Université de Montréal

On the external validity of laboratory experiments

par

Amadou Boly

Département de sciences économiques

Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures

en vue de l'obtention du grade de

Philosophiæ Doctor (Ph.D.)

en sciences économiques

Mai 2009

Université de Montréal

Faculté des arts et des sciences

Cette thèse intitulée :

On the external validity of laboratory experiments

présentée par

Amadou Boly

a été évaluée par un jury composé des personnes suivantes :

Abraham J. Hollander

président-rapporteur

Olivier Armantier

directeur de recherche

Sidartha Gordon

membre du jury

Étienne de Villemeur

examinateur externe

....

représentant du doyen de la FES

Université de Montréal

Faculté des arts et des sciences

Cette thèse intitulée :

On the external validity of laboratory experiments

présentée par

Amadou Boly

a été évaluée par un jury composé des personnes suivantes :

Abraham J. Hollander

président-rapporteur

Olivier Armantier

directeur de recherche

Sidartha Gordon

membre du jury

Soiliou D. Namoro

examinateur externe

....

représentant du doyen de la FES

Sommaire

Définie comme la possibilité de généraliser les résultats obtenus en laboratoire à des situations du monde réel, la validité externe est une préoccupation majeure en économie expérimentale. La présente thèse vise à analyser la généralisabilité des résultats obtenus en laboratoire en comparant une expérience en laboratoire à Montréal (Canada) et une expérience sur le terrain à Ouagadougou (Burkina Faso). La principale différence entre le laboratoire et le terrain est que les participants sur le terrain ne savent pas qu'ils participent à une expérience. Cette thèse comporte trois chapitres.

Le chapitre 1 décrit le design de l'expérience. L'idée principale est de reproduire un contexte académique dans lequel des correcteurs corrigent 20 copies d'examen appartenant à des candidats. Les copies sont numérotées de 01 à 20, la copie 11 venant avec un pot-de-vin dans plusieurs sessions expérimentales. Un tel design permet d'étudier à la fois les incitations au travail en se concentrant sur les 10 premières copies (chapitre 2), et la corruption à travers l'analyse de la décision des correcteurs d'accepter ou non le pot-de-vin (chapitre 3).

Le chapitre 2 analyse les effets du contrôle comme mécanisme d'incitation, dans un contexte où les travailleurs sont motivés par une "mission". Les résultats suggèrent que le contrôle augmente significativement l'effort dans un tel contexte. Afin de tester la robustesse des résultats obtenus en laboratoire, nous les comparons aux résultats de l'expérience conduite sur le terrain à Ouagadougou (Burkina Faso). La direction et la magnitude de tous les effets de traitement sont les mêmes entre le laboratoire à Montréal et le terrain à Ouagadougou.

Le chapitre 3 analyse la validité externe des expériences sur la corruption.

Dans notre expérience, un candidat offre un pot-de-vin à un correcteur afin d'obtenir une meilleure note. Nous trouvons que la direction et la magnitude de la plupart des effets de traitement sont statistiquement indistincts entre le laboratoire et le terrain. En particulier, augmenter la rémunération des correcteurs réduit la probabilité d'accepter le pot-de-vin aussi bien en laboratoire que sur le terrain. Nous identifions également plusieurs micro-déterminants de la corruption.

Mots clés: Expérience en laboratoire, Expérience sur le terrain, Mécanismes incitatifs, Corruption.

Summary

External validity, which is defined as the possibility of generalizing lab results to real-life situations, has been a fundamental concern to experimental economists. This thesis addresses the external validity question by comparing a lab experiment conducted in Montreal (Canada) to a field experiment conducted in Ouagadougou (Burkina Faso). The main difference between the lab and the field is that subjects in the field are not aware they are taking part in an experiment. There are three chapters in this thesis.

Chapter 1 presents the experimental design. The basic idea of the experiment is to reproduce an educational setting in which graders are required to grade 20 exam papers. The papers are numbered and ranked from 01 to 20, with paper 11 coming with a bribe in several experimental sessions. Such a design allows the study of work incentives by focusing on the 10 first papers (chapter 2), and of corrupt behavior by analyzing the grader's decision regarding the bribe offer (chapter 3).

Chapter 2 analyzes the crowding-out effect of monitoring in a mission-oriented work context. We find that monitoring significantly increases graders' effort on average in such a context. To test the robustness of the results obtained in the lab in Montreal, we compare them to those of the field experiment conducted in Ouagadougou. The direction and the magnitude of all treatment effects are found to be fully consistent between the lab in Montreal and the field in Ouagadougou.

Chapter 3 makes an attempt at testing the external validity of corruption experiments by moving from the lab in a developed country, to where corruption may matter the most, the field in a developing country. In this experiment, a candidate proposes a bribe to a grader in order to obtain a better grade. We find the direction and the magnitude of most treatment effects to be statistically indistinguishable between the lab and the field. In particular, increasing the graders' wage reduces in both environments the probability to accept the bribe. We also identify several micro-determinants of corruption.

Key Words: Lab experiments, Field experiments, Incentives, Corruption.

Table des Matières

Liste des Tables

Remerciements

J'aimerais d'abord témoigner toute ma reconnaissance à mon directeur de recherche, Olivier Armantier, pour l'encadrement de ma thèse. Je le remercie infiniment pour sa disponibilité, son soutien et ses conseils.

Je remercie les professeurs du département de Sciences Économiques et mes collègues étudiants, auprès de qui j'ai pu trouver une oreille attentive durant toutes ces années.

Je remercie également le personnel du Département de Sciences Économiques et du Centre Interuniversitaire de Recherche en Économie Quantitative (CIREQ) pour leur disponibilité et leur gentillesse.

Enfin, cette thèse aura été un projet de longue haleine que je n'aurais pu mener à terme sans la présence et le soutien de ma famille et de mes amis. Je leur exprime ici tout ma gratitude.

A ma mère, Sanogo Salimata.

Introduction générale

Laboratory experiments are now a well-established tool in empirical economic analysis. A key advantage of this approach is the possibility to implement exogenous changes in a variable of interest, to study its effect on another one. The internal validity of a lab experiment refers to this ability to make causal inferences by exogenously manipulating a variable. External validity, on the other hand, refers to the ability to generalize lab results to settings outside the lab (see e.g. Loewenstein 1999).[1] The main object of this thesis is to analyze the generalizability of lab results by moving from the lab in a developed country to the field in a developing country. To do so, we conducted the "same" experiment in the lab in Montreal (Canada) and in the field in Ouagadougou (Burkina Faso), in order to compare the results.

External validity has long been a subject of concern in experimental economics.[2] However, its importance arguably depends on the specific goals of the experiment (Schram 2005). For example, Kagel and Roth (1995, p. 22) distinguish three categories of objectives. The first category, "Speaking to Theorists", aims at testing the predictions of theoretical models. In this category, internal validity relative to the theory tested may be considered more important than external validity. The second category, "Searching for Facts", attempts to establish empirical regularities in situations where economic theories are scant. As evidence accumulates, theories about the observed behaviors may be proposed and tested. The third category, "Whispering in the Ears of the Princes" aims at advising policy makers. Because most lab experiments may fall into the

[1] Besides "external validity", other commonly used terms are "parallelism" and "ecological validity".

[2] For a flavor of the debate about external validity, see e.g. Campbell and Stanley (1963), Smith (1982), Levitt and List (2007a, 2007b).

first category, experimental economists have typically put emphasis on internal validity at the expense of external validity (see e.g. Bardsley 2005, Schram 2005). External validity may be an important issue in the last two categories however, as the main objective is to say something about behavior outside the laboratory.

A major obstacle to the generalizability of lab results relates to the artificiality of the setting (Schram 2005).[3] Artificiality may come from the fact that some relevant factors in the real world are omitted in the lab setting which requires simplification for tractability. A proposed strategy to solve this problem may be through "controlled" replication. Indeed, in order to claim that result A cannot happen outside the artificial setting E, one must identify a factor K that is not present in E and is at work in the real world (Guala 2002). Testing this claim would then require to construct a setting E_1 which includes the factor K. However, in certain situations, some relevant factors may happen only in the real world and can hardly be reproduced in lab settings (Bardsley 2005). In those situations, lab experiments may not be the appropriate tool.[4]

To address the concern for external validity, a growing number of field experiments have been conducted.[5] Harrison and List (2004) propose six factors

[3]One of the earliest economic experiments may be traced back to Thurstone (1931), who attempted to derive empirically individual indifference curves. This experiment was subsequently criticized due to the hypothetical nature of the subjects' choices, which relates to the artificiality issue (see e.g. Wallis and Friedman 1942).

[4]For example, studies of tax compliances may lack the authority relationship between the government and the citizens. "People might recognize a civic or legal duty to pay taxes whilst not recognizing a duty to be honest to experimenters in labs, or indeed vice versa" (Bardsley 2005).

[5]For example, departing from traditional lab practices, experimental economists have e.g. recruited non-standard subjects instead of students (e.g. Henrich et al. 2001, Fehr and List 2004, Harrison and List 2008), used framed instead of neutral instructions (e.g. Abbink and Hennig-Schmidt 2006, Alatas et al. 2009, List 2006, Harrison and List 2008), utilized field goods rather than induced valuations (e.g. Kahneman, Knetsch and Thaler 1990, Bateman et al. 1997, Rutström 1998), or given real-effort tasks to subjects instead of abstract chosen

that can be used to determine the field context of an experiment. These factors are: the nature of the subject pool, the nature of the information that the subjects bring to the task, the nature of the commodity, the nature of the task or trading rules applied, the nature of the stakes, and the environment in which the subjects operate. These factors allow distinguishing four broad types of experiment: conventional lab experiments, artefactual field experiments, framed field experiments and natural field experiments. A *conventional lab experiment* uses a standard subject pool of students, an abstract framing, and an imposed set of rules. An *artefactual field experiment* employs non-standard subjects in the laboratory. A *framed field experiment* is conducted with field context in the commodity, the task, or the information set that the subjects can use. A *natural field experiment* take place in the field with the subjects unaware that they are in an experiment. A natural field experiment generates data that is closest to naturally-occurring data, and therefore exhibits the highest degree of external validity.

In this thesis, we analyze the generalizability of lab results to real-world situations by comparing a lab experiment conducted in Montreal (Canada) to a natural field experiment conducted in Ouagadougou (Burkina Faso).[6] The main difference between the lab and the field is that subjects in the field are not aware they are taking part in an experiment. Such a difference is important as subjects in the lab, knowing their decisions will be scrutinized by the experimenter, may alter their behavior. As further explained subsequently, this

numeric effort (e.g. van Dijk, Sonnemans and van Winden 2001, Dickinson and Villeval 2007).

[6] Our subject pool is standard both in the lab and in the field. As a result, it may be argued that the field experiment does not fit exactly under Harrison and List (2004)'s definition of a natural field experiment. Likewise, although the lab experiment is framed and uses real-effort, it may not qualify as a framed field experiment.

thesis also contributes to the literature on work incentives and on corruption. There are three chapters in this thesis.

In chapter 1, we describe the experimental design. This design models a work environment in which graders are recruited to grade 20 exam papers pertaining to candidates. The papers are numbered and ranked from 01 to 20, and graders are required to grade the papers following this precise order. In some of the experimental sessions, exam paper 11 came with a bribe offer. We therefore reproduce a corruption scenario in which a candidate offers a bribe to a grader in order to obtain a better grade. As a result, we obtain two different experiments. The first experiment analyzes the effects of different incentive schemes on work effort by focusing on the 10 first papers graded in the lab and in the field (see chapter 2). The second experiment relates to corruption and analyzes the graders' decision to accept or reject the bribe, as well as the grading of papers 11 to 20 (see chapter 3). In both experiments, we compare the experimental results from the lab and the field to see whether i) the treatment effects in the lab and in the field are the same, ii) the directions of treatment effects are the same, but the magnitudes differ, or iii) the treatment effects in the lab and in the field are different.

Chapter 2 analyzes work incentives. It extends the experimental study of the crowding-out effects of monitoring to a mission-oriented work context, as opposed to profit-oriented. Mission-oriented organizations typically include public and not-for-profit organizations. Different treatments are conducted by varying i) the level of monitoring and punishment, and ii) the wage paid to graders (gift-exchange). In the lab, the results indicate that women and younger subjects are more precise in their grading. We find that monitoring significantly increases the graders' effort relative to the Control treatment. However, increasing monitoring rate and penalties does not raise effort level further. In addition, while monitoring increases effort level, the effects appear to weaken over time. Gift-exchange is found to have no effect on effort. To test their robustness, the results obtained in the lab in Montreal are compared to those obtained in the field in Ouagadougou. The direction and the magnitude of all treatment effects are fully consistent between the two environments.

In Chapter 3, we make an attempt at testing the external validity of corruption experiments. To this end, we moved from the lab in a developed country, to where corruption arguably matters the most, the field in a developing country. We conducted four different treatments, each in the lab in Montreal and in the field in Ouagadougou, by manipulating i) the wage paid to graders, ii) the level of monitoring and punishment, and iii) the amount of the bribe. We find that increasing the wage paid to graders lowers their probability of accepting the bribe, but graders who accept the bribe are more likely to help the briber. Monitoring and possible sanctions appear to have no significant effect on the graders' decision to accept the bribe or to help the briber. The direction and the magnitude of these two treatment effects are similar between the lab and the field. However, increasing the amount of the bribe exacerbates corruption in the field, while it has no effect in the lab. Likewise, when the bribe is increased, accepters in the field are more likely to reciprocate by providing a passing grade to the briber, while no such behavior is observed in the lab. Overall, our results are encouraging as they suggest that lab behavior may generalize to real world situations.

Chapitre 1

Experimental design

1.1 Introduction

The generalizability of lab results to real world situations is an important issue in experimental economics.[1] However, few experimental studies explicitly analyze the generalizability of lab behavior to field situations. To do so, these studies typically compare experiments of different types but with similar design.[2] For instance, Brookshire, Coursey and Schulze (1987) compare demand behavior for a private good in a real-world market setting and in a Vickrey-auction lab setting. Likewise, Rondeau and List (2008) contrast the relative efficacy of challenge and matching gifts in fund-raising using explicitly linked lab and natural field experiments.[3]

Following the approach outlined in the previous paragraph, this thesis compares a lab experiment conducted in Montreal (Canada) with a field experiment conducted in Ouagadougou (Burkina Faso). The experimental design models a work environment in which graders are recruited to grade 20 exam papers pertaining to candidates. The 20 papers are numbered and ranked from 01 to 20, and graders are required to grade the papers in this precise order. In some experimental sessions, exam paper 11 comes with a bribe offer. Consequently, we obtain two distinct experiments. The first experiment studies the effects of various incentive schemes on work effort by focusing on the grading of papers 01 to 10. The second experiment is a corruption experiment. It analyzes the grader's decision to accept or reject the bribe, as well as the grading of papers

[1] According to Levitt and List (2007a): "In order for the laboratory to achieve its full potential as an invaluable empirical tool in economics, we need to understand when, and under what circumstances, we can reliably generalize lab results to naturally occurring markets that are of interest to economists."

[2] See Harrison and List (2004) for a classification of experiments.

[3] For more comparisons of lab and field experiments, see e.g. List (2006), Levitt and List (2007a, 2007b).

11 to 20. Each of the two experiments and its related experimental treatments will be discussed in subsequent chapters.

As noted by Friedman and Sunder (1994, p.11), it is futile to try to replicate every aspect of the field in the lab. Instead, one should try to find the simplest laboratory environment that incorporates some interesting aspects of the field environment. Therefore, in explicitly linking and comparing a lab and a field experiment, most studies focus on the rules of the experimental game and attempt to keep them as similar as possible between the experiments. However, the lab and the field experiment typically differ on various dimensions that may prevent a direct comparison.[4] For example, Antonovics, Arcidiacono and Walsh (2009) reproduced the Weakest Link TV show in the lab, for comparison with the field counterpart. The lab and the field differed regarding e.g. the tasks (questions asked), the stakes and the subject pool. In consequence, lab and field results are compared only through the sign of treatment effects. Likewise, Hennig-Schmidt, Rockenbach and Sadrieh (2008) compare a lab and a natural field experiment with different tasks in the lab and the field. As a result, effort levels in the lab and in the field could not be compared directly.

A distinctive aspect of our design is that we can compare the lab and the field with regard to both the sign and the magnitude of treatment effects. Indeed, by keeping the experimental task identical between the lab and the field, dependent variables of interest (e.g. grades) are measured identically in both environments.[5] As a result, we can pool the data and compare the lab and the field directly, while controlling for observable differences. In the sections

[4] That is, a comparison of both the sign and the magnitude of treatment effects.

[5] Graders in the lab and in the field had to grade the same set of 20 exam papers, in the same order, and following the same grading rules.

below, we describe the experimental procedures with the candidates, followed by the lab and field experiments with graders. Experimental instructions and other materials are presented in the appendices.

1.2 Candidates

Subjects, called "candidates", were recruited to type a text on the computer as it was continuously dictated to them.[6] The text, based on a newspaper article in French, has 290 words and was fitted on two pages.[7] At the beginning of the session, each candidate was assigned to an isolated computer. Instructions were then read aloud, followed by questions. We explained carefully what would, and what would not constitute a mistake. The subjects were also informed that at the end of the dictation, they would not be allowed to spell-check or modify their papers in any way. We told the candidates that we would decide whether their paper would be spell-checked by an experimenter, or by other subjects called "graders". Finally, we explained that a candidate's payment would depend in part on the number of mistakes the graders would report. The lower the number of mistakes reported, the higher the payment.

Each candidate was also asked whether he would be willing to send some of the graders a money offer (explicitly referred to as "a bribe"), accompanied by the following message: "Please, find few mistakes in my exam paper". We

[6] Such a dictation exercise is a classic test in the Francophone schooling system. It is typically conducted with pen and paper. The aim is to evaluate a candidate's spelling abilities, as well as its knowledge of the French grammar. Such a test is administered several times a year to French students between the ages of 8 and 14. In addition, it is one of the requirements to obtain a secretary's diploma, and it is part of the entry-exam to certain civil servant positions. The pages on which a candidate typed the dictated text are referred to as an "exam paper" or simply a "paper".

[7] As the field experiment was conducted in a Francophone country, we only recruited French speaking subjects to conduct the different laboratory and field sessions.

explain to the candidates that if they accepted to do so, then their payoffs may not depend exclusively on the number of mistakes reported. Instead, they may also be affected positively or negatively by each grader's decision to accept or reject the bribe. Finally, the candidates were informed that even if they accepted to do so, we would not necessarily send the message and the bribe to the graders.[8]

We deliberately left the candidates' instructions partly ambiguous. In particular, we did not explain how we would select the papers to be graded by experimental subjects. Likewise, we did not specify the precise way in which the candidates' payoffs would be calculated. We also remained ambiguous about the amount of the bribe that would be proposed to the graders, as well as the exact consequences on the candidate's payoff when a grader accepts or rejects the bribe. The candidates were only informed that they would receive three payments: 20 C$ payable immediately after the conclusion of the typing session, and two additional amounts to be paid respectively three and six months later. This delay allowed us to complete the various grading sessions, both in the lab and in the field. The candidates were told that each of the additional amounts could vary between 20 C$ and 60 C$ depending on the number of mistakes reported, and possibly, on the graders' decisions to accept or reject the bribe.

We conducted two typing sessions at the CIRANO's Bell Laboratory for Experimental Economics located in Montreal. Each session lasted roughly an hour, and included respectively 11 and 12 subjects. All 23 subjects accepted to send a bribe to the graders. On average, the candidates received a total

[8] Immediately after reading the instructions, subjects were given the possibility to leave the laboratory with 10 C$ without having to type the text. None elected to do so.

payment of 70.39 C$, with a maximum of 121.44 C$ and a minimum of 60 C$.

We now describe how we constructed the set of 20 exam papers to be graded by the other experimental subjects. In order to control the distribution of mistakes, we only selected 7 out of the candidates' 23 papers.[9] Out of these 7 papers, we chose a "bribe paper" with 20 mistakes. To complete the set of 20 exam papers, we made up 13 papers with various numbers of mistakes. As the papers would be graded in a specific order, we ordered the set of 20 exam papers in a precise way. First, we decided to place the bribe paper in the 11^{th} position. Second, we arranged so that the first and last set of 10 papers each has a symmetric and roughly identical distribution of mistakes. In particular, they have the same average (15.5), the same median (15.5), and roughly the same standard deviation (6.8 versus 6.7). Third, we decided on a passing grade of 15 mistakes, meaning that if all mistakes were detected and reported, then half the papers would fail.[10] Finally, the exam papers were only identified by a 10-character code combining digits and letters. The first two digits, going from 01 to 20, identified the order in which the graders were asked to grade the papers. For the lab sessions, we only gave the graders the two pages of text. For the field sessions, we added a front page so as to look like a legitimate exam. This front page included in particular the identification code, as well as the instructions given to the candidates.

[9] We eliminated the papers with too many skipped words and too many mistakes. The selection process was made purely on the ground of convenience (i.e. to generate an appropriate distribution of mistakes).

[10] Such a failure rate is common in most exams and admission tests in Francophone countries.

1.3 Lab Graders (Montreal, Canada)

The grading sessions were conducted at CIRANO's Lab in Montreal (Canada). Subjects had to grade the set of 20 papers in a precise order. The graders were provided with an isolated work station, a pen, a report sheet, and an answer book. The sessions had no time limit, and the graders could leave the lab once their task completed.

Graders in the lab were provided with the following information. First, they knew from the start that they were taking part in an experiment. The corruption nature of the experiment, however, was not revealed immediately. Subjects were told at the beginning of the session that they had to grade 20 papers, and that they will receive additional information during the experimental session. Second, the lab graders were informed that some papers had been typed by real subjects, named "candidates", while others had been made up by the experimenters. The exact ratio of real candidates was not specified, and the graders were informed that nothing would enable them to identify whether a paper had been typed by a real candidate or an experimenter. Third, we partially explained to the lab graders how the number of mistakes they report for a paper affects the candidates' payoffs. Namely, if a grader reports more than 15 mistakes, then the paper is not remunerated. In such cases, we asked the graders to check a "Fail" column on the report sheet next to the number of mistakes. If the number of reported mistakes is 15 or less, then the payoff depends on the number of mistakes. The lower the number of mistakes reported, the higher the remuneration for the candidate.

The 20 papers were divided in two packs of 10. After completing the first pack, the graders were given the remaining 10 papers, with additional written instructions to be read privately. These instructions stated that paper 11 had been typed by a real candidate, and that this candidate had accepted to send a message and a money offer to the grader. The instructions then revealed to the grader the exact message and the offer. The graders were free to accept or reject the offer, and the consequences of each decision were explained. If the offer was accepted, then the amount was credited to the grader and debited from the candidate. The grader was then free to decide on the number of mistakes to report on the candidate's paper. The graders knew that paper 11 would be remunerated like any other paper, i.e. according to the number of mistakes reported. If the offer was rejected, then paper 11 was not remunerated. Nevertheless, we instructed the subjects to grade paper 11, as well as the 9 remaining papers. At the end of the session, subjects had to fill a short questionnaire after which they were paid in cash in Canadian Dollars.

1.4 Field Graders (Ouagadougou, Burkina Faso)

The field experiment took place in Ouagadougou (Burkina Faso). Burkina Faso is a landlocked country in West Africa with over 13 million inhabitants, among which 1.4 million live in the capital city Ouagadougou. A former French colony, the country became independent in 1960. Burkina Faso has been categorized by the World Bank as a low income country. In 2007, its real per capita income was 430 US$, compared to an average of 578 US$ for low income countries, and 952 US$ for Sub-Saharan African countries. Formerly called the "Republic of Upper Volta", it was renamed "Burkina Faso" on August 4, 1984, which may be translated into the "Land of Incorruptible People". In 2008, Transparency International ranked Burkina Faso the 80^{th} (out of 180) most corrupt country in the world. All sectors of the economy seem to be affected by corruption. In particular, the educational sector was ranked as the 7^{th} (out of 10) most corrupt public sector in the country in 2006.[11]

To hire graders, we used the service of a local recruiting firm (Opty-RH). Flyers placed around Ouagadougou proposed a part-time job consisting in grading exam papers. The offer stated that the job consisted of two sessions: a grading session lasting half a day, followed a week later by a debriefing session during which graders would be paid. Having a high school diploma and a form of identification were the only documentation required. Interested people were asked to come register in person at the recruiting firm location. Upon registering, graders were given the day, the time, and the location of their two sessions. The field subjects were unaware they were about to participate in an experiment.

The grading sessions took place in a high school located in the center of Ouagadougou. Upon arrival, the subjects were gathered in a large room. Instructions on how to grade the exam papers were read aloud, followed by questions. Each grader was then randomly assigned to a private room where he found an envelope containing the 20 exam papers properly ordered as in the lab, a report sheet, a red pen, and an answer book (i.e. a copy of the text without mistake). No information was given to the graders about the nature of the exam, or the candidates. The graders were explicitly instructed to grade the papers in the proper order. After spell-checking a paper, the graders had to report the number of mistakes both on the front page of the paper, and on

[11]For additional information on the extent of corruption in Burkina Faso, see the "État de la Corruption au Burkina Faso, Rapport 2006" published by the Réseau National de Lutte Anti-Corruption.

the report sheet. Graders were made aware that a candidate would fail the exam when more than 15 mistakes are reported. In such cases, we asked the graders to check the "Fail" column on the report sheet next to the number of mistakes. The graders were also instructed not to leave their room under any circumstance until they were done grading the 20 papers. We told them that we would stop by their room every 15 minutes to answer any potential question. Grading therefore took place behind closed doors, and the graders knew they would be undisturbed except at regular 15 minute intervals. Once their task completed, we gave the graders an "IOU", and reminded them to come back the following week for the debriefing and payment session.

To introduce the bribe, we wrote the candidate's message on an easily removable "post-it", and we taped a banknote with it on the second page of paper 11. We made sure that the message and the money were i) attached securely, ii) not visible unless the exam paper was opened to the second page, and iii) discovered before the grader started spell-checking the paper.[12] When a grader reported the bribe attempt during one of our visits, we asked him to write in bold on the paper "fraud attempt". We then took the banknote and the message, and instructed the grader to spell-check the bribe paper just like any other paper. Note that the instructions given to the graders at the beginning of the session specified that any attempt at fraud by a candidate would be penalized by failure of the exam. This information was also available in bold on the front page of each paper.

In the debriefing sessions, field graders were first informed that they took

[12]Recall that in the field, an exam paper consists of three pages: a front page, plus two pages of text. The bribe and the message were attached to the first page of text. Pictures of the exam paper with the bribe, as well as pictures of the high-school where the experiment took place are available on the author's webpage at www.amaboly.com.

part in an experiment. The nature of the experiment was explained, and information was provided about the objective of the research and the use of the data collected. In particular, we explained that the data would be fully anonymized, and that whatever decisions a subject made during the grading session would be without consequence. After answering all the graders' questions, we asked each subject whether he or she would accept to sign an ex-post consent form giving us the right to use the data we collected on him or her. We informed the subjects that they did not have to sign the consent form, in which case their data would be destroyed. They also knew that refusing to sign the consent form was without consequence on their payment. All subjects, in all of the treatments conducted in the field accepted to sign the ex-post consent form. Finally, the subjects filled a short questionnaire, after which they were paid in cash in return for the "IOU".

1.5 Conclusion

This chapter reports on a real-effort experimental design implemented both in the lab and in the field. As a result, we can compare the lab and the field in order to better understand the generalizability of lab results to real world situations. Relative to the lab experiment, subjects in the field are not aware they are taking part in an experiment. The design can be divided into two distinct experiments, to study work incentives and corrupt behavior. The experiment on work incentives is presented in Chapter 2, while the experiment on corruption is analyzed in Chapter 3.

Annexes

A.1 Instructions pour les candidats

Vous allez participer à une expérience économique. Il y a deux types de participants à cette expérience: nous appellerons le premier type "candidat" et le second type "correcteur". Seuls les candidats participent à la présente session. Vous êtes donc un candidat.

I. Tâche

Si vous acceptez de participer à cette expérience, nous vous dicterons un texte en français que vous devrez écrire en vous servant de l'ordinateur. La dictée comprend 366 mots et durera environ 20 minutes. Aucun temps ne sera alloué pour corriger vos erreurs à la fin de la dictée.

Les dictées seront ensuite corrigées par des correcteurs. Les correcteurs seront soit des expérimentateurs, soit différents groupes de participants qui seront recrutés pour deux séries de sessions expérimentales de correction. La première série se terminera au plus tard le 15 juin 2007 et la seconde série au plus tard le 15 septembre 2007. Comme nous allons vous l'expliquer dans quelques instants, votre rémunération dépendra principalement du nombre de fautes reportées par les correcteurs.

II. Critères de correction

Toutes les copies seront corrigées de la façon suivante :

1. La présentation (mise en page, retour à la ligne, saut de ligne) ne sera pas prise en compte.

2. Seules la précision dactylographique et l'orthographe seront évaluées. Sera compté comme une (1) faute :

- Un mot, une lettre ou une ponctuation (point, virgule...) qui manque.
- Une faute d'orthographe ou de saisie (y compris les accents, tirets...).
- Une minuscule ou une majuscule incorrectement placée dans le texte. Par exemple, une minuscule en début de phrase comptera pour une faute.

D'autre part :

- Un mot comportant plusieurs erreurs compte pour une seule faute. Par exemple " shantter" au lieu de "chanter" comptera pour une seule faute.
- Une même erreur, faite plusieurs fois sur le même mot, compte pour une seule faute. Par exemple, "chantter" comptera pour une seule faute si écrit exactement de cette façon plusieurs fois dans le texte.
- Les nombres doivent être écrits en lettres sauf les années qui doivent être en chiffres.

Notez qu'il est possible que le texte soumis aux correcteurs soit abrégé. Dans ce cas, nous retirerions exactement les mêmes paragraphes pour tous les candidats.

III. Objectif de l'expérience

L'étude à laquelle vous participez fait partie d'un projet de recherche plus général visant à mieux comprendre les problèmes de corruption. Dans la présente expérience, nous souhaitons étudier le comportement des correcteurs lorsqu'un pot-de-vin leur est proposé.

Par conséquent, si vous acceptez de participer à l'expérience, il se peut que nous proposions en votre nom (mais de façon confidentielle) un montant d'argent (c'est-à-dire un pot-de-vin) aux correcteurs en leur demandant de trouver peu de fautes dans votre copie. Dans ce cas, votre rémunération dépendra du nombre de fautes reportées par les correcteurs et de leur décision d'accepter ou non le montant d'argent.

IV. Rémunération

Si vous acceptez de participer à l'expérience, vous obtiendrez une compensation de 10 C$ pour être arrivé à l'heure, ainsi qu'une rémunération de 10 C$ pour avoir dactylographié la dictée. Ce montant de 20 C$ vous sera payé en argent comptant à la fin de la séance expérimentale.

De plus, vous recevrez deux paiements supplémentaires. Le premier paiement vous sera versé le 15 juin 2007 après la première série de sessions expérimentales de correction. Le second paiement vous sera versé le 15 septembre 2007 après la seconde série de sessions expérimentales.

Chacun des deux paiements peut varier entre 20 C$ au minimum et 60 C$ au maximum. Nous ne pouvons pas vous préciser le montant exact de chacun de ces deux paiements car ils dépendront du travail des correcteurs, et possiblement de leurs décisions d'accepter ou non le montant que nous pourrions proposer en votre nom. Cependant, en règle générale, moins vous ferez de fautes dans la dictée, plus votre rémunération sera élevée.

Si vous préférez ne pas participer à l'expérience, vous n'avez pas à accomplir la tâche demandée. Par contre, vous n'obtiendrez pour seul paiement que la compensation de 10 C$ pour être arrivé à l'heure. Ce montant vous sera payé, en argent comptant une fois que nous aurons fini de lire les instructions. Vous serez alors libre de partir.

V. Conditions de participation

En résumé, si vous acceptez de participer à l'expérience, vous acceptez:

- De dactylographier le texte de la dictée.
- Que la longueur de la dictée soit modifiée par les expérimentateurs au besoin.
- Que nous proposions en votre nom (mais de façon confidentielle) un montant d'argent aux correcteurs en leur demandant de trouver peu de fautes dans votre copie.
- De recevoir un paiement de 20 C$ aujourd'hui.
- De recevoir deux autres paiements compris entre 20 C$ et 60 C$, le 15 juin 2007 pour le premier paiement et le 15 septembre 2007 pour le second paiement.

A.2 Instructions pour les correcteurs (Laboratoire)

Vous allez participer à une expérience économique. Il y a deux types de participants à cette expérience: nous appellerons le premier type "candidat" et le second type "correcteur". Seuls les correcteurs participent à la présente session. Vous êtes donc un correcteur.

Au cours de l'expérience, les gains seront mesurés en Unités Monétaires Expérimentales (UME). A la fin de l'expérience, le montant en UME sera converti en $ canadiens. Le taux de change pour les correcteurs est de 12 UME = 1 C$. A cette somme, nous ajouterons 10 C$ pour être arrivé à l'heure à la présente session. Vous serez payé dès la fin de l'expérience, à l'extérieur du laboratoire, en privé et en argent comptant.

I. Tâche

Votre tâche consiste à corriger 2 lots de 10 copies, en suivant les instructions que nous vous donnerons. Une copie est un texte en français comportant des fautes d'orthographe ou de frappe. Une partie des copies a été dactylographiée par de vrais candidats, l'autre partie par les expérimentateurs. Cependant, rien (ni l'ordre, ni le nombre de fautes, ni le type de fautes) ne peut vous permettre d'identifier si une copie a été dactylographiée par un vrai candidat ou par un expérimentateur.

La rémunération des candidats dépendra du nombre de fautes que vous reporterez. Pour les copies avec 15 fautes ou moins, plus le nombre de fautes reportées est petit, plus la rémunération du candidat sera grande. Par contre, les copies comportant plus de 15 fautes ne seront pas rémunérées. La qualité de votre travail de correction est donc très importante car elle peut affecter fortement la rémunération que les candidats recevront.

Après avoir corrigé chaque copie, veuillez additionner et écrire le nombre de fautes sur la première page de cette copie. Nous vous donnerons aussi une "liste de report des fautes" sur laquelle vous devrez reporter le nombre de fautes de chaque copie en face du code d'identification correspondant. Si la copie comporte plus de 15 fautes, veuillez également cocher la colonne "Échec". Enfin, à la fin de l'expérience, nous vous demanderons de remplir un questionnaire.

Vous devez commencer la correction par le lot 1. Une fois la correction du lot 1 terminée, veuillez lever la main. Nous vous remettrons alors le lot 2 ainsi que des informations complémentaires sur le déroulement de la session. Le temps nécessaire pour la correction de chaque lot de 10 copies est estimé à 40 minutes. L'expérience ne devrait donc pas durer plus de 2 heures. Cependant, si vous n'avez pas terminé la correction des lots 1 et 2 au bout des deux heures, vous devrez rester pour terminer votre travail de correction. Dans les lignes suivantes, nous allons vous indiquer comment effectuer la correction des copies. Veuillez suivre ces consignes très précisément lors de votre travail de correction.

II. Critères de correction

1. Un solutionnaire est remis à chaque correcteur. Ce solutionnaire ne comporte aucune faute et doit servir de référence pour corriger les copies.

2. La présentation (mise en page, retour à la ligne, saut de ligne) ne sera pas prise en compte. Vous devez seulement évaluer la précision dactylographique et l'orthographe dans chaque copie.

3. Sur chaque copie, veuillez souligner ou entourer les fautes que vous trouvez.

Sera compté comme une (1) faute :

- Un mot, une lettre ou une ponctuation (point, virgule...) en plus ou en moins.

- Une faute d'orthographe ou de saisie (y compris les accents, tirets...).
- Une minuscule ou une majuscule incorrectement placée dans le texte. Par exemple, une minuscule en début de phrase comptera pour une faute.

D'autre part :

- Un mot comportant plusieurs erreurs compte pour une seule faute. Par exemple "shantter" au lieu de "chanter" comptera pour une seule faute.
- Un mot composé sera considéré comme un seul mot. Par exemple, "après-midi" sera considéré comme un seul mot.
- Une même erreur, faite plusieurs fois sur le même mot, compte pour une seule faute. Par exemple, "chantter" comptera pour une seule faute si écrit exactement de cette façon plusieurs fois dans le texte.
- Les espaces entre deux mots (par exemple, "Les __ instructions") ou entre un mot et une ponctuation (par exemple, "Instructions __ : ") ne sont pas considérés comme des fautes.
- Les nombres doivent être écrits en lettres sauf les années qui doivent être en chiffres.
- Les copies doivent être corrigées exactement dans l'ordre dans lequel vous les recevez.

III. Rémunération et contrôle du travail de correction

La rémunération est de 250 UME pour la correction des lots 1 et 2, soit 12.5 UME par copie. Cependant, les erreurs de correction (une faute non trouvée ou une faute qui n'en est pas une) peuvent entraîner une pénalité monétaire qui sera soustraite des 250 UME.

Nous contrôlerons votre travail afin de déterminer si vous devez être pénalisé. Le contrôle s'effectuera de la façon suivante. Vous tirerez au hasard 5 jetons d'un sac contenant des jetons numérotés de 1 à 20. Chaque numéro correspond à une copie différente. Pour chacune des copies tirées, nous calculerons la différence entre le nombre de fautes que vous avez inscrit sur la "liste de report des fautes" et le nombre de fautes effectivement présentes sur la copie. Nous ne prendrons en compte que la copie la plus mal corrigée, c'est-à-dire celle pour laquelle la différence calculée est la plus grande.

Le tableau ci-dessous donne le montant de la pénalité en fonction de cette différence maximale:

Si la différence maximale est:	Penalité (en UME):
entre 0 et 2 fautes	0
entre 3 et 5 fautes	100
entre 6 et 9 fautes	150
de 10 fautes ou plus	225

Par exemple, si pour les copies tirées au hasard, la différence maximale entre le nombre de fautes reporté sur la "liste de report des fautes" et le nombre de fautes effectif est 7, la pénalité sera de 150 UME. Dans ce cas, votre rémunération sera : $250 - 150 = 100$ UME.

Si vous avez des questions ou problèmes pendant le travail de correction, veuillez nous appeler en silence en levant la main. Par contre, toute communication entre participants est interdite pendant l'expérience. Si vous ne respectez pas ces consignes, nous serons obligés de vous exclure de l'expérience, sans aucun paiement.

A.3 Instructions pour les correcteurs (Suite, Laboratoire)

Les informations ci-dessous sont privées. Veuillez les lire en silence et ne pas les communiquer à d'autres participants. Si vous avez des questions, merci de lever la main. Nous viendrons répondre en privé à vos questions.

La copie 11V250683G que vous allez corriger a été dactylographiée par un vrai candidat, pas par un expérimentateur. Ce candidat a accepté de vous envoyer un message accompagné d'un montant d'argent.

Le message est le suivant : "S'il vous plait, veuillez trouver peu de fautes dans ma dictée". Le montant qui vous est proposé est de 50 UME.

Vous devez décider d'accepter ou non ce montant:

- Si vous acceptez le montant proposé, la copie du candidat sera rémunérée normalement, c'est-à-dire en fonction du nombre de fautes que vous reporterez. D'autre part, le montant de 50 UME sera soustrait des gains du candidat et ajouté à vos gains, et ce quelque soit le nombre de fautes que vous reporterez. En particulier, si vous acceptez le montant et que vous reportez plus de 15 fautes, le candidat ne recevra aucune rémunération pour sa copie (puisqu'il aura échoué), mais nous lui retirerons quand même les 50 UME qui vous seront transférés.

- Si vous refusez le montant proposé, la copie du candidat ne sera pas rémunérée.

Quelque soit votre décision d'accepter ou de rejeter le montant proposé, vous devez corriger la copie du candidat. Cependant, vous êtes libre d'inscrire le nombre de fautes que vous souhaitez sur sa copie. Comme pour toutes les autres copies, le même nombre de fautes devra être inscrit sur la copie et sur la "liste de report des fautes".

Après avoir corrigé la copie 11, veuillez nous indiquer la décision que vous avez prise en cochant l'une des cases sur la "feuille de décision" qui vous a été remise.

Notez bien qu'au moment de contrôler votre travail, nous ne disposerons que de votre "liste de report des fautes". Nous ne saurons donc pas si vous avez accepté ou non le montant proposé par le candidat. Ce n'est qu'après avoir contrôlé votre travail que vous nous remettrez votre "feuille de décision". Nous pourrons alors calculer votre rémunération totale.

A.4 Feuille de décision (Laboratoire)

Numéro de table: xx

Afin de nous indiquer votre décision, veuillez cocher l'une des cases ci-dessous au moment qui vous conviendra le mieux:

☐ "Je refuse le montant"

☐ "J'accepte le montant"

Merci pour votre collaboration !!!

A.5 Instructions pour les correcteurs (Terrain)

I. Conditions et critères de correction

1. Un solutionnaire vous sera remis. Ce solutionnaire ne comporte aucune faute et doit servir de référence pour corriger les copies.

2. Sur la première page de chaque copie, vous trouverez les instructions qui ont été données aux candidats. Veuillez lire attentivement ces instructions afin de bien comprendre ce qui constitue une faute.

3. Vous devez apposer vos initiales et votre signature sur la première page de chaque copie corrigée.

4. Les copies doivent être corrigées exactement dans l'ordre dans lequel vous les recevez.

5. La correction doit se faire sans pause ou sortie de la salle.

6. Nous vous rendrons régulièrement visite (environ toutes les 15 minutes). Au cours de ces visites, vous pourrez nous poser des questions ou reporter tout problème lié à la correction. En aucun cas, vous ne devez sortir de votre salle avant d'avoir terminé votre travail de correction.

7. La durée estimée du travail de correction est d'environ 3 heures.

II. Instructions pour le report des fautes

1. Vous trouverez dans votre salle de classe une "liste de report des fautes" qui comporte les codes d'identification des candidats ainsi qu'une colonne "Échec".

2. Veuillez inscrire votre nom, prénom et signature aux endroits indiqués sur cette "liste de report des fautes".

3. Les copies qui vous seront remises sont anonymes et ne comportent qu'un numéro d'identification.

4. Après avoir corrigé chaque copie, veuillez additionner et écrire le nombre de fautes sur la première page de cette copie. Vous devez reporter le nombre de fautes de chaque copie en face du code d'identification correspondant. Si une copie comporte plus de 15 fautes, veuillez également cocher la colonne "Échec".

III. Rémunération et contrôle du travail de correction

1. Le taux de rémunération est de 5000 FCFA pour 20 copies, soit 250 FCFA par copie. Cependant, les erreurs de correction (une faute non trouvée ou une faute qui n'en est pas une) peuvent entraîner une pénalité monétaire qui sera soustraite des 5000 FCFA. Nous contrôlerons donc votre travail afin de déterminer si vous devez être pénalisé. Le contrôle s'effectuera de la façon suivante. Nous tirerons au hasard 5 de vos copies. Pour chaque copie, nous calculerons la différence entre le nombre de fautes que vous avez inscrit sur la "liste de report des fautes" et le nombre de fautes effectivement présentes sur la copie. Nous ne prendrons en compte que la copie la plus mal corrigée, c'est-à-dire celle pour laquelle la différence calculée est la plus grande. Le tableau ci-dessous donne le montant de la pénalité en fonction de cette différence maximale:

Si la différence maximale est:	Penalité (en FCFA):
entre 0 et 2 fautes	0
entre 3 et 5 fautes	2000
entre 6 et 9 fautes	3000
de 10 fautes ou plus	4500

Par exemple, si pour les copies tirées au hasard, la différence maximale entre le nombre de fautes reporté sur la "liste de report des fautes" et le nombre de fautes effectif est 7, la pénalité sera de 3000 FCFA. Dans ce cas, votre rémunération sera: $5000 - 3000 = 2000$ FCFA.

2. Une fois la correction terminée, veuillez sortir en silence et venir nous voir dans la présente salle. Nous vous prions de bien vouloir laisser le matériel de correction ainsi que les copies dans votre salle de classe.

3. Une reconnaissance de dette vous sera remise. Cette reconnaissance de dette comportera le nom et l'adresse de l'émetteur, ainsi que la date à laquelle vous serez payés.

4. Il est très important que vous arriviez à l'heure indiquée sous peine de ne pas être payé.

Bon travail !!!

A.6 Copie d'examen sur le terrain (première page)

BURKINA FASO		Code d'identification :
Date :		01B080582Y

ÉPREUVE DE DICTÉE AVEC PRISE DACTYLOGRAPHIQUE

Instructions aux candidat(e)s

Seules la précision dactylographique et l'orthographe seront évaluées. La présentation (mise en page, retour à la ligne, saut de ligne) ne sera pas prise en compte. Sera compté comme une (1) faute :

- Un mot, une lettre ou une ponctuation (point, virgule...) en plus ou en moins.
- Une faute d'orthographe ou de saisie (y compris les accents, tirets...).
- Une minuscule ou une majuscule incorrectement placée dans le texte. Par exemple, une minuscule en début de phrase comptera pour une faute.

D'autre part :

- Un mot comportant plusieurs erreurs compte pour une seule faute. Par exemple " shantter" au lieu de "chanter" comptera pour une seule faute.
- Un mot composé sera considéré comme un seul mot. Par exemple, "après-midi" sera considéré comme un seul mot.
- Une même erreur, faite plusieurs fois sur le même mot, compte pour une seule faute. Par exemple, "chantter" comptera pour une seule faute si écrit exactement de cette façon plusieurs fois dans le texte.
- Les espaces entre deux mots (par exemple, "Les__instructions") ou entre un mot et une ponctuation (par exemple, "Instructions__: ") ne sont pas considérés comme des fautes.

- Les nombres doivent être écrits en lettres sauf les années qui doivent être en chiffres.

N.B. : Toute fraude ou tentative de fraude entrainera un échec immédiat.

A.7 Texte de la dictée (Corrigé)

Titre: **Anticiper la fin du pétrole**

On estime à plusieurs milliers de milliards de barils les réserves de pétrole prouvées, soit cent cinquante milliards de tonnes environ. Elles sont très inégalement réparties : près des deux tiers sont situées au Proche-Orient.

Leur évolution ne permet cependant pas de prévoir celle de la production pétrolière, les données relatives aux réserves donnant lieu à de vives controverses entre écoles de pensée.

Les réserves obtenues par des techniques de mise en production modernes et par la réévaluation des réserves de gisements anciens coûtent souvent moins cher à exploiter que celles obtenues par exploration. D'où la limitation de cette activité dans des pays offrant pourtant les meilleures perspectives de découvertes.

Les pessimistes insistent tout d'abord sur le caractère politique des réévaluations de réserves effectuées en 1986 et qui ne correspondent pas à de véritables réserves prouvées.

Pour appuyer leur thèse, ils rappellent que nous disposons enfin d'un accès à l'ensemble des données de tous les bassins pétroliers, ainsi que d'un échantillonnage suffisant pour que des méthodologies prédictives des réserves restant à découvrir soient désormais raisonnablement fiables.

Différentes équipes de spécialistes proposent une vision intermédiaire. Les réserves ultimes de pétrole conventionnel seraient de l'ordre de trois mille milliards de barils, dont mille environ déjà consommés, un peu plus de mille de réserves prouvées, le reste correspondant aux réserves à découvrir.

On peut considérer qu'il existe un continuum de ressources en hydrocarbures : gisements plus difficiles d'accès, plus complexes, plus difficiles à détecter. Ce continuum n'est pas limité aux hydrocarbures d'origine pétrolière : nombreuses sont les recherches sur le développement des techniques de production de carburants à partir du gaz naturel ou à partir du charbon. Il s'étend aussi aux carburants qui sont issus de la biomasse.

Auteurs: Denis Babusiaux et Pierre-René Bauquis[13]

[13] Voir: http://www.monde-diplomatique.fr/2005/01/BABUSIAUX/11803

A.8 Distribution des fautes (copies d'examen)

Rang de la copie	Nombre de fautes
1	15
2	24
3	6
4	25
5	16
6	5
7	20
8	11
9	19
10	14
11	20
12	11
13	12
14	10
15	23
16	16
17	17
18	4
19	27
20	15

Chapitre 2

Monitoring in a mission-oriented context

2.1 Introduction

Standard economic models of incentives consider that agents are likely to shirk because of effort disutility. In order to induce agents to raise their effort level, incentive schemes such as monitoring may be used (see e.g. Alchian and Demsetz 1972). However, some behavioral models suggest that if perceived as hostile or unfair, monitoring may reduce effort by destroying intrinsic motivation (e.g. Frey 1993).[1] Therefore, monitoring may have two opposite effects: a disciplining and a crowding-out effect.

Several experimental studies show that monitoring may indeed reduce overall effort (see e.g. Fehr and Falk 2002, Fehr and Gächter 2002). Such a result may be explained by assuming that agents are heterogenous regarding intrinsic motivation. Some agents are selfish and dislike work effort, while other agents may be intrinsically motivated to work hard.[2] For the first type, monitoring may be effective in eliciting higher effort (Falk and Kosfeld 2006). For the second type however, monitoring may destroy intrinsic motivation, leading to a reduction in effort. This reduction may be large enough to induce a lower overall effort. Most experiments supporting the crowding-out effect of monitoring use a principal-agent design with two decision stages. First, the principal chooses between two incentive schemes, e.g. between monitoring or not monitoring. Then, subject to the chosen incentive scheme, the agent chooses an effort level which determines the profit of the principal. As a result, these experiments may be thought of as studying the profit-oriented sector, which produces private goods that can be valued monetarily and distributed via mar-

[1] Frey and Jegen (2001) provide a survey of the empirical literature.

[2] "Intrinsic motivation" refers to doing something because it is inherently interesting or enjoyable (Ryan and Deci 2000), as opposed to "extrinsic motivation" which refers to incentives coming from outside the person (Frey 1993).

kets.

This chapter extends the analysis of the incentive effects of monitoring to organizations that can be viewed as mission-oriented, as opposed to profit-oriented. Mission-oriented organizations typically include public and not-for-profit organizations (see e.g. Wilson 1989, Tirole 1994, Sheehan 1996). These organizations are generally involved in promoting general welfare through the provision of collective goods such as health, education, or security.[3] Note that some mission-oriented organizations, such as universities or hospitals, may be private but with many other objectives than profit-maximization. It is the nature of the activities that is important, not whether the service is provided publicly or privately (Besley and Ghatak 2003).

In a mission-oriented setup, one may expect monitoring to have disciplining effects when the agent is selfish, as in a profit-oriented setup. However, in a mission-oriented work context, two factors may mitigate the crowding-out effect of monitoring. First, the success of the mission may be valued by the agent more than any monetary income he receives in the process (Besley and Ghatak 2005). Therefore, it may be the case that motivated agents work as hard with as without monitoring. Second, the agent's effort may mainly affects third-parties but not the principal who chooses the incentive scheme.[4] For example, physicians are committed to saving people's life, policemen to protecting citizens, professors to educating students. As a result, motivated

[3]For instance, Rainey and Steinbauer (1999, p. 23) argue that government employees are driven by public service motivation, defined as "a general altruistic motivation to serve the interests of a community of people, a state, a nation, or humankind".

[4]Some public organizations may get their revenues from taxation, but not through market competition. Their existence is typically not at risk when the public official shirks. In the profit-oriented sector, the agent's effort may also affect clients (e.g. an employee at a McDonald dealing with customers), but more importantly, it determines the principal's level of profit.

agents may refrain from reducing effort to avoid hurting these third-parties. In total, overall effort may therefore be higher with monitoring than without monitoring in a mission-oriented setup.

Our experimental design attempts to create a mission-oriented work environment as follows. To begin with, we model a Principal-Agent-Client framework by introducing a third-party (client) besides a principal and an agent (see Banfield 1975). Namely, agents (graders) are hired and paid by the principal (experimenter) to grade exam papers pertaining to clients (candidates). Next, we induce a sense of mission by requiring real-effort from candidates who had to type a text dictated to them. The dictated texts provided us with exam papers. Then, graders' task consisted in finding spelling mistakes in the exam papers. In such a context, the "mission" may consist in providing a fair grading of the candidates' work.[5]

Two main experimental treatments are conducted in the lab. In the Control treatment, subjects are paid a fixed amount for their grading, independent of how they perform the task. In a second treatment called "Monitoring", we introduce a performance-related payment scheme. Specifically, graders are told that they may be imposed a monetary penalty depending on the quality of their grading. Two secondary treatments are also conducted. In the first secondary treatment, we analyze the effects on effort of increasing both the monitoring rate and the monetary penalties. In the second one, a gift exchange treatment is conducted for comparison with monitoring.[6]

[5]Wilson (1989) defines a mission as a culture that is "broadly shared and warmly endorsed" (p. 109). We believe that principles such as honesty and fairness are shared and endorsed by most people in academia. The latter principle upholds that each student should be treated equally, in particular when grading their exams. Fairness in grading is likely to be salient to our experimental subjects as they are mostly students.

[6]Gift exchange in labor markets relates to the fair wage-effort hypothesis (Akerlof 1982),

To test the robustness of the results obtained in the lab in Montreal, we compared them to those of the field experiment conducted in Ouagadougou (Burkina Faso). Relative to the lab, two aspects of the field experiment are noteworthy. First, the field experiment is done unbeknownst to subjects. Second, the field experiment is conducted in a developing country, namely Burkina Faso. These differences allow to check the generalizability of lab results when moving from the lab in a developed country to the field in a developing country.

In the lab, we find that monitoring significantly increases the graders' effort relative to the Control treatment, indicating that monitoring may be effective in raising effort in a mission-oriented setting. However, increasing monitoring rate and monetary penalties does not raise effort level further. Gift-exchange is found to have no effect on effort, suggesting that monitoring (a pecuniary incentive) may be more effective than gift-exchange (a non-pecuniary incentive) in eliciting effort in a mission-oriented setup.[7]

The effect of monitoring is found to depend on the grader's intrinsic motivation. Indeed, to separate motivated graders from selfish graders, we use the graders' decision to reject a bribe offer as a proxy for adherence to the mission of grading exam papers impartially. Graders who rejected the bribe are considered motivated, while those who accepted the bribe are considered selfish. The evidence suggests that motivated graders work as hard with as without monitoring, while monitoring appears to have a significant disciplining effect on selfish graders. The results obtained in the lab are found to be robust. Indeed, although the level of effort differs between the lab in Montreal and the field

which posits that employers pay a wage above the market-clearing wage in order for their employees to reciprocate by exerting higher effort.

[7]For another paper that compares pecuniary and non-pecuniary incentive schemes, see Al-Ubaydli et al. (2008).

in Ouagadougou, we find that the direction and the magnitude of treatment effects are fully consistent between the two environments.

The remainder of this chapter is organized as follows. The related literature is summarized in section 2. Experimental treatments are presented in section 3. Experimental results are analyzed in section 4 and section 5 concludes.

2.2 Related Literature

To the best of our knowledge, no experimental study of the incentive effects of monitoring in the mission-oriented sector is available. We therefore simply discuss experiments on monitoring in profit-oriented setups. Then, we summarize briefly the experimental literature on the generalizability of lab behavior to real world situations.

In a natural field experiment, Nagin et al. (2002) studied the effects of varying monitoring rates in 16 call centers. These authors find results supporting the disciplining effect of monitoring. They also show that a significant proportion of employees exhibit intrinsic motivation by not reacting to variations of monitoring rates. However, several lab experiments have documented that monitoring may in fact reduce effort level. Fehr and Gächter (2002) find that the crowding-out effect of imposing a fine to prevent agents from shirking may be large enough to make, on average, contracts with monitoring less efficient than fixed-price contracts without monitoring. Falk and Kosfeld (2006) also find that the undermining effect of monitoring may outweigh its disciplining effect.[8] These experiments suggest that the crowding-out effects of monitoring

[8]The crowding-out effect of monitoring appears to extend to other contexts than the labor market. Gneezy and Rustichini (2000) show that the introduction of a fine for parents

may outweigh the disciplining effects in a profit-oriented setup. Whether this result extends to a mission-oriented sector is still an open question.

By comparing a lab and a field experiment, this chapter adds to experimental studies on the generalizability of lab behavior to real-world situations (see e.g. List 2006 on social preferences, Benz and Meier 2008 on donation behavior). We are only aware of one labor market lab experiment with an explicit field counterpart. In a gift-exchange experiment, Hennig-Schmidt, Rockenbach and Sadrieh (2008) find similar results between a lab and a natural field experiment, conducted to analyze the effects of own and peer wage variations on workers' effort. However, their tasks in the lab and in the field were different, not allowing for a direct comparison of effort level and treatment effects. In contrast, in our experiment, graders in the lab and in the field had to grade exactly the same set of exam papers and received the same instructions for grading the papers. As a result, we can compare explicitly the direction and the magnitude of treatment effects between the lab and the field.

2.3 Experimental Treatments

In the following section, we describe the experimental treatments. There are four treatments. Each treatment is conducted both in the lab and in the field. In the Control treatment, subjects are paid a fixed amount for their grading, independent of how they perform the task. In the lab, the fixed amount, called a wage hereafter, was 250 Experimental Units (*EU* hereafter) for 20 exam papers or 12.5 *EU* per paper. The conversion rate in the lab was 1 *C$* = 12

arriving late to pick up their children at school, crowded out intrinsic motivation to do so. Monitoring has also been found to undermine trustworthiness (Fehr and Rockenbach 2003), or intrinsic motivation for honesty (Schulze and Frank 2003).

EU. In the field, the wage was set at $5,000$ $FCFA$ for 20 exam papers or 250 $FCFA$ per paper.[9] Note that the lab payoffs are standard while the wage paid in the field has been selected so as to be credible and attractive. Indeed, $5,000$ $FCFA$ represents about $1/6$ of the monthly minimum wage in Burkina Faso.

In the "Monitoring" treatment, we introduced a mechanism aimed at monitoring the accuracy with which the graders perform their task. The monitoring mechanism was explained as follows. We told each grader that we would randomly pick and control 1 of the 20 papers he graded. Then, we would calculate the difference between the number of mistakes reported by the grader and the number of mistakes we found in the paper. This difference would then be used to determine the penalty to be imposed (see Table 2.1).[10] Similarly, in the second monitoring treatment called "High Monitoring", we told each grader that we would randomly pick and control 5 out of the 20 papers graded. Then, we would calculate the difference between the number of mistakes reported by the grader and the number of mistakes we found in the papers. Only the worst graded paper, i.e. the one with the maximum calculated difference, would be considered to determine the monetary penalty (see Table 2.2). Except for the risk of being penalized, the monitoring treatments are identical to the Control treatment.

The "Gift Exchange" treatment is also identical to the Control treatment except that the wage was 40% higher (i.e. 350 EU in the lab and $7,000$ $FCFA$ in the field). In the field, we implemented gift-exchange by providing a direct gift to graders. Specifically, in posters to recruit graders in the field,

[9]The Franc CFA is the currency used in Burkina Faso. In July 2007, the conversion rate was roughly 1 $C\$$ for 442.9 $FCFA$.

[10]While monetary fines are not typically used in the workplace, other more common types of "fines" include verbal warnings, demotion or dismissal (Dickinson 2001).

the announced wage was 5,000 $FCFA$. However, graders were told on the day of the experiment that the amount they would receive had been increased to 7,000 $FCFA$.[11] In the lab, subjects simply received a higher wage compared to the baseline treatment.

In total, 180 (respectively 248) subjects participated in the four treatments conducted in the lab (respectively in the field). More precisely, in the lab (field), 62 (82) subjects participated in the Control treatment, 55 (82) in the Monitoring treatment, 32 (44) in the High Monitoring treatment, and 31 (40) in the Gift Exchange treatment. On average, the total earnings of a lab grader (a field grader) were 31 $C\$$ (6,000 $FCFA$) in the Control treatment, 27.87 $C\$$ (5,060.98 $FCFA$) in the Monitoring treatment, 21.36 $C\$$ (4,545.45 $FCFA$) in the High Monitoring treatment, and 39.25 $C\$$ (8,000 $FCFA$) in the Gift Exchange treatment.[12]

2.4 Experimental Results

In this study, we concentrate on the 10 first papers graded in the lab and in the field. Effort is proxied by the precision of a grader, which is measured as the absolute deviation from the actual number of mistakes in a paper. The higher the absolute deviation, the lower the grader's precision. In the next sections, we use summary statistics and Mann-Whitney tests to analyze the data, followed by a regression analysis.[13] We start by the results obtained in the

[11] This was similar to the surprise approach used in Gneezy and List (2006).

[12] The amounts do not include the bribe. Using the Purchasing Power Conversion table from the UN database, we have that e.g. 31 $C\$ \simeq 5580$ $FCFA$, in 2006. See at http://data.un.org/Data.aspx?d=MDG&f=seriesRowID%3A699.

[13] The reported test statistics are two-sided. Note also that the conclusions of the reported Mann-Whitney tests do not differ from those of t-tests (not reported here).

lab in Montreal, before turning to those obtained in the field in Ouagadougou.[14]

2.4.1 Results from the Lab

Table 2.3, columns "Lab", provides an overview of the graders' performance in the lab. The first column gives the rank of papers as graders were asked to grade the papers in a specific order, and the second column gives the actual number of mistakes of the relevant paper. The following columns show the *per paper* average number of mistakes, reported by treatment and by environment.[15] The "Lab" columns show that graders tend to find fewer mistakes than actually present in an exam paper. Indeed, for papers 01 to 10, the average number of mistakes reported by graders is inferior to the true number of mistakes in all treatments.

Summary statistics on average absolute deviation in the lab in Montreal are given in Table 2.4, columns "Lab". We find that work monitoring increases overall effort. Indeed, relative to the Control treatment, graders' precision is significantly higher at the 1% level in the Monitoring treatment (p-value = 0.001, Mann-Whitney), and at the 10% level in the High Monitoring treatment (p-value = 0.079, Mann-Whitney). While graders' precision is slightly lower in High Monitoring relative to Monitoring, the difference between these treatments is not significant (p-value = 0.432, Mann-Whitney). Finally, giving a higher wage to graders has no significant effect on work effort compared to the Control treatment (p-value = 0.616, Mann-Whitney).

We now turn to a regression analysis to study treatment effects while con-

[14] Data is available from the author on request.

[15] We rely only on the number of mistakes reported by graders. In other terms, we did not check whether the mistakes reported by the graders were the actual mistakes.

trolling for some individual characteristics (see Table 2.5). To do so, we exploit the panel structure of the experimental data to estimate a model of the form:

$$Y_{i,t} = \beta X_{i,t} + w_{i,t} \tag{2.1}$$

where $Y_{i,t}$ is the absolute difference between the number of mistakes reported by subject i for exam paper t $(t = 1, ..., 10)$ and the actual number of mistakes in exam paper t. $X_{i,t}$ is the vector of independent variables. These independent variables include the subject's age (*Age*), the subject's gender (*Female*: 1 if female, 0 otherwise), as well as experimental treatment dummies. We control for the ranking of the exam papers (*Paper Ranking*, 01 to 10) as well as interactions between the ranking and experimental treatments. Exam paper fixed effects are also included in the regression (although not reported). To control for possible grader specific random effects, we model the error term as $w_{i,t} = u_i + \varepsilon_{i,t}$, where $Var(\varepsilon_{i,t}) = \sigma_{\varepsilon}^2$, and $Var\,(u_i) = \sigma_u^2$.[16] The model is estimated by FGLS with clustered standard errors to take into account correlation of unknown form among the 10 observations of each grader.[17]

The estimation results are presented in Table 2.6, column "Lab". Monitoring and High Monitoring significantly increase graders' precision compared to the Control treatment. Yet, precision is not significantly different between Monitoring and High Monitoring. Gift-exchange has no significant effect on graders' effort. The results of the regression regarding treatment effects are therefore consistent with the results of the Mann-Whitney tests. In addition,

[16] Breusch-Pagan LM tests for random effects reject the null hypothesis that $\sigma_u^2 = 0$ in all three regressions of Table 2.6.

[17] The statistical software used is Stata 9. Alternative specifications of equation (2.1) yield results similar to those discussed here. In particular, using two-way random effects or using a count model (Poisson) does not change the results.

Table 2.6 indicates that women and younger subjects are more precise in their grading. The coefficients of the interaction terms between *Paper Ranking* and the Monitoring and High Monitoring treatments are positive and significant. These coefficients suggest that while monitoring increases effort level, the effects may weaken over time.

2.4.2 Robustness Check

In this section, we first discuss the results of the field experiment in Ouagadougou, without pooling the data. Table 2.3, columns "Field", gives an overview of the graders' work. As in the lab, graders find fewer mistakes than actually present in an exam paper.

Table 2.4, columns "Field", provides summary statistics on the average absolute deviation in Ouagadougou. This table shows that the average absolute deviation is lower in the Monitoring and High Monitoring treatment compared to the Control treatment. The differences with the Control treatment are both significant at the 1% level (p-value = 0.002 for Monitoring, and p-value = 0.004 for High Monitoring, Mann-Whitney). However, High Monitoring does not significantly increase effort relative to Monitoring (p-value = 0.475, Mann-Whitney). Table 2.4 also shows that gift-exchange slightly increases absolute average deviation relative to the Control treatment, but the difference is not significant (p-value = 0.756, Mann-Whitney).

Table 2.6, column "Field", presents regression results using only data from the field in Ouagadougou. The dependent variable, the independent variables, and the regression model used are identical to those described in section 4.1. Regarding treatment effects, the regression results are consistent with those obtained using Mann-Whitney tests. Moreover, we find that women are significantly more precise than men, while the subject's age has no explanatory power in the field. Again, the effect of monitoring may weaken over time as suggested by the coefficients of the interaction terms between *Paper Ranking* and each of the two monitoring treatments.

The preceding results show that the direction of treatment effects are the same in the field in Ouagadougou and in the lab in Montreal. We therefore turn to the analysis of the magnitude of the treatment effects. To do so, we pool together the data from the lab and the field. We also include in equation (2.1) a dummy variable *Field* (1 if a field grader, 0 otherwise), as well as interaction terms between experimental treatment dummies and *Field*. Note that the *Field* dummy captures the differences between the lab in Montreal and the field in Ouagadougou, including the fact that subjects in the field are unaware they are in an experiment.

The results are reported in Table 2.6, column "Pooled Data".[18] The *Field* variable shows that the level of absolute deviation differs between the lab in Montreal and the field in Ouagadougou. Indeed, graders in the field are significantly less precise than graders in the lab. To some extent, this is a surprising result as we could have expected graders in the lab not to give their best effort compared to the field where subjects thought they were grading a "real" exam. Once we control for the environment (lab vs. field) and location (Burkina Faso vs. Canada) through the variable *Field*, we find no significant difference between the lab and the field regarding the magnitude of treatment effects. In-

[18]Notice that the results regarding treatment effects obtained with the pooled data are consistent with the results obtained using lab and field data separately.

deed, the coefficients of the interaction terms between experimental treatment dummies and *Field* are all statistically insignificant at conventional levels.

In our experiment, we instructed graders to find all the mistakes in a given exam paper. However, recall that reporting 15 or more mistakes is considered a failure. In the Control treatment, graders may have an incentive to grade seriously only until they find 16 mistakes in a paper, as from there on, the paper is failing "anyway". In contrast, in the monitoring treatments, graders have an incentive to find all the mistakes in an exam paper to avoid being penalized. As a result, the observed effects of monitoring may be mainly driven by papers with 16 or more mistakes, which may have been graded seriously in the monitoring treatments but not in the Control treatment. In an attempt to explore this hypothesis, we divided the sample in two groups of exam papers according to the actual number of mistakes (papers with 15 or less mistakes, and those with more than 15 mistakes). We run a regression for each group with the pooled sample. The results from these regressions (see Table 2.7) are consistent with those presented above.

To summarize, we find that the direction and the magnitude of treatment effects are the same in the lab in Montreal and in the field in Ouagadougou. In our view, this result is remarkable as it suggests that labor market lab behavior in developed countries may generalized to real world situations in developing countries, despite different economic, political and cultural contexts.[19]

[19]For a discussion of conditions under which lab behavior is likely or not to be a close guide of behavior in the field, see Levitt and List (2007a, 2007b).

2.4.3 Discussion

We find the following treatment effects both in the lab in Montreal and in the field in Ouagadougou. First, monitoring increases overall effort, indicating that monitoring may be effective in a mission-oriented setup. This is in contrast to several profit-oriented lab experiments reporting that the crowding-out effects of monitoring may outweigh its disciplining effects (e.g. Fehr and Gächter 2002, Falk and Kosfeld 2006). However, High Monitoring does not increase workers' effort relative to Monitoring, suggesting that monitoring becomes ineffective above a given threshold. Second, gift-exchange does not increase the effort level of graders. This lends some support to experimental studies questioning the robustness of gift-exchange in labor markets (e.g. Engelmann and Ortmann 2002, Gneezy and List 2006, Hennig-Schmidt, Rockenbach, and Sadrieh 2008). Finally, put together, our results suggest that monitoring may be more effective than gift-exchange in eliciting workers' effort in a mission-oriented setup.[20]

To explain the results above, we conjecture that some workers in the mission-oriented sector may be intrinsically motivated to accomplish their "mission" and are not likely to reduce their effort level when monitored. At the same time, monitoring may increase the effort level of agents with low intrinsic motivation, leading to higher overall effort in the two monitoring treatments relative to the Control treatment. Gift-exchange does not affect the effort level of neither an intrinsically motivated agent (because he is already motivated to work hard) nor of a selfish agent (because he is a self-maximizer).

We attempt to explore these conjectures below. To this end, we separate

[20] To some extent, such conclusion is in line with Burgess et al. (2004)'s empirical study which indicates that performance pay may be more cost effective than a general pay rise in a public work context.

motivated graders from selfish graders by using the graders' decision to reject a bribe offer as a proxy for intrinsic motivation. Indeed, self-maximizing behavior should result in accepting the bribe offer, noting in particular that bribe acceptance was arguably without consequence in the lab and in the field.[21] Graders who rejected the bribe are considered motivated, while those who accepted the bribe are considered selfish.[22] We divided our sample accordingly, and run a FGLS regression (see equation 2.1) for each group, pooling the data from the lab and the field together.[23] We use the same variables as in Table 2.6, column "Pooled Data", except for interaction terms between the *Field* variable and experimental treatments.

Table 2.8 presents the regression results, which are broadly supportive of our conjectures. We focus on the treatment effects. Relative to the Control treatment, the Monitoring and High Monitoring treatment do not affect the effort level of motivated graders. They tend to work as hard with as without monitoring. When a majority of workers are motivated, it may therefore be possible to save on incentive schemes such as monitoring. By contrast, compared to the Control treatment, selfish graders work significantly harder both

[21]Outside an experimental framework, bribe acceptance may be extremely difficult to detect in the field in a corruption situation similar to the one we modeled.

[22]64.24% of graders accepted the bribe in the lab, and 46.83% accepted in the field. In total, 54.24% of graders accepted the bribe.

[23]Chronologically, the first 10 papers are graded before graders are aware they will have to make a decision relating to bribery. For this reason, we consider the "Absolute Deviation" as a predetermined (lagged) variable and use it as an explicative variable of bribe acceptance in Chapter 3. For the same reason, bribe acceptance is considered to have no effect of the grading of the 10 first papers. In section 2.5, we attempt to analyze the impact of the level of intrinsic motivation on the effectiveness of monitoring, by dividing graders according to whether or not they accepted the bribe. However, doing so implicitly turns bribe acceptance into a determinant of graders' performance, in contrast to our initial assumptions. As a result, the results in Table 2.8 must only be seen as indicative, and should not be considered as a formal test of whether the effectiveness of monitoring depends on the level of intrinsic motivation.

in the Monitoring and the High Monitoring treatment.[24] Therefore, monitoring appears to have a significant disciplining effect on selfish graders. Finally, the Gift-exchange treatment has no significant effect neither on motivated graders nor on selfish graders.

2.5 Conclusion

This chapter extends the analysis of the disciplining and crowding-out effect of monitoring to mission-oriented organizations. To model a mission-oriented work context, we use a Principal-Agent-Client framework in which graders are hired by the experimenter to grade exam papers pertaining to candidates. We induce a sense of "mission" in performing the experimental task by requiring real-effort from both types of subject.

In the lab in Montreal, we find that monitoring significantly increases effort level relative to the Control treatment. So, even if monitoring had hidden costs, they may not have been sufficient to undermine completely its effectiveness in our mission-oriented setup. However, monitoring becomes ineffective in increasing effort when it is above a certain threshold, suggesting a crowding-out effect. Moreover, our results indicate that the effect of monitoring may weaken over time. It may therefore be interesting to analyze monitoring taking into account the duration of the task in future research. Gift exchange does not increase effort, indicating that monitoring may be a better alternative in a mission-oriented setup.

To assess the generalizability of the results obtained in the lab, a field

[24] Note that the regression coefficients associated to these two treatments are not significantly different.

experiment with a similar design is conducted in Ouagadougou. In addition to the location, a difference between the lab and the field design is that subjects in the field are not aware they are taking part in an experiment. Nevertheless, the direction and the magnitude of treatment effects are congruent between the lab in Montreal and the field in Ouagadougou. Our lab results may therefore generalize to a field setting.

Note that several distinctive features of the mission-oriented sector have not been examined in this chapter, such as performance measurability issues, multi-principals, or multi-tasking (for a more comprehensive discussion, see e.g. Wright 2001 or Dixit 2002). More studies are still needed to better understand the interplay between psychological and pecuniary incentives in the mission-oriented sector, and to determine the conditions under which non-pecuniary incentives such as gift-exchange may have significant impacts on work effort in this sector.

Tables

Table 2.1 Penalties in the Monitoring Treatment		
If the difference is :	Lab (in EU)	Field (in FCFA)
Between 0 and 2 mistakes	0	0
Between 3 and 5 mistakes	50	1000
Between 6 and 9 mistakes	100	2000
10 mistakes or more	200	4000

Table 2.2 Penalties in the High Monitoring Treatment		
If the maximum difference is :	Lab (in EU)	Field (in FCFA)
Between 0 and 2 mistakes	0	0
Between 3 and 5 mistakes	100	2000
Between 6 and 9 mistakes	150	3000
10 mistakes or more	225	4500

Table 2.3 Average Number of Reported Mistakes									
		Treatments							
		Control		Monitoring		High Monitoring		Gift Exchange	
Paper Rank	Actual Number of Mistakes	Lab	Field	Lab	Field	Lab	Field	Lab	Field
1	15	12.194	11.976	13.473	12.817	13.875	13.045	12.968	12.150
2	24	18.984	17.634	20.764	19.378	20.125	19.545	19.419	17.077
3	6	5.887	5.524	5.964	5.744	5.656	5.932	5.774	5.375
4	25	18.323	16.317	20.527	18.207	20.156	18.477	19.290	16.250
5	16	12.597	11.683	13.236	12.659	13.094	12.932	12.935	11.975
6	5	4.145	4.049	4.436	3.976	4.094	4.000	4.194	3.650
7	20	15.855	14.195	17.182	15.024	16.281	15.432	16.000	14.150
8	11	9.435	8.793	10.000	9.037	9.563	9.023	10.129	8.800
9	19	15.403	14.366	16.418	15.378	15.906	14.727	15.645	14.150
10	14	11.790	11.220	12.256	11.707	11.969	11.773	11.710	10.875

Table 2.4
Average Absolute Deviation

	Treatments							
	Control		Monitoring		High Monitoring		Gift Exchange	
	Lab	Field	Lab	Field	Lab	Field	Lab	Field
Obs.	62	82	55	82	32	43	31	39
Mean	3.171	4.032	2.209	3.161	2.522	3.153	3.003	4.082
Median	2.800	3.650	2.200	3.200	2.500	2.800	2.800	4.000
Std. Dev.	1.543	1.804	1.157	1.293	1.501	1.933	1.453	2.06
Min	0.3	0.8	0.2	0.8	0.3	0.5	0.4	0.6
Max	7.4	9.3	5.5	6.3	5.7	10.9	6.8	11.2

Table 2.5
Subject Pool Characteristics

	Age		Gender	
	Lab	Field	Lab	Field
Average	26.612	24.665	0.444	0.129
Standard Deviation	6.549	2.271	0.498	0.336
Min	18	20	0	0
Max	54	30	1	1
No of Obs.	178	248	180	248

Table 2.6
GLS Regression of Grading Quality by Environment
Dependent Variable: Absolute Deviation

Independent Variables	Lab			Field			Pooled Data		
Age	0.044	***	(0.015)	0.011		(0.048)	0.040	***	(0.014)
Female	-0.713	***	(0.204)	-0.751	**	(0.332)	-0.719	***	(0.181)
Monitoring Treatment	-1.305	***	(0.335)	-1.328	***	(0.321)	-1.256	***	(0.275)
High Monitoring Treatment	-1.376	***	(0.361)	-1.490	***	(0.396)	-1.343	***	(0.336)
Gift Exchange Treatment	-0.132		(0.475)	-0.075		(0.479)	-0.030		(0.380)
Paper Ranking	-0.043		(0.029)	-0.034		(0.025)	-0.038	**	(0.019)
Paper Ranking*Monitoring Treatment	0.082	**	(0.042)	0.068	**	(0.031)	0.074	***	(0.025)
Paper Ranking*High Monitoring Treatment	0.123	***	(0.044)	0.111	***	(0.038)	0.116	***	(0.029)
Paper Ranking*Gift Exchange Treatment	0.037		(0.051)	0.004		(0.040)	0.018		(0.032)
Field							0.850	***	(0.287)
Field*Monitoring Treatment							-0.100		(0.337)
Field*High Monitoring Treatment							-0.221		(0.474)
Field*Gift Exchange Treatment							-0.143		(0.488)
Constant	1.841	***	(0.516)	3.120	***	(1.177)	1.725	***	(0.470)
Observations	1780			2478			4258		
Number of graders	178			248			426		
σ_u	1.209			1.608			1.448		

Notes: i) robust standard errors in parentheses, clustered in graders; ii) exam paper fixed effects included; iii) *** p<0.01, ** p<0.05, * p<0.1.

Table 2.7

GLS Regression of Grading Quality by Exams' Type

Dependent Variable: Absolute Deviation

Independent Variables	Actual Mistakes (15 or less)			Actual Mistakes (>15)		
Age	0.026	***	(0.010)	0.054	**	(0.021)
Female	-0.450	***	(0.117)	-0.989	***	(0.265)
Monitoring Treatment	-0.654	***	(0.207)	-1.941	***	(0.427)
High Monitoring Treatment	-0.790	***	(0.232)	-2.035	***	(0.555)
Gift Exchange Treatment	-0.028		(0.314)	0.029		(0.562)
Paper Ranking	-0.019		(0.018)	-0.230	***	(0.036)
Paper Ranking*Monitoring Treatment	0.037		(0.023)	0.128	***	(0.049)
Paper Ranking*High Monitoring Treatment	0.068	**	(0.028)	0.191	***	(0.059)
Paper Ranking*Gift Exchange Treatment	0.028		(0.031)	-0.003		(0.059)
Field	0.372	*	(0.192)	1.328	***	(0.407)
Field*Monitoring Treatment	-0.053		(0.225)	-0.147		(0.485)
Field*High Monitoring Treatment	-0.078		(0.318)	-0.359		(0.680)
Field*Gift Exchange Treatment	-0.174		(0.332)	-0.107		(0.712)
Constant	1.969	***	(0.334)	4.466	***	(0.702)
Observations	2129			2129		
Number of graders	426			426		
σ_u	0.852			2.078		

Notes: i) robust standard errors in parentheses, clustered in graders; ii) exam paper fixed effects included; iii) *** p<0.01, ** p<0.05, * p<0.1; iv) a paper should fail if the actual number of mistakes is strictly more than 15, it should pass otherwise.

Table 2.8 **GLS Regression of Grading Quality by Graders' Type** Dependent Variable: Absolute Deviation		
Independent Variables	Motivated Graders	Selfish Graders
Age	0.064 ** (0.025)	0.028 (0.021)
Female	-1.058 *** (0.279)	-0.548 ** (0.251)
Monitoring Treatment	-0.416 (0.361)	-1.926 *** (0.324)
High Monitoring Treatment	-0.599 (0.371)	-2.101 *** (0.360)
Gift Exchange Treatment	0.280 (0.433)	-0.104 (0.500)
Paper Ranking	0.028 (0.029)	-0.062 *** (0.024)
Paper Ranking*Monitoring Treatment	-0.021 (0.044)	0.117 *** (0.029)
Paper Ranking*High Monitoring Treatment	0.038 (0.043)	0.164 *** (0.038)
Paper Ranking*Gift Exchange Treatment	-0.018 (0.042)	0.011 (0.048)
Field	0.710 ** (0.309)	0.792 *** (0.236)
Constant	0.584 (0.795)	2.297 *** (0.610)
Observations	1620	1918
Number of graders	162	192
σ_u	1.531	1.474

Notes: i) robust standard errors in parentheses, clustered in graders; ii) exam paper fixed effects included; iii) *** p<0.01, ** p<0.05, * p<0.1.

Chapitre 3

Can corruption be studied in the lab?

3.1 Introduction

The micro-determinants of corruption, as well as possible anti-corruption measures have recently been studied in lab experiments conducted in developed countries. If shown to be externally valid (i.e. to be relevant for the real world), then lab experiments could become one of the most effective tools to study corruption. One may wonder, however, whether the lab provides an appropriate setting to study corruption. Indeed, although non-monetary considerations (e.g. moral, ethical, legal or cultural) may be major determinants of corruption, they may be difficult to capture in lab experiments. In the present chapter, we make an attempt at testing whether lab experiments conducted in developed countries can be used to study corruption where arguably it matters the most, the field in developing countries. To do so, we propose to compare the outcomes of a corruption experiment conducted in the lab in Montreal (Canada), and in the field in Ouagadougou (Burkina Faso).

There is ample evidence that corruption is now recognized as one of the most detrimental factors to economic and social development. First, corruption has become a prime concern for major international institutions. In particular, the International Monetary Fund considers that corruption "clearly is detrimental to economic activity and welfare".[1] Similarly, the World Bank declared that it "has identified corruption as the single greatest obstacle to economic and social development".[2] Second, several countries have intensified their fight against corruption. This may be best exemplified with the case of China where for the first time a high ranking official (the previous head of the Food and Drug

[1] http://www.imf.org/external/np/exr/facts/gov.htm.

[2] http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPUBLICSECTORAND GOVERNANCE/EXTANTICORRUPTION/0,,menuPK:384461~pagePK:149018~piPK :149093~theSitePK:384455,00.html).

Administration) was executed on July 10, 2007 after admitting to taking bribes. Likewise, Hong Kong created a legal precedent by implementing a "guilty until proven innocent" approach toward individuals accused of corruption. Finally, as discussed in a subsequent section, the last two decades have witnessed a sharp increase in the production of academic work related to corruption, not only in economics but also in sociology, political science, and law.

There is no comprehensive or widely accepted definition of corruption. It is generally agreed to include such activities as bribery, embezzlement, fraud, nepotism, extortion, or influence peddling. Corruption, however, is occasionally interpreted in a broader sense to encompass any activity lacking integrity, virtue, or moral. The definition of corruption is also sensitive to cultural factors. What may be considered corruption in a country, may simply reflect politeness or traditional gift exchange in a different country. Corruption is not necessarily illegal. For instance, although legal in several countries, lobbying and political contributions are deemed corrupt by most. Likewise, corruption is not necessarily considered immoral. For instance, favoritism toward one's own kind may be perceived as natural and justified. Corruption may be collective or individual, organized or incidental, political or bureaucratic. The latter is the form generally considered in economics where corruption is typically defined as an abuse of public office for private gain. Most agree that although corruption occurs in all countries, it is more prevalent and damaging in developing countries. Corruption affects all sectors of the economy, from tax collection to public contracting, from the justice to the education or health systems. In the present chapter, we adopt the economists' definition of corruption, and we design an experiment related to bribery in the education system.

Due in part to its covered nature, the analysis of corruption has been challenging to economists. In fact, it may be argued that the theoretical and empirical approaches to corruption have been of limited practical impact to understand and combat corruption. Experimental economics, on the other hand, could become the most promising approach to understand the determinants of corruptibility, and test possible anti-corruption measures. Indeed, lab experiments offer the economist the possibility to overcome the unobservability of corrupt activities by generating hard data, while controlling both the environment and the characteristics of the subjects' population. The experimental literature on corruption, however, is in its infancy and its practical relevance will not be fully established as long as the question of external validity remains unaddressed. In other words, we need to test the extent to which the results of corruption experiments obtained in the lab can be extrapolated to real-life situations in the field. As argued by (e.g.) List (2006a) or Levitt and List (2007), a lab and a field experiment on corruption could produce different outcomes as some features of the experiment are unlikely to be exactly the same in the two environments. In particular, the nature of the game played, the stakes, or the subject pool may differ between the lab and the field. Testing the external validity of corruption experiments is therefore necessary and legitimate.

The present chapter may be considered a first step in this direction. Indeed, although the vast majority of experimental labs are located in developed countries, understanding and fighting corruption is considered most relevant to developing countries. It is therefore important to know whether the experimental results obtained in the lab in developed countries can be extrapolated to the field in developing countries. To address this question, we carried out a traditional lab experiment in Montreal, while in Ouagadougou, we conducted what Harrison and List (2004) refer to as a natural field experiment in which the subjects do not know they are participants in an experiment. Our goal is to test whether the direction and/or the magnitude of various treatment effects are comparable when the same experiment is conducted in the lab and in the field. Note that we are not trying to demonstrate that one environment is superior to the other for the analysis of corruption. Like Harrison and List (2004), we believe that the lab and the field can complement each other, even when they produce slightly different outcomes. Our research project may then be interpreted as an effort to identify the dimensions in which the results obtained in the two environments concur or differ when analyzing corruption. We may then understand better how the lab and the field may be used as complementary approaches to study the determinants of corruption, as well as anti-corruption measures.

Moving from the lab to the field without losing too much control is never an easy task. It is even more challenging in the case of corruption, as it is typically considered an illegal and immoral activity. To design our experiment we strived to minimize the possible losses of control under the constraints imposed by the lab and the field. The solution we propose aims at reproducing a corruption scenario in which a candidate offers a bribe to a grader in order to obtain a better grade. In short, subjects in the lab and in the field were asked to grade the same set of 20 exam papers in the same order. The 11^{th} paper came with a money offer and a message stating: "Please, find few mistakes in my exam paper". The key difference between the two environments is that subjects in the field were informed they participated in an experiment only after grading was completed. We conducted four different treatments, each in the lab and in the field, by varying successively i) the amount of the bribe, ii) the wage paid to graders, and iii) the level of monitoring and punishment.

The experimental results indicate that the probability to accept a bribe decreases with the grader's age, religiousness and ability at the grading task. Gender, however, seems to have no significant effect. Controlling for these individual characteristics, we find that increasing the wage paid to graders lowers their probability of accepting the bribe. Monitoring and possible sanctions appear to have no significant effect on the graders' acceptance behavior. The direction and the magnitude of the previous two treatment effects are statistically indistinguishable between the lab and the field. The two environments, however, differ in some dimensions. In particular, we find that increasing the amount of the bribe has no effect in the lab, while it exacerbates corruption in the field.

We also find that graders who accept the bribe are more likely to favor the briber by reporting fewer mistakes. Once we control for individual characteristics, we find that, unlike men, women respond to monitoring by failing the briber more often. In addition, graders who accept the bribe are more likely to help the briber when they receive a higher wage. These two treatment effects are similar in the lab and in the field. There is, however, a notable difference between the two environments: when the bribe is increased, accepters in the field are more likely to reciprocate by providing a passing grade to the briber. No such behavior is observed in the lab.

Finally, we find that accepting the bribe affects how subjects subsequently perform the grading task. Indeed, the grades reported for the remaining nine exam papers by subjects who accept the bribe are significantly less accurate and more inconsistent. In the High Monitoring treatment, however, subjects who accept the bribe do a better job at grading the last exam papers, possibly in an effort to lower their expected penalty. Once again, no significant difference between the lab and the field can be identified.

The remainder of the chapter is organized as follows. Section 2 briefly summarizes the theoretical, empirical and experimental literature related to corruption. The experimental treatments are presented in section 3. In section 4, we discuss the experimental design relative to the study of corruption. The experimental results are analyzed in section 5. Finally, we discuss in section 6 the practical implications of our results.

3.2 Literature Review

We provide here a short summary of the economic literature pertaining to corruption.[3] In doing so, we try to highlight the practical contributions, as well as the possible limitations of the theoretical, empirical, and experimental approaches.

3.2.1 Theoretical and Empirical Approaches

Most of the theoretical literature on corruption builds on Becker (1968) and Rose-Ackerman (1975), who were the first to analyze in a formal setting the economics of crime and corruption respectively. This literature is mainly based on resource allocation and contract theories (Jain 2001). The first major theoretical debate concerned whether corruption can increase welfare. While initially

[3]For more comprehensive surveys of the economic literature on corruption, see Bardhan (1997) or Jain (2001). For a survey focusing more specifically on the theoretical approach, see Aidt (2003). For reviews of the empirical literature, see Lanyi (2004), Dreher and Herzfeld (2005) or Seldadyo and Haan (2006). Finally, for surveys of laboratory experiments on corruption see Dusek, Ortmann and Lizal (2005), as well as Abbink (2006).

seen as "grease-in-the-wheel" (Lui 1985), corruption has been mostly found to reduce welfare by acting as "sand-in-the-wheel".[4] Another major branch of the theoretical literature is concerned with evaluating possible corruption deterrents such as higher wages (Becker and Stigler 1974, Mookherjee and Png 1995, Sosa 2004), increased monitoring (Besley and McLaren 1993, Mookherjee and Png 1995, Acemoglu and Verdier 1998), stiffer penalties (Besley and McLaren 1993, Mookherjee and Png 1995, Acemoglu and Verdier 1998), and increase competition between potential bribees (Rose-Ackerman 1978, Shleifer and Vishny 1993). Most of this literature did not produce unambiguous predictions. As a result, it may be argued that the theoretical approach has been of limited practical impact to combat corruption. In addition, balancing the monetary costs and benefits of corruption may be considered too limiting of an approach (see e.g. Bukovansky 2006). In particular, accounting for non-monetary factors, such as ethical, moral, cultural and religious factors, may be necessary to better understand corruptibility and devise effective anti-corruption measures.

Corruption may typically be considered an illicit and secretive activity. As a result, it is virtually impossible to observe and measure directly. This lack of hard data partially explains the absence of rigorous empirical analyses of corruption prior to the mid 90's. To circumvent the observability problem, surveys aimed at evaluating corruption perceptions have been conducted around the world since the mid 1980's. Starting with the pioneer work of Mauro (1995), and Knack and Keefer (1995), these surveys have been used by economists to study corruption, giving rise to a burgeoning empirical literature. A first

[4] See e.g. Shleifer and Vishny (1993), Banerjee (1997), Bliss and Di Tella (1997), or Ades and Di Tella (1999).

branch of this literature is concerned with the link between corruption and economic activity. The bulk of the research tends to conclude that corruption is harmful, as it appears to hinder growth (Méon and Sekkat 2005), widen income inequality (Gyimah-Brempong 2002), discourage investments (Wei 2000), and cause misallocations of public expenditures (Mauro 1997). A second branch of the empirical literature attempts to identify the causes of corruption. In particular, wealth and corruption have been found to be correlated, but the direction of the relationship appears to be ambiguous.[5] Corruption has also been found to be lower in countries open to foreign trade (Knack and Azfar 2003), with high human capital (Brunetti and Weder 2003), and a higher participation of women to the labour force (Swamy et al. 2001). Finally, several studies find that higher wages reduce corruption (e.g. van Rijckeghem and Weder 2001), while others find no support for this common hypothesis (Rauch and Evans 2000).

The empirical approach to corruption has been criticized on several fronts.[6] First and foremost, corruption indices are believed to suffer from significant measurement errors.[7] Second, the direction of causality, although often far from obvious, cannot be established with reduced form approaches. For instance, regression models cannot test whether corruption exacerbates poverty, or in contrast, whether corruption is more likely to breed in poor countries. Finally,

[5]While most studies find a negative correlation (Hall and Jones 1999, Fisman and Gatti 2002, Serra 2006), some find that wealth and corruption are positively related (e.g. Braun and Di Tella 2004, Frechette 2001).

[6]For a flavor of the debate pertaining to the measure of corruption with survey data, see Golden and Picci (2005) or Kaufmann, Kraay and Mastruzzi (2006).

[7]In particular, Johnston (2001) argues that perception based surveys are skewed toward petty bribery (as it is the most visible to the public), while the economic impact of grand corruption is much greater. Lanyi (2004) also notes that respondents may be reluctant to answer these surveys honestly if they have been directly involved in corrupt transactions. In particular, Murrell and Azfar (2005) estimate that roughly 40% of their sample was reticent in answering questions about corruption.

there is a glaring void in the empirical literature when it comes to testing anti-corruption measures.[8] This void may be explained in part by the lack of micro-level data.

3.2.2 Experimental Approach

The experimental approach to corruption is the most recent, with the first published paper dating back to the beginning of the decade (Frank and Schulze 2000). The bulk of corruption experiments has been conducted in the lab where two forms of corruption have been studied: embezzlement and bribery. While embezzlement experiments use dictator games to study corruptibility in a single decision making, bribery experiments build on the trust game literature to study corruption in a strategic environment.

These lab experiments produced the following results. Barr, Lindelow and Serneels (2004) as well as Jacquemet (2005) find a negative correlation between wages and corruption, while no such treatment effect has been detected in other studies (i.e. Frank and Schulze 2000, Abbink 2002, Schulze and Frank 2003). A deterrence effect is found in staff rotation (Abbink 2004), as well as in monitoring and punishment (Abbink, Irlenbusch and Renner 2002, Schulze and Frank 2003, Barr et al. 2004). The results in Frank and Schulze (2000), as well as Rivas (2006) suggest that women may be less corrupt than men, although Alatas et al. (2009a) suggest that the effect may depend on cultural factors. Likewise, tolerance with respect to corruption may differ across cultures (Cameron et al. 2006) and subject pools (Alatas et al. 2009b). Finally, the use of loaded instructions does not appear to generate a treatment effect

[8] There are a few notable exceptions including Brunetti and Weder (2003), Reinikka and Svensson (2004), or Ferraz and Finan (2008).

(Abbink and Hennig-Schmidt 2006, Barr et al. 2004).

We are only aware of two published field experiments directly related to corruption.[9] As argued below, this void may be explained by the difficulty in designing field experiments on corruption. The first study, conducted by Bertrand et al. (2006) with subjects applying for a driver's license in New Delhi, combines experimental and survey methods to test the "grease-the-wheel" hypothesis. The second study, conducted by Olken (2007) in Indonesia, analyzes the efficacy of audit and grass-roots participation as anti-embezzlement measures in road construction projects.[10]

Although still not fully mature, the experimental approach to corruption may be considered very promising. Indeed, experimental economics has proved to be most useful in three situations, which each applies to corruption. First, when naturally-occurring data are scarce, or do not vary along certain desired dimensions. This is the case with corruption which is difficult to observe directly, and which is rarely observed under different wage, monitoring or punishment regimes. In such cases, lab experiments enable the economist to generate data in order to identify the causal relationship between two variables of interest (e.g., between wages and corruption). Second, to identify the micro-determinants of corrupt behavior. Although crucial to understand and combat corruption, which micro-level factors explain corrupt behavior remain

[9]Some empirical studies have been conducted with data obtained after natural experiments (see e.g. Di Tella and Schargrodsky 2003, or Reinikka and Svensson 2004). We do not consider these to be field experiments as the researchers had no control over the design of the experiments. Recent unpublished field experiments on corruption include Castillo et al. (2008).

[10]Note that in both of these studies, corruption is not observed directly. As a consequence, alternative explanations of the results are possible. For instance, although they provide some evidence to the contrary, Bertrand et al. (2006) cannot exclude that, instead of resorting to corruption (as assumed by the authors), subjects simply exerted more effort to obtain their driver's licence.

essentially an open question that neither the theoretical nor the empirical approach has been able to address adequately. As the economist controls both the experimental design and the characteristics of the subjects' population, the lab provides a unique framework to identify the micro-determinants of behavior. Third, lab experiments have proved to be a useful first step in the area of policy-making when a trial-and-error approach is either too costly or impossible to implement in the field.[11] As argued by (e.g.) Dusek, Ortmann and Lizal (2005) and Abbink (2006), lab experiments could constitute a cost effective "wind tunnel" to test potential policies aimed at curbing corruption.

The experimental literature on corruption, however, is in its infancy, and its practical relevance will not be fully established as long as the question of external validity remains unaddressed. In other words, we need to evaluate the extent to which the results of corruption experiments obtained in the lab can be extrapolated to real-life situations in the field. There are at least four potential reasons why the two environments could produce different results.[12] First, the stakes in the lab might differ from those in the field. In lab experiments, the stakes are essentially limited to "free money" provided by the experimenter. In the field, one of the party is entitled to the money, and the stakes may not be purely monetary. For instance, corruption may lead to imprisonment, physical harm, and in the most extreme cases, death. Similarly, corruption in the field may have moral or social implications that may be difficult to replicate in the

[11] For examples of how experiments have helped in the design of various policies and markets, see Plott (1999), Roth (2002), Klemperer (2004), or Milgrom (2004).

[12] There are several examples in the literature showing that behavior in the lab does not necessarily extend to the field. For instance, List (2006b) finds significant discrepancies between the lab and the field when analyzing the behavior of sportscard dealers. Likewise, Henrich et al. (2001) find that behavior in ultimatum, dictator, and public goods games differs notably from the lab when conducted in the field in developing countries. See also, (e.g.) List (2006a) or Levitt and List (2007), for a discussion of potential factors that may explain why a lab and a field experiment may produce different outcomes.

lab. Second, the game played in the lab and in the field may be different. In the lab, corruption is often modeled as a one shot game played in a context-free environment between anonymous subjects. In the field, corruption may involve repeated interactions between parties who can identify each other, and whose decisions may have life long consequences. Third, the players may be different in the lab and the field. Although efforts have been made to extend the subjects' population, the roles in a lab experiment are typically assigned randomly to a self-selected group of students. In contrast, the distribution of roles in the field may be endogenous. For instance, it is possible that officials have learned to become corrupt, or that they have deliberately chosen their position because they are intrinsically more corruptible. Fourth, subjects in the lab know they are being scrutinized. As a result, Levitt and List (2007) argue that lab subjects may be inclined to make the "moral" choice when morality and wealth are competing objectives, as it is the case with corruption.

3.3 Experimental Treatments

The basic idea behind the experiment is to reproduce a corruption scenario in which a candidate proposes a bribe to a grader in order to obtain a better grade. As further explained in section 4, we concentrate on the graders' behavior in order to maintain as much control as possible over the experiment. In other words, although we have subjects acting as candidates, their role is essentially passive.

We conducted four different grading treatments. Each treatment was conducted both in the lab and in the field. In the Control treatment, subjects were paid a fixed amount for their grading, regardless of how they performed the task. In addition, graders in the Control treatment were offered a bribe. In the lab, the fixed amount was 250 Experimental Units (EU hereafter) and the bribe was 50 EU. The conversion rate in the lab was 1 $C\$ = 12$ EU. In the field, the wage was set at $5,000$ $FCFA$, and the bribe was set at $1,000$ $FCFA$.

The "High Wage" treatment is identical to the Control treatment except that the wage was 40% higher (i.e. $7,000$ $FCFA$ in the field and 350 EU in the lab). The "High Bribe" treatment is identical to the Control treatment except that the amount of the bribe was doubled (i.e. $2,000$ $FCFA$ in the field and 100 EU in the lab). Finally, the last treatment makes an attempt at studying the effect of monitoring. For practical and ethical reasons, we decided against confronting field graders who accepted the bribe. Instead, we introduced a mechanism aimed at monitoring the accuracy with which the graders perform their task. This indirect approach therefore makes it possible to detect and punish corrupt graders when they favor the briber. The monitoring mechanism was explained as follows. We told each grader that we would randomly pick and re-grade 5 of the 20 papers he spell-checked. Then, we would calculate the difference between the number of mistakes reported by the grader and the number of mistakes we found in the paper. Only the worst paper graded would be considered to determine the monetary penalty. More precisely, the penalty imposed in the field was $2,000$ $FCFA$ when the difference was between 3 and 5 mistakes, $3,000$ $FCFA$ when the difference was between 6 and 9 mistakes and $4,500$ $FCFA$ when the difference exceeded 10 mistakes. The penalties imposed in the lab were proportional.[13] Except for the risk of being penalized, the "High

[13]More precisely, the penalty imposed in the lab was 100 EU when the difference was between 3 and 5 mistakes, 150 EU when the difference was between 6 and 9 mistakes and, 225 EU when the difference exceeded 10 mistakes.

Monitoring" treatment is otherwise identical to the Control treatment.

In total, 166 (respectively 125) subjects participated in the four treatments conducted in the field (respectively in the lab). More precisely, in the field (lab), 37 (30) subjects participated in the Control treatment, 40 (31) in the High Wage treatment, 45 (32) in the High Bribe treatment, and 44 (32) in the High Monitoring treatment. On average, the total earning of a field grader (a lab grader) was $6,432.43$ $FCFA$ (33.24 $C\$$) in the Control treatment, $8,375.00$ $FCFA$ (36.44 $C\$$) in the High Wage treatment, $7,155.56$ $FCFA$ (41.31 $C\$$) in the High Bribe treatment, and $4,954.55$ $FCFA$ (23.98 $C\$$) in the High Monitoring treatment.[14]

3.4 Comments on the Experimental Design

To some extent, the design we proposed may be interpreted as the outcome of an optimization problem under constraints: we tried to minimize the possible losses of control when moving from the lab to the field, subject to the constraints imposed by the two environments. We now discuss some of the issues we faced when designing our experiment, and some of the solutions we implemented to address these issues.

Running a field experiment on corruption is complicated by the fact that corruption is an illegal activity. As a result, we must be careful not to ask field subjects to take actions for which they could be prosecuted. To circumvent this problem, we created in Ouagadougou a private and closed environment in which to observe behavior. A second constraint was to find a real life activity which may credibly lend itself to corruption in the field. As mentioned earlier,

[14]These amounts include the bribe.

corruption in education appears to be prevalent in Burkina Faso, and bribing a grader is not uncommon.[15] Another issue was to prevent contamination within the subject pool. In particular, we did not want the word to spread that the grading task was fake, or that bribes were present in the exam papers. We took several actions in order to mitigate this problem: i) we conducted the field experiment in a large city; ii) we tried to recruit a geographically diverse subject pool; iii) we conducted all the sessions within a ten day period; iv) between the sessions in which a bribe was offered, we conducted some additional sessions (not reported here) without bribe; and v) field subjects were informed that they participated in an experiment only after all sessions had been carried out.[16]

Our design also reflects some of the constraints imposed by traditional lab practices. In particular, we did not resort to the use of deception techniques (as defined by Hey 1998). This explains why, although their role is essentially passive, we used real candidates to type the text. As a result, lab graders knew that their decisions could truly impact the financial well being of other subjects. Likewise, following lab conventions we chose to preserve the subjects' anonymity, although it may be argued that the ability to identify the other party is a key feature of real life corruption.

We took a number of measures in an attempt to mitigate the possible losses of control when moving from the lab to the field. In particular, we decided to concentrate exclusively on the graders' decisions. As a result, we were able to control in both environments the amount of the bribe, the distribution of

[15] In fact, a Burkinabe's newspaper ("Le Pays") reported on March 7, 2006 that two students were caught in a bribery attempt comparable to the one in our experiment.

[16] Informal conversations during the debriefing sessions suggest that, until it was revealed to them, the wide majority of field subjects truly believed that they were hired for a real grading task.

mistakes in the 20 exam papers, and the number of mistakes in the bribe paper. We also decided to introduce the bribe with a short written message in order to prevent face-to-face communication and informal bargaining, which could have been influenced by the briber's personal characteristics (e.g. gender, ethnicity). Finally, we had to choose what information to provide the lab and the field graders about the consequences of reporting the bribe attempt. We decided to provide information that we felt was of comparable nature in both environments. Namely, we told the field graders that reporting the bribe would be punished by failure of the exam, and we told the lab graders that this would provide no payoff for the candidate.

Before we conclude this section, we must acknowledge that the experiment we designed does not allow us to tackle entirely the question of external validity for corruption experiments. In particular, as the subjects recruited in the field are not professional graders, we cannot test whether the endogeneity of the subject pool play a role in explaining corrupt behavior. Likewise, our design only allows us to analyze one side of the market, i.e. we observe the behavior of the bribees but not of the bribers. Finally, our one-shot game experiment may not fully capture corruption in the field, which may be learned and may involve repeated interactions. Nevertheless, following Levitt and List (2007), one may argue that our design includes one of the most important features for external validity. Namely, in contrast with the lab, subjects in the field acted without knowing they were participating in an experiment. As a result, our chapter may be considered an important first step in testing the external validity of corruption experiments.

3.5 Experimental Results

In this section, we present the results of the lab and field experiments. To ease the presentation, we divide the analysis in three parts. First, we analyze the grader's decision to accept or reject the bribe. Then, we look at the number of mistakes reported for paper 11, depending on whether or not the grader accepted the bribe. Finally, we test whether accepting the bribe affects the way subjects subsequently perform their grading task for the remaining nine exam papers.

3.5.1 Decision to Accept the Bribe

We start with a brief presentation of descriptive statistics. Then, we conduct an econometric analysis to identify treatment effects, and test for possible differences between the lab and the field. Table 3.1 gives the frequency of the bribe acceptance in the different treatments, both in the lab and in the field. Let us first concentrate on the results obtained in the lab. In the Control treatment, 67% of the subjects accepted the bribe. In other words, nearly one out of three graders essentially refused "free money" despite the fact that i) they did not incur any risk of being caught and ii) accepting the bribe had no negative externality on any other candidate. This rejection rate is slightly higher than in comparable lab experiments.[17] We conjecture two potential explanations for this result. First, unlike previous corruption experiments, our design requires real effort from the briber and the bribee. Second, since most graders are university students, they can personally relate to the grading task, and therefore,

[17] For instance, only 9.4% of the subjects in Frank and Schulze (2000) acted honestly even though corruption had a negative externality on an actual public entity (i.e. a film club). Likewise, the rejection rate was only 13.1% in Cameron et al. (2006) in an environment with negative externalities and possible punishment.

they may be less tolerant toward bribery. Table 3.1 also indicates that in the lab, increasing the wage reduces the probability of accepting the bribe, while monitoring the graders' work or proposing a higher bribe seem to leave this probability unchanged.

As for the field, we can see in Table 3.1 that the probability of acceptance in the Control treatment is slightly below 50%. In contrast with the lab, this relatively high rejection rate may not be attributable solely to intrinsic motivations. Indeed, it is unlikely that all field graders believed that they faced absolutely no risk of being caught. In addition, they may have been under the impression that helping the briber had negative externalities on the other candidates and/or on the institution which administered the exam. Table 3.1 also indicates that in the field, providing a higher wage and monitoring the graders' work seem to lower slightly the probability of accepting the bribe. In contrast, the bribe is accepted more often when the amount proposed is larger.

We now turn to the econometric analysis. To impose as little structure as possible, we adopt a semiparametric approach to model the grader's decision to accept the bribe. More specifically, we specify a binary response model of the form:

$$A_i = \mathbb{I}\left(\beta' X_i + U_i \geq 0\right) \tag{3.1}$$

where A_i is a dummy variable taking the value 1 when grader i accepts the bribe; X_i is a vector of explanatory variables; β is a vector of parameters to be estimated; $\mathbb{I}(\cdot)$ is the indicator function satisfying $\mathbb{I}(z) = 1$ when z is true and $\mathbb{I}(z) = 0$ otherwise; and U_i is an unobserved error term. To estimate β without imposing any distributional assumption on U, we adopt the smoothed maximum score estimator developed by Horowitz (1992) (see also Horowitz

1998 and 2002).[18] This estimator can accommodate arbitrary heteroskedasticity of unknown form, it is asymptotically normal and, under some smoothness conditions, its convergence rate can be made arbitrarily close to $\sqrt{N}$. In other words, in terms of the estimator theoretical properties, there is virtually no cost in using this semiparametric approach versus a more conventional parametric approach.

In addition to treatment and field dummies, we include several individual characteristics to estimate the model in (3.1). In particular, we control for the grader's age and gender. We also include a measure of the grader's religiousness. This was obtained from the post-experiment survey in which we asked the subjects how often they go to a church, a mosque or any other place of worship.[19] This variable has 5 categories, ranging from 0 (never) to 4 (every day). To capture a grader's ability at the grading task, we include two variables measuring the grader's precision and improvement over the first ten exam papers.[20] These measures are valid instruments since the graders (both in the

[18]If the distribution of U is normal (respectively logistic), then (3.1) is the traditional binary probit (respectively, logit) model. Both of these distributional assumptions, however, are rejected by our data. As shown by (e.g.) Horowitz (2002) such distributional mispecifications may lead to severely biased estimates of β. Therefore, we prefer to rely on a more robust semiparametric approach. A well known drawback of the smoothed maximum score estimator, however, is that it does not produce marginal effects. This is a second order issue here since, to address the questions raised in this chapter, we need to correctly identify treatment effects, not marginal effects.

[19]We did not anticipate the survey responses to play such a role in explaining behavior. In hindsight, we should have collected additional personal information about e.g. the subjects' wealth and occupation.

[20]To measure a grader's precision, we averaged over the first ten exam papers the absolute deviation between the number of mistakes he or she reported and the true number of mistakes. The variable "Precision" was then set equal to the opposite of this average. "Precision" is therefore a negative number, and a grader is considered more precise when this variable increases toward 0. To measure a grader's improvement, we calculated for each of the first ten exam papers the absolute percentage deviation between the number of mistakes he or she reported and the true number of mistakes. We then regressed for each grader the opposite of this variable on a constant and the exam number. The variable "Improvement" was then set equal to the estimated slope in this regression. A grader is therefore considered to have improved at the grading task when "Improvement" is positive and large.

field and in the lab) are unaware of the presence of corruption until they reach paper 11. Finally, the econometric model accounts for the time the grader took to complete the grading task.[21]

The descriptive statistics provided in Table 3.2 indicate that the individual characteristics vary markedly between the field and the lab. In particular, observe that although it took the field subjects longer to complete their task, they are slightly less accurate at grading the first 10 papers. Note also that the average lab grader is less religious, and more likely to be a woman. By controlling these individual characteristics, the econometric analysis allows us to disentangle intrinsic differences between the lab and the field, from differences in the composition of the subject pools. In other words, we will be able to test whether two individuals with identical observable characteristics behave differently in each environment. In addition, the econometric analysis will allow us to test which of these individual characteristics may be considered micro-determinants of corruption.

The results of the estimation are reported in Table 3.3.[22] In terms of individual characteristics, we find that an older, a more religious or a more precise grader is significantly less likely to accept the bribe.[23] It is worth noting that, to the best of our knowledge, this is the first study to identify religious fervor,

[21]The parameter β in (3.1) is only identified up to a multiplicative factor. Following the conventional approach, the scale normalization consists here in setting the parameter associated with one of the variables equal to 1. In the estimations that follow, the variable selected for the scale normalization is the time the grader took to complete the grading task.

[22]The estimates presented in the paper are bias corrected. To account for the finiteness of the sample, the standard deviations of the estimates, as well as the distributions of the test statistics have been evaluated by bootstrap.

[23]Note that these results do not appear to be specific to the environment in which the data was collected. In particular, although the value of the parameters are affected, the direction and the significance of the individual characteristics remain unchanged when we estimate the model with the data collected solely in the lab, or solely in the field.

and ability as micro-determinants of corruption. Table 3.3 also indicates that gender does not appear to influence significantly the decision to accept the bribe. This result is somewhat surprising as previous lab experiments suggest that women are less corruptible (Frank and Schulze 2000, and Rivas 2006). It also contrasts with a commonly held belief among practitioners that women are less susceptible than men to accept bribes.[24]

Before testing for possible differences between the lab and the field, we look at general treatment effects. We can see in Table 3.3 that, compared to the Control treatment, increasing the wage paid to graders reduces significantly the probability of accepting the bribe. This result is consistent with the lab experiments of Barr et al. (2004) and Jacquemet (2005), and with several empirical analyses.[25] It is also in line with the views of numerous practitioners and international institutions, who often recommend to pay civil servants up to, or even above their private sector alternative as a mean to deter corruption.[26] Table 3.3 also indicates that proposing a higher bribe and monitoring do not affect significantly the grader's decision to accept the bribe. The first result contrasts with previous lab experiments in which a positive relationship between the bribe level and corruption has been identified (Abbink et al. 2002, Jacquemet 2005). The lack of efficacy of monitoring is not that surprising. Indeed, recall that the type of monitoring we implemented was not aimed at

[24]For instance, the police department in Mexico City decided in 1999 to dispatch women traffic officers at sensitive intersections because they were deemed less corruptible than their male counterparts. See e.g. the August 15, 1999 New York Times article available online at http://query.nytimes.com/gst/fullpage.html?res=940CE7DA1239F936A2575BC0A96F958260&n=Top/Reference/Times%20Topics/Subjects/P/Police%20Brutality%20and%20Misconduct).

[25]See e.g. van Rijckeghem and Weber (2001) or Alt and Lassen (2003).

[26]Singapore and Hong Kong are often presented as successful examples of such a policy. Indeed, these countries are typically ranked among the least corrupt, and they are known to pay high salaries to their civil servants. In particular, the prime minister of Singapore is paid several times more than the U.S. president.

catching the graders who accept the bribe. Instead, it was designed to catch corrupt graders who reciprocate by giving a good grade to the briber.

To identify possible differences between the lab and the field, we included in the econometric model a dummy variable for each treatment conducted in the field. We only find a single significant difference between the two environments. Namely, a higher bribe increases the probability of acceptance in the field, while it appears to have no effect in the lab. This result may indicate that subjects have different price elasticities in the lab and in the field. It may also simply reflect a pure level effect. Indeed, although the bribe is raised by the same factor in the two environments, the amount of the raise is different in the field and in the lab. The lack of statistical differences between the two environments, also implies that the direction and the magnitude of the "High Wage" treatment effect are the same in the lab and in the field. This result is remarkable as it suggests that, at least in some dimensions, corruption experiments conducted in the lab in a developed country and in the field in a developing country are fully consistent. The econometric analysis also reveals that the differences in the acceptance rate between the lab and the field (see Table 3.1) can be explained in large part by the composition of the subject pool in each environment. In other words, once we control for the subjects' observable characteristics, there does not seem to be any genuine difference between the two environments, except in the "High Bribe" treatment.

3.5.2 Decision to Report a Failing Grade

Do graders who accept the bribe tend to favor the briber? To address this question, we concentrate here on the most relevant decision made by the grader:

whether or not to report a failing grade (i.e. more than 15 mistakes) for paper 11.[27] The descriptive statistics in Table 3.1 indicate that graders who accept the bribe tend to fail paper 11 less often, regardless of the treatment or environment. Observe, however, that they do not systematically report a passing grade. Therefore, a non-negligible number of graders act opportunistically by taking the bribe and doing nothing in return.

To confirm that corrupt graders favor the briber, we estimate a binary response model similar to (3.1) in which the dependent variable takes the value 1 when the number of mistakes reported by the grader for paper 11 exceeds 15 mistakes. The results are reported in Table 3.4.[28] Observe first that after controlling for individual characteristics and treatment effects, we can confirm that graders who accept the bribe are significantly less likely to fail the bribe paper, and that their behavior does not vary significantly between the lab and the field.

In terms of the influence of individual characteristics, we find that, all else equal, women and more religious graders are less likely to fail the briber. In contrast, despite playing a role in explaining the decision to accept the bribe, the age of the grader appears to have no effect. Finally, and not surprisingly, the probability to find more than 15 mistakes in paper 11 is positively correlated

[27] An analysis based on the actual grade reported for paper 11 yields essentially the same conclusions.

[28] The econometric models for the decision to accept the bribe and the decision to report a failing grade have been estimated separately. Although this is consistent with the timing of the events, it is also conceivable that a grader checked the bribe paper before accepting the bribe. To test this hypothesis we estimated a bivariate binary response model in which the two decisions are modeled jointly. We find that the correlation between the error terms in each equation is not significantly different from zero, thereby rejecting the hypothesis of a joint decision. This does not imply, however, that we consider the two decisions to be uncorrelated. Indeed, by including a dummy for the bribe acceptance when estimating the decision to report a failing grade, we are only imposing that the error terms in the two models are independent.

with the grader's overall ability. As we shall see in the results presented next, most of these individual effects are consistent across graders, regardless of their decision to accept or reject the bribe.

No general treatment effect emerges from this econometric estimation with pooled data. As shown next, this may be explained by the fact that the treatments essentially affected the subjects who accepted the bribe. Observe, however, that even after controlling for their initial grading precision, women are significantly more likely to fail paper 11 in the High Monitoring treatment. This result is consistent with Frank and Schulze (2000) and Schulze and Frank (2003), who find that women are more responsive to monitoring and punishment. Note also that this behavior is robust as it is not significantly affected by the environment in which the experiment was conducted, and, as shown below, by the grader's decision to accept or reject the bribe.

To gain a better understanding of behavior, we now divide our sample in two groups depending on whether or not the grader accepted the bribe.[29] We can see in Table 3.5 that the estimation results obtained for rejecters are consistent with those just presented.[30] In contrast, the behavior of accepters seems to be influenced by different factors. In particular, among accepters, an older more

[29] We refer to graders who accepted the bribe as "accepters", and graders who reported the bribe as "rejecters". Note also that by splitting our sample we may introduce unobserved heterogeneity. Recall, however, that our semi-parametric estimation method can accommodate arbitrary heteroskedasticity of unknown form. In fact, we find essentially the same results when the models for accepters and rejecters are estimated jointly.

[30] The results obtained for rejecters should be interpreted with some degree of caution. Indeed, although they were instructed to grade paper 11 like any other exam paper, subjects both in the lab and in the field were informed that reporting the bribery attempt would result in failure for the candidate. One may therefore wonder whether the rejecters gave their best effort when grading the bribe paper. Two pieces of evidence seem to refute this hypothesis. First, an econometric analysis indicates that rejecters graded paper 11 with similar accuracy as the other first 10 papers. Second, the grades reported by rejecters for paper 11 in the Control treatment are not significantly different from those reported in an additional treatment we conducted in which no bribe was provided.

able male is more likely to fail the briber. In terms of treatment effects, we find that providing a higher wage decreases the probability that an accepter reports a failing grade. In other words, although graders are less likely to take the bribe when they receive a higher wage, those who accept tend to reciprocate more often by giving the briber a passing grade. Once again, the parameter associated with the "High Wage" dummy variable in the field is not significant. In other words, the direction and the magnitude of the "High Wage" treatment effect is statistically indistinguishable between the lab and the field. The only significant difference between the two environments is related to the effect of a higher bribe. Indeed, compared to the Control treatment, accepters in the field are more likely to reciprocate by providing a passing grade to the briber, while increasing the bribe does not influence the grade reported by accepters in the lab.

3.5.3 Corruption and Subsequent Performance

We now test whether accepting the bribe affects how well a subject subsequently grades the remaining nine exam papers. To do so, we exploit the panel structure of the data collected in the experiment to estimate a model of the form:

$$Y_{i,t} = \alpha' X_{i,t} + U_{i,t} \tag{3.2}$$

where the grading quality is defined as $Y_{i,t} = -|M_{i,t} - M_{0,t}|$, $M_{i,t}$ is the number of mistakes reported by subject i for exam paper t ($t = 12, ..., 20$), and $M_{0,t}$ is the actual number of mistakes in exam paper t.[31] The vector of explanatory variables $X_{i,t}$ includes the two variables measuring the initial ability of grader i

[31] To eliminate possible "exam paper" specific effects, the dependent variable $Y_{i,t}$ has been centered by subtracting its mean calculated over all graders.

over the first ten exam papers (i.e. "Precision" and "Improvement" defined in footnote 20). As we shall see, these variables capture most of the variation in grading quality across subjects.[32] To test whether grader i keeps improving as he or she did over the first ten exam papers, we control for i) the exam number t, and ii) the exam number multiplied by the value of variable "Improvement" for subject i. We also include in $X_{i,t}$ the time the subject took to complete the grading task, as well as various dummy variables for the decision to accept the bribe, the decision to fail the bribe paper, the environment, and the treatments. Finally, to control for possible grader specific random effects, we model the error term as $U_{i,t} = \varepsilon_i + V_{i,t}$, where $Var(V_{i,t}) = \sigma^2$, $Var(\varepsilon_i) = \sigma_a^2$ when subject i is an accepter and $Var(\varepsilon_i) = \sigma_r^2$ when subject i is a rejecter.

The results reported in Table 3.6 indicate that the parameters associated with the variables "Precision" and, to a lesser extent, "Improvement" are highly significant. This therefore confirms that most of the variation in grading quality over last nine papers may be explained by the subjects' initial abilities. The trend parameter is negative and significant, thereby indicating an overall decline in grading quality over the last 9 exam papers. The parameter associated with the variable "Improvement * t" is close to, but significantly greater than zero. In other words, we find a persistence in improvement, whereby (all else equal) subjects whose initial grading improved (respectively deteriorated) over the first 10 exam papers, keep improving (deteriorating) after the bribe paper. Once we control for differences in initial ability, we find that subjects who accept the bribe are significantly less precise when grading the last nine

[32] In particular, the estimations of alternative specifications indicate that accounting for "Precision" and "Improvement" is sufficient to capture general treatment effects, individual characteristics, as well as general differences between the lab and the field. These variables are therefore not included in the model estimated in this section.

exam papers. This lack of precision is even more pronounced among accepters who gave the briber a passing grade. Moreover, the standard error of the individual specific effect is significantly larger for the accepters than for the rejecters (i.e. $\widehat{\sigma}_a > \widehat{\sigma}_r$). In other words, the grading of accepters, and more specifically accepters who helped the briber, becomes more inconsistent and less accurate. To explain this result, we conjecture that accepters may prefer to appear incompetent rather than corrupt. Observe also that accepters do a significantly better grading job over the last nine exam papers in the High Monitoring treatment. In other words, it appears that accepters best respond to monitoring in an effort to lower their expected penalty.[33] Finally, we are once again unable to detect any significant difference between the lab and the field. Indeed, none of the parameters associated with variables controlling for the environment are significantly different from zero.

3.6 Conclusion

As argued by several international institutions (e.g. the IMF or the World Bank), corruption is one of the most detrimental factors currently afflicting the economies of developing countries. Due in part to its secretive nature, economists have had limited success in their effort to understand and combat corruption. Recently, the micro-determinants of corruption, as well as possible anti-corruption measures have been tested in laboratory experiments conducted in developed countries. If shown to be externally valid (i.e. to be relevant for

[33] We are unable to detect a significant difference between the accepters who did or did not fail the briber. This result may be partially explained by the fact that accepters tend to report fewer mistakes for the bribe paper, even when they fail the briber. Doing a better grading job over the last nine exam papers is therefore a best response for both kinds of accepters.

the real world), then laboratory experiments could become one of the most effective tools to study corruption. One may wonder, however, whether the insights gained in the lab in a developed country can be extrapolated to where it matters the most, the field in a developing country.

In an attempt to address this question, we conducted the same corruption experiment in the lab in Montreal (Canada), and in the field in Ouagadougou (Burkina Faso). The key difference between the two environments is that subjects in the field acted without knowing they were participating in an experiment. In short, our design aimed at reproducing a corruption scenario in which a candidate proposes a bribe to a grader in order to obtain a better grade. We conducted four different treatments, each in the lab and in the field, by varying successively i) the amount of the bribe, ii) the wage paid to graders, and iii) the level of monitoring and punishment.

An econometric analysis of the data collected in the lab and in the field reveals several micro-determinants of corrupt behavior. In particular, we find that the probability to accept a bribe decreases with the grader's age, religiousness and ability at the grading task. In addition, our results suggest that women may be more responsive to monitoring and punishment. To the best of our knowledge, this is the first study to identify religiousness and ability as micro-determinants of corruption.

Once we control for these individual characteristics, we find the direction and the magnitude of several treatment effects to be statistically indistinguishable between the lab and the field. In particular, increasing the grader's wage reduces the probability that he will accept the bribe in both environments. In other words, we do not identify any intrinsic difference between the two environments, in the sense that the behaviors of two individuals with identical observable characteristics are not statistically different in the field and in the lab. This is encouraging as it suggests that, at least in some dimensions, the results of corruption experiments conducted in the lab in a developed country carry over to the field in a developing country.

Tables

Table 3.1 Descriptive Statistics									
Treatment		Control		High Wage		High Bribe		High	
Environment*		Lab	Field	Lab	Field	Lab	Field	Lab	Field
# of subjects		30	37	31	40	32	45	32	44
% of graders who took the bribe		0.667	0.487	0.484	0.375	0.656	0.689	0.656	0.409
% of graders who reported more than 15 mistakes for Exam 11	All	0.667	0.649	0.548	0.436	0.656	0.422	0.656	0.698
	Accepters	0.600	0.556	0.400	0.267	0.619	0.355	0.667	0.667
	Rejecters	0.800	0.737	0.688	0.542	0.727	0.571	0.636	0.720

* The lab experiment was conducted in Montreal (Canada), and the field experiment in Ouagadougou (Burkina Faso).

Table 3.2 Subject Pool Characteristics										
	Age		Female		Religiousness		Time (in Min)		Ability	
	Lab	Field	Lab	Field	Lab	Field	Lab	Field	Lab	Field
Average	26.26	24.86	0.41	0.16	0.83	2.68	100.21	140.74	2.96	3.81
Std Dev	6.32	2.24	0.49	0.36	1.06	1.2	17.3	24.93	1.52	1.94
Min	18	20	0	0	0	0	50	70	0.3	0.5
Max	54	33	1	1	3	4	160	223	7.4	11.2

Table 3.3
Binary Response Model for the Decision to Accept the Bribe

Independent Variables	Coefficients		
Age	-0.689	**	(0.349)
Female	-0.752		(0.734)
Religiousness	-1.128	***	(0.437)
Precision	-0.943	**	(0.478)
Improvement	-0.455	**	(0.437)
High Wage Treatment	-1.382	**	(0.684)
High Bribe Treatment	-0.686		(0.653)
High Monitoring Treatment	-0.432		(1.116)
Field* Control Treatment	-1.103		(0.749)
Field*High Wage Treatment	-0.560		(0.694)
Field*High Bribe Treatment	1.431	**	(0.726)
Field*High Monitoring Treatment	0.130		(0.925)
Constant	0.901		(0.667)
Number of Obs.	291		

Significance: *** p<0.01, ** p<0.05, * p<0.1.

Table 3.4
Binary Response Model for the Decision to Report a Failing Grade
(Model Estimated with All Subjects)

Independent Variables	Coefficients		
Age	0.093		(0.408)
Female	-1.932	***	(0.542)
Religiousness	-0.679	**	(0.234)
Precision	-0.692	**	(0.364)
Improvement	0.006	**	(0.285)
Accept	-1.252	**	(0.551)
Accept*Field	0.264		(0.632)
High Wage Treatment	-0.203		(0.527)
High Bribe Treatment	0.229		(0.618)
High Monitoring Treatment	-0.465		(0.546)
Field* Control Treatment	-0.292		(0.695)
Field*High Wage Treatment	-0.708		(0.636)
Field*High Bribe Treatment	-0.134		(0.779)
Field*High Monitoring Treatment	0.584		(0.790)
Female*High Monitoring	2.067	**	1.048
Female*Field*High Monitoring	2.234		1.709
Constant	1.341	**	(0.563)
Number of Obs.	291		

Significance: *** p<0.01, ** p<0.05, * p<0.1.

Table 3.5
Binary Response Model for the Decision to Report a Failing Grade
(Model Estimated with Accepters or Rejecters)

Independent Variables	Accepters	Rejecters
Age	1.275 ** (0.540)	-0.208 (0.366)
Female	-3.123 *** (0.886)	-3.469 * (2.147)
Religiousness	0.301 (0.366)	-1.182 ** (0.454)
Precision	1.572 *** (0.402)	1.139 ** (0.573)
Improvement	0.593 ** (0.265)	0.047 (0.532)
High Wage Treatment	-2.679 ** (0.828)	-1.818 (2.173)
High Bribe Treatment	0.653 (0.752)	-1.338 (2.435)
High Monitoring Treatment	-1.116 (0.825)	-4.562 (3.072)
Field* Control Treatment	-0.356 (0.660)	-2.170 (2.959)
Field*High Wage Treatment	0.500 (2.821)	-2.355 (2.146)
Field*High Bribe Treatment	-1.599 ** (0.745)	-1.387 (2.502)
Field*High Monitoring Treatment	1.413 (1.628)	0.383 (0.762)
Female*High Monitoring	3.990 * (2.277)	4.596 * (2.846)
Female*Field*High Monitoring	-0.537 (2.446)	-
Constant	1.948 *** (0.605)	4.864 * (2.888)
Number of Obs.	159	132

Significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 3.6
Grading Quality after the Bribe

Independent Variables	Coefficients		
Precision	0.598	***	(0.084)
Improvement	0.347	**	(0.177)
t	0.090	**	(0.042)
Improvement*t	0.055	**	(0.020)
Time	-0.282		(0.367)
Accept	-1.684	***	(0.404)
Accept*Pass	-0.785	***	(0.252)
Accept*High Wage Treatment	0.575		(0.369)
Accept* High Bribe Treatment	0.454		(0.336)
Accept*High Monitoring Treatment	-0.934	**	(0.365)
Field	-0.472		(0.358)
Field*Accept	0.172		(0.352)
Field*Pass	-0.045		(0.242)
Constant	4.378	***	(0.773)
σ	2.457	***	(0.110)
σ_accepters	1.219	***	(0.106)
σ_rejecters	0.930	***	(0.128)
Number of Obs.	2619		
Number of Graders	291		

Notes: i)*** p<0.01, ** p<0.05, * p<0.1; ii) the number of observations is the number of graders multiplied by 9, the number of papers after paper 11.

Bibliographie

Introduction Générale

Abbink, K., and H. Hennig-Schmidt (2006): "Neutral Versus Loaded Instructions in a Bribery Experiment," *Experimental Economics*, 9, 103-121.

Alatas, V., L. Cameron, A. Chaudhuri, N. Erkal, and L. Gangadharan (2009): "Subject Pool Effects in a Corruption Experiment: A Comparison of Indonesian Public Servants and Indonesian Students," *Experimental Economics*, 12(1), 113-132.

Bardsley, N. (2005): "Experimental Economics and the Artificiality of Alteration," *Journal of Economic Methodology*, 12, 239-251.

Bateman, I., A. Munro, B. Rhodes, C. Starmer, and R. Sugden (1997): "Does Part-Whole Bias Exist? An Experimental Investigation," *Economic Journal*, 322-332.

Campbell, D., and J. Stanley (1963): *Experimental and Quasi-Experimental Designs for Research.* Rand McNally Chicago.

Dickinson, D., and M. C. Villeval (2007): "Does Monitoring Decrease Work Effort? The Complementarity between Agency and Crowding-out Theories," *Games and Economic Behavior*, 63, 56-76.

Fehr, E., and J. A. List (2004): "The Hidden Costs and Returns of Incentives - Trust and Trustworthiness among CEOs," *Journal of the European Economic Association*, 2, 743-771.

Guala, F. (2002): "On the Scope of Experiments in Economics: Comments on Siakantaris," *Cambridge Journal of Economics*, 26, 261-267.

Harrison, G. W., and J. A. List (2004): "Field Experiments," *Journal of Economic Literature*, 42, 1009-1055.

Harrison, G. W., and J. A. List (2008): "Naturally Occurring Markets and Exogenous Laboratory Experiments: A Case Study of the Winner's Curse," *Economic Journal*, 118, 822-843.

Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, and R. McElreath (2001): "In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies," *American Economic Review*, 91, 73-78.

Kagel, J. H., and A. E. Roth (1995): *Handbook of Experimental Economics.* Princeton, NJ: Princeton University Press.

Kahneman, D., J. L. Knetsch, and R. H. Thaler (1990): "Experimental Tests of the Endowment Effect and the Coase Theorem," *Journal of Political Economy*, 98, 1325-1348.

Levitt, S. D., and J. A. List (2007a): "Viewpoint: On the Generalizability of Lab Behaviour to the Field," *Canadian Journal of Economics*, 40, 347-370.

Levitt, S. D., and J. A. List (2007b): "What Do Laboratory Experiments Measuring Social Preferences Tell Us About the Real World," *Journal of Economic Perspectives*, 21, 153-174.

List, J. A. (2006): "The Behavioralist Meets the Market: Measuring Social Preferences and Reputation Effects in Actual Transactions," *Journal of Political Economy*, 114, 1-37.

Loewenstein, G. (1999): "Experimental Economics from the Vantage-Point of Behavioural Economics," *Economic Journal*, 109, 25-34.

Rutström, E. (1998): "Home-Grown Values and Incentive Compatible Auction Design," *International Journal of Game Theory*, 27, 427-441.

Schram, A. (2005): "Artificiality: The Tension between Internal and External Validity in Economic Experiments," *Journal of Economic Methodology*, 12, 225-237.

Smith, V. L. (1982): "Microeconomic Systems as an Experimental Science," *American Economic Review*, 72, 923-955.

Thurstone, L.L. (1931): "The Indifference Function," *Journal of Social Psychology*, 2, 139-167.

van Dijk, F., J. Sonnemans, and F. van Winden (2001): "Incentive Systems in a Real Effort Experiment," *European Economic Review*, 45(2), 187-214.

Wallis, W.A., and M. Friedman (1942): "The Empirical Derivation of Indifference Functions", in *Studies in Mathematical Economics and Econometrics in Memory of Henry Schultz* ed. by O. Lange, F. McIntyre, and T.O. Yntema, 175-189. Chicago, University of Chicago Press.

Chapitre 1

Antonovics, K., P. Arcidiacono, and R. Walsh (2009): "The Effects of Gender Interactions in the Lab and in the Field," *Review of Economics and Statistics*, 91, 152-162.

Brookshire, D., D. Coursey, and W. Schulze (1987): "The External Validity of Experimental Economics Techniques: Analysis of Demand Behavior," *Economic Inquiry*, 25, 239-250.

Friedman, D. and S. Sunder (1994): *Experimental Methods: A Primer for Economists.* Cambridge University Press, UK.

Harrison, G. W., and J. A. List (2004): "Field Experiments," *Journal of Economic Literature*, 42, 1009-1055.

Hennig-Schmidt, H., B. Rockenbach, and A. Sadrieh (2008): "In Search of Workers' Real Effort Reciprocity - A Field and a Laboratory Experiment," Forthcoming in *Journal of the European Economic Association.*

Levitt, S. D., and J. A. List (2007a): "Viewpoint: On the Generalizability of Lab Behaviour to the Field," *Canadian Journal of Economics*, 40, 347-370.

Levitt, S. D., and J. A. List (2007b): "What Do Laboratory Experiments Measuring Social Preferences Tell Us About the Real World," *Journal of Economic Perspectives*, 21, 153-174.

Levitt, S. D., and J. A. List (2008): "Field Experiments in Economics: The Past, the Present, and the Future," *European Economic Review*, 53(1), 1-18.

List, J. A. (2006): "Field Experiments: A Bridge between Lab and Naturally Occurring Data," *Advances in Economic Analysis & Policy*, 6, Article 8.

Rondeau, D., and J. A. List (2008): "Matching and Challenge Gifts to Charity: Evidence from Laboratory and Natural Field Experiments," *Experimental*

Economics, 11, 253-267.

Chapitre 2

Akerlof, G. A. (1982): "Labor Contracts as Partial Gift Exchange," *Quarterly Journal of Economics*, 97(4), 543-569.

Alchian, A. A., and H. Demsetz (1972): "Production, Information Costs, and Economic Organization," *American Economic Review*, 62(5), 777-795.

Al-Ubaydli, O., S. Andersen, U. Gneezy, and J. A. List (2008): "For Love or Money? Comparing the Effects of Non-Pecuniary and Pecuniary Incentive Schemes in the Workplace," Working Paper, George Mason University.

Banfield, E. C. (1975): "Corruption as a Feature of Governmental Organization," *Journal of Law and Economics*, 18, 587-605.

Benz, M., and S. Meier (2008): "Do People Behave in Experiments as in the Field? Evidence from Donations," *Experimental Economics*, 11, 268-281.

Besley, T. J., and M. Ghatak (2003): "Incentives, Choice, and Accountability in the Provision of Public Services," *Oxford Review of Economic Policy*, 19, 235-249.

Besley, T. J., and M. Ghatak (2005): "Competition and Incentives with Motivated Agents," *American Economic Review*, 95, 616-636.

Burgess, S., C. Propper, M. Ratto, and E. Tominey (2004): "Incentives in the Public Sector: Evidence from a Government Agency," Working Paper, Leverhulme Centre for Market and Public Organisation.

Dickinson, D. L. (2001): "The Carrot vs. the Stick in Work Team Motivation," *Experimental Economics*, 4(1), 107-124.

Dixit, A. (2002): "Incentives and Organizations in the Public Sector: An Interpretative Review," *Journal of Human Resources*, 37, 696-727.

Engelmann, D., and A. Ortmann (2002): "The Robustness of Laboratory Gift Exchange: A Reconsideration," Working Paper, CERN-EI.

Falk, A., and M. Kosfeld (2006): "Distrust - The Hidden Cost of Control," *American Economic Review*, 96(5), 1611-1630.

Fehr, E., and A. Falk (2002): "Psychological Foundations of Incentives," *European Economic Review*, 46(4-5), 687-724.

Fehr, E., and S. Gächter (2002): "Do Incentive Contracts Crowd Out Voluntary Cooperation?," Working Paper No. 34, Institute for Empirical Research in Economics, University of Zurich.

Fehr, E., and B. Rockenbach (2003): "Detrimental Effects of Sanctions on Human Altruism," *Nature*, 422, 137-140.

Frey, B. S. (1993): "Does Monitoring Increase Work Effort? The Rivalry with Trust and Loyalty," *Economic Inquiry*, 31, 663-670.

Frey, B. S., and R. Jegen (2001): "Motivation Crowding Theory," *Journal of Economic Surveys*, 15(5), 589-611.

Gneezy, U., and J. A. List (2006): "Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments," *Econometrica*, 74(5), 1365-1384.

Gneezy, U., and A. Rustichini (2000): "A Fine Is a Price," *Journal of Legal Studies*, 19, 1-18.

Hennig-Schmidt, H., B. Rockenbach, and A. Sadrieh (2008): "In Search of Workers' Real Effort Reciprocity - A Field and a Laboratory Experiment," Forthcoming in *Journal of the European Economic Association.*

Levitt, S. D., and J. A. List (2007a): "Viewpoint: On the Generalizability of Lab Behaviour to the Field," *Canadian Journal of Economics*, 40, 347-370.

Levitt, S. D., and J. A. List (2007b): "What Do Laboratory Experiments Measuring Social Preferences Tell Us about the Real World," *Journal of Economic Perspectives*, 21(2), 153-174.

List, J. A. (2006): "The Behavioralist Meets the Market: Measuring Social Preferences and Reputation Effects in Actual Transactions," *Journal of Political Economy*, 114, 1-37.

Nagin, D. S., J. B. Rebitzer, S. Sanders, and L. J. Taylor (2002): "Monitoring, Motivation, and Management: The Determinants of Opportunistic Behavior in a Field Experiment," *American Economic Review*, 92(4), 850-873.

Rainey, H. G., and P. Steinbauer (1999): "Galloping Elephants: Developing Elements of a Theory of Effective Government Organizations," *Journal Public Administration Research Theory*, 9, 1-32.

Ryan, R. M., and E. L. Deci (2000): "Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions," *Contemporary Educational Psychology*, 25(1), 54-67.

Schulze, G. G., and B. Frank (2003): "Deterrence versus Intrinsic Motivation: Experimental Evidence on the Determinants of Corruptibility," *Economics of Governance*, 4, 143-160.

Sheehan, R. M. (1996): "Mission Accomplishment as Philanthropic Organization Effectiveness: Key Findings from the Excellence in Philanthropy Project," *Nonprofit and Voluntary Sector Quarterly*, 25, 110-123.

Tirole, J. (1994): "The Internal Organization of Government," *Oxford Economic Papers*, 46, 1-29.

Wilson, J. Q. (1989): *Bureaucracy: What Government Agencies Do and Why They Do It.* New York: Basic Books.

Wright, B. E. (2001): "Public-Sector Work Motivation: A Review of the Current Literature and a Revised Conceptual Model," *Journal of Public Administration Research and Theory*, 11, 559-586.

Chapitre 3

Abbink, K. (2002): "Fair Salaries and the Moral Costs of Corruption," CeDEx Working Paper 2002-5, University of Nottingham.

Abbink, K. (2004): "Staff Rotation as an Anti-Corruption Policy: An Experimental Study," *European Journal of Political Economy*, 20(4), 887-906.

Abbink, K. (2006): "Laboratory Experiments on Corruption," in *The Handbook of Corruption*, ed. by S. Rose-Ackerman. Edward Elgar Publishers, Cheltenham, UK, and Northampton, US.

Abbink, K., and H. Hennig-Schmidt (2006): "Neutral versus Loaded Instructions in a Bribery Experiment," *Experimental Economics*, 9(2), 103-121.

Abbink, K., B. Irlenbusch, and E. Renner (2002): "An Experimental Bribery Game," *Journal of Law, Economics, and Organization*, 18(2), 428-454.

Acemoglu, D., and T. Verdier (1998): "Property Rights, Corruption and the Allocation of Talent: A General Equilibrium Approach," *Economic Journal*, 108(450), 1381-1403.

Ades, A., and R. Di Tella (1999): "Rents, Competition, and Corruption," *American Economic Review*, 89(4), 982-993.

Aidt, T. S. (2003): "Economic Analysis of Corruption: A Survey," *Economic Journal*, 113, 632-652.

Alatas, V., L. Cameron, A. Chaudhuri, N. Erkal, and L. Gangadharan (2009a): "Gender, Culture and Corruption: Insights from an Experimental Analysis," *Southern Economic Journal*, 75, 663-680.

Alatas, V., L. Cameron, A. Chaudhuri, N. Erkal, and L. Gangadharan (2009b): "Subject Pool Effects in a Corruption Experiment: A Comparison of Indonesian Public Servants and Indonesian Students," *Experimental Economics*, 12(1), 113-132.

Alt, J. E., and D. D. Lassen (2003): "The Political Economy of Institutions and Corruption in American States," *Journal of Theoretical Politics*, 15(3), 341-365.

Banerjee, A. V. (1997): "A Theory of Misgovernance," *Quarterly Journal of Economics*, 112(4), 1289-1332.

Bardhan, P. (1997): "Corruption and Development: A Review of Issues," *Journal of Economic Literature*, 35, 1320-1346.

Barr, A., M. Lindelow, and P. Serneels (2004): "To Serve The Community Or Oneself: The Public Servant's Dilemma," Working Paper, University of Oxford.

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

Becker, G. S., and G. J. Stigler (1974): "Law Enforcement, Malfeasance, and Compensation of Enforcers," *Journal of Legal Studies*, 3(1), 1-18.

Bertrand, M., S. Djankov, R. Hanna, and S. Mullainathan (2006): "Obtaining a Driving License in India: An Experimental Approach to Studying Corruption," *Quarterly Journal of Economics*, 122(4), 1639-1676.

Besley, T., and J. McLaren (1993): "Taxes and Bribery: The Role of Wage Incentives," *Economic Journal*, 103(416), 119-141.

Bliss, C., and R. Di Tella (1997): "Does Competition Kill Corruption?," *Journal of Political Economy*, 105(5), 1001-1023.

Braun, M., and R. Di Tella (2004): "Inflation, Inflation Variability, and Corruption," *Economics and Politics*, 16(1), 954-1985.

Brunetti, A., and B. Weder (2003): "A Free Press Is Bad News for Corruption," *Journal of Public Economics*, 87(7-8), 1801-1824.

Bukovansky, M. (2006): "The Hollowness of Anti-Corruption Discourse," *Review of International Political Economy*, 13, 181-209.

Cameron, L., A. Chaudhuri, N. Erkal, and L. Gangadharan (2006): "Propensities to Engage in and Punish Corrupt Behavior: Experimental Evidence from Australia, India, Indonesia and Singapore," Working Paper, University of Melbourne.

Castillo, M., R. Petrie, M. Torero, and A. Viceisza (2008): "Corruption in the Mail Sector: A Field Experiment in Peru," Working Paper, Georgia Institute of Technology.

Di Tella, R., and E. Schargrodsky (2003): "The Role of Wages and Auditing during a Crackdown on Corruption in the City of Buenos Aires," *Journal of Law and Economics*, 46(1), 269-292.

Dreher, A., and T. Herzfeld (2005): "The Economic Costs of Corruption: A Survey and New Evidence," Working Paper 0506001, Public Economics, EconWPA.

Dusek, L., A. Ortmann, and L. Lizal (2005): "Understanding Corruption and Corruptibility Through Experiments: A Primer," *Prague Economic Papers*, 14(2), 147-162.

Ferraz, C., and F. Finan (2008): "Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes," *Quarterly Journal of Economics*, 123, 703-745.

Fisman, R., and R. Gatti (2002): "Decentralization and Corruption: Evidence Across Countries," *Journal of Public Economics*, 83, 325-345.

Frank, B., and G. G. Schulze (2000): "Does Economics Make Citizens Corrupt?," *Journal of Economic Behavior & Organization*, 43(1), 101-113.

Frechette, G. R. (2001): "An Empirical Investigation of the Determinants of Corruption: Rent, Competition, and Income Revisited," Working Paper, Ohio State University.

Golden, M. A., and L. Picci (2005): "Proposal for a New Measure of Corruption, Illustrated with Italian Data," *Economics and Politics*, 17(1), 37-75.

Gyimah-Brempong, K. (2002): "Corruption, Economic Growth, and Income Inequality in Africa," *Economics of Governance*, 3(3), 183-209.

Hall, R. E., and C. I. Jones (1999): "Why Do Some Countries Produce So Much More Output Per Worker Than Others?," *Quarterly Journal of Economics*, 114(1), 83-116.

Harrison, G. W., and J. A. List (2004): "Field Experiments," *Journal of Economic Literature*, 42, 1009-1055.

Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, and R. McElreath (2001): "In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies," *American Economic Review*, 91(2), 73-78.

Hey, J. D. (1998): "Experimental Economics and Deception: A Comment," *Journal of Economic Psychology*, 19, 397-401.

Horowitz, J. (1992): "A Smoothed Maximum Score Estimator for the Binary Response Model," *Econometrica*, 60, 505-531.

Horowitz, J. (1998): "Bootstrap Methods for Median Regression Models," *Econometrica*, 66, 1327-1351.

Horowitz, J. (2002): "Bootstrap Critical Values for Tests Based on the Smoothed Maximum Score Estimator," *Journal of Econometrics*, 111, 141-167.

Jacquemet, N. (2005): "Corruption as Betrayal: Experimental Evidence on Corruption Under Delegation," Working Paper 0506, Groupe d'Analyse et de Theorie Economique (GATE).

Jain, A. K. (2001): "Corruption: A Review," *Journal of Economic Surveys*, 15, 71-121.

Johnston, M. (2001): "The Definitions Debate: Old Conflicts in New Guises," in *The Political Economy of Corruption*, ed. by A. K. Jain, pp. 11-31. Routledge, London.

Kaufmann, D., A. Kraay, and M. Mastruzzi (2006): "Measuring Corruption: Myths and Realities," Working Paper, World Bank.

Klemperer, P. (2004): *Auctions: Theory and Practice.* Princeton: Princeton University Press.

Knack, S., and O. Azfar (2003): "Trade Intensity, Country Size and Corruption," *Economics of Governance*, 4(1), 1-18.

Knack, S., and P. Keefer (1995): "Institutions and Economic Performance: Cross-Country Tests Using Alternative Institutional Measures," *Economics and Politics*, 7(3), 207-227.

Lanyi, A. (2004): "Measuring the Economic Impact of Corruption: A Survey," Working Paper 04/04, IRIS, University of Maryland.

Levitt, S., and J. A. List (2007): "What Do Laboratory Experiments Measuring Social Preferences Tell Us about the Real World," *Journal of Economic Perspectives*, 21(2), 153-174.

List, J. A. (2006a): "Field Experiments: A Bridge between Lab and Naturally Occurring Data," *Advances in Economic Analysis & Policy*, 6 (2).

List, J. A. (2006b): "The Behavioralist Meets the Market: Measuring Social Preferences and Reputation Effects in Actual Transactions," *Journal of Political Economy*, 114, 1-37.

Lui, F. T. (1985): "An Equilibrium Queuing Model of Bribery," *Journal of Political Economy*, 93(4), 760-781.

Mauro, P. (1995): "Corruption and Growth," *Quarterly Journal of Economics*, 110(3), 681-712.

Mauro, P. (1997): "The Effects of Corruption on Growth, Investment, and Government Expenditure: A Cross-Country Analysis," in *Corruption and the Global Economy*, ed. by K. A. Elliott, pp. 83-107. Institute for International Economics, Washington, D.C.

Milgrom, P. (2004): *Putting Auction Theory to Work.* Cambridge: University of Cambridge Press.

Méon, P. G., and K. Sekkat (2005): "Does Corruption Grease or Sand the Wheels of Growth?," *Public Choice*, 122(1), 69-97.

Mookherjee, D., and I. P. L. Png (1995): "Corruptible Law Enforcers: How Should They Be Compensated?," *Economic Journal*, 105(428), 145-159.

Murrell, P., and O. Azfar (2005): "Identifying Reticent Respondents: Assessing the Quality of Survey Data on Corruption and Values," Working Paper 05-001, University of Maryland.

Olken, B. A. (2007): "Monitoring Corruption: Evidence from a Field Experiment in Indonesia," *Journal of Political Economy*, 115(2), 200-249.

Plott, C. R. (1999): "Policy and the Use of Laboratory Experimental Methodology in Economics," in *Uncertain Decisions Bridging Theory and Experiments*, ed. by L. Luini, pp. 293-315. Kluwer Academic Publishers, Boston.

Rauch, J. E., and P. B. Evans (2000): "Bureaucratic Structure and Bureaucratic Performance in Less Developed Countries," *Journal of Public Economics*, 75(1), 49-71.

Reinikka R., and J. Svensson (2004): "The Power of Information : Evidence from a Newspaper Campaign to Reduce Capture," Policy Research Working Paper Series 3239, World Bank.

Rivas, F. (2006): "An Experiment on Corruption and Gender," Working Paper 0806, Universitat Autonoma de Barcelona.

Rose-Ackerman, S. (1975): "The Economics of Corruption," *Journal of Public Economics*, 4(2), 187-203.

Rose-Ackerman, S. (1978): *Corruption: A Study in Political Economy.* New York: Academic Press.

Roth, A. E. (2002): "The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics," *Econometrica*, 70(4), 1341-1378.

Schulze, G. G., and B. Frank (2003): "Deterrence versus Intrinsic Motivation: Experimental Evidence on the Determinants of Corruptibility," *Economic governance*, 4, 143-160.

Seldadyo, H., and J. de Haan (2006): "The Determinants of Corruption: A Literature Survey and New Evidence," Working Paper, University of Groningen.

Serra, D. (2006): "Empirical Determinants of Corruption: A Sensitivity Analysis," *Public Choice*, 126(1), 225-256.

Shleifer, A., and R.W. Vishny (1993): "Corruption," *Quarterly Journal of Economics*, 108(3), 599-617.

Sosa, L. A. (2004): "Wages and Other Determinants of Corruption," *Review of Development Economics*, 8(4), 597-605.

Swamy, A., S. Knack, Y. Lee, and O. Azfar (2001): "Gender and Corruption," *Journal of Development Economics*, 64(1), 25-55.

van Rijckeghem, C., and B. Weder (2001): "Bureaucratic Corruption and the Rate of Temptation: Do Wages in the Civil Service Affect Corruption, and by How Much?," *Journal of Development Economics*, 65(2), 307-331.

Wei, S.-J. (2000): "How Taxing is Corruption on International Investors," *Review of Economics and Statistics*, 82(1), 1-11.