoly along the lines of the previous section (with $N$ equal to 2). In treatment LOW, the unconstrained collusive quantity—which is the smallest of the three unconstrained benchmark quantities—could be maintained for only one period. In treatment HIGH, firms had a larger stock; as a result the unconstrained collusive benchmark could be maintained for up to five periods. Table 2.1 gives an overview of the parameters corresponding to the two treatments.\footnote{Besides stock there were two other parameters which differed between treatments. These were fixed costs and the conversion rate of experimental points to Euros. They were changed to create similar incentives in all treatments; they did not affect any of the benchmarks in any way.}

We propose that the experimentally induced variation in stock size results in a shift of relative focus between strategic behavior and dynamic optimization. We assume that producers in both treatments are computationally constrained, so that they cannot fully take both aspects into account. Moreover, we assume that the dynamic optimization aspect is both more salient and more cost-beneficial in treatment HIGH than in treatment LOW.\footnote{Relative to the static Nash level, producing according to the dynamic Nash level yields a 61\% higher revenue for treatment LOW compared to a 48\% higher revenue in treatment HIGH.} In other words, we argue that producers in treatment HIGH will pay less attention to dynamic optimization than producers in treatment LOW; conversely producers in treatment HIGH will pay more attention to strategic behavior.

To test this idea we use a panel regression to estimate what aspects of the decision problem producers take into account; we will refer to the estimate as producers’ production function for convenience reasons. With regard to the dynamic optimization aspect, a dynamically optimized strategy requires the extraction decision to be a function of the remaining resource stock. In terms of the estimated production function, producers who take dynamic optimization into account should—all other things being equal—extract a larger quantity of resources for higher levels of their resource stock. To the extent that producers in treatment HIGH pay less attention to dynamic optimization, these producers should then be less likely to condition their production decision on their remaining resource stock. This leads to the following hypothesis:

**Hypothesis 1A:** Producers in treatment HIGH condition their production decision less strongly on their own stock than producers in treatment LOW.
With regard to the strategic behavior aspect, a dual line of reasoning holds. In particular, for producers who behave strategically the extraction decision should be a function of the expected production level of the other producer on the market. To the extent that producers in treatment HIGH pay more attention to strategic behavior, we expect firms in treatment HIGH to be more likely to base their production decision on the expected production level of the other producer on the market. This leads to the following hypothesis:

**Hypothesis 1B:** Producers in treatment HIGH condition their production decision more strongly on the expected production level of the other producer than producers in treatment LOW.

If producers in treatment HIGH indeed focus less on the dynamic optimization aspect than producers in treatment LOW, this should also be visible in market production levels. In particular, a producer who pays no heed to the dynamic optimization aspect cannot produce according to a dynamic benchmark. Instead, the only available alternative is to produce according to an unconstrained benchmark. Since unconstrained benchmarks have a higher extraction rate than their dynamic counterparts, producers who adopt an unconstrained benchmark will push up average production levels away from the Nash level. This will also pull down prices and scarcity rents, leading to a failure of the Hotelling rule. To the extent that producers in treatment HIGH are more likely to produce according to a static benchmark, production levels should thus be further away from the Nash benchmark in treatment HIGH. This leads to the following hypothesis:

**Hypothesis 2:** Producers in treatment HIGH are more likely to overproduce relative to the Nash benchmark than producers in treatment LOW.

### 2.5 Experimental Design

The experiment was computerized using PHP/MySQL and consisted of two stages: the prologue and the main part (see table 2.1). We realized that the nonrenewable resource problem would initially be difficult for many participants to tackle. Since
we did not want the Hotelling rule to fail because of a lack of understanding, we instituted a prologue that helped participants get to know the nonrenewable resource oligopoly problem in a stepwise way. The first phase of the prologue familiarized participants with the dynamic optimization aspect and the second phase familiarized them with strategic behavior. All in all, the prologue lasted for approximately 65 minutes. The prologue was identical for both treatments; treatment variation took place in the main part only. The main part consisted of a nonrenewable resource duopoly; it will be the focus of the analysis in the next section. It lasted for approximately 65 minutes as well, bringing the total duration of the experiment to 2 hours and 30 minutes including a questionnaire and payment.

The first phase of the prologue consisted of a nonrenewable resource monopoly. Participants first received a set of instructions and check-up questions (all instructions, questions and questionnaires are reprinted in appendix B). Once every participant had finished these, the experiment moved to a 15 minute practice stage. In this practice stage each participant represented a resource owner with a limited stock of resources to be allocated over a total of 6 time periods, along the lines of the model of section 3 with \( N = 1 \). Thus, participants were monopolists, which allowed them to learn about the the dynamic optimization aspect without having to worry about strategic behavior. In every time period, each participant decided how much of her resource to extract in the current period and how much to save for the remainder. After making a decision, the participant moved on to the next period where she again had to decide how much of her resource to extract. The resulting decision problem is non-trivial because of discounting; we incorporated discounting into the experiment by explicitly introducing an interest rate, such that income earned in earlier periods would be worth more.

After the sixth and final time period, participants were informed of their total income, which was calculated by adding profits and interest incomes from

\[ \text{As a bonus, the prologue also allowed us to compare behavior in the prologue to behavior in the main part of the experiment.} \]
\[ \text{Without discounting, all three benchmarks would collapse into extracting one sixth of the stock in every period.} \]
\[ \text{A possible alternative sometimes used in the literature is to use a stochastic ending mechanism. However, explicitly incorporating an interest rate avoids issues of risk aversion, the gambler's fallacy (see e.g. [Terrel 1994] and keeps all rounds comparable (same number of periods), whilst also staying close to the theoretical framework. In any case, [Brown, Flinn, and Schotter 2011] show that both mechanisms may yield very similar results.} \]
all periods and subtracting a fixed cost. This ended the first practice round; participants could then immediately proceed to the next practice round. During practice time participants could go through as many practice rounds of the monopoly set-up as they liked. Thus, each participant had the time to check many possible production paths; as a result we expected most to get to know at least the basic rule of dynamic optimization in a nonrenewable resource context - which is to produce more in early periods than at the end. To make sure that every participant put in sufficient effort during practice, we also included a fully incentivized round after practice. All in all, the first phase of the prologue took approximately 35 minutes.

The next phase of the prologue consisted of an unconstrained duopoly. As a result, phase two allowed participants to familiarize themselves with the presence of another producer on the market (the strategic behavior aspect) without having to worry about dynamic optimization. In particular, we expected that phase two would teach participants at least the basic rule of Cournot oligopoly - which is that (up to a point) increasing production in a given period will increase your profits, but decrease the profits of the other producer on the market. Like phase one, phase two started off with a new set of instructions and questions. Participants then went through three rounds of the unconstrained oligopoly set-up; in every round participants were matched with one other participant, so that there were always two active producers on every market. Each round consisted of six periods as in the previous phase of the prologue; each market moved on to the next period once both producers had made their production decision. The experiment then moved forward towards the next round once all markets had finished all six periods; all three rounds contributed to final earnings. In total, phase two of the prologue lasted approximately 30 minutes.

After everyone had finished the prologue, the experiment moved on to the main part, which forms the basis for the analysis presented in the next section. All participants received a final set of instructions and questions and then went through ten incentivized rounds of the nonrenewable resource oligopoly framework of section 3. The main part thus incorporated both the dynamic optimization aspect and the strategic behavior aspect. Each participant represented a resource owner with a limited stock of resources (as in the first phase of the prologue);

\[22\text{On average, participants went through 26 practice rounds, with a minimum of 10 and a maximum of 53.}\]
moreover participants were paired so that there were always two active producers on every experimental market (as in the second phase of the prologue).

During the decision process the decision screen gave participants access to the production decisions of both participants as well as the price levels in preceding periods and remaining resource stocks of both participants. After the sixth and final time period, participants were informed of their total income, which was calculated by adding profits and interest incomes from all periods and subtracting a fixed cost. Once all other pairs were done as well, the experiment moved on to the next round, where the same set-up was repeated. In every round, participants were matched to a different participant in their matching group. In total, there were ten rounds, each of which was incentivized.

In general, we realized that the decision problem would be quite hard for many participants for a large time horizon or a large number of competitors, even after having familiarized them with both aspects of the nonrenewable resource problem in the prologue. Hence, we stuck to a relatively simple set-up by limiting the number of periods to six and market size to two firms. Moreover, all participants had access to an on-screen calculator which allowed them to compute the profits and interest incomes for any period and any production level of themselves and the other producer.

One final thing to note about the main part is that in every *even round* participants were asked in every period to indicate how much they expected the other firm to produce in that period. Any strategic production decision directly depends on the expected production strategy of other producers; an advantage of experiments is that expectations can be elicited directly. Predictions were incentivized; at the end of the experiment, participants received a payment.

---

23 An example of a decision screen is given at the end of Appendix B.
24 Matching groups consisted of between 6 and 10 participants, depending on the number of participants in the session. Participants could only be matched to participants from their matching group. Moreover, participants could never face the same participant twice in succession. Finally, participants never learned the identity of the other participant.
25 Indeed, we had previously run a pilot Van Veldhuizen (2009) where we had 10 periods and a group size of three and found that a small number of participants occasionally took a very long time (sometimes nearly 10 minutes) to make a single production decision. Since 98% of all decisions in the pilot were made within 90 seconds, we decided to limit the decision time per period to two minutes.
26 However, elicited expectations are not uncontroversial in the literature. In particular, they might suffer from a false consensus or reciprocity effect (see e.g. Croson 2000) and even the elicitation procedure itself may change behavior in a round (see e.g. Gächter and Renner 2010). We come back to this issue in appendix A.
depending on the accuracy of one randomly determined prediction. For this purpose, we asked one subject to come forward and roll a die to determine the prediction round and period that would be used to determine payment. Prediction income was then computed using a linear scoring rule, where a unit deviation from the actual value would reduce earnings by 20 cents, from a maximum of five to a minimum of zero Euros.

After finishing the last round of the main part, participants received an overview of their earnings over the whole experiment. They were then asked to fill out a questionnaire, which consisted of some background questions, some questions relating to the way they played in the experiment as well as the shortened version of the Stanford Time Perspective Inventory (D’Alessio et al., 2003)—a questionnaire related to time preferences. After finishing the questionnaire, participants collected their earnings and were kindly requested to leave the laboratory.

2.6 Results

The experiment was conducted in February 2010 at the CREED laboratory of the University of Amsterdam. Participants were recruited using an online registration system. Most participants were students coming from various disciplines, with the largest fraction (58%) studying economics. In total, there were 6 sessions (3 for treatment HIGH and 3 for treatment LOW) in which a total of 136 subjects took part (72 for treatment HIGH, 64 for treatment LOW). On average, participants earned 29.27 euros.

In this section, we first take a brief look at the results of the prologue to check if participants were able to independently understand both strategic behavior and dynamic optimization. We then estimate their production function to gain insight into which aspect producers paid most attention to when making their production decision. The next subsection then brings together the prologue and the main part as a second way to investigate what aspect producers paid most attention to. Finally, we examine if differences in the production function are reflected by market outcomes as well.
Figure 2.4: Monopoly Results

Notes. The top panel of the figure plots the time series of the average observed scarcity rent and the optimal scarcity rent conditional on the average remaining resource stock. The lower panel plots the smoothed density of the deviation from the optimal scarcity rent divided by the optimal quantity $\frac{\lambda_t - \lambda^*}{\sigma_q}$ (Epanechnikov kernel, bandwidth = .028). We use deviations from the optimal scarcity rent since the optimal scarcity rent differs depending on the period and the remaining resource stock. Deviations are weighted by the optimal quantity to take into account that a unit deviation from the optimum should be given more weight in periods where the expected optimal production is already quite low. Period 6 is omitted from the lower panel since in all cases the optimal decision is to fully extract the remaining resource stock.
2.6. RESULTS

Figure 2.5: Unconstrained Oligopoly Results

Notes. The top panel of the figure plots the time series of the average observed market quantities as well as the three benchmarks. The lower panel plots the smoothed density of observed market (Epanechnikov kernel, bandwidth = 4) over all periods; C, N and W represent the Cournot, Nash and Walras quantities.
Table 2.2: Unconstrained Oligopoly Production Function

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>Quantity in period ( t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other producer quantity in ( t - 1 )</td>
<td>.1512*** (.0386)</td>
</tr>
<tr>
<td>Observations</td>
<td>2040</td>
</tr>
<tr>
<td>Adj. ( R^2 )</td>
<td>.1640</td>
</tr>
</tbody>
</table>

* p<0.1%; ** P<0.05%; *** p<0.01%

Notes. This table displays a fixed effects regression of period \( t \) quantity on other producer quantity in period \( t - 1 \). Fixed effects are included for both individuals and time periods; time-individual fixed effects are omitted because of possible multicollinearity. Standard errors are clustered by individual and reported in parentheses below the associated coefficient. P-values are calculated using two-sided t-tests. The regression uses data from all three rounds; in every round the first period is omitted since lagged quantities are only available from period two onwards.

2.6.1 Prologue

The purpose of the first phase of the prologue (monopoly) was to familiarize participants with the dynamic optimization aspect of the nonrenewable resource problem. In particular, we expected participants to learn at least the basic rule of dynamic optimization, which dictates that scarcity rents should be monotonically nondecreasing over time. This expectation is supported by the data: scarcity rents were monotonically nondecreasing for 90% of our participants (or 123/136). Moreover, 89% (121/136) displayed a significant positive time trend in scarcity rents. Furthermore, 87% (118/136) exceeded the earnings corresponding to a constant production schedule. In fact, the median participant was within 5 cents (or 4%) of the maximum (theoretical) pay-off. However, as figure 2.4 shows, participants display a small tendency to exhaust the resource prematurely: on average the scarcity rent was 4.80% lower than the optimal level (\( t(136)=-1.95, p<0.1 \)). This echoes the findings of Brown, Chua, and Camerer (2009) who find that in a dynamic savings experiment, participants also tend to save too little. On the whole, however, most producers were very close to the optimum, suggesting that they managed to achieve at least a basic understanding of dynamic optimization.

\[^{27}\] The trend was estimated using a linear regression of scarcity rent on a constant and a linear time trend; significance was obtained using a two-sided t-test with a significance level of 5%.

\[^{28}\] More accurately, their earnings exceeded the earnings corresponding to the extraction of 47,47,47,47,46,46 in periods 1 to 6 respectively.
2.6. RESULTS

The purpose of phase two of the prologue (unconstrained oligopoly) was to familiarize participants with the strategic behavior aspect. As the top panel of figure 2.5 shows, market production levels were quite close to the Nash benchmark on average. In terms of the distribution, there is also a small peak at the collusive quantity; moreover at the producer level, production levels are more dispersed than at the market level. On the whole this is not unlike what is commonly observed in oligopoly experiments (see e.g. Engel 2010). To check if participants understood the strategic behavior aspect, we use a panel regression to estimate their production function. Producers display evidence of strategic behavior if they condition their production strategy on the expected production strategy of the other producer. Since we did not elicit expectations in the prologue, we proxy for expectations using last period’s other producer quantity. Table 2.2 documents the results of the regression. On average, participants increased their production if their rival previously produced a high quantity. All in all, the finding that average production levels are close to the Nash level and that participants production decisions are correlated to last period’s other producer quantity suggest that most participants also gained some understanding of the strategic behavior aspect.

2.6.2 Producer Focus in the Main Part

We now turn to producer focus in the main part, where both strategic behavior and dynamic optimization were possible. We investigate the degree to which producers pay attention to either dynamic optimization or strategic behavior by estimating their production function. For this purpose we estimate the following panel regression:

\[ q_{rit} = \beta \cdot \mathbb{E}[q_{rjt}] + \gamma_1 \cdot S_{rit} + \gamma_2 \cdot S_{rjt} + T_t + \delta_i + \epsilon_{rit} \]  

\[ (2.3) \]

\(^{29}\)At the market level, production levels were not significantly different from the Nash benchmark either overall or in any individual period. At the producer level the average quantity over all periods was slightly lower than Nash (F(1,135)=4.10, p<0.05). The average quantity was also significantly smaller in period 1 (p<.01), period 2 (p<.01) and period 3 (p<.1), whilst being significantly higher in period 6 (p<.05).

\(^{30}\)The producer level figure is available upon request.

\(^{31}\)This is a good proxy to the extent that participants based their expectations on what the other firm produced in the previous period. See appendix A for evidence that suggests that this is indeed the case.
This equation posits that producer $i$’s quantity in period $t$ of round $r$ is a function of both strategic and dynamic optimization variables. $E[q_{rjt}]$ is producer $i$’s prediction for the quantity of producer $j$ (the other producer); this variable represents the degree to which producer $i$ pays attention to the strategic behavior aspect. $S_{rit}$ is the producer $i$’s stock, which represents the dynamic optimization aspect. $S_{rjt}$ is the other producer’s stock; this variable is relevant only if both dynamic optimization and strategic behavior play a role, as in the dynamic Nash benchmark. Finally, the regression also includes time fixed effects ($T_t$) and individual fixed effects ($\delta_i$) to correct for differences between people and periods.

Throughout the analysis, standard errors are clustered by producer. Note also that the inclusion of the prediction variable means that the analysis will only use data from rounds where predictions were requested (i.e. all even rounds).

Table 2.3 displays the results of the regression for both treatments. For the dynamic optimization variable stock, the results are in line with hypothesis 1A. In particular, the coefficient for stock is larger for treatment LOW in absolute size and has a lower p-value. The difference in coefficients is significant at the 1% level. Moreover, the coefficient for stock is not significantly different from the Nash coefficient for stock in treatment LOW, whereas it is significantly lower at the 1% level for treatment HIGH (the optimal coefficients are .37 and .33).

---

32 Since producers were asked to make predictions and production decisions simultaneously, there are legit concerns that predictions may be endogenous. This issue is addressed in appendix A both using an Instrumental Variables approach and using lagged other producer quantity as an indirect measure of expectations.

33 We do not include individual specific time fixed effects since that would greatly increase the number of parameters per individual, which would put too much strain on the data. It is possible to include round fixed effects, but these are never significant in any treatment and including them does not affect the coefficient estimates; hence we do not include them here.

34 It would have also been possible to do the clustering by matching groups. However, the resulting standard errors are smaller and perhaps less reliable because of the relatively small number of matching groups per treatment. Hence we stick to the more conservative estimate.

35 Another feasible regressor would have been lagged quantity, since it is not infeasible that a producer’s production decision would be partially influenced by his previous production decision even after correcting for the other variables. However, including lagged quantity as a regressor would have made the the model dynamic, which would have made unbiased inference very difficult. For the same reason we also excluded lagged predictions (which may be endogenous to lagged production, see appendix A). We also excluded lagged other producer quantity, since its p-value always exceeds .2 and including would have resulted in the removal of the first period of every round from the analysis.

36 All comparisons between coefficients in table 2.3 are based on a pooled regression with data from all treatments, that includes all the variables (and fixed effects) of table 2.3 as well as interaction terms between a treatment dummy and these variables. The difference in coefficients between treatments is significant only if the corresponding interaction terms are.
respectively). All in all, these results suggest that producers were indeed less mindful of the dynamic optimization aspect in treatment HIGH, in line with hypothesis 1A.

When it comes to the strategic behavior variable “predicted other producer quantity”, results are less clear cut. On the one hand, the results are in the direction predicted by hypothesis 1B: a change in predicted other producer quantity had a larger effect in treatment HIGH than in treatment LOW. On the other hand, the difference between treatments is small and not significant at conventional levels.\textsuperscript{37}

The final variable -the other producer’s stock- should only be significant for producers who adopt a dynamic Nash strategy that incorporates both dynamic optimization and strategic behavior. This variable is significant only in treatment LOW (in the direction predicted by the Nash benchmark), which suggests that producers in treatment LOW were more likely to adopt a dynamic Nash strategy. This is in line with the finding that dynamic optimization behavior appears more strongly in treatment LOW whereas there is little evidence for differences in the level of strategic behavior.

\subsection*{2.6.3 Comparing the Prologue and the Main Part}

Thus we have seen that in terms of the production function, the dynamic optimization aspect seemed to be more important in treatment LOW, whereas there was no pronounced pattern for the strategic behavior aspect. Another way to examine producer focus in the main part is to correlate behavior in the main part to behavior in the prologue. Specifically, if hypotheses 1A and 1B are correct, the degree to which behavior correlates between the prologue and the main part could also depend on the treatment. Behavior in treatment LOW should then be most correlated to behavior in the monopoly phase of the prologue, whereas behavior in treatment HIGH should be most correlated to the unconstrained oligopoly phase.

To test this idea we correlate indicators of behavior and success in the main part with similar indicators from the prologue. From the monopoly part of the prologue we take the difference between the first period scarcity rent and last

\textsuperscript{37}We also ran two separate sessions where production in the main part was unconstrained. Repeating the regression of table\textsuperscript{2.3} for these sessions gives a higher coefficient (.3941, with S.E. .0793) for predicted other producer quantity, although the difference in coefficients with treatment LOW and treatment HIGH was also not significant.
Table 2.3: Main Part Production Function

<table>
<thead>
<tr>
<th></th>
<th>Dependent Variable: Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LOW</td>
</tr>
<tr>
<td>Predicted other producer quantity</td>
<td>.3050*** .0809</td>
</tr>
<tr>
<td>Stock</td>
<td>.4269*** .0422</td>
</tr>
<tr>
<td>Other producer stock</td>
<td>−.0884*** .0250</td>
</tr>
<tr>
<td>Observations</td>
<td>1600</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td>.7436</td>
</tr>
</tbody>
</table>

* significant at 10%; ** significant at 5%; *** significant at 1%

Notes. This table contains the results of a panel regression of quantity on the predicted quantity of the other producer, stock and the stock of the other producer. Period and producer fixed effects are also included but not reported; standard errors are clustered at the producer level. Predictions were only elicited in even rounds; moreover, the final period is omitted from the analysis since all benchmarks are trivially equal to the remaining resource stock in the final period. Thus, the number of observations per individual in all treatments is equal to 25, from 5 rounds with 5 observations each. P-values are calculated using two-sided t-tests.

period scarcity rent (or dispersion) as an indicator of dynamic optimization. For the unconstrained oligopoly part of the prologue we take the first period production quantity as a measure of the intention to produce cooperatively. Moreover we take a measure of income in both parts and correlate that with main part income to see if success is also correlated between the prologue and the main part.38

Table 2.4 shows the resulting correlation coefficients. Firstly, participants with a high dispersion in the monopoly part of the prologue also had a high dispersion in the main part, but only in treatment LOW.39 Participants who were successful in the monopoly part of the prologue were more successful in treatment LOW in

38 In comparing the unconstrained oligopoly with the main part we correlated overall income in the unconstrained oligopoly to overall income in the main part. However, for the monopoly part there are several complications: (a) because of practice almost all participants were able to do well in the monopoly part of the prologue; as a result there was little variation in the production schedule used in the incentivized round. Moreover, (b) starting from round two participants in the main part could also adopt the production schedule they learned from the other producer in a preceding round. To solve the first problem, we constructed a dummy variable which was equal to one only if the participant managed to run a profit in at least one of his first three practice rounds (the results are similar if we take the first four, five or six practice rounds instead). This split the sample roughly in half, since 57% of participants managed to run a positive profit in one of the first three practice rounds. To address possible learning considerations we used only the first round of the main part.

39 The results are identical if we use the differential between the highest and lowest scarcity rent instead.
Table 2.4: Correlations between the Prologue and the Main Part

<table>
<thead>
<tr>
<th></th>
<th>LOW</th>
<th>HIGH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main part dispersion &amp; Monopoly dispersion</td>
<td>.4019***</td>
<td>.0430</td>
</tr>
<tr>
<td>Main part first period quantity &amp; Oligopoly first period quantity</td>
<td>.1640</td>
<td>.1743</td>
</tr>
<tr>
<td>Main part income &amp; Monopoly income</td>
<td>.3431***</td>
<td>.1922</td>
</tr>
<tr>
<td>Main part income &amp; Unconstrained oligopoly income</td>
<td>.0236</td>
<td>.0350</td>
</tr>
<tr>
<td>Observations</td>
<td>64</td>
<td>72</td>
</tr>
</tbody>
</table>

* significant at 10%; ** significant at 5%; *** significant at 1%

Notes. This table contains Pearson correlations. Main part dispersion is the difference between the first period and sixth period scarcity rents in the main part averaged over all rounds. Monopoly dispersion is the difference between the first and sixth period scarcity rents in the monopoly part. Main part (oligopoly) first period quantity is the quantity produced in the first period of the main part (unconstrained oligopoly part) averaged over all 10 (3) rounds. For the third correlation, main part income is the total income over 6 periods in the main part for the first round only. Monopoly income is a dummy variable that indicates if the producer achieved a positive profit in at least one of his first three trial rounds. For the fourth correlation, main part income is the total income over 6 periods in the main part averaged over all 10 rounds. Unconstrained oligopoly income is the total income over 6 periods in the oligopoly part averaged over all three rounds.

terms of income as well. On the other hand, there is no significant correlation for either the intention to behave cooperatively (first period quantity) or for success in the second part of the prologue (oligopoly) for either treatment. Thus there seems to be some evidence that behavior and success in the monopoly phase are correlated to behavior and success in treatment LOW (but not treatment HIGH); however there is no correlation between the unconstrained oligopoly phase and treatment HIGH (or treatment LOW). Overall, these findings are in line with the results of the previous section; there appears to be a difference between treatments for the importance of the dynamic optimization aspect, but not for the strategic behavior aspect.

2.6.4 Market Outcomes in the Main Part

In the previous sections we saw that the dynamic optimization aspect seemed to be less important in treatment HIGH. If hypothesis 2 is true, producers in treatment HIGH should then be more likely to overproduce relative to the Nash benchmark. Figures 2.6 to 2.9 and table 2.5 give an overview of scarcity rents.

\[\text{\textsuperscript{40}}\text{It is important to restate that when we refer to the Nash, Walras or Collusive benchmarks in this section, we refer to the dynamic benchmarks that depend on the current state of the market (i.e. the period/stock level combination). In particular, since in period 1 different markets will}\]
Notes. The figure plots the time series of the average observed scarcity rent as well as the benchmark Nash, Collusive and Walras scarcity rent with respect to the unconstrained Nash price. This is equivalent to subtracting a fixed number ($p^N_U = 124$) from the observed and benchmark prices respectively (i.e. $p^i - p^N_U$, where $p^i$ is the observed, Nash, Collusive or Walras price respectively). Here we use the dynamic benchmarks that depend on the current state of the market.
2.6. RESULTS

Figure 2.7: Treatment LOW Density Plots

Notes. The top panel plots the smoothed density (Epanechnikov kernel, bandwidth = .26) of the deviation from the Nash quantity weighted by the distance between the Collusive quantity and the Nash quantity (i.e. \( \frac{q_t - q^N_t}{q^N_t} \)) in periods one to five. The lower panel does similarly but then for producer quantity (bandwidth=.28). We look at deviations since from period two onwards the three benchmarks are different for every market; we use weights since a unit deviation from the Nash quantity means more in periods where the three benchmarks are closer together. C, N and W represent Cournot, Nash and Walras levels respectively; the Cournot and Nash levels are -1 and 0 by the normalization whereas the Walras level differs between markets (average =1.91, always bigger than zero). Period 6 is omitted since all benchmarks are equal to the resource stock in this period.
Notes. The figure plots the time series of the average observed scarcity rent as well as the benchmark Nash, Collusive and Walras scarcity rent with respect to the unconstrained Nash price. This is equivalent to subtracting a fixed number ($p_{NU}^N = 124$) from the observed and benchmark prices respectively (i.e. $p^i - p_{NU}^N$, where $p^i$ is the observed, Nash, Collusive or Walras price respectively). Here we use the dynamic benchmarks that depend on the current state of the market.
2.6. RESULTS

Figure 2.9: Treatment HIGH Density Plots

Notes. The top panel plots the smoothed density (Epanechnikov kernel, bandwidth = .22) of the deviation from the Nash quantity weighted by the distance between the Collusive quantity and the Nash quantity (i.e. $\frac{q_t - q^N_t}{q^N_t - q^C_t}$) in periods one to five. The lower panel does similarly but then for producer quantity (bandwidth=.19). We look at deviations since from period two onwards the three benchmarks are different for every market; we use weights since a unit deviation from the Nash quantity means more in periods where the three benchmarks are closer together. C, N and W represent Cournot, Nash and Walras levels respectively; the Cournot and Nash levels are -1 and 0 by the normalization whereas the Walras level differs between markets (average =2.11, always bigger than zero). Period 6 is omitted since all benchmarks are equal to the resource stock in this period.
Table 2.5: Main Part Market Quantity and Benchmarks

<table>
<thead>
<tr>
<th>Period</th>
<th>Treatment LOW (N=320)</th>
<th>Treatment High (N=360)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average Quantity</td>
<td>Nash</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>101.70 (2.192)</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>80.75 (2.322)</td>
<td>83.04 (2.691)</td>
</tr>
<tr>
<td>3</td>
<td>64.65 (1.338)</td>
<td>67.23 (2.184)</td>
</tr>
<tr>
<td>4</td>
<td>45.975 (1.149)</td>
<td>50.28** (1.270)</td>
</tr>
<tr>
<td>5</td>
<td>29.67 (1.924)</td>
<td>32.97*** (.737)</td>
</tr>
<tr>
<td>6</td>
<td>17.21 (2.646)</td>
<td>17.27 (.37)</td>
</tr>
</tbody>
</table>

* significant at 10%; ** significant at 5%; *** significant at 1%

Notes. This table compares the observed average market quantity per period to the Collusive and Walras and (feedback) Nash benchmark for every treatment. Here we use the dynamic benchmarks that depend on the current state of the market. For this purpose, all three benchmarks are calculated for every data point conditional on period, producer stock and other (feedback) Nash benchmark for every treatment. Here we use the dynamic benchmarks that depend on the current state of the market. For this purpose, all three benchmarks are calculated for every data point conditional on period, producer stock and other producer stock and then summed over both producers on the market. In the first period, all benchmark quantities are integers since all producers have the same resource stock and only integer amounts can be produced. In the final period, fully exhausting the resource is always the optimal strategy regardless of benchmark and treatment. Standard errors are clustered at the matching group level. Significance is determined using two-sided t-tests.
and production levels in the main part. For treatment HIGH, average production levels are higher than the Nash level in all periods; as a result the average scarcity rents are lower. For treatment LOW, the scarcity rent is indistinguishable from the (feedback) Nash level in the first three periods; in periods 4 and 5 scarcity rents are actually higher than Nash and indistinguishable from the Collusive level. All in all, the Nash Hotelling rule consistently fails in treatment HIGH because of overproduction, whereas in treatment LOW it describes the data quite well. This is exactly what one would expect if producers paid less attention to the dynamic optimization aspect in treatment HIGH (hypothesis 2).

Thus producers in treatment LOW are closer to the Nash benchmark than producers in treatment HIGH on average. Figures 2.7 and 2.9 show that a similar pattern holds for the distribution as a whole. The peak of the distribution of both treatments lies close to the Nash level both at the producer level and at the market level, but it falls slightly towards the Collusive side of the Nash level in treatment LOW and towards overproduction in treatment HIGH. In terms of market (producer) production levels, 70.5% (62.4%) of all production levels are larger than the (feedback) Nash benchmark in treatment HIGH, whereas the respective percentages are 28.3% (30.5%) for treatment LOW.

The finding that in treatment HIGH producers are more likely to overproduce should also be reflected by their earnings. To see if this is indeed the case, we compare average realized earnings to potential earnings in the Nash benchmark and weigh this using the distance between the Nash benchmark and the Collusive benchmark. This results in the normalized earnings index $Y_{\text{norm}} = \frac{Y - Y_N}{Y_C - Y_N}$, where $Y_N$ is the theoretical Nash profit, $Y_C$ is the theoretical Collusive profit and $Y$ is actual income. The average index is equal to -6.70 for treatment HIGH and -1.72 for treatment LOW, the difference is significant at the 1% level (Mann-Whitney, $z(64,72)=6.422$). In other words, in terms of the index earnings were significantly lower in treatment HIGH. This is in line with the finding that producers were

---

41 Since scarcity rents and prices are an affine transformation of market quantities, the test statistics for market quantities, prices and scarcity rents are identical.

42 Note also that Collusive benchmark can be a very good dynamic strategy especially if both producers adhere to it.

43 It is not possible to compare realized earnings levels directly because of difference in stock levels, fixed costs and conversion rates.
more likely to overproduce with respect to Nash in treatment HIGH.

In summary, we have seen that producers paid more attention to the dynamic optimization aspect in treatment LOW and that this induced them to produce closer to the Nash benchmark than producers in treatment HIGH. At the same time, it is worthwhile to point out that on average producer behavior was quite close to the Nash level in both treatments. As a result, scarcity rents are actually increasing in both treatments, in line with the Hotelling rule and contrary to the findings of most previous empirical studies. Importantly, however, producers in treatment HIGH deviate most from the Nash benchmark and do so in the direction predicted by hypothesis 2. Thus, even producers who -based on the evidence of treatment LOW- should be able to approximate the Nash time path almost perfectly still overproduce if their stock levels are relatively high.

2.7 Discussion

This chapter has used the methods of experimental economics to investigate a possible cause for the failure of the Hotelling rule. In the experiment, we have seen that when resource stocks were large (as in treatment HIGH), producers chose to partially ignore the dynamic component of their production decision and overproduced with respect to the Nash Hotelling rule. However, when resource stocks were small (as in treatment LOW), producers extracted close to the optimal Nash Hotelling level.

In terms of real world markets, the relative abundance of many non-renewable resources may have induced producers to overextract, leading to the failure of the Hotelling rule. At the same time, we do not believe that this is the whole story; indeed in the experiment scarcity rents are still increasing over time in both treatments. A full explanation of the failure of the Hotelling rule may also require other elements, including for example the discovery of new deposits, capacity constraints or technological progress. In future work, it could be worthwhile to use the experimental approach to try to disentangle the relative explanatory

---

Interestingly, average earnings were below the Nash level in both treatment HIGH and treatment LOW, despite the finding that in treatment LOW average scarcity rents were never significantly below the Nash level. Recall, however, that although on average the Nash scarcity rent was quite well approximated by the data, many individual producers still overproduced or underproduced, which resulted in lower earnings. The upper and lower panels of figures 2.7 and 2.9 indeed show substantial heterogeneity in observed production levels.
power of these elements. Indeed, we believe that experimental data can serve as a complement to field data to aid the profession in gaining a better grasp of the mechanisms driving the behavior of producers on nonrenewable resource markets.
Appendix 2.A Re-Examining Predictions

In this section we take a closer look at the expectations elicitation procedure we used in the experiment. Recall that in every even round participants were asked in every period to indicate to predict the other firm’s production in that period. Directly asking participants what they expect the other producer to extract gives rise to at least two possible concerns. For one, asking for predictions directly may lead to a false consensus effect (see e.g. Croson (2000). In the context of this experiment, participants might base their expectation of the other firm’s production on their own production level. Since we are interested in the opposite effect, this means that we have to take possible reverse causality into account when investigating producer dynamics by means of expectations. Moreover, another problem with eliciting expectations directly is that the elicitation procedure itself may change behavior in a round (see e.g. Gächter and Renner (2010).

With field data, reverse causality issues tend to be addressed using an instrumental variables (IV) approach. A good instrument should (a) have a causal impact on the instrumented variable (period t predictions) and (b) be uncorrelated to the dependent variable (period t production). We can exploit the panel structure of the data and use lagged variables as instruments. One variable variable that meets this criteria is lagged other producer quantity.

Table 2.6 gives the results of the IV approach; we also repeat the regression of table 2.3 without period one to make sure that any differences are not due to to the removal of the first period. The first stage estimates show that last period’s other producer quantity is indeed strongly correlated to this period’s prediction in all two treatments. The estimated coefficients for the prediction variable in stage 2 are approximately half the size of the coefficients for OLS. As a consequence, they are no longer significantly different from zero at the 5% level. This suggests that the false consensus effects may have had a sizable impact on the predictions that participants made. At the same time, using the IV approach does not affect the conclusions with respect to hypotheses 1A and 1B; hypothesis 1B still does

45There is a correlation between producer quantity and lagged other producer quantity, but it disappears if corrected for current period prediction. This suggests that any effect lagged other producer quantity has on current production runs through predictions, which makes it a good instrument. Also, the impact of lagged other producer quantity on current prediction is credibly causal, since it is impossible for a producer to base his on the other producer’s next period prediction.
### Table 2.6: Main Part Production Function with Instrumental Variables

<table>
<thead>
<tr>
<th></th>
<th>LOW</th>
<th></th>
<th></th>
<th>HIGH</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Stock in $t$</td>
<td>-.0285</td>
<td>.0173</td>
<td></td>
<td>.0446</td>
<td>.0126</td>
</tr>
<tr>
<td>Other producer stock in $t$</td>
<td>.3097***</td>
<td>.0143</td>
<td></td>
<td>.1461***</td>
<td>.0123</td>
</tr>
<tr>
<td>Other producer quantity in $t - 1$</td>
<td>.4048***</td>
<td>.0216</td>
<td></td>
<td>.5721***</td>
<td>.0249</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Dependent Variable: Quantity in $t$ (IV 2nd stage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted other producer quantity in $t$</td>
<td>.1397*</td>
</tr>
<tr>
<td>Stock in $t$</td>
<td>.3131***</td>
</tr>
<tr>
<td>Other producer stock in $t$</td>
<td>-.0430*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Dependent Variable: Quantity in $t$ (OLS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted other producer quantity in $t$</td>
<td>.2224***</td>
</tr>
<tr>
<td>Stock in $t$</td>
<td>.3130***</td>
</tr>
<tr>
<td>Other producer stock in $t$</td>
<td>-.0584***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Observations</th>
<th>Adj. R-Squared (IV stage 1)</th>
<th>R-Squared (IV stage 2)</th>
<th>Adj. R-Squared (OLS)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1280</td>
<td>.7437</td>
<td>.6949</td>
<td>.6797</td>
</tr>
<tr>
<td></td>
<td>1440</td>
<td>.5820</td>
<td>.3763</td>
<td>.3512</td>
</tr>
</tbody>
</table>

* significant at 10%; ** significant at 5%; *** significant at 1%

**Notes.** This table provides the result of an instrumental variables or two stage least squares regression and an OLS regression for comparison. The first stage of IV consists of a regression of predicted other producer quantity on stock, other producer stock and lagged other firm quantity (the instrument). The second stage of IV is a regression of quantity on the predicted other producer quantity, stock and the stock of the other producer. Both stages also contain period and producer fixed effects; standard errors are clustered at the producer level. For OLS, we repeat the regression of table 2.6 with period 1 removed. Predictions were only elicited in even rounds; moreover the first period is omitted because of the use of a lagged variable. The final period is omitted from the analysis since all benchmarks are trivially equal to the remaining resource stock in the final period. Thus, the number of observations per individual in all treatments is equal to 20, 5 rounds with 4 observations each. P-values are calculated using two-sided t-tests.
### Table 2.7: Main Part Production Function in Rounds with and without Expectation Elicitation

<table>
<thead>
<tr>
<th></th>
<th><strong>Dependent Variable: Quantity in ( t )</strong></th>
<th><strong>LOW</strong></th>
<th><strong>COEFF.</strong></th>
<th><strong>STD ERR.</strong></th>
<th><strong>HIGH</strong></th>
<th><strong>COEFF.</strong></th>
<th><strong>STD ERR.</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Other producer quantity in ( t - 1 )</td>
<td>.0122</td>
<td>.0324</td>
<td>.0277</td>
<td>.0412</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stock in ( t )</td>
<td>.2815***</td>
<td>.0217</td>
<td>.1795***</td>
<td>.0281</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other producer stock in ( t )</td>
<td>.0019</td>
<td>.0111</td>
<td>-.0144</td>
<td>.0189</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictround x other producer quantity in ( t - 1 )</td>
<td>.0452</td>
<td>.0481</td>
<td>.0537</td>
<td>.0717</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictround x stock</td>
<td>.0210</td>
<td>.0151</td>
<td>-.0490</td>
<td>.0334</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictround x other producer stock</td>
<td>-.0014</td>
<td>.0171</td>
<td>.0285</td>
<td>.0242</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>2560</td>
<td></td>
<td>2880</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-Squared</td>
<td>.6676</td>
<td></td>
<td>.3347</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*significant at 10%; **significant at 5%; ***significant at 1%

**Notes.** This table provides the result of a regression of quantity on lagged other producer quantity, stock and the stock of the other producer as well as the same variables interacted with a dummy (predictround) which is equal to one for all rounds where predictions were made. The regression also contains period and producer fixed effects (the former are also interacted with the predictround dummy); standard errors are clustered at the producer level. The first period is omitted because of the use of a lagged variable; the final period is omitted since all benchmarks are then trivially equal to the remaining resource stock. Thus, the number of observations per individual in all treatments is equal to 40, 10 rounds with 4 observations each. P-values are calculated using two-sided t-tests.
not receive any support from the data, whereas hypothesis 1A strongly does. Indeed, for the dynamic optimization variables stock and other producer stock, the coefficient estimates of stage 2 are very close to the results of OLS.

To examine if the elicitation procedure itself also influenced behavior in the experiment, we can compare rounds with expectations to rounds where expectations were not elicited. For this purpose we repeat the regression of table 2.3 but include a dummy for prediction rounds and interact this dummy with all regressors. Since we have no data on predictions for rounds where predictions were not asked, we use last period’s other firm quantity as a proxy. The first stage regression in table 2.6 showed that this variable is highly correlated to the prediction variable and may as such be a useful proxy. Table 2.7 gives the results of this regression. The most important thing to note is that none of the interaction terms are significant. This suggests that the elicitation procedure itself did not significantly affect producer behavior in this experiment.

\[\text{[46]}\]

\[\text{[46]}\] We also examined if the distribution of quantities differed between rounds with and without predictions; however we found no differences.
Appendix 2.B  Experimental Instructions

This section contains the instructions and questions used in treatment HIGH of the experiment for both the prologue and the main part. Part I and part II refer to phase one and phase two of the prologue respectively, whereas part III refers to the main part. The instructions for treatment LOW were identical except that stock, fixed costs and the conversion rate were different in the main part (part III). An example of a decision screen is provided at the end of this appendix.

Introduction Part I

In part 1 of this experiment, you are the manager of a firm. In particular, you will have to decide on the quantity that your firm is going to produce. Your firm is the only active firm on the market (i.e. it is a monopolist) and as such only your decision determines the market price. In this part of the experiment, the minimum market price is 0 and the maximum market price is 372. Moreover, increasing your quantity by 1 will lower the price by 2. Quantity and price in turn determine revenue according to the following formula:

\[ \text{Revenue} = \text{Price} \times \text{Quantity} \]

The payment you receive at the end of the experiment will be based on total revenue. This part of the experiment consists of several rounds. Each round in turn consists of 6 periods. In each period you have to decide what quantity your firm is going to produce. After 6 periods, the round will end and your pay-off over the round will be determined. After this a new round will start, which will again have 6 periods.

You may have noticed that there is a calculator at the bottom of the screen. It can be used to calculate what will be the price and revenue level if you pick a certain quantity. You can now go to the next page of the instructions. Note that you can always return to this page later by clicking on the blue headers at the top of this page (only the pages you have already been to can be accessed).

The Payment Mechanism

In this experiment, revenue earned in earlier periods is more valuable than revenue earned in later periods. One way to think about this is that your revenue will
be put on a bank account, where it will earn 10% interest per period. The final column in the calculator describes exactly how much a given level of revenue earned in a certain period will be worth in terms of End Income. For example, a quantity of 54 produced in period 1 yields a revenue of 14256, which will yield an end income of 22959 at the end of the round. The same numbers can also be accessed using the calculator by filling in 54 for first period quantity.

At the end of every round, your firm will calculate its total end income by adding up end income levels of all 6 periods. However, your firm also has a fixed cost equal to 98000, which will have to be paid at the end of every round. This fixed cost is unavoidable and will have to be paid regardless of the amount you produced over the round. Think of this amount as the total cost of maintaining a factory over the whole round. As a result, your payment at the end of a round will be determined according to the following formula:

$$\text{Payment} = \text{Total End Income} - \text{Fixed Cost}$$

At the end of the experiment, the points from all rounds will be converted into euros at a rate of 2000 points per euro, such that 1000 points are worth 0.50 euros. Be assured though that over the whole experiment it will not be possible to lose money. A negative pay-off over a round can be compensated by a positive pay-out in another round, in another part of the experiment, or by the show-up fee (7 euros), with a minimum possible pay-off of zero over the whole experiment. Note, however, that this is very unlikely to happen.

**Resource Stocks**

One thing we have neglected so far is the production process. Producing one unit of your firm’s good will require one unit of resource. Over each round, only 280 units of resource are available, so that at most 280 units can be produced. Thus, consuming one unit of resource in the first period means that you will not be able to use it in any of the following periods. At the start of each new round, your resource stock will be replenished.

One thing to note about the resource stock is that you do not have to use it all. Likewise, it is not necessary to produce in every single period. At the same time, it is also very well possible to use up your whole resource stock and produce
in all periods. Before we start the actual experiment, you will have some time to practice to familiarize yourself with this set-up.

As a final comment, be sure that you understand the difference between period and round, and also between revenue, end income and payment. One round consists of 6 periods. Similarly, revenue is what your firm earns every period, end income is what your firm’s revenue of a period will be worth at the end of the round (i.e. after taking interest into account) and payment is the total amount of points you get at the end of a round.

To make sure you understand these ideas, we have a few questions for you.

**Question 1**

The first question is about the price mechanism. Suppose your firm is going to produce 58. What will be the market price? (Integer between 0 and 372). Tip: use the calculator (any period will do, since the price does not depend on the period)!

**Question 2**

Which of the following statements is true? If you produce 55 in both period 2 and period 3, you will have...

1. Different prices, revenues and end incomes in both periods
2. The same price and end income in both periods and a higher revenue in period 3
3. The same price and end income in both periods and a higher revenue in period 2
4. The same price and revenue in both periods and a higher end income in period 2
5. The same price and revenue in both periods and a higher end income in period 3
6. The same price, revenue and end income in both periods
2.B. EXPERIMENTAL INSTRUCTIONS

Question 3
Which of the following statements is false? It is possible to...

1. Produce zero in some periods
2. Produce something in all periods
3. Over all periods produce less than your stock
4. Over all periods produce as much as your stock
5. Over all periods produce more than your stock

Question 4a

Now suppose your firm has arrived in period 5 (out of 6) and still has a stock of 26. However, you are doubting between two different options. Option A entails producing 19 in period 5 and 7 in period 6. Option B would be to produce 7 in period 5 and 19 in period 6. What option would yield the highest end income?

Question 4b

Suppose you have indeed decided to produce 19 in period 5 and 7 in period 6. What end income is your firm going to earn?

End of Instructions

You are now ready to start the experiment. By clicking on the next link, you will go to a practice session as soon as everyone has finished the instructions. The results you obtain during practice will not count towards your pay-out at the end of the experiment. Practice time will last for approximately 10 minutes; during this time you can work through as many rounds as you like. After the practice session has ended you will move on to the part where earnings will be paid out. This part is identical to practice, except that there will be only 1 round. All results obtained in the practice session will be saved and made available during the real experiment, so use practice time to familiarize yourself as well as possible with the set-up.
One more thing to note is that the bottom right corner of the screen will show a timer. You can see an example of the timer in the bottom of right of this screen. The timer indicates the amount of time you have left to make a decision in the current period. In this part of the experiment, you will have a maximum of 40 seconds to make your decision. If you fail to make your decision in time, you will automatically produce zero and move on to the next period. The timer is reset in every period, regardless of how much time you spent in the preceding period. Finally, note that it in many cases only a small fraction of the required time might be needed to make a decision.

**Instructions Part II**

We will now start with the second part of this experiment. In this part of the experiment, you will be the manager of a firm, like in part I. However, several other aspects have changed. For one, you now face competition from one other firm. For another, you will have an unlimited amount of resources to produce with. Moreover, fixed costs will be slightly higher. These changes will be explained in greater detail below and on the next page.

**Dealing with other firms**

Firstly, you will now face competition from another firm. The decisions for the other firm will be made by another participant of this experiment. The other firm you face will be the same in every period of the same round but will change in each new round. Moreover, decisions will be made simultaneously, so that you will not know the other firm’s production level until after the end of the period, just like the other firm will not know your production level. Similarly, you will not know with whom you will be matched, like others will not know with whom they will be matched. Anonymity is ensured.

As a result of the presence of the other firm, the effect of your quantity on market price has changed. In particular, a one unit increase in production by either you or the other firm will now lower the market price by 1. Moreover, the price will now be between 0 and 360. The calculator has been changed and will now be able to also take the decisions of the other firm into account. You will be able to practice with the new situation in one of the exercises.
Resources and Earnings

Another difference between part II and part I is that you will no longer have a limited stock of resources. As a result, producing a high amount in an early period will no longer limit your production in later periods. The final difference with the first part is that fixed costs are now equal to 100000. Other than that, the payment mechanism in this part of the experiment is identical to the mechanism used in the first part. Your payment after each round is still determined using the following formula:

\[
\text{Payment} = \text{Total End Income} - \text{Fixed Cost}
\]

Your payment will be converted into euros at a rate of 10000 points per euro, such that 1000 points are worth 0.10 euros. Moreover, interest will still be equal to 10%. Before going to the experiment, we would like to ask two checkup questions.

Question 1

This question will make use of the following table. The table can be read as follows: the left column contains your production decision (in red). The top row contains the production decision of the other firm (in blue). The cells in the table indicate what level of revenue your firm will earn (again in red) for the associated combination of production levels by your firm and the other firm. Moreover, the cells also contain the revenue that the other firm will earn (in blue). For example, to look up your revenue in case the other firm produced 60 and your firm produced 120, you would have to go right from 120 and down from 60, where you would find that you would earn a revenue of 216 and the other firm would get 108.

There are a few more things to note about the table. For one, note that the numbers in the table are revenues, which are equal to end income only in the last period. For another, it is important to note that although the set-up for this question is identical to the set-up used in the experiment itself, we have chosen only a few values as examples. For example, in the actual experiment you would also be able to produce 119 or 121 (or any other amount). Moreover, we have underlined the quantity that will give you the highest revenue keeping the other firm’s production level constant. In some cases, there may be two such quantities in the table; however, this holds in the table only because of the particular values
used for this example. Finally and importantly, the last two digits have been removed from the revenue numbers in the table. Thus, for example 180 would actually be 18000, and the latter is what you would have to fill in below.

<table>
<thead>
<tr>
<th></th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>180</td>
</tr>
<tr>
<td>Self</td>
<td></td>
</tr>
<tr>
<td>180</td>
<td>0</td>
</tr>
<tr>
<td>150</td>
<td>45</td>
</tr>
<tr>
<td>120</td>
<td>72</td>
</tr>
<tr>
<td>90</td>
<td>81</td>
</tr>
<tr>
<td>60</td>
<td>72</td>
</tr>
<tr>
<td>30</td>
<td>45</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Suppose the other firm produced 60. What revenue would you earn if you produced 120?

Now suppose the other firm produced 90. Firstly, what quantity should you produce to get the highest revenue for you? What quantity should you produce to get the highest revenue for the other? And finally, what quantity would yield the highest combined revenue? Hint: sometimes multiple answers may be possible.

**Question 2**

In the previous question, you were asked what quantity would yield the highest revenue for your firm if the other firm produced 90. In the table, there were two correct answers: 120 and 150. However, in the experiment itself it will also be possible to pick any integer quantity between 120 and 150. What production level will yield the highest end income for your firm if you are allowed to pick any integer production level? Tip: use the calculator!
End of Instructions

You are now ready to start the experiment. By clicking on the next link, you will go to the experiment as soon as everyone has finished reading the instructions. In this part of the experiment, there will be no time to practice: you will immediately go on to the part where your earnings will be paid out. In total, there will be 3 rounds. In every round you will be matched with a different firm. Because there are no practice rounds, the timer will be set to 1 minute and 40 seconds per period in the first round and to 40 seconds per period in later rounds. Thus, you will have slightly more time in the first round to make a decision. If you fail to make your decision in time, you will automatically produce zero and move on to the next period. The timer is reset in every period, regardless of how much time you had left after the preceding period. Finally, note that in many cases only a small fraction of the required time might be needed to make a decision.

Instructions Part III

We will now start with the third and final part of this experiment. In many ways, this part will be a mix between part I and part II. In particular, you will still be the manager of a firm. Moreover, in this part you will have to deal with one other firm, like in part II. At the same time, you will only have a limited stock of resources available for production, like in part I. Finally, fixed costs will be slightly different from either of these parts. These changes will be explained in greater detail below and on the next page.

Firstly, as in part I you will have only a limited stock of resources available for production. To be more precise, you will have a total resource stock of 480 in every round. As in part I, you do not have to use up all your resources. Likewise, it is not necessary to produce in every single period. At the same time, it is also very well possible to use up your whole resource stock and produce in all periods. At the start of each new round, your resource stock will be replenished, as before. In total, this part of the experiment will consist of 10 rounds.

Moreover, there is one other active firm on the market, like in part II. Once again the other firm you face will be the same in every period of the same round but will change in each new round. You will not know with whom you will be matched, like others will not know with whom they will be matched. Anonymity
is ensured.

As a result of these changes, the effect of your quantity on market price has changed. Like in part II, a one unit increase in production by either you or the other firm will lower the market price by 1. However, the resulting prices are slightly different; in particular, prices are now between 0 and 372. The calculator has been updated to take this into account.

Expectations and Earnings

The payment mechanism in this part of the experiment is almost identical to the mechanism used in the first two parts. Your payment after each round is still determined using the following formula:

\[ \text{Payment} = \text{Total End Income} - \text{Fixed Cost} \]

Your payment will be converted into euros at a rate of 10000 points per euro, such that 1000 points are worth 0.10 euros. Moreover, the interest rate will still be equal to 10%. The only difference is that fixed costs will be equal to 113000.

There is one new thing about this part of the experiment though. In every even round (2, 4, etc.), you are also asked to predict the production of the other firm. It will also be made clear before the start of the round whether or not you have to make predictions during that round and it will also be clear from the decision screen. During rounds where predictions are asked you will have slightly more time to make your decision. Other than that, the decision screen will be very similar to the previous parts of the experiment.

At the end of the experiment, we will randomly pick one prediction you made during the experiment and pay you an additional amount of money based on its accuracy. For a perfect prediction, you will earn 5 euros. If you make an error you will earn 5 euros minus the error times 20 cents. Thus, if you make an error of 25, you will earn 0 (and if you make a bigger error you will still earn 0). Take predictions seriously, since they will earn you extra money at the end of the experiment. Before going to the experiment, we would like to ask you one more check-up question.
Question 1

Suppose that you still have a stock remaining of 93. Suppose also that you are in period 5 (out of 6). Your goal is to allocate the remaining stock optimally over the two remaining periods. Suppose now that the other firm is going to produce 56 in period 5 and 37 in period 6. You now have to choose between implementing two possible production plans. Plan A will entail producing 40 in period 5 and 53 in period 6, whereas Plan B will entail producing 53 in period 5 and 40 in period 6. How much end income will your firm earn with each plan? What will thus be the optimal plan to implement?

End of Instructions

You are now ready to start the experiment. By clicking on the next link, you will go to the experiment as soon as everyone has finished reading the instructions. In this part of the experiment, there will be no time to practice: you will immediately go on to the paid-out part. In total, there will be 10 rounds. In every round you will be matched with a different firm. Because there are no practice rounds, the timer will be set to 1 minute and 40 seconds per period in the first round and to 40 seconds per period in later rounds. When you have to make a prediction, the time will be increased by 20 seconds. Thus, you will have slightly more time in the first round and in prediction rounds. If you fail to make your decision in time, you will automatically produce zero and move on to the next period. The timer is reset in every period, regardless of how much time you had left after the preceding period. Finally, note that in many cases only a small fraction of the required time might be needed to make a decision.
Appendix 2.C  

Your Decision in Period 3 out of 6 (Round 2 out of 10)

So far you have earned 52902 points.

You still have 300 units of resource remaining.

How much do you want to produce this period? (Integer between 0 and 186)
How much do you expect the other to produce this period? (Integer between 0 and 186)

This Round’s History:

<table>
<thead>
<tr>
<th>Period</th>
<th>Own</th>
<th>Other</th>
<th>Expected Other</th>
<th>Price</th>
<th>Revenue</th>
<th>End Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>88</td>
<td>85</td>
<td>90</td>
<td>199</td>
<td>17512</td>
<td>25639</td>
</tr>
<tr>
<td>1</td>
<td>92</td>
<td>96</td>
<td>91</td>
<td>184</td>
<td>16928</td>
<td>27263</td>
</tr>
<tr>
<td>Remaining</td>
<td>300</td>
<td>299</td>
<td>-</td>
<td>-</td>
<td>Totals</td>
<td>52902</td>
</tr>
</tbody>
</table>

Figure 2.10: Decision Screen
Chapter 3

The Influence of Wages on Public Officials’ Corruptibility: a Laboratory Investigation

Previous studies have proposed a link between corruption and wages in the public sector. This chapter investigates this link using a laboratory experiment. In the experiment, public officials have the opportunity to accept a bribe and can then decide between a neutral and a corrupt action. The corrupt action benefits the briber but poses a large negative externality on a charity. The results show that increasing public officials’ wages greatly reduces their corruptibility. In particular, experienced low wage public officials accept 91% of bribes on average, whereas high wage public officials accept 38%. Moreover, high wage public officials are less likely to choose the corrupt action.

\[1 \text{This chapter is based on Van Veldhuizen (2012). I would like to thank Adrian de Groot Ruiz, Ernst Fehr, Guillaume Frechette, Jona Linde, Theo Offerman, Elke Renner, David Rojo Arjona, Andy Schotter, Joep Sonnemans and seminar participants at the University of Amsterdam, the Tinbergen Institute, the University of Maastricht, the 2011 ESA World Meeting, the 2011 CREED-CEDEX-UEA meeting in Nottingham and the 2011 CREED-CESS meeting in New York for helpful comments and suggestions. I am especially grateful to Francisca Then for pointing me towards the study of experiments on corruption and helping me in running the experiment. Financial support from the University of Amsterdam Research Priority Area in Behavioral Economics is gratefully acknowledged.}\]
3.1 Introduction

Corruption is a significant problem in large parts of the world. As a consequence, fighting corruption has at least ostensibly become a primary goal for many of the world’s governments. One possible policy instrument that has prompted considerable debate is the level of public official compensation. Theory (starting with Becker and Stigler [1974]) suggests that increasing the wages of public officials should reduce their corruptibility. If this relationship holds, it provides governments with a policy instrument that falls directly under their control and would therefore be relatively easy to implement.

However, empirical evidence for this relationship has been hard to come by. Since corrupt activities are usually not incorporated into official accounts, direct measures of corruption are only rarely available. As a consequence, empirical studies on corruption have often been forced to rely on indirect measures of corruption. Additionally, even if more direct measures of corruption are available, establishing causality may still be problematic. Laboratory experiments allow us to obtain a more direct measure of corruption as well as establish causality. This chapter presents the results of a laboratory experiment in which participants in the role of public officials either accept or reject a bribe and then decide between a neutral and a corrupt action. The corrupt action benefits the briber but poses a large negative externality on a charity. In the experiment, I exogenously vary public officials’ wages and find that increasing public officials’ wages reduces their corruptibility. In particular, it makes experienced public officials 53 percentage points less likely to accept a bribe on average and reduces the number of corrupt actions by 27 percentage points.

The remainder of this chapter is organized as follows. The next section provides a discussion of related studies as well as possible mechanisms that can explain the link between wages and corruption. Section three then provides an overview of the bribery model that forms the basis of the experiment. Section four covers the design of the experiment and section five explains the experimental hypotheses.

---

2 See Svensson [2005] for an overview of several studies that find a detrimental effect of corruption on economic performance.

3 As just one example, in recent years the Indian government has repeatedly emphasized the importance of fighting corruption, see for example http://www.nytimes.com/2011/02/17/world/asia/17india.html.
3.2 Background

Previous studies have suggested at least two reasons why increasing public officials’ wages could reduce the level of corruption. Firstly, increasing public official salaries may increase the expected monetary costs of corruption. A wage increase will reduce the relative value of the wage a public official could expect to earn in the private sector. Under the right combination of monitoring and punishment, the expected loss from corruption for public officials will increase, inducing them to behave less corruptly (This is the mechanism suggested by Becker and Stigler, 1974; see also Olken, 2007).

Increasing public officials’ wages may also increase the nonmonetary or ‘moral’ costs of corruption. There are at least three reasons why nonmonetary costs of corruption could be increasing in the wage of public officials. Firstly, public officials may perceive a high wage as more fair, making it more (morally) costly for them to go against the government’s wishes by behaving corruptly; this idea is similar to the fair wage-effort hypothesis (Akerlof and Yellen, 1990; see also Van Rijckeghem and Weder, 2001). Secondly, there may be a social norm condoning side payments for low wage public officials but not for high wage public officials (Fisman and Miguel, 2007). Thirdly, inequality averse public officials may be more willing to increase their income through corruption if their wage is lower than the comparison wage (Fehr and Schmidt, 1999; Abbink, 2005).

However, field studies have produced little evidence in favor of the link between corruption and public sector salaries. I found four directly relevant empirical studies, which are also discussed in Svensson (2005). These studies are Rauch and Evans (2000), Treisman (2000), Van Rijckeghem and Weder (2001) and Di Tella and Shargrodsky (2003). Of these four, the first two find no robust evidence; the latter two find a small negative association. However, as Svensson argues, an additional mechanism applies if public officials’ utilities are a concave function of money. Having a large salary will then decrease their marginal utility of money, decreasing the attractiveness of accepting bribes.
the first three studies are based on cross-country data that make it hard to establish causality; moreover they use ranked data rather than absolute levels to measure corruption. Di Tella and Schargrodsky (2003) make use of exogenous variation in the audit probability in the city of Buenos Aires, which increases the risks involved in corruption and does not directly affect the relative wage of public officials.

In response to this apparent difficulty in acquiring high quality data, the last decade has seen a large increase in the number of laboratory experiments in the area of corruption. Lab experiments can be used as a substitute for field data when field data are not available or of low quality - as is often the case in the area of corruption. Even if good field data are available, lab experiments can serve a complementary role by presenting an environment with a level of control and noiselessness that cannot be achieved in the field. Starting with Frank and Schulze (2000) and Abbink, Irlenbusch, and Renner (2002), corruption experiments have investigated issues ranging from the effect of staff rotation (Abbink 2004), culture (Barr and Serra, 2010) Cameron et al. (2009) and intermediaries (Drugov, Hamman, and Serra 2011) to the effect of different voting systems (Azfar and Nelson 2007); see Abbink (2006) for an overview.

Laboratory experiments have also previously been used to investigate the influence of public officials’ wages on their corruptibility. Abbink (2005) investigates the link between wages and corruption by varying the wage of public officials with respect to the wage of a third party and found no effect. Armantier and Boly (2008) compare the results of a framed lab and field experiment in which participants have to grade homeworks. In one of the homework sets, graders receive a bribe accompanied by a request to be lenient in grading. They find that increasing graders’ wages decreases their corruptibility, although this effect is significant only in the lab with a large set of controls. Azfar and Nelson (2007) find that higher wages decrease the corruption of an executive party in a public choice experiment but have no effect on the corruptibility of an attorney general. Finally, Jacquemet (2005) studies a three player corruption game with delegation and finds that corruption actually increases in the wage of the public officials.\footnote{Jacquemet conjectures that this is caused by the fact that being corrupt is costly in the experiment, so that high wage public officials can more easily afford to be corrupt. Barr, Lindelow, and Serneels (2009) also document a link between public officials’ wages and corruption in a laboratory experiment. However, in this study the monitoring rate is endogenously determined.}
Overall, the laboratory evidence on the link between wages and corruption thus appears to be rather mixed as well.

One possible reason why previous studies examining the link between wages and corruption have yielded mixed findings is that they have paid only limited attention to the question of what constitutes an appropriate reference wage. Indeed, both monetary and nonmonetary considerations require a reference wage to determine what wage should be regarded as a ‘high’ wage and what wage constitutes a ‘low’ wage. Previous field studies have tended to take aggregate level variables as reference wages, such as for example the average wage in the manufacturing sector (e.g. Van Rijckeghem and Weder 2001). However, previous work in both psychology and economics suggests that people compare themselves to individuals who are similar to them, whom they often interact with and who are salient at the moment a comparison is made (see e.g. Festinger 1954, Buunk and Mussweiler 2001, Suls, Martin, and Wheeler 2002, Sweeney and McFarlin, 2004 or see Linde and Sonnemans 2012 for a recent application in economics).

For income comparisons colleagues or other people encountered in work environments are the most likely reference points. A typical economist for example may compare her wage to economists of a similar age, working in the same field and possibly at the same or similar level universities. For public officials on the verge of deciding whether or not to take a corrupt action, people in the work environment are either colleagues or people that require their service (i.e. potential bribers). At the time of a bribe, the focus on the potential briber is likely to be particularly strong since public officials are in direct personal contact with bribers at the time a bribe takes place. Moreover, through bribing public officials and bribers can influence each others’ incomes, making the income comparison between bribers and public officials especially salient. By contrast, aggregate variables such as the average wage in the private sector are abstract (and possibly unknown) and therefore not likely to be salient enough to be the object of an income comparison.

This suggests that the salience of the reference wage should be regarded as an important matter also in experimental studies. In particular, the use of different reference wages may explain the mixed results reported in previous studies. Of the aforementioned experimental studies, only Abbink (2005) explicitly introduces and increasing in the public official’s wage; hence it becomes impossible to separate the effect of wages on corruptibility from the effect of monitoring.
a reference wage in the experiment. He varies the reference wage by varying public officials’ wages with respect to the wages of a third party and finds no effect. However, the third party in this study had no role in the experiment other than to absorb negative externalities. Hence, the third party may not have been very salient to the public official at the time of the bribery decision and may thus not have served as a reference point. Armantier and Boly (2008), Azfar and Nelson (2007) and Jacquemet (2005) do not explicitly address what constitutes the appropriate reference wage in their experiments, which may explain the mixed findings between these studies.

Thus, this chapter contributes to the experimental literature by studying the relationship between the relative wage of public officials and their corruptibility using a salient (and perhaps more natural) reference wage in the experiment. Additionally, it introduces a new way of implementing corruption in the lab by deducting money from a charity fund every time public officials make a corrupt decision; the latter point will be discussed in greater detail in the next section.

3.3 The Bribery Model

To study bribery in an experimental context, I use an adapted version of the experimental bribery game (Abbink, Irlenbusch, and Renner, 2002). The experimental bribery game describes a situation in which a citizen (or firm) can use a bribe to attempt to convince a public official to select a favorable action (or policy) to implement. This reflects for example situations where a citizen needs to acquire a driver’s license or needs a construction permit to build a new home. Importantly, the action that is favorable to citizens imposes a negative externality on society, as is the case for instance if the citizen is an incapable driver or wants to build a new house in a protected forest area.

The experiment is a repeated game of 25 periods. In the stage game (displayed in figure 3.1), the citizen (C) decides whether to offer a transfer (or bribe) of a nonnegative integer amount $t$ to the public official (P). If a positive transfer

---

6 The third party was performing a useful task, but not one that was related to the experimental situation the public official and the potential briber were partaking in.

7 Additionally, since the third party consisted of laboratory subjects not contributing to the group income, some public officials may have felt that the third party deserved a punishment for not being productive.
3.3. THE BRIBERY MODEL

Notes. In the game tree, C represents the citizen, who moves first. P represents the public official, N is nature (or chance) and S is society (i.e. the charity). The gallows represents the possibility of disqualification.

has been offered (i.e. if \( t > 0 \)), the public official decides whether to accept or reject the transfer. If the public official has decided to accept the offer, there is a small probability (.003) that both players are caught and receive a punishment. To mimic the possibly large fines and job loss associated with getting caught in the corrupt act in practice, the punishment in the experiment is set to the largest feasible level. This means that players who are caught are disqualified from the experiment, which means that they lose all their earnings in the current and preceding periods and are not allowed to participate in subsequent periods.\(^8\)

Provided that players have not been disqualified, the public official can then choose between two alternatives G and B. Here G is a status quo action and B is a corrupt alternative. Option B is a genuinely corrupt option, since choosing it

---

\(^8\)In the experiment, disqualified participants still received a show-up fee of 7 euros. With the probability of punishment set to .003, pairs with positive transfers in all 25 periods had a probability of of \( 1 - .997^{25} = .072 \) of being disqualified.
will take money away from a good cause (a charity). However, a selfish citizen strongly prefers option B to option G to represent the gains to corruption. Option B is slightly less favorable to the public official to represent the idea that she will need to exert some effort to justify a ‘corrupt’ choice to her superiors.

Allowing the cost of corruption to be imposed on a charity represents a new approach in the literature. Using a charity as the victim of corrupt behavior reflects the way corruption imposes negative externalities on society in the field. In particular, the same way that corruption is almost universally regarded as a bad thing, not many people would condone taking money away from a charity. By contrast, previous experimental studies have largely imposed negative externalities on other laboratory subjects, which may not have such clear negative moral connotations. For example, if a participant in an experiment expects other participants to be corrupt, he may actually feel that they deserve to have money taken away from them.

Returning to the game tree in figure 3.1, note that the subgame perfect Nash equilibrium (for selfish preferences) of the stage game is for the public official to always choose option G and for no transfers to take place. As the last mover the selfish public official will always choose option G - the option that gives her the highest payoff. As a consequence, the citizen knows that he should not offer a transfer, since offering a transfer can only lower his payoff.\footnote{Technically this holds only if the citizen expects the public official to accept the transfer with positive probability, otherwise the citizen will be indifferent between proposing and not proposing a transfer.} Moreover, Abbink, Irlenbusch, and Renner (2002) use a mathematical induction argument to show that the stage game result holds for all periods in a repeated game as well.

The experiment uses two treatments varying with respect to the wage of the public official. Figure 3.1 gives the pay-offs associated with both treatments. The public official’s wage is either equal to the income of the briber in the status quo option G (treatment LOW) or higher (treatment HIGH).

### 3.4 Experimental Design

The experiment was conducted with 76 participants over four sessions in June 2010 and June 2011 at the CREED laboratory of the University of Amsterdam. Participants were recruited using an online recruitment procedure. The vast
majority were students, with the largest fraction (52%) from the economics department.

The experiment itself was computerized using PHP/MySQL. Upon entering the laboratory, subjects were randomly assigned to a computer terminal and received a set of instructions. As part of the instructions, participants went through a set of questions to make sure they fully understood the instructions. The instructions and questions are reproduced in appendix A.

At this point, it is worthwhile to emphasize that the experiment avoided corruption-related words like bribe, citizen or public official. Instead, the experiment referred to the citizen and the public official as player 1 and player 2 respectively. Moreover, the bribe was called a transfer and actions B and G were called options X and Y. Note, however, that Abbink and Hennig-Schmidt (2006) found no evidence of a framing effect on the results in a bribery experiment that also builds on Abbink, Irlenbusch, and Renner (2002).

After finishing the check-up questions participants were asked to choose a charity for the current session. At the beginning of every session, a substantial sum of money (5000 experimental points or 50 Euros) was reserved for a single charity. As part of the instructions, participants were told that every time any public official in the current session chose option B (the corrupt option), this would lower the charity fund by 30 points. At the end of the instructions, participants were asked to pick one charity from a list of five charities that are well-known in the Netherlands. These were UNICEF, the Dutch Red Cross, the World Wildlife Foundation, Cliniclowns and the Prins Bernhard Cultuurfonds. They could also specify another charity of their choosing, although they were told that including a controversial charity could lead to the payment being awarded to another charity instead. At the end of the session, the charity choice of one randomly determined

---

10Relative to a fixed charity, allowing participants to select from multiple charities made it possible for them to select a charity that fit better with their personal tastes. With a fixed charity, it would have been possible for at least some participants to not care about the one charity in the experiment at all; allowing participants to pick their own charity decreased this chance. Since choosing a certain charity increased the chance that this charity would be picked, each participant had the incentive to pick his or her preferred charity.

11Of these five charities, the first three are well-known internationally as well. The Cliniclowns are an organization of Dutch clown doctors, who seek to help alleviate some of the stress in seriously ill, hospitalized young children. The Prins Bernhard Cultuurfonds sponsors a wide range of cultural activities in the Netherlands, such as theater, art and the conservation of architectural monuments.
participant was implemented.\footnote{12}

After every participant had finished the instructions and check-up questions and chosen a charity, the experiment started. Every session consisted of 25 periods. Before the first period, every participant was told their role (citizen or public official). Their role remained fixed over the whole experiment and public officials were matched to the same citizen for all 25 periods.\footnote{13}

Every period in the experiment consisted of five stages (see figure 3.1). In the first stage, citizens decided whether or not to offer a bribe. Conditional on offering a bribe, they could specify the size of the bribe in stage 2. In stage 3, public officials decided whether or not to accept the proposed bribe. Conditional on accepting the bribe, stage 4 consisted of a random draw that determined disqualification; disqualified subjects were immediately notified and asked to fill out an unrelated questionnaire for the remainder of the experiment. Finally, in stage 5 public officials had to choose between options G and B. Note that many pairs skipped stages 2, 3 and/or 4 in several periods. For example, citizens who did not offer a bribe skipped stages 2, 3 and 4. The decision screen displayed all possible moves by both players and indicated at what stage the players had currently arrived. The decision screen is reproduced in appendix B.

Every period ended only after all pairs had finished stage 5; for all pairs the waiting screen between periods displayed the results of all preceding periods for the given pair. After 25 periods the experiment ended and one subject was randomly picked to roll a die to determine the winning charity. Participants then received an overview of their earnings and were asked to fill out a questionnaire. The questionnaire contained background questions, motivational questions, a questionnaire related to corruption taken from Rabl and Kühlmann (2008) and a psychological questionnaire relating to aggression (Buss and Perry 1992). Upon finishing the questionnaire, participants were paid their earnings (including a show-up fee of 7 euros) and were kindly requested to leave the laboratory.

Participants earnings ranged from 14.14 to 23.70 euros with an average of 17.63 euros. Charities earned between 20.60 and 41.90 euros, with an average of

\footnote{12}The number of participants that chose UNICEF, the Red Cross, the WWF, the Cliniclowns, the Prins Bernhard Cultuurfonds and another charity was equal to 34, 15, 14, 6, 1 and 6 respectively. None of the 6 alternative charities were too controversial to exclude. The winning charities were the Red Cross (once) and the WWF (three times).

\footnote{13}In other words the experiment used a partners design. See Abbink 2004 for an experimental analysis of the effect of using a partners or a strangers design in a bribery experiment.
3.5. **HYPOTHESES**

31.40 euros. In total every session lasted approximately 75 minutes (15 minutes for the instructions, 30 minutes for the decision problem and 30 minutes for the questionnaire plus payment). Since no feedback from other pairs was given to participants, the number of independent observations is equal to 38.

### 3.5 Hypotheses

This study examines the relationship between an increase in the relative wage of public officials and their corruptibility. There are at least two reasons for high wage public officials to be more reluctant to accept a bribe. Firstly, they may face higher non-monetary costs of corruption. Inequality averse public officials will for example note that accepting a bribe in treatment LOW may decrease disadvantageous inequality (a good thing), whereas accepting a bribe in treatment HIGH will increase advantageous inequality (a bad thing). Public officials who care about status can guarantee themselves a higher income level than the briber without accepting a bribe in treatment HIGH, whereas in treatment LOW a large bribe is necessary to guarantee a higher income level than the briber. Secondly, public officials in treatment HIGH have more to lose from accepting a bribe if caught (i.e. a higher monetary cost). Both monetary and non-monetary mechanisms lead to the following hypothesis.

**Hypothesis 1:** Public officials are less likely to accept a bribe in treatment HIGH than in treatment LOW.

If the monetary and nonmonetary costs of corruption are increasing in the public official’s relative wage, then the frequency of corrupt (B) choices should decrease as well. The bribery relationship is a reciprocal relationship between a briber and a public official. Public officials who want to continue a bribery

---

14In the remainder of this chapter I will focus mostly on the behavior of public officials. The reasons for deemphasizing citizens are that citizen behavior (a) is not directly relevant to the link between the wages and corruptibility of public officials, (b) is less interesting in scope (only a transfer offer) and (c) crucially depends on how citizens expect public officials to behave (in contrast to public officials, who already know the behavior of the citizen by the time they have to make their decisions).

15A possible third mechanism -that public officials’ utility functions are concave in money- is unlikely to have a large bite in the experiment since for small amounts it is reasonable to assume that utility functions are approximately linear.
relationship should pick option B after accepting a bribe to reciprocate the briber. If hypothesis 1 holds, public officials in treatment HIGH are less likely to accept bribes. Thus, there is less reason for them to maintain the bribery relationship, which means they are less likely to pick option B. This leads us to the second hypothesis.

**Hypothesis 2:** Public officials are less likely to choose the corrupt option B in treatment HIGH than in treatment LOW.

### 3.6 Results

This section presents the results of the experiment. Before moving on to a test of hypotheses 1 and 2, it is important to recall that the Nash equilibrium prediction of the model of section 3.3 is that no bribery will take place. Thus, both hypothesis 1 and hypothesis 2 can only have a bite if a significant number of participants fail to stick to the Nash equilibrium. Figure 3.2 shows that in almost all (34/38) pairs transfer/bribe proposals occurred at least once. Moreover, for many pairs transfer proposals were present in a substantial number of rounds; the median number of rounds a bribe was offered is equal to 9 (out of 25). Though somewhat less frequent, B choices also occurred in a large majority of pairs (28/38); the median number of rounds a B decision was made is equal to 3.5.

In the remainder of this section I will report the results for both the whole sample and for periods 11 to 25; I include the latter to minimize the noise generated by participants who are still trying to learn the game. Note, however, that investigating public officials’ corruptibility is only possible for public officials who have been offered at least one bribe. In four pairs (three in treatment LOW, one in treatment HIGH) no bribe was ever offered and these pairs can thus not be incorporated into the analysis.

---

16 In the remainder of this section, I will use the terms transfer and bribe interchangeably.
17 Because of random assignment, whether public officials were ever offered a bribe is random for the whole sample; therefore it is not a problem for any statistics that apply to all periods. For periods 11-25, however, one could worry that attrition may be non-random since bribers may be induced to stop bribing by the public official’s behavior in the preceding periods. In particular, it may be that the results reported in this section overstate the actual wage effect if bribe rejecting public officials are more likely to drop out in treatment LOW and/or bribe accepting officials are more likely to drop out in treatment HIGH. However, neither scenario
3.6. RESULTS

Figure 3.2: Incidence of Transfers and B Choices

Notes. The left bar plots the distribution of the fraction of periods a positive bribe (or transfer) was offered for each pair. The right bar plots the fraction of periods a B choice was made, again for each pair.

3.6.1 Bribe Acceptance

Hypothesis 1 suggests that public officials in treatment HIGH should be less likely to accept bribes. Figure 3.3 shows that this is indeed the case. Public officials in treatment LOW accept on average 80% of proposed bribes (91% for periods 11-25), whereas public officials in treatment HIGH accept 44% of bribes (38% in periods 11-25). This difference is statistically significant for the whole sample (Mann-Whitney; $N_{LOW} = 17, N_{HIGH} = 17$, $z=3.109$, $p=.002$) and for periods 11 to 25 (Mann-Whitney; $N_{LOW} = 12, N_{HIGH} = 15$, $z=3.653$, $p=.000$).\(^{18}\) Thus, the evidence is in line with hypothesis 1: increasing public officials’ wages reduces the acceptance rate of bribe offers.\(^{19}\)

This result by itself does not tell us why public officials chose to accept fewer

\(^{18}\text{Probit regressions with clustered standard errors by pair yield similar results; } z=-3.28, p=.001 \text{ for all periods and } z=-3.89, p=.000 \text{ for periods 11-25). In what follows I will only report the results of regressions if they lead to a different conclusion than the Mann-Whitney test.}\)

\(^{19}\text{In total, 120 bribes were accepted over all sessions; no pair was actually disqualified in the experiment. The probability of no disqualifications with 120 bribes is equal to } (1-.003)^{120} = .697.\)
Figure 3.3: Fraction of Accepted Transfers

Notes. The figure plots the cumulative distribution of transfer acceptance rates by pair for both treatments. The upper panel reports the results for all periods, the lower panel reports the results for periods 11-25.
3.6. RESULTS

Table 3.1: Motivations for Transfer Acceptance

<table>
<thead>
<tr>
<th></th>
<th>LOW</th>
<th>HIGH</th>
<th>Difference</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Own payoff</td>
<td>4.30</td>
<td>3.89</td>
<td>−.41</td>
<td>.562</td>
</tr>
<tr>
<td>Charity payoff</td>
<td>3.15</td>
<td>4.95</td>
<td>1.80</td>
<td>.004***</td>
</tr>
<tr>
<td>Player 1’s payoff</td>
<td>3.95</td>
<td>3.95</td>
<td>.00</td>
<td>.952</td>
</tr>
<tr>
<td>Observations</td>
<td>20</td>
<td>18</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* significant at 10%; ** significant at 5%; *** significant at 1%

Notes. This table gives the average response to three questions in the post-experimental questionnaire. These questions were “In deciding to accept player 1’s transfer offer the charity’s/my own/player 1’s pay-off was an important factor.” Answers were reported on a Likert Scale ranging from 1 to 7. The reported p-values are calculated using Mann-Whitney tests.

bribes in treatment HIGH. One possible reason is that public officials in treatment HIGH were offered smaller transfers. This would require (a) that citizens in treatment HIGH indeed offered smaller transfer and (b) that public officials were more likely to reject smaller transfer offers. However, the data provide little evidence for either claim. Indeed, although average proposed transfer size is slightly lower in treatment HIGH (9.57) than in treatment LOW (11.39), the difference between treatments is significant at the 10% level only for the whole sample (Mann-Whitney; \(N_{LOW} = 17, N_{HIGH} = 17\), \(z=1.671, p=.095\)) and not significant for periods 11 to 25 (Mann-Whitney; \(N_{LOW} = 12, N_{HIGH} = 15\), \(z=1.199, p=.230\)). Importantly (and perhaps surprisingly) public officials are not more likely to reject smaller transfer offers. These two findings combined suggest that the difference in acceptance rates between treatments cannot be explained by differences in transfer amounts.

Possible supplementary evidence for the motivations of public officials comes from the post-experimental questionnaire. In the questionnaire, public officials answered the following questions: “In deciding to accept player 1’s [i.e. the citizen’s] transfer offer the charity’s/my own/player 1’s pay-off was an important factor.” Table 3.1 reports the results of the three questions by treatment. Strikingly, for public officials in treatment HIGH, avoiding damage to the charity was named as the most important factor in deciding (not) to accept bribes, whereas for public officials in treatment LOW, the charity was the least important factor. This finding is not consistent with a monetary cost explanation, since for a monetary

---

20In a probit regression of the transfer acceptance decision on transfer amount and a constant, the p-value for transfer amount equals .268 for the whole sample and .464 for periods 11-25.
cost explanation the payoff of the charity is irrelevant. It does however fit with the idea that nonmonetary costs are increasing in public officials’ wages, since high wage public officials care relatively less about their own payoff and the payoff of the briber.

### 3.6.2 G and B Choices

The difference in transfer acceptance rates is also reflected by the percentage of G and B choices in each treatment. Figure 3.4 gives an overview of the percentage of B choices conditional on a transfer having been proposed. For the whole experiment, the percentage of B choices is 15 percentage points lower in treatment HIGH; this difference is not significant (Mann-Whitney; $N_{LOW} = 17$, $N_{HIGH} = 17$, $z=.975$, $p=.330$). For periods 11 to 25 the difference becomes larger (27 percentage points) and significant at the 10% level (Mann-Whitney; $N_{LOW} = 12$, $N_{HIGH} = 15$, $z=1.876$, $p=.061$). The difference seems to become more pronounced in later periods; indeed if for instance periods 16-25 are taken instead of periods 11-25 the difference is significant at the 1% level. Thus, to some extent the number of B choices seems to reflect the difference in bribe acceptance rates described above, although the effect is smaller (27 versus 53 percentage points for periods 11-25).

### 3.7 Robustness: Bribery without Monitoring

Thus far we have seen that increasing public officials’ wages greatly decreases the fraction of transfers they accept and slightly decreases the number of corrupt (B) choices they make. This tells us that within the current experimental setting (positive monitoring rate, large penalty to the charity), increasing public officials’ wages reduces their corruptibility. This section describes the results of additional

---

21 Incidentally, there are no gender differences for either the transfer acceptance decision or the number of B choices. This is true for the two treatments taken separately as well as jointly. The only gender difference appears with citizens; female citizens are more likely to attempt to bribe a public official than male citizens ($p=0.025$).

22 The reason the difference between the number of B choices is smaller than the difference between transfer rates is due to two factors. For one, not all accepted transfers lead to B choices; the number of accepted transfers leading to G choices is equal to 32.4% for treatment LOW and 26.3% for treatment HIGH. For another, the fraction of B choices taken after rejected transfers is not equal to zero (it is equal to 6.7% and 13.3% for treatments LOW and HIGH respectively).
3.7. ROBUSTNESS: BRIBERY WITHOUT MONITORING

Figure 3.4: Fraction of B Choices

Notes. The figure plots the cumulative distribution of the percentage of B choices by pair for both treatments. The upper panel reports the results for all periods, the lower panel reports the results for periods 11-25.
sessions that explore the robustness of these findings to setting the monitoring rate to zero. A zero monitoring rate is also of practical interest, since monitoring activities in practice are costly and often subject to corruption themselves; they should as such only be maintained if necessary to reduce corruption levels.

Setting the monitoring rate to zero removes the possibility of disqualification from the experiment and thus removes monetary costs considerations from public officials. Thus, to the extent that monetary costs were relevant with a monitoring rate of .003, we should expect a smaller treatment effect with a monitoring rate of zero. However, monetary costs were already quite small even with monitoring. Indeed, with monitoring the only predicted difference for risk neutral public officials is that they should accept all bribes larger than 3 in treatment LOW and all bribes larger than 5 in treatment HIGH. In the experiment, however, only 5% of proposed bribes were equal to 3 or 4. Moreover, we saw that bribe size does not affect the probability that a transfer is accepted. Thus with risk neutrality monetary costs can only explain a small fraction of the total difference between treatments.

Setting the monitoring rate to zero may also affect the nonmonetary costs of corruption. In particular, a positive monitoring rate may be a signal to public officials that accepting a transfer is not a moral or normative thing to do. Without monitoring this signal disappears, which could induce public officials to behave more corruptly.

To investigate the influence of setting the monitoring rate to zero I ran an additional four sessions in June 2011. These sessions were identical to the sessions described in the previous sections, except that the monitoring rate was equal to zero instead of .003. In total, 84 subjects took part in these sessions, earning between 13.42 and 21.88 euros. Charities earned between 11.00 and 34.10 euros, with an average of 25.50 euros.

To analyze the influence of monitoring on the influence of wage increases

\[ \text{In the experiment, average earnings over all periods for public officials in treatment LOW and HIGH were 943 and 1420 points respectively. Thus, accepting a single bribe leads to an expected loss from disqualification equal to } 0.003 \times 943 = 2.83 \text{ and } 0.003 \times 1420 = 4.26 \text{ for treatments LOW and HIGH respectively. Risk-neutral public officials should only accept bribes that exceed the expected loss from disqualification.} \]

\[ \text{Introducing risk aversion would predict stronger differences, although with small probabilities risk seeking is more commonly observed than risk aversion. See e.g. Tversky and Kahneman, 1992 and Abbink, Irlenbusch, and Renner, 2002 for evidence that subjects underestimate disqualification probabilities in a bribery game.} \]
3.7. ROBUSTNESS: BRIBERY WITHOUT MONITORING

Figure 3.5: Transfer Acceptance Rate with and without Monitoring

Notes. The figure plots the cumulative distribution of transfer acceptance rates by pair for both sessions with monitoring and sessions without monitoring. The top panel displays the results for all periods and the lower panel displays the results for periods 11-25.
Table 3.2: Overview of Transfer Acceptance Rates

<table>
<thead>
<tr>
<th></th>
<th>LOW</th>
<th>HIGH</th>
<th>Difference</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monitoring</td>
<td>.91</td>
<td>.38</td>
<td>−.53</td>
<td>.000***</td>
</tr>
<tr>
<td>No Monitoring</td>
<td>.97</td>
<td>.79</td>
<td>−.18</td>
<td>.016**</td>
</tr>
<tr>
<td>Difference</td>
<td>.06</td>
<td>.41</td>
<td>.35</td>
<td>.013**</td>
</tr>
<tr>
<td>P-value</td>
<td>.368</td>
<td>.002***</td>
<td>.013**</td>
<td></td>
</tr>
</tbody>
</table>

* significant at 10%; ** significant at 5%; *** significant at 1%

Notes. This table gives the mean transfer acceptance rates conditional on a transfer being offered. That is, first I computed the average acceptance probability for every pair and then averaged these probabilities over all pairs for every treatment. The reported p-values for the difference estimators are calculated using Mann-Whitney tests. The p-value for the difference-in-difference estimator (.013) is calculated using an OLS regression of the transfer acceptance decision on a treatment dummy, a dummy for monitoring and an interaction of the two dummies; the p-value corresponds to the p-value of the interaction term. Only periods 11 to 25 are used, for OLS standard errors are clustered at the pair level.

on corruptibility, I compare the results of these sessions with the results of the previous section. Figure 3.3 and table 3.2 give the transfer acceptance rates for both treatment HIGH and treatment LOW for sessions with monitoring and sessions without monitoring. Relative to sessions with monitoring, the difference in transfer acceptance rates in sessions without monitoring falls from .36 to .09 in all periods (p=0.020) and from .53 to .18 in periods 11 to 25 (p=0.013). Table 3.2 shows that this change is driven almost exclusively by the greater corruptibility of HIGH wage public officials in sessions without monitoring. As a consequence, the difference in the percentage of B choices falls from 27 percentage points to 4 percentage points in periods 11-25 and from 15 percentage points to 1 percentage point in all periods. Finally, table 3.3 shows that self-reported care for the charity in treatment HIGH drops to the level of treatment LOW in sessions without monitoring, whereas it was substantially higher in sessions with monitoring.

All in all this suggests that the evidence presented in the previous section is only partially robust to the removal of monitoring. In fact, it suggests that both monitoring and a high wage are necessary to decrease corruption. However, this result may also be due to a ceiling effect in the transfer acceptance rates in treatment LOW. Even with monitoring public officials in treatment LOW accepted 91% of proposed transfers on average in periods 11-25, meaning there was hardly any scope for the transfer acceptance rate to increase. On the other hand, with monitoring the average acceptance rate was only 38% in treatment
### Table 3.3: Importance of Charity in Transfer Acceptance

<table>
<thead>
<tr>
<th></th>
<th>LOW</th>
<th>HIGH</th>
<th>Difference</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monitoring</td>
<td>3.15</td>
<td>4.95</td>
<td>1.80</td>
<td>.004***</td>
</tr>
<tr>
<td>No Monitoring</td>
<td>3.78</td>
<td>3.48</td>
<td>−.30</td>
<td>.565</td>
</tr>
<tr>
<td>Difference</td>
<td>.63</td>
<td>−1.47</td>
<td>2.10</td>
<td>.015 **</td>
</tr>
<tr>
<td>P-value</td>
<td>.236</td>
<td>.035 **</td>
<td>.015 **</td>
<td></td>
</tr>
</tbody>
</table>

* significant at 10%; ** significant at 5%; *** significant at 1%

**Notes.** This table gives the average response to the question “In deciding to accept player 1’s transfer offer the charity’s pay-off was an important factor” in the post-experimental questionnaire. Answers were reported on a Likert Scale ranging from 1 to 7. The reported p-values for the difference estimators are calculated using Mann-Whitney tests. The p-value for the difference-in-difference estimator (.015) is calculated using an OLS regression of the transfer acceptance decision on a treatment dummy, a dummy for monitoring and an interaction of the two dummies; the p-value corresponds to the p-value of the interaction term.

HIGH, leaving a lot of room for the transfer acceptance rate to increase.

### 3.8 Discussion

In this study, I have investigated the link between public officials’ wages and their corruptibility. The results show that increasing the wage of public officials reduces their corruptibility. In particular, experienced low wage public officials accept 91% of bribes, whereas experienced high wage public officials accept only 38%. Moreover, high wage public officials are 27 percentage points less likely to choose the corrupt option. A robustness check suggests that a positive monitoring rate may be necessary for higher salaries to affect the corruptibility of public officials.

All in all, these results provide greater support for a link between wages and bribery than previous experimental studies. The contrast with Abbink (2005) is particularly illuminating since its experiment is based on the same bribery model and also has a positive level of monitoring. The difference in findings suggests that the reference wage is important; if a third party is used as a reference wage as in Abbink, the relative wage of the public official does not seem to matter. By contrast, if the briber is used as the reference wage as in this study, there is a large and statistically significant effect.

This suggests that in empirical studies that investigate the link between wages and corruption, it is important to use an appropriate reference wage. In particular,
if the reference wage that is used in the study is not used as a reference wage by most public officials, one will not find a relationship between wages and corruption even such a relationship does exist with the appropriate reference wage. Taken together with the results of Abbink, this study could thus provide an explanation for the mixed results reported in both laboratory and field studies.

As the robustness check showed, a positive monitoring rate seems to be necessary for high wages to decrease corruption. At the same time, even the positive monitoring rate used in the experiment was very small. This suggests that the exact size of the monitoring rate may not be important for small monitoring rates; rather the presence of monitoring may serve as a signal to public officials that accepting bribes is not a morally acceptable thing to do. Even if nonmonetary costs are important, a small but positive level of monitoring may thus be necessary to reduce corruption.

For future experimental work, several extensions are possible. It may for example be interesting to vary the wage of public officials within the same session. To the extent that the wages of colleagues can also serve as reference wages, it may be expected that public officials with wages that are higher than both colleagues and bribers will be even less likely to accept bribes. Another possibility would be to allow public officials to solicit bribes rather than have them wait for bribers to offer one, as in Barr and Serra (2010). These extensions may help provide additional insights on the conditions that need to be met for the link between wages and corruptibility to appear.
Appendix 3.A  Instructions

Welcome to the CREED laboratory. Please read the following instructions carefully.

In today’s experiment, there are two types of participants: Player 1 and Player 2. Your type will be randomly drawn after everyone has finished the instructions. You will then also be randomly matched to a player of the other type. Both your type and the player you are matched with will remain unchanged throughout the experiment.

All in all the experiment consists of 25 periods. The payment you receive at the end of the experiment depends on the decisions you make. Moreover, you will be able to earn money for a charity. The currency of the experiment is the experimental franc. At the end of the experiment, all francs you earned will be converted into euros at a rate of 100 francs per euro, such that 1000 francs are worth 10 euros. You will also receive a show-up fee of 7 euros.

Decision Situation

Every period in this experiment consists of 5 stages, which will always take place in the following order:

Stage 1: Transfer or no Transfer

Player one decides whether or not he wants to transfer an amount to player two. If he does, then the period is continued with stage 2. If player one decides not to transfer an amount, then the period continues with stage five.

Stage 2: The Amount to Be Transferred

Player one decides on the amount to be transferred to player two. The transferred amount can be any whole number greater than zero. The period then continues with stage 3.

25These instructions are the instructions for both public officials and citizens in the LOW wage treatment with monitoring. In the instructions for treatment HIGH, only the payoffs mentioned in stage 5 change. In sessions without monitoring, stage IV is omitted and stage V is called stage IV instead.
Stage 3: Acceptance or Rejection of the Transfer

Player two then decides whether he accepts or rejects the proposed transfer. If player two decides to accept the transfer, the proposed amount is removed from player 1’s credit and added to player 2’s credit. The period then continues with stage 4. If player two rejects the transfer, then the credits remain unchanged. The period is then continued with stage four.

Stage 4: Possibility of Getting Disqualified

If player 2 decided to accept the transfer in stage 3, a number out of the range from 0 to 999 is randomly drawn. If the number is 0, 1 or 2, then both player 1 and player 2 are disqualified. That means that the experiment ends for these two players and all their previous earnings are canceled. (At the end of the experiment, both players receive only their show-up fee.) The two disqualified participants fill in a questionnaire until the experiment has ended. For the other participants, the experiment continues normally. If the randomly drawn number is 3, 4, ..., 998, or 999, the period is continued with stage 5 (see next page).

Stage 5: Player 2 Chooses Between X and Y

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Player 1</td>
<td>36</td>
<td>56</td>
</tr>
<tr>
<td>Player 2</td>
<td>36</td>
<td>30</td>
</tr>
<tr>
<td>Charity</td>
<td>0</td>
<td>-30</td>
</tr>
</tbody>
</table>

Player 2 chooses one of the alternatives X or Y. If player 2 selects alternative X, then his credit is increased by 36 and the credit of player 1 is increased by 36 (as in the table above). The credit of the charity remains unchanged. If player 2 selects alternative Y, then his credit is increased by 30 and the credit of player 1 is increased by 56. The credits of the charity are decreased by 30 francs.

There will be only one charity for this experiment. The charity starts off with a total of 5000 francs, which is equal to 50 euros. The final donation depends on the decisions made by the participants in the experiment. The donation will be strictly anonymous; no mention will be made of either the UvA, CREED or any participant of this experiment.
After stage 5, the period has ended. Overall pay-offs are the sum of all changes of credits during the 5 stages of the period.

The Pay-Offs

The decision situation will be repeated for 25 periods. You receive your earnings at the end of the experiment, where the exchange rate is 1 euro for 100 francs. In addition you will receive a show-up fee of 7 euros.

Question 1

Suppose you are player 2 and player 1 has proposed a transfer of 8. If you accept, what will be your (player 2’s) pay-off if you choose option X? What will be player 1’s pay-off in this case? What will be the pay-offs for option Y? TIP: look up the values for X and Y on one of the previous pages or on the printout of the instructions.

Question 2

In this experiment, there are a total of 20 participants, such that there are 10 pairs. Suppose that in a certain period there are 5 pairs in which player 2 chooses option Y. How many francs will the charity lose in this period?

Charities

For this experiment, we have selected a total of five charities. At the end of the experiment, we will pick the charity selected by one randomly determined person. Thus, the likelihood that a charity is picked is proportional to the number of people that picked this charity. For example, a charity chosen by six people will be three times more likely to be picked than a charity chosen by two people. If you would like to support another charity, you can select option ‘F: Other Charity’ and type the name of the charity in the text box. We must emphasize that a self-chosen charity will only be paid out if it passes a ‘fit-and-proper-charity’ test. For example, organizations like AlQaeda or your best friend’s holiday fund will be considered invalid charities. If an invalid option is drawn, we will redraw until a valid charity has been selected.
A. **UNICEF:** Created by the United Nations General Assembly on December 11, 1946, to provide emergency food and health care to children in countries that had been devastated by World War II. Presently, its activities include promoting children’s rights, and securing worldwide visibility for children threatened by poverty, disasters, armed conflict, abuse and exploitation. UNICEF was awarded the Nobel Peace Prize in 1965.

B. **WWF/WNF:** Founded on September 11, 1961, its official mission is “to halt the destruction of our environment”. Currently, the WWF focuses on restoring populations of 36 species (including elephants and tunas) as well as conserving 25 globally important ecoregions (including the Amazon Forest).

C. **Red Cross:** Founded on February 9, 1863, its official mission is “to stand for the protection of the life and dignity of victims of international and internal armed conflicts.” Amongst its activities, it attempts to organize nursing and care for those who are wounded on the battlefield; it also supervises the treatment of prisoners of war.

D. **Cliniclowns:** Founded in 1992, its goal is to cheer up severely sick or handicapped children to help them recuperate from their ailments. Its most important activity is to send clowns to visit children’s wards to cheer up the children, but it has also started a theater tour for children with multiple disabilities.

E. **Prins Bernhard Cultuurfonds:** Founded in 1940 by Prince Bernhard of the Netherlands, its goal is to support projects that work to preserve Dutch cultural and natural heritage. Its activities include awarding prices and scholarships to talented musicians, poets and other artists. On average, it supports 4000 projects per year.
Appendix 3.B  Screenshot

Figure 3.6: Decision Screen
Chapter 4

Willpower Depletion and Framing Effects

This chapter investigates whether depleting people’s cognitive resources (or ‘willpower’) affects the degree to which they are susceptible to framing effects. Recent research in social psychology and economics has suggested that willpower is a resource that can be temporarily depleted and that a depleted level of willpower is associated with self-control problems in a variety of contexts. In this study, we extend the willpower depletion paradigm to framing effects and argue that willpower depletion should increase framing effects. To test this idea we ran an experiment in which we depleted participants’ willpower and subsequently had them take part in a series of framing tasks, including a framed prisoner’s dilemma, an attraction effect task, a compromise effect task and an anchoring task. However, we found no evidence that framing effects were indeed more prevalent in willpower depleted participants than in controls.

4.1 Introduction

Every day people are subjected to temptations that require self-control to be resisted. A recent literature in social psychology has emphasized that exercising self-control requires willpower and that people’s willpower should be regarded as

1This chapter is based on De Haan and Van Veldhuizen (2012). We would like to thank Gary Charness, Joep Sonnemans and other colleagues from the Center of Research in Experimental Economics and Political Decision-making (CREED) and the University of Amsterdam for their helpful comments. Financial support from the University of Amsterdam Research Priority Area in Behavioral Economics is gratefully acknowledged.
a limited resource (see for example Muraven and Baumeister 2000; Baumeister et al. 2008, or see Hagger et al. 2010 or Bucciol, Houser, and Piovesan 2010). In particular, this literature argues that exercising self-control can temporarily ‘deplete’ the willpower resource. Once someone’s willpower has been temporarily depleted, this will make him or her be more likely to yield to subsequent temptations.

The resource (or willpower) depletion literature has recently started to get a foothold in economics. Most commonly, willpower depletion has been related to self-control, procrastination and affiliated concepts. Recent theoretical models that have explicitly incorporated willpower depletion include Ozdenoren, Salant, and Silverman (2011); Ali (2011) and Fudenberg and Levine (2012). The empirical literature is slightly sparser, although Bucciol, Houser, and Piovesan (2011a,b) find that a willpower depleting activity reduces subsequent productivity in both children and adults. Another related paper is Burger, Charness, and Lynham (2011), who investigate the effect of willpower depletion on procrastination and find that depleted subjects are more likely to postpone filling out a questionnaire for a day.

In this paper we extend the economic literature on willpower depletion to another aspect of individual decision making. In particular, we use a laboratory experiment to investigate if depleting people’s willpower increases their susceptibility to framing effects. Framing effects (or context effects) occur when people’s decisions are affected by seemingly irrelevant aspects of decision problems (see e.g. Tversky and Kahneman 1981; Levin, Schneider, and Gaeth 1998; Druckman 2001; De Martino et al. 2006). A popular explanation of framing effects argues that they are the result of the use of simplified decision rules, heuristics or ‘system 1’ processes that selectively process only a limited number of (possibly irrelevant) details of the decision problem (Tversky and Kahneman 1974, 1981). System 1 processes can be overruled by higher level or ‘system 2’ processes, but this requires willpower (Pocheptsova et al. 2009). As a consequence, willpower depletion should increase susceptibility to framing effects.

In the experiment, we use a five-minute version of the well-known ‘Stroop’ task (Stroop 1935) to deplete the willpower of half our participants. The Stroop task is followed by one of five secondary tasks. These tasks are an ‘attraction’ effect

---

[1] Alternative terms used for resource depletion and willpower depletion are ego depletion, cognitive depletion and psychological depletion, amongst others.
task, a ‘compromise’ effect task, an ‘anchoring’ effect task, a framed prisoner’s dilemma and a cognitive task. The first four tasks test for different types of framing effects, whereas the cognitive task allows us to test for differences in cognitive performance. This sequence of a Stroop task followed by a secondary task is repeated five times, such that participants go through all five secondary tasks exactly once.

By comparing framing effects in all four framing tasks between depleted and non-depleted participants, we can investigate if depleted individuals are indeed more susceptible to framing effects. The experiment also allows us to address at least three additional questions of interest. Firstly, the cognitive task allows us to investigate the effect of willpower depletion on cognitive performance. Secondly, the repeated nature of our set-up allows us to trace the effects of depletion over successive tasks. Thirdly, making participants go through four different framing tasks allows us to investigate if participants who are susceptible to one framing effect are more susceptible to other framing effects as well.

All in all, however, we find little evidence that depleted participants are more susceptible to framing. Framing effects appear in the attraction, compromise and anchoring tasks but do not differ between depleted and non-depleted participants. Furthermore, we do not find a depletion effect on the performance of participants in the cognitive task or differential willpower depletion effects over successive tasks; we also find little evidence of between-task correlations in framing effects. In many ways, these results are quite striking since the depleting effects of the Stroop task have been documented in many studies and especially since the effect of depletion on the attraction effect and compromise effect has already been documented by Pocheptsova et al. (2009). This suggests it may be important from a methodological perspective to critically examine the conditions under which willpower depletion is likely to have an effect on performance and under which it is not.

The structure of the rest of the paper is as follows. In section 2 we present the design, where we describe each task in greater detail. In section 3 we describe the research questions, after which we provide the results in section 4 and conclude in

Pocheptsova et al. (2009) describe a series of willpower depletion experiments, two of which examine the compromise effect and one of which examines the attraction effect. To deplete participants they use different methods in different experiments, including the Stroop task for one of the compromise effect experiments. We will say more about the differences and similarities between our study and Pocheptsova et al. (2009) in the following sections.
4.2 Experimental Design

The goal of this study was to investigate the effect of willpower depletion on participants’ susceptibility to framing effects. We therefore conducted an experiment where subjects first participated in a task that either depleted or did not deplete their cognitive resources or willpower. Upon completing this task participants then went through a secondary task in which framing effects had previously been shown to play a role. In the remainder of this section, we first discuss the depletion task and then discuss all five secondary tasks individually, followed by a short description of the experimental procedure.

4.2.1 Depletion Task

For the depletion task, we used the Stroop (1935) task. Originally created to test psychological interference theory, the Stroop task has since been used in many applications (see MacLeod 1991 for an overview of the first 55 years of applications). Importantly, the Stroop test has also been used as a way to deplete participants’ willpower in several studies (see for example Webb and Sheeran 2003; Pochetsova et al. 2009; Burger, Charness, and Lynham 2011). We used a computerized version of the Stroop task, in which participants were asked to indicate the font color a color name was printed in. There were five possible font colors and color names: blue, red, yellow, orange and purple. We adopted two different versions (or treatments) of the Stroop task to experimentally vary the level of willpower depletion in participants. For the control treatment, the font colors were always identical to the color names. For example, the word ‘blue’ would always be printed in blue letters. For the depletion treatment the font color and color name were randomly matched, so that they were identical in only on average one third of all cases.

The depletion treatment of the Stroop task has been argued to lead to depletion in the following way. When a color word is written in a different font color, our initial impulse is to read the semantic meaning of the word. Naming the font color instead requires us to override our initial tendency to read the color word. In terms of the willpower paradigm, overriding our initial (system 1) tendency to
say the color word is a form of self-regulation (system 2) that requires willpower to be undertaken. As a consequence, taking part in the depletion treatment of the Stroop task lowers the amount of remaining willpower that can be used in subsequent tasks. An alternative but closely related way to look at it is that overriding our initial tendencies is a cognitively demanding activity, which leaves fewer cognitive resources for subsequent tasks. Importantly, the control version of the Stroop task requires no self-regulation and fewer cognitive resources, such that control participants should be less depleted than participants in the depletion treatment.

Over the course of the experiment, every participant went through five Stroop tasks; each participant faced the same version (control or depletion) of the Stroop task every time. For each Stroop task, participants faced a random sequence of words drawn by the computer, with randomly matched font colors for the depletion treatment. The color words appeared in the middle of the computer screen; subjects could indicate the color of the word by pressing the corresponding key on the keyboard. After they had given their response, they received a new word after pressing the spacebar and waiting for approximately 0.55 seconds. Each Stroop task lasted precisely five minutes. Incentives were such that participants received one cent for each correct response and were deducted two cents for each incorrect answer. Feedback on the number of correct and incorrect answers was only provided at the end of the experiment.

4.2.2 Secondary Tasks

After every depletion task, participants took part in one of five secondary tasks. Four of these tasks were framing tasks, and one was a cognitive task. The order of the tasks was randomized between participants to minimize potential ordering effects. Each secondary task took six minutes in total; the first two minutes were reserved for an instructions screen, the remaining four minutes were reserved for the task itself.

---

4 The time for a new word to appear was uniformly random between 0.3 and 0.8 seconds.
5 A screenshot of the Stroop task can be found in Appendix C.
CHAPTER 4. WILLPOWER DEPLETION AND FRAMING EFFECTS

Attraction Effect Task

The attraction effect or asymmetric dominance effect (Huber, Payne, and Puto 1982) occurs in decision problems when adding an asymmetrically dominated alternative to a choice set increases the share of the dominating alternative. For example, Huber, Payne, and Puto (1982) asked subjects to choose between three types of cars that differed on two dimensions. Two of the options were roughly equally attractive, with one option having a better value on one dimension (e.g., price) and the other option having a better value on the other dimension (e.g., fuel efficiency). The third (or ‘decoy’) option was similar to one of the first two options, but was strictly dominated by this option. In the car example, the decoy option was both less fuel efficient and more expensive than one of the first two options. Huber, Payne, and Puto demonstrated that people are more inclined to choose the alternative that dominates the decoy option. The attraction effect has been replicated in a large number of studies in a variety of contexts.\(^4\)

The attraction effect task in this experiment was based on Herne (1999), who used a lottery task that allows for an incentivized test of the attraction effect. In the attraction effect task, participants had to choose between three binary lotteries which varied along two dimensions: probability of winning and prize. Of the three lotteries, two lotteries (say lotteries \(x\) and \(y\)) had equal expected value; one lottery had a higher prize and the other lottery a higher probability of winning. The third lottery was a decoy lottery which was dominated on both the probability and the prize dimension by either lottery \(x\) or lottery \(y\) (and not the other). The attraction effect occurs if participants are more likely to choose the lottery that dominates the decoy. Herne (1999) showed that many people are indeed susceptible to the attraction effect in this task. In total, participants faced five sets of lotteries consisting of three lotteries each; see table 4.12 in Appendix A.\(^7\) We used a between subjects design, so that for half the participants lottery \(x\) was the dominating option and for the other half it was lottery \(y\). The outcome of the lotteries was only revealed at the end of the experiment. The expected value of lotteries \(x\) and \(y\) was equal to 45 cents in all choice menus; the expected value of the decoy lottery was 36 cents.

\(^4\)Ok, Ortoleva, and Riella (2011) mention more than 20 studies documenting the attraction effect in economics, psychology, marketing and political science.

\(^7\)The lotteries are identical to those used by Herne (1999) except that all prizes were multiplied by 1.5.
Compromise Effect Task

The compromise effect (Simonson, 1989; Simonson and Tversky, 1992) occurs in decision problems when a given alternative is chosen more often when it is presented as the middle alternative compared to when it is an extreme alternative. For example, in Simonson and Tversky, subjects had to make a hypothetical choice between three of four calculator batteries varying in expected life (in hours) and the probability of corrosion. They found that subjects were more likely to pick the battery with the second lowest probability of corrosion if this battery was the middle option, i.e., if the choice set included the battery with the lowest probability of corrosion as well. The compromise effect has been supported empirically by a range of studies (see e.g. Bernatzi and Thaler, 2002; Busemeyer et al., 2007; Müller, Kroll, and Vogt, 2010, 2011).

The compromise effect task we used in this experiment was also based on Herne (1999). Similar to the attraction effect task, participants had to choose between three lotteries varying in two dimensions. Two lotteries (say $x$ and $y$) were identical for all participants, whereas the third (or decoy) lottery differed between decision frames. In one frame, a decoy lottery was included that made lottery $x$ the middle option, whereas in the other frame the decoy lottery that was included made lottery $y$ the compromise. Herne (1999) provides evidence that suggests that some people may be susceptible to the compromise effect in this task. As with the attraction effect task, participants faced five different sets of lotteries and we used a between subjects design. Table 4.13 in the appendix shows the lotteries we used in the experiment. The outcome of the lotteries was only revealed at the end of the experiment. For all lotteries (including the decoys), the expected value was 45 cents.

Anchoring Task

The anchoring effect task was based on Tversky and Kahneman (1974), who asked their subjects to estimate various statistics stated in percentages, including for example the percentage of African countries in the United Nations. They then randomly drew a number between 0 and 100 by spinning a wheel of fortune in their subjects’ presence and asked their subjects to specify if the statistic to

---

8 As with the attraction effect task, the lotteries are identical to those used by Herne (1999) except that all prizes were multiplied by 1.5.
be estimated was higher or lower than the randomly drawn number. Subjects were then also asked to give a precise estimate of the statistic. Even though participants knew that the randomly drawn number provided no information on the true value of the statistic, participants who were subjected to a larger number (or ‘anchor’) on average still provided a higher answer to the second question. Since \cite{Tversky and Kahneman} the anchoring effect has been consistently replicated both in a similar setting and in several generalizations (see e.g. \cite{Epley and Gilovich, 2001; Mussweiler and Strack, 2001; Ariely, Loewenstein, and Prelec, 2003; or see Furnham and Boo, 2011 for a recent literature review).

The anchoring set-up we used in this experiment was very similar to that of \cite{Tversky and Kahneman}. Participants had to answer five rounds of two questions. In each round, the first question asked participants if a particular statistic was larger or smaller than a randomly drawn integer between 0 and 1000, the second question then asked them to estimate the true value of this statistic. An anchoring effect occurred if this estimate was correlated with the randomly drawn number (or anchor). All five statistics were chosen such that all participants were expected to have at least some idea of their values, but would be unlikely to know any value exactly. For example, one of the statistics asked participants for the distance between Paris and the Dutch city of Eindhoven in kilometers.\footnote{The other statistics were: the highest measured top speed of the fastest road car in the world in kilometers per hour, the number of years ago the Dutch city of Alkmaar received city rights, the height of the tallest building on earth in meters and the number of inhabitants of Vatican City.}

The random integer appeared on screen as if generated by a slot machine to emphasize its randomness. A correct answer to the first question yielded 25 cents. For the estimate of the statistic, participants earned 1 euro minus one cent times the difference between their answer and the true value of the statistic (with a minimum of zero). No feedback on the correct answers was provided during the experiment, although all answers and earnings were provided at the end.

**Prisoner’s Dilemma Task**

Framing effects have also been shown to occur in prisoner’s dilemmas. In the experiment we used a symmetric prisoner’s dilemma task, for which we varied the framing in two ways. Firstly, we followed \cite{Ross and Ward} in labeling the
prisoner’s dilemma as either a ‘community game’ or a ‘banker game’. Ross and Ward found that subjects were more likely to choose the cooperative option when the game was presented as a community game.

Secondly, we used the decomposed prisoner’s dilemma framing of Pruitt (1967). Prisoner’s dilemmas are typically presented to subjects by stating the payoffs for both players that correspond to each outcome. Pruitt instead presented prisoner’s dilemmas to his subjects by stating the payoffs as a function of a subject’s choice. Two prisoner’s dilemmas can have identical payoffs for outcomes but different payoffs for choices. Prisoner’s dilemmas 1a and 1b of table 4.1 -taken from Pruitt (1967)- give an example of two such prisoner’s dilemmas. Pruitt showed that the percentage of cooperative choices varies as a function of how the choices were represented, even if the corresponding outcomes were identical.

As with all secondary tasks, the prisoner’s dilemma task started with two minutes of instructions time. During the instructions, subjects were told that they would be anonymously matched to another participant in the experiment for the prisoner’s dilemma only, and that they would only learn the choices of the other participant at the end of the experiment. During the instructions, the banker-vs-community frame was already visible in the page heading; half the participants were assigned to each label. The banker or community heading persisted into the four minutes of decision time. Participants played the three prisoner’s dilemmas displayed in table 4.1. The sequence was either 1a/2/1b or 1b/2/1a. Prisoner’s dilemmas 1a and 1b were the two different representations of the same prisoner’s dilemma representing the (decomposition) framing of Pruitt, allowing us to compare cooperation rates within subjects. Prisoners dilemma 2 was different from the other two in terms of payoffs and was placed in between to make the similarity between dilemmas 1a and 1b less obvious.

Cognitive Task

Finally, we included a cognitive task to check for differences in cognitive functioning between depleted and control participants. The task itself consisted of adding three two-digit integers, which has been used in several studies before (e.g. Slooff and Van Praag 2010). The three numbers were presented vertically to make it

\footnote{Ross and Ward (1995) used ‘Wall Street game’ rather than ‘banker game’. Since ‘Wall Street’ does not translate into Dutch very easily, we elected to use a word with similar connotations (‘banker’) instead.}
4.2.3 Procedure

The experiment was computerized using PHP/MySQL software and conducted at the CREED laboratory of the University of Amsterdam. The experiment started with an initial set of instructions that explained the Stroop task. As part of the instructions, subjects had to answer two check-up questions and had to perform 10 practice trials of the Stroop task to familiarize themselves with the Stroop interface. The instructions also contained some information about the general structure of the experiment. Specifically, participants were told that the Stroop task would be repeated five times and that each repetition would be followed by a different task. They were also told that the other tasks would be explained later and that they would have five minutes for every Stroop task and six minutes for each other task. Apart from the practice trials for the Stroop task, all instructions were identical regardless of whether subjects were in the depletion or control treatment. A copy of the initial instructions was provided to participants as well, so that they could refer back to the instructions over the course of the experiment if necessary.\footnote{A translation of all instructions can be found in Appendix B.}

The first Stroop task started after all participants in a session had finished the initial instructions. After the five minutes for the first Stroop task had expired,
4.3. RESEARCH QUESTIONS

participants moved on to their first secondary task. Each participant received a short set of instructions on screen; they had exactly two minutes to read these instructions and four minutes to complete the task. During all secondary tasks, the remaining time was displayed on a clock on the bottom right of the screen. After the fifth and final secondary task had been completed, the computer matched the choices of two participants for the prisoner’s dilemma task and calculated the outcome of the lotteries for the attraction and compromise tasks. Participants then received a detailed overview of their results including their earnings separately for each task. After having reviewed their earnings, they were asked to fill out a short questionnaire that contained background questions, five questions per task asking if they thought the task had been interesting, difficult, tiresome, fun or boring as well as several general questions related to participants’ willpower and emotional states.\footnote{These questions asked them how frustrated, happy, satisfied, irritated, content, disappointed and tired they felt. We also asked them how hungry they were, how often they practiced sport, drank alcohol, smoked and how long they had slept the previous night and we asked them how likely they would be to study, smoke, do something annoying or fill out important forms immediately after the experiment.} After filling out the questionnaire, participants received their earnings and could leave the laboratory. All in all, the experiment lasted for approximately an hour and a half including the instructions, questionnaire and payment.

4.3 Research Questions

In this study, we seek to investigate the effects of willpower depletion on participants’ susceptibility to framing effects. In particular, we argue that depleted participants are more susceptible to framing effects than control participants. In this section, we go into what this hypothesis implies for the four framing tasks. We also discuss three additional research questions related to the cognitive task, changes in depletion effects over successive tasks and correlations between different framing effects.

We expect depleted participants to be more susceptible to framing effects in the attraction effect task. Pocheptsova et al. (2009) argue that the attraction effect is the result of a simple and largely automated (system 1) heuristic that can potentially be overwritten by a more thoughtful (system 2) decision process.
However, overriding the simple heuristic requires willpower or cognitive effort. To the extent that depleted participants have less willpower available, they should then be less likely to override the simple heuristic and thus be more susceptible to the attraction effect. Pocheptsova et al. (2009) depleted participants using an attention regulation task wherein participants in the depletion treatment were asked not to look at phrases that appeared at the bottom of a video screen and control participants received no such instructions. They then asked participants to make a hypothetical choice between three apartments to investigate the attraction effect. They find a strong attraction effect for depleted participants, but no attraction effect for controls. In line with these results, we expect depleted participants to be more susceptible to the attraction effect in this study as well.

For the compromise effect, the argument is slightly more subtle. Simonson and Tversky (1992) argue that the compromise effect is the result of an extended notion of loss aversion. In compromise effect tasks with three options, the middle option has a small advantage (on one dimension) and a small disadvantage (on the other dimension) relative to the extreme options, whereas both extreme options have a large advantage and a large disadvantage with respect to the other extreme option. If disadvantages loom larger than advantages, this should make the middle option most attractive. If reference dependence and loss aversion are lower level (system 1) decision processes that can be overruled by exercising willpower or cognitive effort (as argued by Kahneman 2003), we expect depleted participants to be more susceptible to the compromise effect. We get a similar prediction if we regard “choosing the middle option” as a simple heuristic (as argued by for example Bettman, Luce, and Payne 1998) that can be overruled by a more sophisticated decision process. Thus, we expect depleted participants to be more susceptible to the compromise effect.

It must be noted, however, that Pocheptsova et al. (2009) found exactly the opposite effect. They investigated the effect of depletion on the compromise effect using two separate experiments. In one experiment they depleted participants using the video task described above, in the other experiment they used a 40 trial non-incentivized version of the Stroop task. They found that the compromise effect is actually weaker in depleted participants. They argue that this makes sense if the compromise effect is the result of a higher level decision strategy that requires significant cognitive resources to be used. We favor the idea that the
compromise effect is based on lower level processes, but acknowledge that the opposite argument can also be made.

We expect depleted participants to be more susceptible to the anchoring effect as well. As first suggested by Tversky and Kahneman (1974), an anchoring effect appears when participants use the anchoring-and-adjustment heuristic but adjust insufficiently. People who use this heuristic make estimates by starting from an initial value that is subsequently adjusted to yield the final estimate. When this adjustment is insufficient, different starting values may lead to different final estimates. We expect depleted participants to have fewer cognitive resources or less willpower available to override the anchoring-and-adjustment heuristic; as a consequence depleted participants should be more susceptible to the anchoring effect.

For the prisoner’s dilemma task we also expect depleted participants to be more susceptible to both framing effects. Calculating the payoffs associated with the four possible outcomes based on the payoffs associated with choices takes cognitive effort, which depleted subjects may not be willing or able to provide. On the other hand, basing a choice on the community-versus-banker title or on the payoffs associated with choices (as opposed to outcomes) does not require any computation. Thus we predict that both framing effects will appear more strongly in depleted participants than in controls.

We can exploit the structure of this experiment to investigate at least three more questions of interest. Using the cognitive task, it is possible for us to investigate the effect of willpower depletion on cognitive performance. If cognitive resources are in fact depleted in the Stroop task as well, depletion should also lead to reduced cognitive performance.

The repeated nature of the experiment allows us to investigate if and how the effect of depletion on framing effects changes over successive tasks. Repeatedly administering the Stroop task can have several effects. For example, it is possible that participants in the depletion treatment get more strongly depleted over every repetition of the Stroop task, leading to a large difference with controls in later rounds. Conversely, if control participants are also being slightly depleted by every Stroop task and depleted participants are fully depleted in the first Stroop task, the difference between depleted participants and controls could also grow smaller in later tasks.
Finally, we can use the fact that participants took part in multiple framing tasks to examine if participants who display a framing effect on one framing tasks are also more susceptible to a framing effect on another framing task. We can also investigate if such correlations between tasks are larger for depleted or control participants.

4.4 Results

We ran six sessions of the experiment in November 2010, in which a total of 104 subjects participated. Half of these subjects were assigned to the depletion treatment, the other half were controls. Sixty-five percent of participants were male, with an average age of 21.17 years and the majority (68%) studying economics and/or business. Mann-Whitney tests on social demographic variables show no significant difference in characteristics between the two treatment groups.

In the remainder of the section we first present the results of the Stroop task as a manipulation check. We subsequently present the results separately for each of the four framing tasks, where we both investigate if the framing effect occurred at all and if it occurred to a different degree for depleted participants. We then look at the cognitive task, followed by an overview of how the size of the depletion effects develop over time and a look at correlations between tasks. For all statistical tests, each participant is treated as an independent observation.

4.4.1 Stroop Task

As a manipulation check it is useful to first compare the results of depleted and control subjects on the Stroop task. Recall that participants in the depletion treatment did the incongruent version of the Stroop task, which was more difficult than the control version. Thus, we expected depleted participants to perform worse on the Stroop task overall. Table 4.2 and figure 4.1 show that this is indeed the case; depleted participants earned less on the Stroop task overall than control participants. The second and third row of table 4.1 show that this difference is driven exclusively by depleted participants having a slower response time; the number of mistakes does not differ between treatments.\footnote{Interestingly, participants in our study responded almost four times as quickly as participants in Pocheptsova et al. (2009). This effect could be due to learning, increased incentives and...} Moreover, according...
Figure 4.1: The Distribution of Earnings in the Stroop Task (in Cents)

*Notes.* The figure plots the smoothed density (Epanechnikov kernel) of the earnings of participants over all five Stroop tasks separately for each treatment. The bandwidth was equal to 25.2 for the depletion treatment and 28.5 for the control treatment.
Table 4.2: Stroop Task Statistics

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Depletion</th>
<th>Difference</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average total earnings</td>
<td>944.06 (73.50)</td>
<td>824.19 (61.69)</td>
<td>119.87</td>
<td>p&lt;0.0001</td>
</tr>
<tr>
<td>Average total mistakes</td>
<td>13.13 (12.21)</td>
<td>13.52 (7.83)</td>
<td>0.39</td>
<td>p=0.2917</td>
</tr>
<tr>
<td>Average reaction time</td>
<td>0.747 (0.081)</td>
<td>0.960 (0.107)</td>
<td>0.213</td>
<td>p&lt;0.0001</td>
</tr>
</tbody>
</table>

Notes. The table gives the total earnings, the number of mistakes and the average reaction time in the Stroop task separately for both treatments; the bracketed numbers are standard deviations. P-values are calculated using Mann-Whitney tests.

to the questionnaire participants in the depletion treatment found the Stroop task more exhausting, more interesting, more fun, less boring (at the 5% level or better) and more difficult (at the 10% level) than participants in the control treatment. All in all, this is in line with previous studies and suggests that the Stroop manipulation was successful.

4.4.2 Attraction Effect Task

For the attraction task, participants went through five decision problems, in each of which they had to choose between three lotteries. One of these lotteries was always the decoy lottery, which had both a higher risk and lower earnings than one of the other lotteries. The attraction effect predicts that participants are more likely to choose the lottery that dominates the decoy lottery. To investigate if this was the case, we compare the fraction of times the dominating lottery was chosen over all five decision problems to the fraction of times the non-dominating lottery was chosen. In line with the attraction effect and Herne (1999), table 4.3 shows that the dominating lottery was picked more often than the non-dominating lottery; this difference is significant for both the control treatment and the whole sample.

In terms of the questionnaire variables, we also found a positive correlation between Stroop income and the answer to the question “in general as a person I would describe myself as easily influenced” (p=.003, Pearson correlation). Students of economics performed better on the Stroop task in the control treatment (p=.03, Mann-Whitney) but not in the depletion treatment (p=.3, Mann-Whitney). Males were better than females at the Stroop task in the control treatment (p=.046, Pearson correlation) but not in the depletion treatment (p=.68). Participants who performed well on the Stroop task in the depletion treatment were less likely to report they were going to study after the experiment (p=.048), more likely to exercise (p=.045), and were heavier smokers (p=.023).
### 4.4. RESULTS

#### Table 4.3: Attraction Effect Task Statistics

<table>
<thead>
<tr>
<th></th>
<th>Decoy</th>
<th>Dominating</th>
<th>Non-Dominating</th>
<th>P-value (Attraction Effect)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control subjects</td>
<td>1.5%</td>
<td>56.9%</td>
<td>41.5%</td>
<td>.002</td>
</tr>
<tr>
<td>Depleted subjects</td>
<td>1.9%</td>
<td>53.1%</td>
<td>45.0%</td>
<td>.126</td>
</tr>
<tr>
<td>All subjects</td>
<td>1.7%</td>
<td>55.0%</td>
<td>43.3%</td>
<td>.001</td>
</tr>
</tbody>
</table>

Notes. The table gives the percentage of lottery sets participants chose the dominating, non-dominating and decoy lotteries respectively; this uses the results from all five lotteries. To compute the p-values, we take the fraction of dominating choices and non-dominating choices for every individual and use a Wilcoxon test to check if they are equal. Thus, the number of independent observations is 52 per treatment.

However, the attraction effect does not differ between treatments (Mann-Whitney, p-value=0.2994). This finding stands in stark contrast to Pocheptsova et al. (2009), who found that depleted participants were 37 percentage points more likely to choose the dominant option than control participants. Here, depleted participants are actually 4 percentage points less likely to choose the dominant lottery, with a 95% confidence interval of [-11,3] percentage points. All in all, we find evidence of an attraction effect, but this effect does not differ between depleted participants and controls.

#### 4.4.3 Compromise Effect Task

For the compromise effect task, participants also went through five decision problems with three lotteries, similar to the attraction effect task. We used a between-subjects design where the two main lotteries (say x and y) were always the same and the third lottery (or ‘decoy’) differed between participants. The decoy was either strictly riskier or strictly safer than both other lotteries, but had the same expected value. The compromise effect predicts that participants are more likely to choose option x or y if it is the middle option. To investigate if this was the case, we compared the percentage of trials the ‘compromise’ lottery was chosen and compared this percentage to the percentage of trials the other main (non-decoy) lottery was chosen. Table 4.4 shows that participants were more likely to choose the non-compromise lottery than the compromise lottery. Thus, we find a reverse compromise effect, which is significant (with a p-value of .002 for the whole sample). The compromise option was also chosen significantly less often than the decoy (Wilcoxon, p-value<0.001), whereas the difference between the percentage of decoy choices and non-compromise choices is not significant.
Table 4.4: Compromise Effect Task Statistics

<table>
<thead>
<tr>
<th>Decoy</th>
<th>Compromise</th>
<th>Non-Compromise</th>
<th>P-value (Comp. Effect)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control subjects</td>
<td>42.7%</td>
<td>23.5%</td>
<td>33.9%</td>
</tr>
<tr>
<td>Depleted subjects</td>
<td>38.5%</td>
<td>24.2%</td>
<td>37.3%</td>
</tr>
<tr>
<td>All subjects</td>
<td>40.6%</td>
<td>23.9%</td>
<td>35.6%</td>
</tr>
</tbody>
</table>

Notes. The table gives the percentage of trials participants chose the compromise, non-compromise and decoy lotteries respectively; this uses the results from all five lotteries. To compute the P-values, we take the fraction of compromise choices and non-compromise choices for every individual and use a Wilcoxon test to check if they are equal. Thus, the number of independent observations is 52 per treatment.

Table 4.5: Compromise Effect Task Statistics by Riskiness

<table>
<thead>
<tr>
<th>Compromise</th>
<th>Most risky choice</th>
<th>Least risky choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control subjects</td>
<td>23.5%</td>
<td>27.7%</td>
</tr>
<tr>
<td>Depleted subjects</td>
<td>24.2%</td>
<td>25.0%</td>
</tr>
<tr>
<td>All subjects</td>
<td>23.9%</td>
<td>26.4%</td>
</tr>
</tbody>
</table>

Notes. The table gives the percentage of trials participants chose the compromise, most risky and least risky lotteries respectively; this uses the results from all five lotteries.

The finding of a reverse compromise effect is quite surprising, since it is contrary to the results of the studies we previously mentioned including Herne (1999), who used the same lottery task. However, this apparent puzzle can be explained by risk attitudes. Table 4.5 shows that the least risky lottery was chosen in nearly half of the decision problems, whereas the middle option and the most risky option were both chosen approximately 25% of the time. The least risky lottery was chosen significantly more often than either alternative (Wilcoxon, p-value < .0001), which suggests that at least some choices were driven by risk aversion and that risk aversion is more common than risk seeking, which is intuitive.

Herne (1999), however, used a within-subjects design that allows her to look only at the small number of subjects (at most 12 and sometimes as few as 2) who (a) never chose the decoy and (b) did not choose the same lottery regardless of the decoy. She reports a compromise effect (at the 10% level or better) among these subjects even if there are only 2 observations (table 4.4 choice set e, p = .0784).

We also observe a preference for the least risky lottery in the attraction effect task, where the least risky lottery is chosen 62.7% of the time and the riskiest option is chosen 35.6% of the time.

Interestingly, we find no gender differences in the percentage of trials the risk averse option was chosen for either the attraction effect task or the compromise effect task. This stands in contrast with a recent review of Croson and Gneezy (2009), who concluded that most studies suggest that women are on average more risk averse.
The finding that many participants tended to choose the least risky lottery does not change the fact that several participants still picked the compromise. Thus, in principle it would still be possible for the compromise to be chosen more often in the depletion treatment by subjects who were not influenced by risk attitudes. However, this does not seem to be the case (Wilcoxon, p-value=0.9868), which contrasts with Pocheptsova et al. (2009) who found an effect in two separate experiments.\footnote{Note, however, that Pocheptsova et al. (2009) used a hypothetical decision task in which risk attitudes do not matter. To sum up, we find a reverse compromise effect which is driven by risk aversion and does not differ between treatments.}

4.4.4 Anchoring Effect Task

For the anchoring effect task, participants had to answer five rounds of questions. In each round, the first question asked participants if a particular statistic was larger or smaller than a randomly drawn number, the second question then asked them to estimate the true value of this statistic. An anchoring effect occurred if the estimate of the true value was correlated with the randomly drawn number (or anchor). We used OLS to investigate whether the value of the anchor affected the answer to the second question. In the regression, we also controlled for the true answer and included a dummy for depleted participants. Moreover, we interacted both the true answer and the anchor with the dummy to test if anchoring effects differed between depleted and control participants.

Table 4.6 displays the results of this regression. Two things are apparent. Firstly, there is strong evidence of an anchoring effect (p=.000); increasing the anchor by 100 increases the estimate of the statistic by 28.\footnote{The anchoring effect is significant at the 1\% level for three statistics taken individually; these are the city rights, the car speed and the Vatican city statistic. The anchoring effect is significant at the 10\% level for the tallest building statistic, and not significant (p=.186) for the distance between Eindhoven and Paris.} Secondly, there is no evidence that the size of the anchoring effect differed between depleted and control participants. If anything, the anchoring effect was slightly smaller (.202 instead of .280) among depleted participants; however this difference is not statistically significant (p=.342).
Table 4.6: Anchoring Effect Task Estimates

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Std Error</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anchor</td>
<td>0.28</td>
<td>0.06</td>
<td>0.000</td>
</tr>
<tr>
<td>True answer</td>
<td>0.20</td>
<td>0.07</td>
<td>0.003</td>
</tr>
<tr>
<td>Depleted</td>
<td>−52.50</td>
<td>59.87</td>
<td>0.380</td>
</tr>
<tr>
<td>Depleted × Anchor</td>
<td>−0.08</td>
<td>0.08</td>
<td>0.342</td>
</tr>
<tr>
<td>Depleted × True answer</td>
<td>0.11</td>
<td>0.09</td>
<td>0.240</td>
</tr>
<tr>
<td>Constant</td>
<td>243.25</td>
<td>42.28</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Notes. This table gives the results of a linear regression of the estimate of the true value of the statistic (i.e. the answer to the second question in the anchoring task) on a constant, a treatment dummy (‘Depleted’), the random anchor, the true answer and interaction terms between the treatment dummy and the other two variables. The regression uses the data from all five statistics in the anchoring task. Standard errors are clustered at the participant level; thus there are 104 independent observations.

Table 4.7: Prisoner’s Dilemma Task Statistics

<table>
<thead>
<tr>
<th>% Cooperation</th>
<th>PD 1a</th>
<th>PD 1b</th>
<th>PD 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control Participants</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Community label (n=26)</td>
<td>30.8%</td>
<td>46.2%</td>
<td>19.2%</td>
</tr>
<tr>
<td>Banker label (n=26)</td>
<td>46.2%</td>
<td>50.0%</td>
<td>26.9%</td>
</tr>
<tr>
<td>Depleted Participants</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Community label (n=33)</td>
<td>54.5%</td>
<td>42.4%</td>
<td>30.3%</td>
</tr>
<tr>
<td>Banker label (n=18)</td>
<td>33.3%</td>
<td>50.0%</td>
<td>16.7%</td>
</tr>
</tbody>
</table>

Notes. This table gives the percentage of cooperative choices for the three prisoner’s dilemma tasks (PDs) separately for both treatments. Prisoner’s dilemmas 1a and 1b were either the first or third prisoner’s dilemma participants received, the second prisoner’s dilemma was always the same for all participants, see table 1 for more details.

4.4.5 Prisoner’s Dilemma Task

In the prisoner’s dilemma task, participants had to choose between a cooperative and a non-cooperative choice in three separate prisoner’s dilemmas. Recall that there are two framing effects that could occur. First, there was a between-subject difference in labeling: half the participants had the game presented to them with the label ‘community game’ and the other half were told they were playing the ‘banker’ game. Second, there was a within-subject varying in framing à la Pruitt (1967). In particular, all participants were presented with the same prisoner’s dilemma in two different representations as in table 4.1.

Table 4.7 gives the cooperation rates for the two composition frames (prisoner’s dilemmas 1a and 1b) and the two labels separately for depleted and control
participants. Within either depleted or control participants, label framing effects are observed if cooperation rates differ between rows and composition framing effects are observed if cooperation rates differ between the first two columns. Overall, prisoner’s dilemma 1a leads to slightly higher cooperation rates than prisoner’s dilemma b in three of four cases. Similarly, the banker frame leads to slightly higher cooperation rates than the community frame in three of four cases. However, none of the differences in cooperation rates are statistically significant either overall or within specific subgroups based on willpower depletion, label and/or composition frame.

To investigate the difference in framing effects between depleted and control participants, we ran a set of probit regressions where we regressed the indicator for a cooperative choice on a depletion dummy, a framing dummy and the interaction between framing and depletion. For the label framing, we ran this regression separately for all three prisoner’s dilemmas as well as for all three combined. Similarly, for the composition framing we ran this regression for prisoner’s dilemmas 1a and 1b separately as well as the combined data. Table 4.8 reports the results for the only regression that yielded a statistically significant treatment effect. The results suggest that changing the label from ’community’ to ’banker’ in prisoner’s dilemma 1a has a positive effect on the share of social choices in the control treatment and a net negative effect in the depletion treatment. This treatment-framing interaction effect is significant at the 10% level. However, the effect is small and the main framing effect is significant for neither depleted nor control participants. Moreover, there was no significant depletion effect in the other 6 specifications we tested. Thus, there is little evidence that willpower depleted participants are differently affected by framing effects than controls.

---

20For one participant in the banker/depleted group, the prisoner’s dilemma task did not set up correctly due to a software error, so her results were dropped from the analysis of the prisoner’s dilemma task.
21Note that participants were statistically significantly less cooperative in prisoner’s dilemma 2, but this is the result of different payoffs rather than a framing effect.
22Interestingly, depleted participants were on average more cooperative than controls in this specification. However, this effect did not appear in any of the other 6 specifications we tested and was also not robust to removing the interaction term from the regression.
Table 4.8: Prisoner’s Dilemma 1a Probit Estimates

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Std Error</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Community frame</td>
<td>0.16</td>
<td>0.14</td>
<td>0.26</td>
</tr>
<tr>
<td>Depleted</td>
<td>0.56</td>
<td>0.23</td>
<td>0.01</td>
</tr>
<tr>
<td>Community frame×Depleted</td>
<td>−0.37</td>
<td>0.20</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Notes. This table gives the results of a probit regression of a dummy for a cooperative choice on a constant, a framing dummy, a treatment dummy (‘Depleted’) and the interaction between the treatment and framing dummies. The reported results are marginal effects. The regression uses the data from prisoner’s dilemma 1a.

Table 4.9: Cognitive Task Statistics

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Depletion</th>
<th>Difference</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive task earnings</td>
<td>193.65 (72.84)</td>
<td>195.00 (73.26)</td>
<td>1.35</td>
<td>p=0.9637</td>
</tr>
<tr>
<td>Number of exercises completed</td>
<td>22.52 (7.01)</td>
<td>22.00 (7.67)</td>
<td>0.52</td>
<td>p=0.6231</td>
</tr>
<tr>
<td>Number of mistakes</td>
<td>1.58 (1.76)</td>
<td>1.25 (1.41)</td>
<td>0.33</td>
<td>p=0.4437</td>
</tr>
</tbody>
</table>

Notes. The table gives the earnings, the number of completed exercises and the number of mistakes in the cognitive task separately for both treatments; the bracketed numbers are standard deviations. P-values are calculated using Mann-Whitney tests.

### 4.4.6 Cognitive Task

Examining the cognitive task allows us to investigate the effect of the Stroop task on subsequent cognitive performance. Recall that in this task participants had to solve addition problems consisting of three two-digit numbers. Table 4.9 shows the average earnings in cents from the cognitive task per treatment. The difference between the performance of willpower depleted and control participants is not significant. The number of completed problems and the number of mistakes also did not differ between treatments.

### 4.4.7 Depletion over Successive Tasks

Our design also allows us to investigate how the effects of willpower depletion on framing develop over successive repetitions of the Stroop task (or rounds). Before turning to the secondary tasks, we first examine if performance on the Stroop task also changed over successive repetitions. Table 4.10 shows that participants do worse in the first Stroop task in both treatments and that performance in general increases slightly for each repetition of the Stroop task. Importantly, the difference in performance between treatments stays roughly constant and is
Table 4.10: Stroop Task Earnings per Round

<table>
<thead>
<tr>
<th>Order</th>
<th>Control</th>
<th>Depletion</th>
<th>Difference</th>
<th>Mann-Whitney P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>177.33 (16.37)</td>
<td>148.92 (13.16)</td>
<td>28.41</td>
<td>p&lt;0.0001</td>
</tr>
<tr>
<td>2</td>
<td>188.73 (15.92)</td>
<td>164.83 (14.26)</td>
<td>23.90</td>
<td>p&lt;0.0001</td>
</tr>
<tr>
<td>3</td>
<td>190.56 (17.23)</td>
<td>168.73 (14.02)</td>
<td>21.83</td>
<td>p&lt;0.0001</td>
</tr>
<tr>
<td>4</td>
<td>193.23 (14.90)</td>
<td>170.40 (13.60)</td>
<td>22.83</td>
<td>p&lt;0.0001</td>
</tr>
<tr>
<td>5</td>
<td>194.21 (16.31)</td>
<td>171.31 (14.56)</td>
<td>22.90</td>
<td>p&lt;0.0001</td>
</tr>
</tbody>
</table>

Notes. The table gives the earnings for the Stroop task per treatment separately for each successive repetition of the Stroop task (or round); the bracketed numbers are standard deviations.

significant in all repetitions of the Stroop task (p<0.0001).

To trace framing effects over successive rounds, we need to use an individual-specific measure of the size of the framing effect for each task. For the anchoring task, we use the coefficient for the anchor variable in the regression of table 4.6 estimated at the individual level. For the compromise and attraction task we use the fraction of times the compromise and dominating options were chosen respectively. Finally, for the prisoner’s dilemma we include an dummy for whether a participant changed her choice between prisoner’s dilemmas 1a and 1b. We also include performance on the cognitive task (in euros) to check for differences in cognitive performance over different rounds. We then take the difference between depleted and control participants on these respective measures to check if depletion had a different effect in different rounds.

Figure 4.2 plots the difference in the size of the framing effects between depleted and control participants. Overall, no clear time patterns can be discerned from the picture. Moreover, none of the depletion effects is significant in any round for any task (Mann-Whitney, p-value=0.101 for the cognitive task in round 1). However, not finding any differences between rounds is perhaps not surprising given the lack of a depletion effect overall.

### 4.4.8 Correlations between Tasks

Finally, our data also allow us to investigate if participants’ behavior is correlated between each of the six experimental tasks. To investigate this question, we use

\[23\]To be precise, we estimate a regression of the estimate of the true value of the statistic on a constant, the random anchor and the true answer. The depletion variable dummy not vary at the individual level and is therefore omitted.
Figure 4.2: Depletion Effects over Successive Rounds

Notes. This figure gives the effects of willpower depletion on behavior in successive rounds of the game. A positive value indicates that depleted participants were more susceptible to the framing effect in this round (for framing tasks) or were more successful (for the cognitive task).

Table 4.11: Correlations between Tasks

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th>Depleted</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>corr.</td>
<td>p-value</td>
<td>corr.</td>
</tr>
<tr>
<td>Stroop/Cognitive</td>
<td>0.126</td>
<td>0.201</td>
<td>0.388</td>
</tr>
<tr>
<td>Stroop/Prisoner’s Dilemma</td>
<td>0.149</td>
<td>0.132</td>
<td>0.427</td>
</tr>
</tbody>
</table>

Notes. This table gives the Pearson correlation coefficients between performance in the Stroop task and the cognitive task and between performance in the Stroop task and a dummy that indicates whether a participant change their choice between prisoner’s dilemmas 1a and 1b.
the same measures mentioned in the previous section to measure framing effects in the framing tasks. We correlate these measures at the individual level with each other as well as overall income on the cognitive task and Stroop tasks.

Table 4.12 displays the correlation coefficients for which we found a significant effect at the 5% level or better among either the whole sample or only the depleted or control participants. Most intuitive is the positive correlation between performance on the Stroop task and performance on the cognitive task. Interestingly, this correlation is present only among depleted participants. This suggests that the depletion version of the Stroop task is also a cognitively demanding task (whereas the control task is not) so that subjects who excel at cognitive tasks perform well both in the summation task and in the depletion version of the Stroop task. There is also a correlation between the Stroop task and the prisoner’s dilemma. Depleted participants who performed better on the Stroop task were also more likely to change their choice between prisoner’s dilemmas 1a and 1b. Both these effects are significant at the 1% level. However, of all 45 correlations we computed, only two are significant at the 5% level or better. Thus, overall we find little evidence that behavior is correlated between the six experimental tasks.

4.5 Discussion

In this study we investigated the effect of willpower depletion on participants’ susceptibility to framing effects. Framing effects appear in the attraction, compromise and anchoring tasks but do not differ between depleted and non-depleted participants. Furthermore we do not find a depletion effect on the performance of participants in the cognitive task or differential depletion effects over time. In many ways, these results are quite striking since the depleting effects of the Stroop task have been documented in many studies and in particular since the effect of depletion on the attraction effect and compromise effect has already been documented by Pocheptsova et al. (2009).

Our failure to find differences in framing effects between depleted and control participants could be due to our implementation of the Stroop task. In particular, it is possible that the Stroop task did not manage to more strongly deplete the

\footnote{In total, we computed 45 correlations. All results are identical if we use Spearman correlations or OLS instead. For the Prisoner’s dilemma variable (which is a binary variable), the results are identical if we use a t-test, Mann-Whitney test or probit regression as well.}
willpower of participants in the depletion treatment. In this light it is useful to compare our implementation of the Stroop task with previous studies. Most strikingly, to our knowledge this is the first study that uses an incentivized version of the Stroop task to deplete participants’ willpower. Incentivizing the Stroop task may have increased the motivation of control participants to do well. The congruent Stroop task is less interesting than the incongruent version and doing well on the congruent Stroop task requires participants to maintain focus and concentration. Hence, it is possible that control participants were also depleted by the Stroop task.

Our failure to find differences in framing effects could also be because of the secondary tasks used in the experiment. However, it is important to note that we found a framing effect in the anchoring, attraction and compromise tasks that was significant at the 1% level or better. Moreover, in all four secondary framing tasks there was sufficient scope for framing effects to vary between treatments. However, a crucial difference with most previous studies (such as Pocheptsova et al., 2009; Burger, Charness, and Lynham, 2011; Muraven and Baumeister, 2000) is that participants in our study were also incentivized to do well on the secondary task. Comparing our results to the literature, one possibility is that depletion effects disappear when participants are incentivized to do well on a secondary task.

If our finding that willpower depletion does not increase framing effects is caused by our implementation of the Stroop task, this suggests that the willpower depletion effects of the Stroop task are not robust to adding incentives, which would add an important insight to the willpower depletion methodology. If instead willpower depletion effects disappear when incentives are added to secondary tasks, this could indicate that willpower depletion is only relevant in contexts where the stakes are low or nonexistent. To the extent that economic behavior is typically subject to real incentives, willpower depletion may then not be as relevant to economic behavior as suggested by previous studies. We leave it to future research to investigate under what conditions the effect of willpower depletion on framing is or is not likely to appear.

\[25\]

Additionally, the average number of trials for a single Stroop task (180) larger than the number of trials participants went through in Pocheptsova et al. (2009) (40); however, it is not considered to be particularly high in the literature. For example, Burger, Charness, and Lynham (2011) use 250 trials. Since we repeated the Stroop task five times, the total number of trials in our experiment is larger; however, note that we also found no treatment effect looking only at the results of the first secondary task for any of the four framing tasks.
### Appendix 4.A  Lotteries of the Compromise and Attraction Effect Tasks

#### Table 4.12: Lotteries of the Attraction Effect Task

<table>
<thead>
<tr>
<th>lottery</th>
<th>framing</th>
<th>lottery 1</th>
<th>lottery 2</th>
<th>lottery 3</th>
<th>lottery 1</th>
<th>lottery 2</th>
<th>lottery 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>40</td>
<td>80</td>
<td>35</td>
<td>112</td>
<td>56</td>
<td>103</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>40</td>
<td>80</td>
<td>75</td>
<td>112</td>
<td>56</td>
<td>48</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>65</td>
<td>30</td>
<td>70</td>
<td>56</td>
<td>150</td>
<td>64</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>25</td>
<td>30</td>
<td>70</td>
<td>144</td>
<td>150</td>
<td>64</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>40</td>
<td>75</td>
<td>80</td>
<td>112</td>
<td>48</td>
<td>56</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>40</td>
<td>35</td>
<td>80</td>
<td>112</td>
<td>103</td>
<td>56</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>70</td>
<td>45</td>
<td>50</td>
<td>64</td>
<td>80</td>
<td>90</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>70</td>
<td>65</td>
<td>50</td>
<td>64</td>
<td>56</td>
<td>90</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>30</td>
<td>25</td>
<td>60</td>
<td>150</td>
<td>144</td>
<td>75</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>30</td>
<td>50</td>
<td>60</td>
<td>150</td>
<td>72</td>
<td>75</td>
</tr>
</tbody>
</table>

**Notes:** This table gives the ten lotteries used in the attraction effect task. Every participants went through five lotteries; half the participants went through the lotteries corresponding to framing 1, the other half went through the lotteries of framing 2. The probability of winning for each lottery is expressed in percentages, the prize is expressed in cents. Participants who did not win the prize got zero instead.
### Table 4.13: Lotteries of the Compromise Effect Task

<table>
<thead>
<tr>
<th>lottery</th>
<th>framing</th>
<th>lottery 1</th>
<th>lottery 2</th>
<th>lottery 3</th>
<th>lottery 1</th>
<th>lottery 2</th>
<th>lottery 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>40</td>
<td>80</td>
<td>30</td>
<td>112</td>
<td>56</td>
<td>150</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>40</td>
<td>80</td>
<td>90</td>
<td>112</td>
<td>56</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>30</td>
<td>70</td>
<td>80</td>
<td>150</td>
<td>64</td>
<td>56</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>30</td>
<td>70</td>
<td>20</td>
<td>150</td>
<td>64</td>
<td>225</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>50</td>
<td>30</td>
<td>20</td>
<td>90</td>
<td>150</td>
<td>225</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>50</td>
<td>30</td>
<td>60</td>
<td>90</td>
<td>150</td>
<td>75</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>90</td>
<td>80</td>
<td>70</td>
<td>50</td>
<td>56</td>
<td>64</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>60</td>
<td>80</td>
<td>70</td>
<td>75</td>
<td>56</td>
<td>64</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>50</td>
<td>30</td>
<td>40</td>
<td>90</td>
<td>150</td>
<td>112</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>50</td>
<td>60</td>
<td>40</td>
<td>90</td>
<td>75</td>
<td>112</td>
</tr>
</tbody>
</table>

*Notes:* This table gives the ten lotteries used in the compromise effect task. Every participants went through five lotteries; half the participants went through the lotteries corresponding to framing 1, the other half went through the lotteries of framing 2. The probability of winning for each lottery is expressed in percentages, the prize is expressed in cents. Participants who did not win the prize got zero instead.
Appendix 4.B  Experimental Instructions

This section contains the experimental instructions, separately for each task. The headings used in this section are different from the ones used in the experiment. The initial instructions were simply called ‘instructions’ in the experiment; they explained the structure of the experiment as well as the Stroop task. All secondary tasks had the heading “Instructions for Round 2/4/6/8/10”, with the exception of the prisoner’s dilemma, which had the heading “Instructions for Round 2/4/6/8/10: Community/Banker game”.

4.B.1 Initial Instructions

Welcome to the CREED laboratory. Today’s experiment will consist of ten rounds; every round will last between 5 and 6 minutes. Your earnings will be determined by the results of your choices in this experiment. This experiment makes use of a point system; 100 points translate to 1 euro at the end of the experiment. The number of points you earn in any round will only become known to you at the end of the experiment, so you will not receive any feedback about your earnings during the rounds.

On the next page you will receive the instructions for the first round of the experiment. The instructions for the other rounds will be given immediately preceding those rounds. Using the navigation tool at the top of the screen you can go back to any of the previous pages of the instructions. You will also receive a printed version of these instructions.

Round 1: Name the color

During the first round of the experiment a sequence of words will be shown on the screen. These words will be printed in the colors yellow, blue, purple, orange or red. Your task is to indicate the color the word is printed in. Each correct color will earn you 1 point, while each incorrect color will cost you 2 points. In other words, the more colors you name correctly, the more points you earn. This task will last for 5 minutes, after which you will automatically continue to next task.

[26]The original instructions (in Dutch) are available on request.
You can indicate the color of your choice using the keyboard. The relevant keys are g (for yellow), r (red), p (purple), o (orange) and b (blue). The key-color combinations will also be visible at the bottom of the screen throughout the task. Be aware: if you press any other key than the key corresponding to the correct color, this will be counted as an incorrect assignment. This also holds for keys that do not refer to any color. On the next page you will have the opportunity to practice the task for 10 rounds with no payoff consequences.

4.B.2 Compromise/Attraction Effect Task

In this task the procedure is as follows. You are going to make a choice between 3 lotteries. These lotteries will vary in the amount of points you can win and the chance that you will win this amount of points. Below an example of a lottery choice menu is displayed. Each lottery will have only 2 possible outcomes. The lengths of the green and yellow parts of the rectangle symbolize the chances of either a green or yellow outcome. The exact probability and the amount of points you will win in this event are also printed in both the green and yellow parts of the rectangle.

In total you will make 5 lottery choices; each time you make a choice between 3 lotteries. After the instruction time has run out you will be automatically directed to the first lotteries; in other words there are no practice rounds for this task. This task itself will last for 4 minutes; During the task you can see how much time is left to answer on the clock in the lower right corner of the screen.

![Figure 4.3: Example Used for the Compromise and Attraction Effect Tasks](image)

---

27 The Dutch word for yellow is ‘geel’, hence the key for yellow was ‘g’.
4.B.3 Anchoring Effect Task

In this task the procedure is as follows: First a random number between 0 and 1000 is drawn. Then you will be asked to answer two questions. The first question is a yes/no question, in which a correct answer will earn you 25 points. To answer the second question you will have to enter a number; you will earn an amount of points depending on how close your answer is to the correct answer. Answering exactly correctly will earn you 100 points but each step you are removed from the answer will cost you 1 point, up to earning a minimum of 0 points in this task. For example if the difference between your answer and the correct answer is 36, then you will earn 64 points in this task.

In this task there will be a total of 5 question rounds, each containing 2 questions. After the instruction time has run out you will be automatically directed to the first question round; in other words there are no practice rounds for this task. This task itself will last for 4 minutes; In case you fail to give answers to some questions in time you will earn 0 points for these questions. During the task you can see how much time is left to answer on the clock in the lower right corner of the screen.

Example:

The random number is: 273

1. Are there a larger or smaller number of mountain gorillas living in the wild (as of 2007)?
   - Larger
   - Smaller

2. How many mountain gorillas are there living in the wild (as of 2007)?
   Amount: 600

Figure 4.4: Example Used for the Anchoring Effect Task

4.B.4 Prisoner’s Dilemma Task

For this task you will be randomly matched to another participant. The identity of the participant you will be matched with will stay hidden to you and similarly
CHAPTER 4. WILLPOWER DEPLETION AND FRAMING EFFECTS

your identity will remain unknown to the other participant. The rest of the procedure is as follows: You will be presented with several situations where you have to choose between one of two options. A choice for a particular option will have consequences both for the amount of points you will earn for this task as well as the point earnings of the matched participant. The other participant faces the exact same choice task with the same consequences for you and the other participant. Your point total will be calculated based on your choices and the other participant’s choices. Your earnings for this round will be made known to you at the end of this experiment.

In total there will be 3 choice situations with 2 possible choices each. After the instruction time has run out you will be automatically directed to the first choice situation, so there is no practice time for this task. This task itself will last for 4 minutes; in case you fail to make a choice in time you and the participant you are matched with will earn 0 points for said choice situation. During the task you can see how much time is left to answer on the clock in the lower right corner of the screen.

<table>
<thead>
<tr>
<th>Points:</th>
<th>You</th>
<th>The Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Option 1:</td>
<td>+50</td>
<td>+50</td>
</tr>
<tr>
<td>Option 2:</td>
<td>+60</td>
<td>-40</td>
</tr>
</tbody>
</table>

Figure 4.5: Example Used for the Prisoner’s Dilemma Task

4.5 Cognitive Task

For this task the procedure will be as follows: You will be shown 3 two-digit numbers on the screen. Your task is to calculate the sum of these three numbers. For each correct answer you will receive 10 points, while for each incorrect answer 10 points will be deducted. After the instruction time has run out you will be automatically directed to the first problem set, so there is no practice for this task. This task itself will last for 4 minutes. During this time you can do as many calculations as you want. During the task you can see how much time is left to answer on the clock in the lower right corner of the screen.
Example of a possible exercise:

What is the sum of the following numbers?

Number A: 16
Number B: 72
Number C: 23

A + B + C =

Your Answer: 111

Figure 4.6: Example Used for the Cognitive Task
Appendix 4.C  Screenshot of the Stroop Task

Figure 4.7: Screenshot of the Stroop Task

Notes: This figure is a screenshot of the Stroop task we used in the experiment, translated from Dutch to English. The capital letters next to the colors at the bottom are the keys used to indicate a given color in the experiment; a G was used for yellow since the Dutch word for yellow is 'Geel'.
Chapter 5

Peers at Work in the Lab

In an influential study, Mas and Moretti (2009) show that workers are influenced by the productivity of those coworkers who can observe them during the production process. This study attempts to replicate this finding in a laboratory experiment that was specifically designed to capture the fundamental characteristics of the production process of Mas and Moretti (2009). However, we find no evidence of peer effects, which suggests that the findings of Mas and Moretti (2009) may not be as general as often claimed.

5.1 Introduction

Peer effects are potentially relevant in production processes whenever workers interact with coworkers. Introducing a highly productive worker into a shift may for example lead to social spillovers by allowing coworkers to learn better production strategies or encouraging them to increase their effort levels. If such social spillovers play a role, this has important consequences for the choice of production environment (e.g. individual vs team production) as well as the level of worker compensation (i.e. workers should be remunerated based on their production as well as their social spillovers). In recent years, these observations have led to a growing literature that investigates peer effects in production processes both theoretically (starting with Kandel and Lazear (1992)) and empirically (Falk and

---

1This chapter is based on Van Veldhuizen, Oosterbeek, and Sonnemans (2012). We are grateful to seminar participants at the University of Amsterdam, the Tinbergen Institute, the 2012 CeDEx-CREED-CBESS Meeting in Norwich and the 2012 ESA World Meeting for helpful comments. Financial support from the University of Amsterdam Research Priority Area in Behavioral Economics is gratefully acknowledged.
A particularly influential empirical study is Mas and Moretti (2009, abbreviated as “M&M, 2009” in the remainder of this chapter), which investigates peer effects among cashiers in a large US supermarket chain. The study shows that cashiers increase their effort levels when a highly productive coworker joins their shift. This study is of particular interest, since its data contain the exact spatial orientation of the workers in every shift. This allows the authors to differentiate between two possible peer effect mechanisms: social pressure and prosocial behavior. The former occurs when workers are averse to being caught exerting low effort by a highly productive worker. Prosocial preferences occur when workers are willing to reciprocate a highly productive worker by increasing their own effort. The authors argue that social pressure is only relevant for cashiers who are in the line of sight of the highly productive worker, whereas for prosocial behavior observability is not important. Mas and Moretti (2009) find that cashiers are influenced by the productivity of a worker that observes them but not by the productivity of workers that they observe, suggesting that peer effects in their sample are driven by social pressure.

By differentiating between two different peer effects mechanisms, M&M (2009) provide a first attempt to open the ‘black box’ of peer effect mechanisms. This is an important first attempt in the literature, which could be relevant for several reasons. Firstly, different peer effect mechanisms may have different policy implications. For example, if peer effects are driven by social pressure, firms should make sure that highly productive workers are able to watch over (and thus put social pressure on) as many of their coworkers as possible. On the other hand, if peer effects are driven by learning effects, firms should try to facilitate knowledge spillovers between workers.

Secondly, different peer effect mechanisms may counteract each other. For example, introducing a productive worker into a shift increases the effort levels of coworkers who are affected by social pressure, but decreases the effort of coworkers who think it is no longer necessary for them to work hard. If both mechanisms are active simultaneously, the net effect is likely to be smaller than the forces of the mechanism separately, which underestimates the potential of peer effects to improve worker productivity if only positive mechanisms are facilitated. Thirdly and relatedly, understanding different peer effect mechanisms makes it possible to more accurately extrapolate empirical results from one sample to another. In
5.1. INTRODUCTION

particular, coefficient estimates are more likely to generalize to samples in which similar mechanisms play a role.\(^2\) For example, \(\text{M&M (2009)}\)'s coefficient estimates should be more likely to generalize to production settings where social pressure is possible.

However, the degree to which \(\text{M&M (2009)}\)'s coefficient estimates generalize to other settings also depends on whether the setting of \(\text{M&M (2009)}\) reflects a larger class of production processes. Many subsequent studies have assumed that this is indeed the case (Ellingsen and Johannesson 2008; Fehr, Goette, and Zehnder 2009; Moretti 2011; Charness and Kuhn 2011). For example, Ellingsen and Johannesson (2008) state that \(\text{M&M (2009)}\) shows that “low-productivity workers put in more effort when observed by high-productivity peers.” Thus, \(\text{M&M (2009)}\) is implicitly assumed to have high external validity.

Whether this assumption is warranted should ultimately be an empirical question, however. Replications of \(\text{M&M (2009)}\) in other settings could inform us on whether their results can indeed be extrapolated to other contexts. However, in practice such replications are difficult to do because of the level of detail required by the data. Replicating the results of \(\text{M&M (2009)}\) would require a data set that contains information about the relative spatial positioning of different workers as well as a detailed individual-specific measure of production. When such data are not available in the field, laboratory experiments provide a different means to approximate the key characteristics of \(\text{M&M (2009)}\). To the extent that the findings of \(\text{M&M (2009)}\) reflect fundamental aspects of human behavior that also apply outside the original context, they should be replicable in a laboratory environment as well.

In this study, we replicate the setting of \(\text{M&M (2009)}\) in a laboratory experiment. Our experiment closely follows the set-up of \(\text{M&M (2009)}\) in a number of important ways. Participants in the role of workers are members of teams and are not financially dependent on the effort of other workers but instead face a higher workload if their team members exert lower effort. Workers perform a real effort task, we vary worker observability and we obtain a direct measure of (permanent) productivity. We vary the extent to which permanent productivity is visible to coworkers in the experiment, which allows us to differentiate between peer effects based on contemporary production and baseline productivity, something that

\(^2\)See Carrell, Sacerdote, and West (2011) for an interesting case in point.
M&M (2009) are by their own admission unable to do.

In line with M&M (2009), we expected workers to work harder when partnered with more productive coworkers who can observe them during the production process. However, we find that this result does not generalize to the laboratory context. We observe no evidence of peer effects despite the finding that workers are well aware of the production and productivities of their coworkers when this information is available and have the ability to both increase and decrease their effort levels. This suggests that M&M (2009)’s findings may not be as general as has been assumed.

The remainder of this chapter is structured as follows. We first give a brief overview of the results of M&M (2009). We then describe the design and hypotheses of the experiment, which are based on M&M (2009). The next section gives the results, after which we end with a brief discussion.

5.2 Overview of Mas and Moretti (2009)

M&M (2009) estimate peer effects among a two year sample of 394 cashiers working for a US-based supermarket chain. As a measure of productivity, they use the average number of items scanned by a cashier over a 10 minute interval.\(^3\) By the authors’ admission, this abstracts away from several potentially important aspects of performance (such as quality of service and absenteeism), though it does provide a precise estimate of cashiers’ production speeds. For every 10 minute interval, the authors know exactly which workers were on duty and at which cash register they were working. This allows them to identify the workers’ spatial orientation, which they use to define the observing and observable set.

At this point, it is useful to first give a few definitions of terms used by M&M (2009). The focal worker is the worker in a shift whose behavior is being analyzed. The observing set consists of coworkers who are facing the focal worker. The observable set consists of coworkers whom the focal worker is facing.

The empirical estimates are based on variations on the following equation

\(^3\)To be precise, within each 10 minute interval M&M (2009) include only periods when transactions are taking place. Thus, their productivity estimate is the number of items scanned per 10 minutes divided by the time the cashier was involved in a transaction.
5.2. OVERVIEW OF MAS AND MORETTI (2009)

(equation 6 in M&M, 2009).

\[ \Delta y_{itcs} = \alpha + \beta \Delta \bar{\theta}_{-itcs} + \pi \Delta N_{tcs} + e_{itcs} \]  

Here, \( \Delta y_{itcs} \) is the production of worker \( i \) in 10 minute interval \( t \), at date \( c \), at store \( s \) and relative to worker \( i^\prime s \) production in the previous 10 minute interval. \( \Delta \bar{\theta}_{-itcs} \) is the change in the average permanent productivity of coworkers, \( \Delta N_{tcs} \) is the change in the number of workers on duty and \( e_{itcs} \) is the error term. Peer effects are captured by the coefficient \( \beta \); a positive coefficient indicates that workers increase their production speed when the average permanent productivity of their coworkers increases. The key independent variable -average permanent productivity of coworkers- is not directly observed and needs to be constructed using a separate estimation procedure that corrects for possible influences of arbitrary social interactions.

Note that equation (1) assumes that peer effects operate through permanent productivity. However, in practice workers may also be influenced by the contemporaneous effort of their coworkers. M&M (2009) are by their own admission not able to empirically distinguish between these two mechanisms. An attempt to estimate a model that includes contemporaneous effort as an independent variable would have led to a reflection problem (Manski, 1993). M&M (2009) argue that their estimates are likely a combination of both the contemporaneous and permanent productivity effects.

M&M (2009)’s main findings are presented in the first three columns of table 5.1. The first column shows the results of the primary estimate of equation 1. The coefficient indicates that increasing the average permanent productivity of the coworkers by 10% increases the production of focal worker \( i \) by 1.5%. Column 2 shows that these peer effects only appear for changes in the average permanent productivity of coworkers who can observe the focal worker (the observing set). This suggests that peer pressure is more important than prosocial preferences. Column 3 shows that peer effects are specific to workers with below average productivity.

M&M (2009) report several additional estimates. These show inter alia that peer effects persist over time and are larger for workers who occupy registers.

\footnote{The reflection problem is the identification problem that arises when trying to identify a causal relationship between two individuals (or groups) that are observed simultaneously.}
Table 5.1: Main Results of [Mas and Moretti (2009)]

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference in log productivity of the focal worker between $t$ and $t-1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta$ Average coworker permanent productivity</td>
<td>.15</td>
<td>.17</td>
<td>.01</td>
<td>.24</td>
</tr>
<tr>
<td></td>
<td>(.02)</td>
<td>(.02)</td>
<td>(.02)</td>
<td>(.03)</td>
</tr>
<tr>
<td>$\Delta$ Average coworker permanent productivity in observing set</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.03)</td>
<td>(.03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta$ Average coworker permanent productivity in observable set</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.03)</td>
<td>(.03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta$ Presence worker in observing set</td>
<td>.031</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.003)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta$ Presence worker in observable set</td>
<td></td>
<td></td>
<td>-.030</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(.003)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>1,718,052</td>
<td>1,649,916</td>
<td>1,718,052</td>
<td>1,732,941</td>
</tr>
</tbody>
</table>

*Notes.* This table displays the results of four regressions taken from [Mas and Moretti (2009)]. Columns (1), (2), (3) and (4) are column (1) of table 2, column (1) of table 6, column (1) of table 3 and column (3) of table 6 in [Mas and Moretti (2009)] respectively. For column (3), the coefficient for above average workers is the total coefficient for above average workers, which is the sum of the two coefficients reported by [Mas and Moretti (2009)] in column 1 of table 3. The observing set consists of workers who are facing the focal worker. The observable set consists of workers whom the focal worker is facing. The bracketed numbers are standard errors.
in the close proximity of the focal worker. Most relevant for this study are the results reprinted in column 4 of table 5.1, which show that adding a worker to the observing set increases the focal worker’s productivity by 3.1%. Interestingly, adding a worker to the observable set decreases the focal worker’s productivity by almost the same percentage.

5.3 Experimental Design

The goal of this study is to see if the results of M&M (2009) can be replicated in a laboratory experiment. The experiment consisted of two stages. In the first (or baseline) stage, participant worked alone, allowing us to obtain a measure of their permanent (or baseline) productivity in the absence of peer effects. In the second (or production) stage, participants worked in teams of four, allowing us to investigate the impact of peer effects. In the remainder of this section we first discuss the baseline stage, followed by a discussion of the production stage and a brief comment on the procedure at the end of the experiment.

5.3.1 Baseline Stage

The experiment was computerized using PhP/MySQL. In the experiment, participants had to perform a production task that consisted of adding three two-digit numbers. We chose this task since it is easy to understand, captures the essential features of the production process described by M&M (2009) and results in sizable differences in productivity between participants, which allows us to examine differences between low productivity and high productivity workers. The three numbers appeared on the computer screen together with information about whether the answer to the previous exercise was correct and the cumulative number of successfully completed exercises up to that point. The sequence of numbers used in the exercises was randomly generated before the first session of the experiment, so that it was identical for all participants; we used a separate sequence for the baseline stage and the production stage.

Upon entering the laboratory, participants were welcomed and assigned to a random computer. They received the instructions for the baseline stage of the experiment on screen; the instructions included a single check-up question. After everyone had finished the instructions, the baseline stage started. In the baseline
stage, participants worked individually for 4 minutes and were paid 10 Euro cents for every correct answer they provided.\footnote{An English translation of all instructions and two screenshots can be found in the appendix; the original Dutch version of the instructions is available upon request.}

### 5.3.2 Production Stage

After the baseline stage, the experiment moved to the production stage, for which participants received additional instructions and check-up questions. After all participants had finished the instructions and check-up questions, the production stage started. In the production stage, there were three different treatments. In two treatments (BASELINE TEAM and TEAM), participants were randomly grouped into teams of four. They were told that as a team they had to solve a number of exercises somewhere between 750 and 1150 (the actual number was 829). There was no longer a fixed production time; participants received a fixed fee of 10 euros for their participation in the production stage, regardless of the number of exercises they had solved.

Importantly, they were also told that during the production stage they might receive information about the number of exercises solved by one or more of their teammates. The left part of their computer screen contained an overview of their team. Figure 5.1 gives the team overview used for treatment BASELINE TEAM. An arrow going from one participant to another indicates that this participant could see the number of exercises solved by the other in the production stage. For example, in our set-up participant B could see the number of exercises solved by participant A.

The team structure we used allows us to compare four different information perspectives. Participant A knew the number of exercises solved by him could be seen by one team member, whereas participant D knew he could see the number of exercises solved by one team member. Participant B knew he could both see one team member and be seen by another team member and participant C knew he could neither see nor be seen by another participant. The structure of the team remained fixed for the duration of the experiment.

The difference between treatments BASELINE TEAM and TEAM is that participants in treatment BASELINE TEAM also learned the number of exercises solved by their team members in the \textit{baseline} part of the experiment (their perma-
You are participant B

A: 17
46

B: 23
55

C: 22
?

D: 31
?

Figure 5.1: Team Overview

Notes. The figure gives the team overview used in treatment BASELINETEAM of the experiment. The numbers above the squares are baseline productivities. These are visible for all team members in treatment BASELINETEAM and are absent in treatment TEAM. The numbers inside the squares are the cumulative production levels in the production stage; these are known only for the participant him/herself and for participants he or she can see (as indicated by the arrows).
They learned the permanent productivity for all participants in their team, even for those for whom they did not know the number of exercises solved in the production stage. As a consequence, treatment BASELINE TEAM is closest to M&M (2009) in that it allows peer effects to work through permanent productivity as well as contemporaneous productivity; treatment TEAM only allows peer effects to work through contemporaneous productivity. Note that there is no reflection problem in either treatment, since information flows go only in one direction: for example worker B can be influenced by the contemporaneous productivity of worker A but not vice versa (there is no mirror).

Finally, we also ran an individual treatment in which participants individually had to solve between 188 and 288 exercises (the actual number was 207). Participants in treatment INDIVIDUAL never got any feedback about the performance of other participants in the experiment and were also allowed to leave the experiment after having solved the required number of exercises. We included this treatment to check if organizing workers into teams per se changed their productivity.

5.3.3 End of the Experiment

For each participant, the production stage ended after she (treatment INDIVIDUAL) or she and her team (treatments BASELINE TEAM and TEAM) had completed the required number of exercises. After finishing their final exercise, participants received an overview of their earnings and were asked to fill out a questionnaire. The questionnaire contained several demographic questions, a self-monitoring questionnaire (Snyder, 1974) and questions about the experiment. After finishing the questionnaire, participants could collect their payment and leave the laboratory, even if other participants were still solving exercises or working on the questionnaire.

5.4 Hypotheses

M&M (2009) find that increasing the average permanent productivity of the set of coworkers increases workers’ production speeds. Importantly, they find that this effect only appears when the change applies to coworkers who are in the focal worker’s observing set and that this effect only appears among low-productivity workers. For the experiment, these findings directly translate into the following
5.4. *HYPOTHESES*

hypotheses:

| Hypothesis 1 (peer effects): Increasing the average permanent productivity of the set of coworkers increases the production speed of the focal worker. |
| Hypothesis 2 (observability): Increasing the average permanent productivity of the set of *observing coworkers* increases the production speed of the focal worker; increasing the average permanent productivity of the set of *observable coworkers* has no effect. |
| Hypothesis 3 (ability): Increasing the average permanent productivity of the set of coworkers increases the production speed of a low productivity focal worker, but not of a high productivity focal worker. |

All three hypotheses are testable using treatment BASELINETEAM. Hypotheses 1 and 3 could in principle also apply to treatment TEAM. Although permanent productivity is not directly available to workers in treatment TEAM, workers B and D do know the contemporaneous production of one team member; they could infer the permanent productivity of this team member from his contemporaneous production speed. Hypothesis 2 can only be tested using treatment BASELINETEAM, since permanent productivity levels of observing coworkers are not known and cannot be indirectly inferred in treatment TEAM.

Our design also allows us to investigate the effect of having either one or no coworkers in the observing or observable set. It also makes it possible to examine aspects that lie outside the scope of [M&M (2009)]. Since our set-up avoids the reflection problem, we can check if differences in contemporaneous production of observable workers affect the production speed of the focal worker. We can also investigate to what extent workers are heterogeneous in the way they are influenced by the contemporaneous productivity of their coworkers. By comparing treatments BASELINETEAM and TEAM, we can investigate if knowing the permanent productivity of team members is a necessary ingredient for peer effects to appear. By comparing BASELINETEAM and TEAM with the INDIVIDUAL treatment, we can see if organizing workers into teams per se changes their productivity.
5.5 Results

We ran 9 sessions in February and April 2012, in which a total of 188 subjects participated (84 in TEAM, 84 in BASELINE TEAM and 20 in INDIVIDUAL). Participants had an average age of 22.5, 38% reported they studied economics and 58 percent were male.

In the remainder this section, we give an overview of the results. First, we examine evidence of peer effects and investigate if the strength of peer effects is greater for workers in the observing set and for low productivity workers (as per hypotheses 1-3). Second, we investigate several additional effects, some of which could not be estimated using the data of M&M (2009) and look at differences between treatments. Since the results show little evidence of peer effects, we will also investigate possible reasons why peer effects did not appear in our data.

5.5.1 Peer Effects

Hypothesis 1 states that workers’ production speed is increasing in the average permanent productivity of their coworkers. As a measure of production speed, we take the average number of exercises solved by the worker per minute in the production stage. As a measure of permanent productivity, we take the number of exercises solved by the worker in the baseline stage.

Table 5.2 shows the result of an OLS regression of worker production speed on average coworker permanent productivity, in which we also correct for the number of exercises solved by the focal worker in the baseline stage. The results in column 1 show that increasing average coworker permanent productivity by 10% increases the production speed of the focal worker by .02%. This is a much smaller percentage than M&M (2009)'s estimate (1.5%, as in column 1 of table 5.1) and is neither an economically nor a statistically significant number. The results in columns 2 and 3 shows that a similar story applies to both treatments taken separately.

As an alternative measure of permanent productivity, we also considered taking only a subset of the baseline, for example only the last two minutes. We chose the overall baseline production since it was most highly correlated (r=.78) to production speed in the production stage.
Table 5.2: Peer Effects Estimates

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log average coworker permanent productivity</td>
<td>.002</td>
<td>-.030</td>
<td>.011</td>
</tr>
<tr>
<td>permanent productivity</td>
<td>(.083)</td>
<td>(.120)</td>
<td>(.128)</td>
</tr>
<tr>
<td>Log focal worker permanent productivity</td>
<td>.678</td>
<td>.590</td>
<td>.767</td>
</tr>
<tr>
<td>permanent productivity</td>
<td>(.044)</td>
<td>(.065)</td>
<td>(.063)</td>
</tr>
<tr>
<td>Constant</td>
<td>.657</td>
<td>.841</td>
<td>.497</td>
</tr>
<tr>
<td></td>
<td>(.172)</td>
<td>(.238)</td>
<td>(.288)</td>
</tr>
</tbody>
</table>

Sample Observations: all BASELINETEAM TEAM 168 84 84

Notes. This table displays the results of three OLS regressions; the numbers in parentheses are standard errors.

5.5.2 Observability

Thus we find no evidence that peer effects are relevant at the aggregate level. However, it is possible that this obscures the fact that peer effects are active more locally. Indeed, by hypothesis 2 peer effects should be larger with respect to coworkers who can observe the focal worker. Table 5.3 displays the results of three regressions that examine if this is indeed the case. The regressions examine if the focal worker’s production speed is affected by the permanent productivity of the observing coworker, the observable coworker or both. In all cases, we use only the results for BASELINETEAM, since permanent productivity is only available to workers in this treatment.

The results (column 1) show that increasing the permanent productivity of the observing coworker by 10% decreases the production speed of the focal worker by .1%. This effect is much smaller than the effect found by [M&M (2009)] (1.7%), goes in the opposite direction and is not significant. For observable coworkers, the effect is larger than for [M&M (2009)] (.48% versus .1%), although it is also not significant and also goes in the opposite direction. Column 3 shows that if both the effect of observing and the effect for observable coworkers are estimated simultaneously, the resulting estimates are small, negative and not significant as well.
Table 5.3: Peer Effects Estimates by Observability

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log average production speed per minute of the focal worker</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log permanent productivity (observing set)</td>
<td>-.010</td>
<td>-.035</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.122)</td>
<td>(.261)</td>
<td></td>
</tr>
<tr>
<td>Log permanent productivity (observable set)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-.048</td>
<td>-.027</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.125)</td>
<td>(.268)</td>
<td></td>
</tr>
<tr>
<td>Log focal worker permanent productivity</td>
<td>.459</td>
<td>.508</td>
<td>.461</td>
</tr>
<tr>
<td></td>
<td>(.121)</td>
<td>(.126)</td>
<td>(.247)</td>
</tr>
<tr>
<td>Constant</td>
<td>1.04</td>
<td>.988</td>
<td>1.10</td>
</tr>
<tr>
<td></td>
<td>(.275)</td>
<td>(.283)</td>
<td>(.674)</td>
</tr>
<tr>
<td>Sample Observations</td>
<td>Players A&amp;B</td>
<td>Players B&amp;D</td>
<td>Player B</td>
</tr>
<tr>
<td></td>
<td>42</td>
<td>42</td>
<td>21</td>
</tr>
</tbody>
</table>

Notes. This table displays the results of three OLS regressions. The regressions use data from treatment BASELINETEAM only. The numbers in parantheses are standard errors.

5.5.3 Ability

Thus far we have found no evidence of peer effects at the aggregate level or separately for observing or observable coworkers. By hypothesis 3, one reason for the lack of effect could be that peer effects only appear among low productivity workers. To investigate if this is the case, we repeat the regression of table 5.2 separately for low and high productivity workers (using a median split on baseline productivity).

Table 5.4 shows the results of these regressions. Increasing the average permanent productivity of coworkers by 10% increases the production speed of high productivity workers by 1.63%, whereas it reduces the production speed of low productivity workers by .64%. Thus, if anything high productivity workers appear more likely to be affected by peer effects, although neither coefficient is significant at conventional levels. The estimates are similar if we only look at the data from treatment BASELINETEAM (as in columns 3 and 4).
5.5. RESULTS

Table 5.4: Peer Effects Estimates by Ability

<table>
<thead>
<tr>
<th></th>
<th>Dependent Variable: Log average production speed per minute of the focal worker</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Log average coworker permanent productivity</td>
<td>-0.064</td>
</tr>
<tr>
<td>Constant</td>
<td>2.56</td>
</tr>
<tr>
<td>Productivity low</td>
<td></td>
</tr>
<tr>
<td>Productivity high</td>
<td></td>
</tr>
<tr>
<td>Sample all</td>
<td>84</td>
</tr>
</tbody>
</table>
| Notes. This table displays the results of four OLS regressions; the numbers in parantheses are standard errors.

5.5.4 Additional Effects

Our data also allow us to investigate several additional effects. In particular, [M&M (2009)](as reprinted in column 4 of table 5.1) show that adding a worker to the observing set increases production speed, whereas adding a worker to the observable set decreases production speed. In this study, the size of the observable set differs per worker type within each group as per figure 5.1. Using the coefficient estimates of [M&M (2009)](as reprinted in column 4 of table 5.1), this suggests that the ranking of production speeds should be A ≈ B > C > D.

Given the lack of evidence for any peer effects obtained thus far, it is perhaps not surprising that production speeds do not seem to differ systematically between different player types. Figure 5.2 shows that in treatment BASELINETEAM production speeds do not differ systematically between player types. For treatment TEAM, players C produce at a faster speed than players B at the 10% level, but this effect disappears if adjusted for baseline productivity levels.

Using our data we can also investigate the effect of differences in the contemporaneous productivity of coworkers on the production speed of the focal worker. This can be done for the two worker types (B and D) that can see the contemporaneous productivity of at least one coworker. However, table 5.5 shows that the average contemporaneous production speed of the coworker in the
Figure 5.2: Production Speeds

Notes. The figure plots the average number of exercises solved per minute for each player type for treatments BASELINE TEAM and TEAM. Note that the declining production speeds after approximately minute 25 are due to a selection effect: slower workers are more likely to remain in the experiment after minute 25.
Table 5.5: Peer Effects Estimates in Contemporaneous Production

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>Value</th>
<th>Std. Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log average production speed per minute of the focal worker</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log average contemporaneous production speed (observable set)</td>
<td>-0.057</td>
<td>0.080</td>
</tr>
<tr>
<td>Log focal worker permanent productivity</td>
<td>0.625</td>
<td>0.082</td>
</tr>
<tr>
<td>Constant</td>
<td>0.842</td>
<td>0.194</td>
</tr>
</tbody>
</table>

Sample: Players B&G
Observations: 84

Notes. This table displays the results of a single OLS regression. This regression uses data from both treatment TEAM and treatment BASELINETEAM. The numbers in parantheses are standard errors.

The observable set does not influence the production speed of the focal worker.

The previous estimate exploits between-worker variation in the production speed of observable coworkers. However, our data also allow us to use within-worker variation in the production speed of coworkers. In particular, the data allow us to see if the focal worker’s production speed in minute \( t \) is influenced by the production speed of observable coworkers in minute \( t \) or minute \( t - 1 \). To see if this is the case, we ran a fixed effects regression in which we allowed the focal worker’s production speed to depend on the current and lagged production speed of observable workers. Note that these variables are exogenous, since the observable worker does not know the production speed of the focal worker. We also allow for a linear trend to correct for possible learning effects. We also compute the Arellano-Bond estimator to allow the focal worker’s production speed in minute \( t \) to depend on her production speed in minute \( t - 1 \).

Table 5.6 shows that the production speed of the focal worker is positively affected by the production speed of observable workers in the current minute and in the preceding minute. The latter effect is significant at the 10% level though not very robust; it disappears if the first minute of data or the coworker production speed minute in \( t \) variable is omitted. Thus, there is little evidence that workers are influenced by the contemporaneous production of observable coworkers.

\footnote{Note that since the number of exercises solved by coworkers was continuously updated, workers were also aware of the production speed of observable coworkers in the same minute.}
Table 5.6: Peer Effects Panel Estimates

<table>
<thead>
<tr>
<th>Dependent Variable: Production speed of the focal worker in minute $t$</th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Production speed in $t$ (observable coworker)</td>
<td>0.016</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Production speed in $t - 1$ (observable coworker)</td>
<td>0.032</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Production speed in $t - 1$ (focal worker)</td>
<td></td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.024)</td>
</tr>
<tr>
<td>Time trend</td>
<td>0.019</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>Average Fixed Effect</td>
<td>5.67</td>
<td>5.42</td>
</tr>
<tr>
<td></td>
<td>(0.164)</td>
<td>(0.293)</td>
</tr>
<tr>
<td>Sample Players B&amp;D</td>
<td>2566</td>
<td>2482</td>
</tr>
<tr>
<td>Sample Players B&amp;D</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes. This table displays the results of two regressions. The first column gives the results of a fixed effects regression estimated using OLS, the second column gives the Arellano-Bond estimator. We removed the last minute from the sample since workers did not work for the whole minute. Unlike previous specifications we do not use logs, since several workers did not solve any exercise during one or more minutes. The numbers in parentheses are robust standard errors.
Table 5.7: Peer Effects Estimates by Self-Monitoring

| Dependent Variable: Log average production speed per minute of the focal worker |
|-------------------------------|-----------------|-----------------|-----------------|-----------------|
|                               | (1)             | (2)             | (3)             | (4)             |
| Log average coworker permanent productivity                | .049            | -.117           |                 |                 |
| Log permanent productivity (observing set)               |                 | -.091           | .070            |                 |
| Log focal worker permanent productivity                   |                 |                |                 |                 |
| Constant                                                |                 | .762            | .836            | 1.04            | 1.10            |
| Self-monitoring                                         | low             | high            | low             | high            |
| Sample                                                   | all             | all             | Players A&B     | Players A&B     |
| Observations                                             | 45              | 39              | 24              | 18              |

Notes. This table displays the results of four OLS regressions; the numbers in parentheses are standard errors. The regressions use only data from treatment BASELINE TEAM.

It could also be that the lack of a correlation between the production speed of the focal worker and the production speed of the coworker is due to worker heterogeneity. In particular, a longstanding research tradition in social psychology suggests that people differ in the degree to which they self-monitor to ensure they maintain a desired public image. Highly self-monitoring individuals could be expected to be more susceptible to peer effects, particularly so if social pressure is relevant. However, table 5.7 show that peer effects appear among neither low self-monitoring nor high self-monitoring workers (as measured by the self-monitoring questionnaire of [Snyder 1974]) even if we only look at the effect of coworkers in the observing set (columns 3-4). Thus, individual heterogeneity in self-monitoring does not seem to explain the lack of peer effects observed in the study.

It could also be that the absence of peer effects on the aggregate is due to other forms of worker heterogeneity. For example, it is possible that some workers decrease their production speed when coworkers work faster, whereas other workers have prosocial preferences and increase their production speed if coworkers produce more quickly. To check this, we re-estimate the regression of column 1 of table 5.6 at the individual level. In 23 out of 84 regressions, either current observable worker production (in 19 cases) or lagged worker production...
(in 4 cases) is significant at the 10% level or better, with 19 positive and 4 negative coefficients. The finding that there are only a few negative coefficients suggests that individual heterogeneity is not important. Rather, in line with the results of table 5.6, it appears that only a minority of participants show any evidence of peer effects with respect to observable coworkers. Thus, we find no evidence that the lack of a significant peer effect in table 5.6 is caused by individual heterogeneity.

5.5.5 Comparing Treatments

So far we have focused on investigating if the findings of M&M (2009) also appear in the laboratory. Comparing treatments TEAM and BASELINETEAM also allows us to see if providing workers with information on the baseline productivity of their peers influences their productivity. Additionally, comparing treatment INDIVIDUAL with both group treatments allows us to see if putting workers in groups rather than making them work individually influences their production speed as well.

Figure 5.3 gives the time series for all three treatments. As column 1 of table 5.8 shows, workers in treatment BASELINETEAM work at a lower production speed than workers in treatment TEAM, whereas the differences with treatment INDIVIDUAL are not significant. However, the difference between TEAM and BASELINETAM is no longer significant when we correct for differences in permanent productivity (column 2). Column 3 shows that this difference is significant at the 10% level for workers who are not observable even after correcting for baseline productivity, whereas the interaction term in column 4 shows that the treatment difference appears only for workers who are highly productive. Column 5 shows that only the latter effect remains if both effects are included simultaneously.

5.5.6 Why no Peer Effects?

So far we have seen that peer effects do not seem to play a role in either treatment TEAM or treatment BASELINETEAM. We have also seen that the lack of an

\footnote{\textsuperscript{8}However, note that in table 5.6 the lagged coworker production speed seemed to have a stronger effect, whereas in this case it is coworker production speed in the same minute that is more often significant.}

\footnote{\textsuperscript{9}An alternative interpretation of this coefficient is that production speed is less strongly correlated to permanent productivity for workers in treatment BASELINETEAM.}
5.5. RESULTS

Figure 5.3: Production Speeds by Treatment

Notes. The figure plots the number of exercises solved per minute for treatments BASELINETEAM, TEAM and INDIVIDUAL. The figure displays only the first 20 minutes, since after minute 20 the first participants had finished all exercises, meaning that in subsequent periods the treatments would no longer be directly comparable.
Table 5.8: Peer Effects Estimates by Treatment

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BASELINE TEAM</strong></td>
<td>-.114</td>
<td>-.036</td>
<td>-.076</td>
<td>.276</td>
<td>.224</td>
</tr>
<tr>
<td></td>
<td>(.047)</td>
<td>(.048)</td>
<td>(.042)</td>
<td>(.147)</td>
<td>(.154)</td>
</tr>
<tr>
<td><strong>INDIVIDUAL</strong></td>
<td>-.012</td>
<td>-.018</td>
<td>-.035</td>
<td>-.019</td>
<td>-.032</td>
</tr>
<tr>
<td></td>
<td>(.075)</td>
<td>(.030)</td>
<td>(.053)</td>
<td>(.048)</td>
<td>(.052)</td>
</tr>
<tr>
<td>Log focal worker permanent</td>
<td>.679</td>
<td>.681</td>
<td>.769</td>
<td>.766</td>
<td></td>
</tr>
<tr>
<td>productivity</td>
<td>(.042)</td>
<td>(.042)</td>
<td>(.059)</td>
<td>(.059)</td>
<td></td>
</tr>
<tr>
<td>Focal worker observable</td>
<td>-.032</td>
<td>-.026</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.042)</td>
<td>(.042)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Focal worker observable x BASELINE TEAM</td>
<td>.081</td>
<td>.067</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.060)</td>
<td>(.060)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log permanent productivity</td>
<td>-.179</td>
<td>-.168</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x BASELINE TEAM</td>
<td>(.082)</td>
<td>(.083)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>1.90</td>
<td>.677</td>
<td>.689</td>
<td>.513</td>
<td>.533</td>
</tr>
<tr>
<td></td>
<td>(.033)</td>
<td>(.078)</td>
<td>(.083)</td>
<td>(.108)</td>
<td>(.113)</td>
</tr>
<tr>
<td>Sample</td>
<td>All</td>
<td>All</td>
<td>All</td>
<td>All</td>
<td>All</td>
</tr>
<tr>
<td>Observations</td>
<td>188</td>
<td>188</td>
<td>188</td>
<td>188</td>
<td>188</td>
</tr>
</tbody>
</table>

*Notes.* This table displays the results of five OLS regressions. The numbers in parantheses are standard errors.
Table 5.9: Recall of Production Stage

<table>
<thead>
<tr>
<th>Overall production of player</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Player A’s estimate</td>
<td>.93</td>
<td>.12</td>
<td>-.26</td>
<td>.07</td>
</tr>
<tr>
<td>Player B’s estimate</td>
<td>.89</td>
<td>.85</td>
<td>.21</td>
<td>.17</td>
</tr>
<tr>
<td>Player C’s estimate</td>
<td>-.11</td>
<td>-.26</td>
<td>.93</td>
<td>.27</td>
</tr>
<tr>
<td>Player D’s estimate</td>
<td>-.12</td>
<td>.70</td>
<td>.44</td>
<td>.91</td>
</tr>
</tbody>
</table>

Notes. This table displays Pearson correlation coefficients. 6 workers did not fill in the recall questions and were thus omitted from the sample, leaving 162 observations.

Table 5.10: Recall of Permanent Productivity

<table>
<thead>
<tr>
<th>Permanent productivity of player</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Player A’s estimate</td>
<td>.94</td>
<td>.82</td>
<td>.48</td>
<td>.71</td>
</tr>
<tr>
<td>Player B’s estimate</td>
<td>.96</td>
<td>.98</td>
<td>.94</td>
<td>.83</td>
</tr>
<tr>
<td>Player C’s estimate</td>
<td>.85</td>
<td>.87</td>
<td>1</td>
<td>.82</td>
</tr>
<tr>
<td>Player D’s estimate</td>
<td>.70</td>
<td>.78</td>
<td>.94</td>
<td>.95</td>
</tr>
</tbody>
</table>

Notes. This table displays Pearson correlation coefficients. 4 workers did not fill in the recall questions and were thus omitted from the sample, leaving 80 observations.

Overall effect does not seem to be the result of positive and negative peer effects canceling each other out in the average. In this section, we explore two possible alternative explanations for the lack of peer effects observed in the experiment. Firstly, it could also be that workers are so focused on their own work that they did not pay any attention to the production speeds or permanent productivities of their coworkers. Secondly, it could be that workers are unable to increase or decrease their production speeds.

To investigate the former line of reasoning, we asked workers after the experiment to recall both the overall production and permanent productivities (in BASELINETEAM) of all their coworkers. These recall questions were not incentivized and were not announced until after the experiment had ended. Nevertheless, table 5.9 shows that worker recall of the overall production of observable coworkers was very good (the correlation is between .7 and .9 and significant at the 1% level in all cases). Moreover, table 5.10 shows that workers in BASELINETEAM also did well at recalling the permanent productivity of their coworkers; all correlations are positive, large and significant at the 1% level. This suggests that workers were well aware of both the permanent productivity and (when applicable) the current production of their coworkers. Thus the lack of peer effects observed in
Table 5.11: Variation in Production Speed

<table>
<thead>
<tr>
<th></th>
<th>Sample</th>
<th>All</th>
<th>BASELINETEAM</th>
<th>TEAM</th>
<th>INDIVIDUAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation of production speed</td>
<td></td>
<td>.451</td>
<td>.495</td>
<td>.425</td>
<td>.372</td>
</tr>
<tr>
<td>Weighted standard deviation of production speed</td>
<td></td>
<td>.412</td>
<td>.436</td>
<td>.402</td>
<td>.350</td>
</tr>
<tr>
<td>Observations</td>
<td></td>
<td>188</td>
<td>84</td>
<td>84</td>
<td>20</td>
</tr>
</tbody>
</table>

Notes. This table gives the average standard deviation of worker production speed relative to the average production speed per exercise. For the second row, the standard deviation is weighted by the production speed of the respective worker relative to the average production speed in the experiment.

the study is not due to a lack of knowledge of the production and productivity of the coworkers.

To see if workers were able to sufficiently adjust their production speed, we investigate if workers were able to change their production speed between exercises. We do not look directly at minutes spent per exercise, since some exercises were more difficult than others. Instead, we look at worker production speed per exercise divided by the average production speed for the respective exercise (among all workers). Table 5.11 shows that the average standard deviation in worker production speed is around 40-50%, which suggests that workers had sufficient scope to change their production speed when required.

5.6 Discussion

Overall, the results of Mas and Moretti (2009) do not generalize to the experiment discussed in this study. In particular, we find no evidence of peer effects and also see no evidence that workers are more likely to be influenced by the productivity of coworkers in their observing set. We have also shown that this result is not due to individual heterogeneity and appears despite the finding that workers were well aware of the characteristics of their coworkers and were able to substantially vary their production speed.

It is important to note that we designed the experiment specifically to capture

\footnote{More anecdotally, we observed on several occasions that workers simply stopped working for a few minutes, suggesting that decreasing the production speed was quite possible as well.}
the most fundamental elements of the production process discussed by Mas and Moretti (2009). In particular, we used a (not directly incentivized) repetitive real effort task where the number of observable and observing coworkers were carefully controlled. The fact that the experiment fails to replicate their results suggests that the findings of Mas and Moretti (2009) may not be so general as subsequent studies have assumed. In particular, their findings might be specific to certain settings or may be dependent on less fundamental aspects of the production process used in their data (which were not captured by the experiment).

On a more general level, this suggests that there may be an important role for experiments to replicate the results of influential empirical papers. To the extent that empirical results reflect fundamental aspects of human behavior, they should replicate to a wide variety of contexts, including the laboratory. Laboratory experiments can be particularly valuable in cases when it is difficult to replicate a study in the field.
Appendix 5.A  Instructions

Welcome to this experiment. During the experiment you are not allowed to communicate with other participants. If you have a question, please raise your hand. One of the experimenters will then come to your cubicle to answer your question.

Today’s experiment consists of two parts; part two will take considerably more time than part one. Part two of the experiment will be explained after you have finished part one of the experiment. Your income will be determined on the basis of your results in the experiment. You will also receive a show-up fee of 7 Euros.

Please read through the following instructions carefully. As part of the instructions you will be asked a practice question to test your understanding of the instructions. When you have correctly answered this question, the experiment will move on. Using the navigation bar at the top of your screen it will be possible to return to previous pages during the instructions and practice question.

Instructions Part One

In part one of the experiment the procedure will be as follows. The computer screen will display three two-digit numbers (as in the example below). Your task is to calculate the sum of these three numbers. For every correct answer you will receive 10 Euro cents. An incorrect answer does not earn you any money; after an incorrect answer you will automatically go on to the next exercise. This part of the experiment will take up 4 minutes in total; during these 4 minutes you can do as many exercises as you want. The clock in the lower right corner of the screen tells you how much time you have left. The number of exercises you have answered both correctly and incorrectly is displayed above the current exercise; the (+1) indicates if the previous exercise was answered correctly or incorrectly.

Practice Question

Hank has finished 11 exercises, providing the correct answer to 8 and an incorrect answer to 3. How many Euro cents has Hank earned?
**Example of a possible exercise:**

Correct Answers: 1 (+1)
Incorrect Answers: 0

What is the sum of the following numbers?

Number A: 16  
Number B: 72  
Number C: 23

\[ A+B+C= \]

Your Answer: 111

---

**Instructions End**

You are now ready for part 1 of the experiment. By pressing the link below you will reach a waiting screen. The first part of the experiment starts as soon as all the others have also finished the instructions. On the waiting screen you can read back the text of the instructions.

**Instructions Part Two**

Like in part one, your task in part two will be to add three two-digit numbers. However, during this part of the experiment you will form a team with three other persons. The experiment will last until you and the other three people in your team have provided a correct answer to a fixed number of exercises. This fixed number of exercises will be somewhere between 750 and 1150 exercises. For this part of the experiment both you and all other team members will get a fixed payment of 10 Euros.\[1\]

---

\[1\] These are the instructions for treatment BASELINETEAM. The instructions for the other treatments are available on request.
As soon as you and your team have solved the required number of exercises, the experiment will be over for your team after a short questionnaire. One of the experimenters will come to your cubicle to pay out your earnings for the experiment. After payment you can leave the laboratory, even if the other teams are not done yet.

Information

During the experiment the left side of your screen provides an overview of your team, comparable to the figure displayed below. Each of the squares A, B, C and D represents one of the team members; your square will be colored in orange (in the example below you are participant B). Within the figure, the blue numbers above the squares indicate how many exercises each participant has solved in part 1 of the experiment. Thus, in the example below, participant A has solved 18 exercises, participant B 22, participant C 35 and participant D 21.

The figure also contains arrows between some participants. In the example below there is an arrow from participant D to participant B, an arrow from participant B to participant A and an arrow from participant A to participant B. These arrows represent information flows. An arrow from one participant to another indicates that this participant is able to see the number of solved exercises of the other participant in part two up to that point. Only the number of correct answers will be counted. The number of solved exercises by other participants will be displayed in red letters within the square that corresponds to said participant.

In the example below you are participant B and can see how many exercises you have solved (117) and how many exercises participant A has solved (125). In the example below you have no information about the number of exercises solved by participants C and D, who therefore have a “?” in their corresponding square. This means that you will at no stage get to know the number of exercises solved by participants C and D (not even after the experiment).

Finally, note that both participant A and participant D can see how many exercises you have solved up to that point. Participant C, however, does not know how many exercises you have solved and will at no stage get to know this number (not even after the experiment).

The figure below only represents an example of a who-sees-who; the who-sees-who that will be used in the experiment (which can have fewer, more or different
arrows) will be announced after the instructions. However, the who-sees-who will remain the same during the experiment; id est, both the arrows, your participant letter and the participant letter of the other participants will remain the same for the whole of the experiment. The composition of your team will not change during the experiment either.
Example Screen
Check-Up Question 1

The figure below gives an example of a who-sees-who. Indicate for every team member for which team member he or she can see the number of solved exercises. Also indicate for every team member who can see the number of exercises solved by them.

<table>
<thead>
<tr>
<th>Participant</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant A knows the number of exercises solved by:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participant B knows the number of exercises solved by:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participant C knows the number of exercises solved by:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participant D knows the number of exercises solved by:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>This participant knows the number of exercises solved by participant A:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>This participant knows the number of exercises solved by participant B:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>This participant knows the number of exercises solved by participant C:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>This participant knows the number of exercises solved by participant D:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Check-Up Question 2

The figure below gives an example of a who-sees-who (the same one as in the previous question). You are participant C, therefore you have all the information that participant C has access to. Indicate for all participants how many exercises they have solved in part one of the experiment. When possible, indicate for every participant how many exercises this participant has solved so far in this part of the experiment (part two). If the number of solved exercises is not known, do not fill in anything.

![Who-sees-who diagram]

Check-Up Question 3

Finish the sentence: this part of the experiment ends when you/your team mates/you and your team mates/everybody in the experiment have given the correct answer to 750/1150/a fixed number between 750 and 1150 exercises.

Check-Up Question 4

Which participants will know after the experiment how many exercises you have answered correctly?
• All participants in the experiment

• All participants who during the experiment could see the number of exercises you solved.

• All participants of which during the experiment you could see the number of exercises solved.

• Nobody

Instructions End

You are now ready to start part two of the experiment. By pressing the link below you will arrive at a waiting screen. Part two of the experiment will start as soon as all participants have finished the instructions. On the waiting screen you can read back the instructions of this part of the experiment as well.
Appendix 5.B Screenshots

Correct Answers: 1 (+1)
Incorrect Answers: 0

What is the sum of the following numbers?

Number A: 22
Number B: 36
Number C: 75

A+B+C=

Your Answer: 

Submit

Figure 5.4: Screenshot of the Baseline Phase
5.B. SCREENSHOTS

Figure 5.5: Screenshot of the Production Phase
Bibliography


157


Van Rijckeghem, Caroline and Beatrice Weder. 2001. “Bureaucratic corruption and the rate of temptation: do wages in the civil service affect corruption, and


Samenvatting (Summary in Dutch)

Dit proefschrift is gebaseerd op de resultaten van experimenteel economisch onderzoek. Centraal staat de methodologie van deze subdiscipline, waarin gebruik gemaakt wordt van laboratoriumexperimenten om keuzeprocessen na te bootsen. Laboratoriumexperimenten worden in de wetenschap al eeuwenlang gebruikt om causale verbanden tussen twee variabelen vast te stellen, maar in de economische wetenschap hebben experimenten pas in de laatste decennia echt aan populariteit gewonnen. Economische experimenten zijn vooral geschikt in twee soorten situaties, namelijk situaties waarbij het ingewikkeld is om vast te stellen hoe causale verbanden lopen en situaties waarbij te weinig goede data voorhanden zijn voor een gedegen analyse.

In dit proefschrift wordt de methode van de experimentele economie toegepast op vier uiteenlopende onderwerpen. Elk van de vier bijbehorende laboratoriumexperimenten volgt een vergelijkbare procedure. Ongeveer 20 proefpersonen nemen in een computerruimte deel aan een experimentele taak die een economisch beslissingsprobleem vertegenwoordigt. Er zijn hierbij steeds meerdere versies van het experiment, waarbij tussen de verschillende versies alleen de waarde van één onderzoeksvariabele gevarieerd wordt. Door de resultaten van de verschillende versies met elkaar te vergelijken is het dan mogelijk om de causale invloed van deze variabele op menselijk gedrag te achterhalen.

Hoofdstuk 2 beschrijft de resultaten van een laboratoriumexperiment dat zich toesptit op markten voor niet hernieuwbare grondstoffen, zoals aardolie en kolen. Volgens de economische theorie zouden de prijzen van deze grondstoffen jaarlijks moeten toenemen met een percentage dat gelijk is aan de rente. Dit resultaat -de zogeheten Hotelling regel- komt echter niet goed overeen met de werkelijkheid.
Van de meeste grondstoffen zijn de (voor inflatie gecorrigeerde) wereldwijde marktprijzen in de laatste decennia namelijk niet of nauwelijks gestegen. Hoewel een aanzienlijk aantal verklaringen voor deze empirische puzzel is aangedragen, is het wegens een gebrek aan goede data over bijvoorbeeld productiekosten en marktstructuren lastig gebleken deze verklaringen ook empirisch te toetsen.

Het experiment dat centraal staat in hoofdstuk 2 moet in deze lacune voorzien. In dit hoofdstuk stellen Joep Sonnemans en ik dat grondstofbezitters met een relatief kleine voorraad grondstoffen de Hotelling regel relatief goed zouden moeten volgen. Grondstofbezitters met een relatief grote voorraad grondstoffen zouden daarentegen meer van de Hotelling regel moeten afwijken. Om te testen of dit inderdaad het geval is moeten proefpersonen in de rol van grondstofbezitters in het experiment hun grondstoffen over een aantal perioden verdelen. Een aantal proefpersonen hebben hierbij de beschikking over een grote hoeveelheid grondstoffen, terwijl de rest een kleine hoeveelheid grondstoffen tot zijn beschikking heeft. De resultaten tonen aan dat proefpersonen met een kleine voorraad grondstoffen zich inderdaad beter aan de Hotelling regel houden. Op basis hiervan concluderen wij dat we verwachten dat de Hotelling regel de werkelijkheid beter beschrijft in situaties waarin grondstofbezitters relatief weinig grondstoffen overhebben.

Hoofdstuk 3 beschrijft de resultaten van een experiment waarin wordt gekeken naar het verband tussen corruptie en ambtenarensalarissen. Theoretische modellen zoals die van Becker and Stigler (1974) stellen dat hogere ambtenarensalarissen leidden tot minder corruptie. Het is echter buitengewoon lastig gebleken deze theoretische uitkomst empirisch te toetsen, en wel om twee redenen. Ten eerste zijn corrupte activiteiten illegaal en ontbreken ze dus in officiële macro-economische statistieken, waardoor er een gebrek is aan goede data. Ten tweede is het zelfs met goede data vaak nog moeilijk om te bepalen hoe causale verbanden lopen. Maken hoge ambtenarensalarissen bijvoorbeeld ambtenaren minder corrupt, of krijgen minder corrupte ambtenaren als beloning een hoger salaris?

Hoewel corruptie in de praktijk moeilijk meetbaar is, kan in een laboratorium een beslissingssituatie worden gecreëerd die het wel mogelijk maakt om corruptie te meten. In het experiment krijgen deelnemers in de rol van ambtenaar de mogelijkheid een steekpenning te accepteren en moeten ze vervolgens kiezen tussen een neutrale en een corrupte optie. Hierbij verhoogt de corrupte optie de inkomsten van de omkoper (een andere proefpersoon in het experiment), maar
verlaagt de corrupte optie de inkomsten van een goed doel (b.v. UNICEF). Er zijn twee versies van het experiment: in versie 1 is het salaris van de ‘ambtenaar’ relatief hoog en in versie 2 is het salaris relatief laag. De resultaten van het experiment tonen aan dat hogere ambtenaresalarissen binnen de experimentele omgeving leiden tot minder corrupt gedrag. Voor zover deze resultaten generaliseren naar het dagelijks leven, laat dit zien dat hogere salarissen wel degelijk corruptie kunnen bestrijden.

Hoofdstuk 4 vindt zijn oorsprong in een recente serie onderzoeken uit de sociale psychologie, waarin wordt gesteld dat wilskracht beschouwd moet worden als een beperkte energiebron, die bij overmatig gebruik tijdelijk uitgeput kan raken. In dit hoofdstuk onderzoeken Thomas de Haan en ik met behulp van een laboratoriumexperiment of mensen wier wilskracht tijdelijk is uitgeput eerder beïnvloed worden door zogeheten framing effects. Er is sprake van een framing effect als iemand’s beslissing beïnvloed wordt door schijnbaar irrelevante aspecten van een beslissingsprobleem. Een bekend voorbeeld is het ankereffect, waarbij mensen het antwoord op een vraag (bijvoorbeeld “wat is de afstand tussen Eindhoven en Parijs?”) laten afhangen van een willekeurig en totaal niet gerelateerd getal (het anker). Hoewel framing effecten een krachtig effect op beslissingen kunnen hebben, zijn mensen ook in staat ze te beperken door een bewustere, minder automatische, keuze te maken; hier is echter wel voldoende wilskracht voor nodig.

In het experiment wordt bij de helft van de proefpersonen de wilskracht uitgeput door een experimentele taak, terwijl de andere helft een controletaak te verwerken krijgt waarbij de wilskracht intact blijft. Vervolgens nemen de proefpersonen deel aan een serie van vijf framing effect taken, waaronder een ankertaak, een attractieffect taak, een compromiseffect taak en een gevangenendilemma. Onze verwachting was dat mensen wier wilskracht experimenteel was uitgeput meer door framing effecten zouden worden beïnvloed. De data geven echter geen indicatie dat framing effecten vaker voorkomen onder mensen met uitgeputte wilskracht. Dit suggereert dat verder onderzoek nodig is om uit te zoeken onder welke voorwaarden wilskrachtuitputting framing effecten wel en niet beïnvloedt.

Hoofdstuk 5 tenslotte beschrijft de resultaten van een onderzoek waarin Hessel Oosterbeek, Joep Sonnemans en ik een invloedrijke studie van Mas and Moretti (2009) in het laboratorium reproduceren. Mas and Moretti (2009) laten in hun studie zien dat cassières harder werken als ze bekeken worden door productieve
collega’s (een zogeheten peer effect). Een aanzienlijk aantal hieropvolgende studies heeft deze resultaten overgenomen zonder daarbij de kanttekening te plaatsen dat het nimmer is aangetoond dat deze resultaten ook generaliseren naar andere situaties. Onderzoek naar de generaliseerbaarheid van hun uittkomsten in andere omgevingen is echter niet zo eenvoudig te realiseren omdat hiervoor zeer gedetailleerde data nodig zijn. Daarom gebruiken wij in dit hoofdstuk een laboratoriumexperiment om de belangrijkste aspecten van de oorspronkelijke setting van Mas and Moretti (2009) te benaderen. Voor zover de resultaten van Mas and Moretti (2009) het gevolg zijn van fundamentele aspecten van menselijk gedrag zouden deze resultaten ook in een laboratoriumexperiment naar voren moeten komen.

Proefpersonen in het experiment vertegenwoordigen arbeiders die lid zijn van een productieteam. Hoewel hun beloning vaststaat kunnen ze door harder te werken ervoor zorgen dat hun teamgenoten minder werk hoeven te verrichten en eerder naar huis mogen. Conform de resultaten van Mas and Moretti (2009) verwachtten we dat onze proefpersonen harder zouden werken naarmate ze samenwerkten met productievere collega’s die hun observeerden. Onze resultaten zijn echter niet in lijn met de resultaten van Mas and Moretti (2009). We vinden geen enkele indicatie dat proefpersonen worden beïnvloed door peer effects ondanks het feit dat alle proefpersonen bijzonder goed op de hoogte zijn van de productiviteit van hun teamgenoten en in staat zijn als ze dat willen ook harder of langzamer te werken. Dit suggereert dat de resultaten van Mas and Moretti (2009) wellicht niet zo eenvoudig naar andere situaties generaliseren als vaak is aangenomen.
The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

499. M.H.A. RIDHWAN, Regional Dimensions of Monetary Policy in Indonesia
500. J. GARCA, The moral herd: Groups and the Evolution of Altruism and Cooperation
501. F.H. LAMP, Essays in Corporate Finance and Accounting
502. J. SOL, Incentives and Social Relations in the Workplace
503. A.I.W. HINDRAYANTO, Periodic Seasonal Time Series Models with applications to U.S. macroeconomic data
504. J.J. DE HOOP, Keeping Kids in School: Cash Transfers and Selective Education in Malawi
505. O. SOKOLINSKIY, Essays on Financial Risk: Forecasts and Investor Perceptions
506. T. KISELEVA, Structural Analysis of Complex Ecological Economic Optimal Management Problems
507. U. KILINC, Essays on Firm Dynamics, Competition and Productivity
508. M.J.L. DE HEIDE, R&D, Innovation and the Policy Mix
509. F. DE VOR, The Impact and Performance of Industrial Sites: Evidence from the Netherlands
510. J.A. NON, Do ut Des: Incentives, Reciprocity, and Organizational Performance
511. S.J.J. KONIJN, Empirical Studies on Credit Risk
512. H. VRIJBURG, Enhanced Cooperation in Corporate Taxation
513. P.ZEPPINI, Behavioural Models of Technological Change
514. P.H.STEFFENS, It’s Communication, Stupid! Essays on Communication, Reputation and (Committee) Decision-Making
515. K.C. YU, Essays on Executive Compensation - Managerial Incentives and Disincentives
516. P. EXTERKATE, Of Needles and Haystacks: Novel Techniques for Data-Rich Economic Forecasting
517. M. TYSZLER, Political Economics in the Laboratory
518. Z. WOLF, Aggregate Productivity Growth under the Microscope
520. P.R. KOSTER, The cost of travel time variability for air and car travelers
521. Y.ZU, Essays of nonparametric econometrics of stochastic volatility
522. B.KAYNAR, Rare Event Simulation Techniques for Stochastic Design Problems in Markovian Setting
523. P. JANUS, Developments in Measuring and Modeling Financial Volatility
524. F.P.W. SCHILDER, Essays on the Economics of Housing Subsidies
525. S. M MOGHAYER, Bifurcations of Indifference Points in Discrete Time Optimal Control Problems
526. C. AKMAKLI, Exploiting Common Features in Macroeconomic and Financial Data
527. J. LINDE, Experimenting with new combinations of old ideas
528. D. MASSARO, Bounded rationality and heterogeneous expectations in macroeconomics
529. J. GILLET, Groups in Economics
530. R. Legerstee, Evaluating Econometric Models and Expert Intuition
531. M.R.C. BERSEM, Essays on the Political Economy of Finance
532. T. WILLEMS, Essays on Optimal Experimentation
533. Z. GAO, Essays on Empirical Likelihood in Economics
534. J. SWART, Natural Resources and the Environment: Implications for Economic Development and International Relations
535. A. KOTHIYAL, Subjective Probability and Ambiguity
536. B. VOOGT, Essays on Consumer Search and Dynamic Committees
537. T. DE HAAN, Strategic Communication: Theory and Experiment
538. T. BUSER, Essays in Behavioural Economics
539. J.A. ROSERO MONCAYO, On the importance of families and public policies for child development outcomes
540. E. ERDOGAN CIFTCI, Health Perceptions and Labor Force Participation of Older Workers
541. T.WANG, Essays on Empirical Market Microstructure
542. T. BAO, Experiments on Heterogeneous Expectations and Switching Behavior
543. S.D. LANSDORP, On Risks and Opportunities in Financial Markets
544. N. MOES, Cooperative decision making in river water allocation problems
545. P. STAKENAS, Fractional integration and cointegration in financial time series
546. M. SCHARTH, Essays on Monte Carlo Methods for State Space Models
547. J. ZENHORST, Macroeconomic Perspectives on the Equity Premium Puzzle
548. B. PELLOUX, the Role of Emotions and Social Ties in Public On Good Games: Behavioral and Neuroeconomic Studies
549. N. YANG, Markov-Perfect Industry Dynamics: Theory, Computation, and Applications