THE ECONOMIC JOURNAL

The Economic Journal, **123** (*December*), 1168–1187. Doi: 10.1111/ecoj.12019 © 2013 The Author(s). The Economic Journal © 2013 Royal Economic Society. Published by John Wiley & Sons, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

COMPARING CORRUPTION IN THE LABORATORY AND IN THE FIELD IN BURKINA FASO AND IN CANADA*

Olivier Armantier and Amadou Boly

We investigate the external validity of corruption experiments by conducting the same experiment in three different environments: a laboratory in a developed country, a laboratory in a developing country and the field in a developing country. In the experiment, a candidate proposes a bribe to a grader to obtain a better grade. We find the direction and magnitude of several treatment effects to be statistically indistinguishable across the three environments. In particular, increasing the graders' wage reduces the probability of accepting the bribe but promotes reciprocation. Our results therefore provide evidence that laboratory experiments on corruption can have empirical relevance.

Corruption is now recognised as one of the most detrimental factors to economic and social development.[1] Because of its secretive nature, however, the analysis of corruption has been challenging to applied economists. To circumvent the absence of hard data, public perception surveys have been used since the mid 1990s to conduct empirical analyses. This approach, however, has been criticised in part because of the potential for significant measurement errors (Golden and Picci, 2005; Kaufmann *et al.*, 2006). As an alternative, laboratory experiments on corruption have recently been conducted.[2] This approach offers two key advantages: first, corrupt behaviour is unambiguously observed at the individual level; second, the researcher controls both the institutional environment and the characteristics of the subjects' population. As a result, the researcher is in a unique position to test possible corruption deterrents (e.g. higher wages, monitoring and punishment) and to identify microdeterminants of corruption (e.g. gender, religion). Laboratory experiments have therefore the potential to become a powerful tool to understand and fight corruption.

* Corresponding author: Armantier Olivier, Federal Reserve Bank of New York, 33 Liberty Street, New York, NY 10013, USA. Email: olivier.armantier@ny.frb.org.

We thank the editor, as well as four anonymous referees for helpful comments. We are grateful to O. Al-Ubaydli, J. Engle-Warwick, G. Gaudet, G. Harrison, J. Heroux, A. Hollander, J. List, C. Montmarquette, S. Daw Namoro. We also thank seminar and conference participants at the Federal Reserve Banks of New York and Boston, SUNY Stony Brook, UC Santa Barbara, NYU, the ESA meeting in Tucson, the CSAE Conference in Oxford, the IMEBE in Alicante, the New Techniques in Development Workshop in Canberra, the CEA in Vancouver and the SCSE in Montebello. We thank the recruiting firm Opty-RH for its help with hiring subjects in Burkina Faso. Finally, the authors gratefully acknowledge financing from the CRSH. As always, all remaining errors are ours. The views expressed here do not necessarily reflect those of the Federal Reserve Bank of New York, the Federal Reserve System or the United Nations Industrial Development Organization.

[1] In particular, the World Bank states on its website that it 'has identified corruption as the single greatest obstacle to economic and social development' (http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPUBLICSECTORANDGOVERNANCE/EXTANTICORRUPTION/0,,contentMDK:21540659~menuPK:384461~pagePK:148956~piPK:216618~theSitePK:384455,00.html).

[2] Laboratory experiments conducted in developed countries include Frank and Schulze (2000), Abbink *et al.* (2002), Schulze and Frank (2003), Abbink (2004), Rivas (2008), Barr and Serra (2009). Laboratory experiments conducted in developing countries include Alatas *et al.* (2009*a*,*b*), Cameron *et al.* (2009), Barr *et al.* (2009). We are aware of only three field experiments, all conducted in developing countries, directly related to corruption (Bertrand *et al.*, 2006; Olken, 2007; Castillo *et al.*, 2011).

Although promising, the experimental literature on corruption is in its infancy and its practical relevance will not be fully established as long as the question of external validity remains unaddressed.[3] Because laboratory experiments sometimes fail to incorporate some relevant features from the field (e.g. appropriate stakes, environment, population), laboratory and field experiments do not always produce identical results (Harrison and List, 2004; List, 2006). Of particular interest here is the issue of scrutiny. Indeed, a typical difference between laboratory experiments and the field is that subjects know upon entering the laboratory that their behaviour will be monitored and analysed by an experimenter. Because of this scrutiny, it has been hypothesised (Hoffman *et al.*, 1994; Hoffman *et al.*, 1996; or Levitt and List, 2007) that laboratory subjects may be more inclined to make the 'moral' choice when morality and wealth are competing objectives, as it is the case with corruption. In addition, although non-monetary considerations (e.g. moral, ethical, legal) may be major determinants of corruption, they may be difficult to capture in the laboratory.

More generally, one may wonder how attitudes towards corruption vary across countries and cultures. While the empirical literature has now clearly shown that the level of corruption is related to socio-economic factors (e.g. education levels, religious denominations, economic freedom, the strength of the judicial system),[4] little is known about whether corrupt behaviour in different countries has the same determinants and responds to the same stimuli. In other words, although differences in the level of corruption may be expected when conducting the same experiment in two different countries, it is difficult to predict whether or not the same treatment effects will emerge. This issue is not only interesting in itself, but it is also particularly important to establish the practical relevance of laboratory experiments on corruption. Indeed, although understanding and fighting corruption are considered a priority for developing countries, laboratory experiments have been conducted primarily in developed countries where most laboratories are located. So, even if laboratory experiments can be shown to have empirical relevance for the field, we need to know where these laboratory experiments should be conducted.

This study raises two questions: First, do laboratory experiments on corruption conducted in a developed and a developing country produce similar outcomes? Second, to what extent can the results of corruption experiments obtained in the laboratory be extrapolated to the field? In an attempt to address these questions, we carried out a corruption experiment in three different environments: a laboratory in a developed country, a laboratory in a developing country and the field in a developing

[3] The issue of external validity is not specific to laboratory experiments. As argued by Falk and Heckman (2009) or Kessler and Vesterlund (2011), similar concerns should be raised about the generalisability of any result based on data collected in a specific context. In other words, the question of external validity applies also to field experiments and even empirical studies conducted with non-experimental data. Furthermore, we are not arguing that to be relevant, laboratory experiments necessarily need to be fully replicable in the field. In particular, like Camerer (2011), we believe that while it is necessary to establish the external validity of experiments that aim to inform real-world policies (like most corruption experiments), the problem of external validity may be considered less relevant for experiments designed to test theories.

[4] See for example La Porta *et al.* (1999), Fisman and Gatti, (2002), or Kunicova and Rose-Ackerman, (2005).

country. The experimental design essentially reproduces a corruption scenario in which a candidate offers a bribe to a grader to obtain a better grade. In short, subjects were asked to grade 20 examination papers. The 11th paper came with a money offer and a message saying: 'Please, find few mistakes in my examination paper'. To determine whether subjects behave differently when they know they are being observed (the scrutiny hypothesis), subjects in the field were informed they had participated in an experiment only after grading was completed. We conducted four different treatments in each of the three environments by varying factors susceptible to promote or deter corruption:

(*i*) the amount of the bribe,
(*ii*) the wage paid to graders and
(*iii*) the level of monitoring and punishment.

We then tested whether the direction and the magnitude of the observed treatment effects differed across the three environments.[5]

The remainder of the article is organised as follows. The design of the experiment is presented in Section 1. The experimental results are analysed in Section 2. Finally, we discuss the practical implications of our results in Section 3.

1. Experimental Design

To conduct the same laboratory experiment in two different countries does not present any major methodological challenges. Moving from the laboratory to the field, however, is much more difficult, especially when studying corruption, an illegal, covert and immoral activity. In particular, we had to find a real-life activity that could

(*i*) credibly lend itself to corruption in the field without asking subjects to take actions for which they could be prosecuted, and
(*ii*) be replicated in the laboratory.

In addition, we tried to maintain as much control as possible over the experimental task to facilitate the comparison of behaviour between treatments and between the laboratory and the field. Finally, because of the inherent differences between the laboratory and the field, we had to adjust the way the experiment was implemented in each environment. In this Section, we describe the experimental task we designed to meet these challenges, and we discuss how the experiment was implemented in the laboratory and in the field.

[5] In other words, we test both what Camerer (2011) defines as the 'qualitative' and 'quantitative' external validity of corruption experiments. In short, qualitative external validity implies that the direction of a causal effect generalises beyond the laboratory, while quantitative external validity also requires the causal effect to be of similar magnitude inside and outside the laboratory. Both Camerer (2011) and Kessler and Vesterlund (2011) argue that attention should be focused primarily on qualitative external validity as the objective of most laboratory experiments is to investigate comparative statics. Note also that, to the best of our knowledge, this is the first study to address the external validity question by conducting the same experiment in the laboratory and in the field, as well as in a developed and in a developing country.

1.1. *The Task*

The experimental task given to subjects in all three environments consisted in spellchecking 20 *dictées*. A dictée is a classic examination in the Francophone schooling system, whereby candidates type a text as it is continuously dictated to them.[6] The text, based on a newspaper article in French, has 290 words and fits in two pages.[7] To type some of the examination papers, we recruited 23 subjects (called 'candidates' hereafter) in Montreal whose pay-offs depended

(*i*) on the number of mistakes subsequently identified by subjects in the laboratory and field experiments, and
(*ii*) on their decision about whether or not to offer graders a bribe.

As the role of the candidates is essentially passive, we refer the reader to Appendix A where details are provided about the candidates and the typing sessions.

To control the distribution of mistakes in the set of papers to be graded, we selected seven of the candidates' papers with a realistic distribution of mistakes (i.e. without too many skipped words, blank spaces, obvious misspellings or mistakes). Of these seven papers, we chose a 'bribe paper' with 20 mistakes. Note that this bribe paper had been typed by a candidate who elected to bribe the graders. To complete the set of 20 examination 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 examination papers in a precise way. First, we decided to place the bribe paper in 11th position.[8] Second, we made sure that the first and last set of 10 papers each had a symmetric and roughly identical distribution of mistakes with an average of 15.5 mistakes. Third, we decided on a passing grade of 15 mistakes, meaning that if all mistakes were detected and reported, then half the papers (including the bribe paper) would fail. Note that such a failure rate is common in most examinations and admission tests in Francophone countries. Finally, the examination papers were identified only 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 laboratory sessions, we gave the graders only the two pages of text. For the field sessions, we added a front page so that it would look like a legitimate examination. This front page included in particular the identification code, as well as the instructions given to the candidates.

[6] The aim is to evaluate one's spelling ability and knowledge of French grammar (many words in French have the same pronunciation but different spelling). Such a test is administered several times a year to students between the ages of 8 and 14. In addition, it is one of the requirements for obtaining a secretary's diploma, and it is part of the entry examination to several civil servant positions. Although traditionally conducted with pen and paper, a *dictée* may also be typed. In particular, secretary's examinations in Burkina Faso are still often typed on a typewriter and spell checked by graders as in our experiment.

[7] A copy of the materials used for the experiment (i.e. the original text, the candidates' examination papers, as well as the instructions given to the candidates, the laboratory graders, and the field graders) may be found on the website of one of the authors, https://sites.google.com/site/olivierarmantier/.

[8] Observe that by focusing exclusively on the graders' behaviour, we were able to fix the amount of the bribe, the distribution of mistakes in the 20 examination papers, the position of the bribe paper in the set of 20 examination papers and the number of mistakes in the bribe paper. Maintaining this amount of control over the experiment makes it easier to compare the graders' decision to accept the bribe across treatments and across the three environments.

1.2. *The Controlled Field Experiment (Ouagadougou)*

As the experimental design was essentially shaped by what is feasible in the field, we first describe how the controlled field experiment was implemented.[9] The controlled field experiment took place in Ouagadougou, the capital of Burkina Faso. This former French colony located in West Africa is classified as a 'low-income country' by the World Bank and is known to suffer from a serious corruption problem.[10] In Burkina Faso, part-timers are regularly hired to grade examinations.[11] Following this practice, we asked a local recruiting firm (Opty-RH) to place flyers around Ouagadougou proposing a part-time job as an examination grader during what is known as the 'national exams' period'. In addition to a form of identification, the graders had to possess a university diploma or a proof of enrolment at a university. People interested in the part-time job were invited to register in person at the recruiting firm location. After validating their credentials, they were randomly assigned to a treatment and a session. At no point, however, were the subjects informed they were about to participate in an experiment.

The grading sessions took place in a high school located in the centre of Ouagadougou.[12] Upon arrival, the subjects were gathered in a large room. We read aloud instructions on how to grade the examination papers and answered any questions. Each grader was then randomly assigned to a private room where he found an envelope containing the 20 examination papers properly ordered, a report sheet, a red pen and an answer book (i.e. a copy of the text without mistakes). No information was given about the nature of the examination 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 the report sheet. When reporting more than 15 mistakes, the graders had 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 had finished grading the 20 papers. We told them we would stop by their room every 15 minutes precisely 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.

[9] We summarise here the main features of the controlled field experiment. For a more comprehensive description of the experimental procedures and a discussion of the issues raised by the implementation of the protocol in the field, we refer the reader to Armantier and Boly (2011).

[10] The corruption index reported by Transparency International for Burkina Faso has systematically been well below the median (3.0 of 10 in 2011). Note also that bribing a grader is not uncommon in Burkina Faso. For instance, a Burkinabe's newspaper (*Le Pays*) reported on 7 March 2006 that two students were caught in a bribery attempt similar to the one in our experiment.

[11] In fact, roughly 50% of the field subjects we interviewed reported having previously taken part in similar part-time grading jobs. Note also that, to close the loop, we could have conducted the same field experiment in Montreal. We elected not to do so because it would have been difficult to convince subjects in Montreal that they were hired for a real grading task. Indeed, unlike Burkina Faso, hiring part-timers to grade examinations, and in particular *dictées*, is extremely uncommon in Canada.

[12] Note that every feature of the experiment (from the recruiting, to the design of the task, to the way the grading sessions were conducted) was chosen so as to appear like a regular part-time grading job in Burkina Faso. None of the subjects, even those with prior experience with such part-time grading jobs, questioned the credibility of the experiment. However, given the fact that the experiment took place in a controlled environment, it should not be considered a pure field experiment.

To prevent face-to-face communication and informal bargaining, we introduce the bribe by handwriting 'Please, find few mistakes in my examination paper' on an easily removable 'post-it' and taped it with a banknote 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 examination paper was opened to the second page, and
(*iii*) discovered before the grader started spell checking the paper.[13]

When a grader reported the bribe attempt during one of our visits, we asked him to write in bold 'fraud attempt' on the paper. We took the banknote and the message, and instructed the grader to spell check the bribe paper just like any other paper. Note that the graders were told in the instructions that any attempt at fraud in the examination would be penalised by failure. In other words, the graders knew the implications of reporting the bribe. Finally, the field subjects were informed that they had taken part in an experiment only after grading was completed. In this debriefing session, explanations were provided about the nature of the research and the experimental protocol. Subjects were also informed that they did not grade an actual examination and that their decisions were anonymous. After answering questions and addressing concerns, the subjects were paid in cash.[14]

1.3. *The Laboratory Experiments (Montreal and Ouagadougou)*

We conducted the same laboratory experiment in Montreal and in Ouagadougou. The laboratory experiment in Ouagadougou was conducted roughly a year after the field experiment with a different set of subjects. The recruiting procedure in Ouagadougou was essentially the same for the laboratory and field experiments. Namely, the same recruiting firm posted flyers around town and we imposed the same educational restrictions. The only difference is that the flyers used for the laboratory experiment specified that we were looking for subjects to participate in an economic experiment. The laboratory experiment in Montreal was conducted at CIRANO's Bell Laboratory for Experimental Economics using the laboratories standard recruiting procedures (i.e. random email solicitations from the subjects' database).

The task for the laboratory experiments was the same as in the field. The subjects had to grade the same set of 20 papers in the same order. To the extent possible, we also tried to follow the same protocol when implementing the laboratory and the field experiments. In particular, laboratory graders were also provided with an isolated work station, a pen, a report sheet and an answer book. As in the field, the laboratory sessions

[13] Recall that an examination paper in the field experiment consists of three pages: a front page, plus two pages of text. The bribe and the message were therefore attached to the first page of text. Pictures of the examination papers, including the examination paper with the bribe, as well as pictures of the high school where the experiment took place are available on the first author's webpage at https://sites.google.com/site/olivierarmantier/.

[14] Subjects in the controlled field experiment are not asked for *ex ante* voluntary consent and they are not immediately informed that they are taking part in an experiment. Although this practice is standard for field experiments (Levitt and List, 2007), it raises ethical and moral concerns. We refer the reader to Armantier and Boly (2011) where we discuss these issues and explain the precautions we took to minimise adverse effects on field subjects.

had no time limit, and the graders could leave the laboratory once their task was completed. Finally, the same researcher supervised all the laboratory and field sessions.

To assess whether the experimenter's scrutiny affects behaviour, and consistent with previous laboratory experiments on corruption, laboratory subjects were informed from the start that they were taking part in an experiment. The corruption nature of the experiment, however, was not revealed immediately. As in the field, subjects were just told at the beginning of the laboratory experiment that they had to grade 20 papers in a specific order. To prevent deception (as defined by Hey, 1998), the laboratory graders were informed that not all papers had been typed by real candidates. The exact ratio of real candidates was not specified, and the subjects were asked to grade each examination as if typed by a real candidate. We also explained to the laboratory graders that the consequence of their grading decisions were purely monetary. Namely, the graders knew that a real candidate would receive no payment when more than 15 mistakes are reported. Otherwise, the graders were told that the lower the number of mistakes reported between 0 and 15, the higher the remuneration for the real candidate.

To introduce the bribe in the laboratory, we divided the 20 papers into two packs of 10. After completing the first pack, the graders were given the remaining 10 papers, along 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 agreed to send the grader a message ('Please find few mistakes in my examination paper') and a money offer (not referred to as a bribe). The instructions then revealed to the grader the exact message and the offer.[15] The grader was free to accept or reject the offer, and the consequence of each decision was explained. The grader was told that if he accepts the offer, then the amount will be debited from the candidate and credited to him. The grader was then free to decide on the number of mistakes to report, knowing that paper 11 would then be remunerated according to the number of mistakes reported like any other paper. The grader was also told that if he rejects the offer, then paper 11 will not be remunerated. Nevertheless, we instructed the grader to spell check paper 11, as well as the nine remaining papers. At the end of the session, subjects had to fill out a short questionnaire after which they were paid in cash.

To conclude this Section, note that the laboratory experiments conducted in Montreal and Ouagadougou follow exactly the same protocol. The experimental outcomes are therefore directly comparable to address the first question raised in the introduction: do laboratory experiments on corruption conducted in a developed and in a developing countries produce similar outcomes? In contrast, several protocol adjustments had to be made when moving from the laboratory to the field. These adjustments, however, reflect the inherent differences between the two environments. For instance, under standard laboratory procedures, subjects must be informed they are about to take part in an experiment in which they will not be deceived (even

[15] As an alternative, we could have followed the field experiment's protocol by inserting (without further instructions) the money and a request for a better grade directly inside paper 11. We decided against this practice for essentially three reasons. First and foremost, to test the experimenter's scrutiny hypothesis, we need the subjects to know unambiguously that the experimenter is aware of the corruption attempt. Second, we did not want to deceive laboratory subjects. Third, it is inconsistent with the way previous corruption experiments have been conducted in the laboratory.

temporarily). Laboratory subjects in corruption experiments are also provided with explicit instructions explaining the exact consequences of their actions and that these consequences are purely monetary. Our objective here is not to test how behaviour may be affected by each of the adjustments made when moving from the field to the laboratory. Instead, we argue that the different protocol adjustments should be considered jointly to capture the fundamental difference between the laboratory and the field. In other words, we believe that the comparison of our laboratory and field experiments is informative to address the second question raised in the introduction: to what extent the results of corruption experiments obtained in the laboratory can be extrapolated to the field?

1.4. *The Experimental Treatments*

We conducted four treatments in each of the three environments, that is, in the laboratory in Montreal, in the laboratory in Ouagadougou and in the field in Ouagadougou. Each of the treatments conducted in each environment consisted of three sessions with different groups of subjects. In the control treatment, the subjects were paid a fixed amount (called a wage hereafter) regardless of how they performed the grading task. In addition, graders in the control treatment were offered a bribe. In the laboratory, the wage was 250 experimental units (EU hereafter) and the bribe 50 EU. The conversion rate was C$1 = 12 EU in Montreal and 20 FCFA = 1 EU in Ouagadougou.[16] In the field, the wage was 5,000 FCFA and the bribe 1,000 FCFA. Three features of the design are worth noting. First, the bribe-to-wage ratio is the same in the laboratory and field experiments. Second, the pay-offs for the laboratory and the field experiments conducted in Ouagadougou are identical. Third, because we did not want the field subjects to know they were participating in an experiment, we selected the wage so that it would not raise suspicion: 5,000 FCFA roughly corresponds to the amount our subjects would earn in a similar part-time job of grading.

Following the economic literature on corruption, we selected treatment variables that have long been studied for their ability to promote or deter corruption: bigger bribes (Rose-Ackerman, 1975; Abbink *et al.*, 2002), higher wages (Becker and Stigler, 1974; van Rijckeghem and Weder, 2001; Sosa, 2004) and monitoring and punishment (Rose-Ackerman, 1978; Mookherjee and Png, 1995; Schulze and Frank, 2003). 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 laboratory). 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 laboratory). Finally, the last treatment makes an attempt at studying the effect of monitoring and punishment. For obvious practical and ethical reasons, we decided against confronting field graders who accepted the bribe. Instead, we introduced a mechanism aimed at

[16] The Franc CFA (Franc Communauté Financière Africaine) is the currency used in Burkina Faso. The conversion rate was roughly *C*$1 for 400 FCFA at the time the laboratory and field experiments were conducted. It is difficult to compare the wages paid in Montreal and in Ouagadougou. The wages are larger in Montreal based on the official exchange rate, but those paid in Ouagadougou provide more purchasing power.

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 favour the briber. The monitoring mechanism was explained in the instructions as follows. We told each grader that we would randomly pick and regrade 5 of the 20 papers he spell checked. Only the worst of the five papers regraded was considered for the monetary penalty. More specifically, we calculated the difference between the number of mistakes reported in the worst regraded paper and the actual number of mistakes. The penalty imposed in the laboratory was 100 EU when the difference was between three and five mistakes, 150 EU when the difference was between six and nine mistakes and 225 EU when the difference exceeded 10 mistakes. The penalties imposed in the field were proportional. Except for the risk of being penalised, the 'monitoring' treatment is identical to the control treatment.

To conclude this section, note that, while the data collected in the laboratory in Montreal and in Ouagadougou have not been studied elsewhere, the data from the controlled field experiment constitute a subset of the data previously analysed in Armantier and Boly (2011).[17] The two papers, however, address different questions. While the object of Armantier and Boly (2011) is to identify factors that may promote or deter corruption in the field in Ouagadougou, this study assesses the external validity of corruption experiments by comparing the behaviour of subjects facing the same corruption scenarios in the laboratory and in the field, as well as in the developed and in a developing country.

2. Experimental Results

A total of 422 subjects participated in the experiments, with an average of 141 subjects per environment (see Table 1), 35 subjects per treatment (see Table 2) and 12 subjects per session. Note that there are marked differences between the subjects in Montreal and in Ouagadougou. In particular, we can see in Table 1 that subjects in

Table 1

Characteristics of the Subjects in each Environment

	N	Age	Female	Religiousness	Time (in Min)	Precision (in %)	Improvement
Field Ouagadougou	164	24.933	0.152	2.671	140.396	−23.125	−0.045
		(2.381)	(0.361)	(1.254)	(26.674)	(12.135)	(1.350)
Lab Ouagadougou	133	24.872	0.195	2.496	164.707	−18.355	0.358
		(3.014)	(0.398)	(1.241)	(34.905)	(10.137)	(1.591)
Lab Montreal	125	26.264	0.408	0.832	100.208	−18.990	−0.068
		(6.324)	(0.493)	(1.063)	(17.303)	(9.991)	(1.493)

Notes. Averages and standard deviations in parenthesis. A cell shaded in dark grey (light grey) indicates that the variable has a distribution significantly different at the 5% level (10% level) than in the laboratory in Ouagadougou according to Wilcoxon rank-sum test

[17] Two additional treatments are analysed in Armantier and Boly (2011): A 'no bribe' treatment to test whether being offered a bribe affects grading, and a 'high monitoring' treatment to test whether monitoring and punishment can crowd out intrinsic motivations for honesty.

Table 2

Descriptive Statistics

Column number			1	2	3	4	5
				Average No. of mistakes reported for paper 11		% of Graders who report fewer mistakes for paper 11 than for paper 7	
	Treatment	No. of Subjects	% of Graders who accept the bribe	Accepters	Rejecters	Accepters	Rejecters
Field Ouagadougou	Control	36	50.0	14.7	16.4	27.8	22.2
	High bribe	45	68.9**	13.1	15.4	71.0**	35.7
	High wage	39	35.9*	12.9	15.7	64.3**	24.0
	Monitoring	44	40.9	15.7	15.9	33.3	26.9
Lab Ouagadougou	Control	33	48.5	16.4	17.3	25.0	23.5
	High bribe	33	66.7**	15.2	16.6	72.7**	36.4
	High wage	33	36.4*	14.9	17.5	57.1**	38.1
	Monitoring	34	41.2	17.5	16.5	28.6	25.0
Lab Montreal	Control	30	66.7	14.9	16.7	35.0	30.0
	High bribe	32	65.6	14.8	15.6	42.9	27.3
	High wage	31	48.4**	13.9	16.3	60.0**	43.8
	Monitoring	32	65.6	15.0	15.5	33.3	25.5

Notes. Environment effects: Within each treatment, we test for differences across environments (e.g. do graders in the High bribe treatment accept the bribe more often in the laboratory in Ouagadougou compared to field in Ouagadougou?). A cell shaded in dark grey (light grey) indicates that the variable has a distribution significantly different at the 5% level (10% level) than in the same treatment conducted in the laboratory in Ouagadougou according to a Fisher–Boschloo exact test of proportions.

Treatment effects: Within each environment, we test for the presence of treatment effects (e.g. do graders in the laboratory in Ouagadougou accept the bribe more often in the High bribe treatment compared to the Control treatment?). **(*) indicates that the variable has a distribution significantly different at the 5% level (10% level) than in the Control treatment conducted in the same environment according to a Fisher–Boschloo exact test of proportions.

Ouagadougou were significantly younger, more religious, less likely to be a woman and took longer to complete the grading task.[18] To measure ability, we also calculated each subject's grading precision and improvement over the first 10 examination papers.[19] We can see in Table 1 that, on average, subjects in each of the three environments missed roughly 20% of the spelling mistakes in the first 10 examination papers and did not exhibit any significant grading improvement.

2.1. *The Decision to Accept the Bribe*

We report in Table 2 (column 1) the frequency of bribe acceptance in each environment and each treatment. Observe first that the results obtained in the laboratory and in the field in Ouagadougou are remarkably similar in all four treatments. In particular, roughly 50% of the Ouagadougou subjects accepted the bribe in the control treatment. In other words, half the graders essentially refused 'free money' despite the absence of explicit risks and negative externalities.[20] Table 2 also indicates that, compared with the control treatment, the probability of accepting the bribe in Ouagadougou is lower when graders receive a higher wage and when their work is monitored, but only the first effect is statistically significant. In contrast, laboratory and field subjects in Ouagadougou accepted the bribe significantly more often when the bribe was larger.

Compared with Ouagadougou, the laboratory experiment conducted in Montreal produced higher probabilities of acceptance in almost all treatments. In particular, observe in Table 2 (column 1) that the rejection rate in the control treatment is significantly smaller in Montreal than in Ouagadougou, as only one in three Montreal subjects rejected the bribe.[21] Despite the differences in the level of bribe acceptance,

[18] The measure of religiousness was obtained from the post-experiment questionnaire in which we asked subjects how often they go to a church, a mosque or any other place of worship. This variable has five categories, ranging from 0 (never) to 4 (every day).

[19] Formally,

$$Precision_i = \frac{1}{10}\sum_{t=1}^{10}\Delta_{i,t}, \text{ where } \Delta_{i,t} = -100\left|\frac{R_{i,t} - M_t}{M_t}\right|$$

is a percentage deviation, $R_{i,t}$ is the number of mistakes reported by subject i for paper t and M_t is the true number of mistakes in paper t. Note that $Precision_i \leq 0$, and grader i is considered more precise when $Precision_i$ increases towards 0. To obtain the improvement measure, we estimate for each subject i the regression model $\Delta_{i,t} = \alpha_{0,i} + \alpha_{1,i}t$ (for $t = 1,\ldots,10$), and we set the variable $Improvement_i$ equal to the estimated slope $\hat{\alpha}_{1,i}$. Grader i is then considered to have improved at the grading task when $Improvement_i$ is positive and large. Finally, note that these measures of ability are valid exogenous variables as the graders both in the laboratory and in the field were unaware of the presence of corruption until they reached paper 11.

[20] This 50% rejection rate is substantially higher than in comparable laboratory experiments; e.g. the rejection rate is 9.4% in Frank and Schulze (2000) and 13.1% in Cameron *et al.* (2009). We conjecture that framing may explain this difference, as our subjects are (or used to be) university students and therefore, they may be less tolerant towards the type of corruption implemented in our experiment.

[21] The higher frequency of bribe acceptance in Montreal contrasts with the fact that Canada is universally considered less corrupt than Burkina Faso. In particular, Transparency International consistently ranks Canada (Burkina Faso) in the top 10 (outside the top 100) least corrupt countries. Cameron *et al.* (2009) obtained a similar effect in their experiment as they found higher levels of corruption in Singapore (one of the least corrupt countries) than in Indonesia. We conjecture that the higher bribe acceptance in Canada may reflect the fact that laboratory subjects in Montreal are not as aware of the negative consequences of corruption.

Table 3

The Decision to Accept the Bribe

	Environment					
	Field Ouagadougou		Lab Ouagadougou		Lab Montreal	
Variable	Parameter	Marginal effect	Parameter	Marginal effect	Parameter	Marginal effect
Constant	0.006		0.044		0.882**	
	(0.233)	—	(0.229)	—	(0.387)	—
Female	−0.572	−0.196	0.170	0.072	−0.484	−0.126
	(0.395)	(0.135)	(0.322)	(0.137)	(0.316)	(0.083)
Age	−0.368**	−0.149	−0.083**	−0.035	−0.097***	−0.038
	(0.048)	(0.020)	(0.036)	(0.015)	(0.027)	(0.010)
Time	−0.019**	−0.009	−0.009**	−0.005	−0.032***	−0.011
	(0.009)	(0.003)	(0.004)	(0.002)	(0.007)	(0.003)
Religiousness	−0.479***	−0.161	−0.306***	−0.116	−0.708***	−0.192
	(0.102)	(0.034)	(0.110)	(0.043)	(0.155)	(0.041)
Precision	−0.249**	−0.120	−0.674***	−0.250	−0.491**	−0.178
	(0.122)	(0.059)	(0.178)	(0.065)	(0.192)	(0.069)
Improvement	−0.104	−0.040	−0.750***	−0.258	−0.523***	−0.191
	(0.084)	(0.032)	(0.149)	(0.052)	(0.126)	(0.046)
High bribe treatment	0.732**	0.293	1.169***	0.384	−0.239	−0.084
	(0.331)	(0.132)	(0.395)	(0.131)	(0.420)	(0.147)
High wage treatment	−0.741**	−0.295	−0.863**	−0.314	−1.023***	−0.352
	(0.327)	(0.130)	(0.382)	(0.139)	(0.406)	(0.141)
Monitoring treatment	−0.251	−0.099	−0.345	−0.139	−0.343	−0.121
	(0.320)	(0.125)	(0.351)	(0.141)	(0.418)	(0.146)
Ln L	−76.109		−65.610		−50.698	
N	164		133		125	

Notes. In each environment, the subject of reference is a man in the control treatment. The standard deviations are robust and clustered at the session level. Significance: *10%, **5%, ***1%. A cell shaded in dark (light) grey indicates that the estimated parameter is significantly different at the 5% level (10% level) than in the laboratory in Ouagadougou according to a Likelihood ratio test.

the direction of some of the treatment effects is similar in the laboratory experiments conducted in Montreal and Ouagadougou. In particular, higher wages appear to curb bribe acceptance significantly in both locations. In contrast with the laboratory experiment conducted in Ouagadougou, however, proposing a bigger bribe to Montreal subjects does not seem to affect their decision to accept the bribe.

The cross-country differences in bribe acceptance documented in Table 2 may reflect intrinsic differences in attitudes towards corruption in Montreal and Ouagadougou, or simply differences in the composition of the subject pools in each location. To disentangle the two effects better, we now estimate a probit model for a grader's decision to accept the bribe in which we control for the subjects, characteristics.[22]

[22] To simplify the presentation of the results, in Table 3, we report the outcomes of probit regressions estimated separately with the data collected in each environment. To test formally for differences between the laboratory in Ouagadougou, the laboratory in Montreal and the field in Ouagadougou, we also estimate joint probit models with the data collected in different environments. The results of this joint estimation is reported in the online Appendix Table D1.

Observe first in Table 3 that the same individual characteristics shape corrupt behaviour in all three environments. Specifically, we find that an older, more religious, more competent or more deliberate grader is significantly less likely to accept the bribe. In addition, only some of these individual effects, indicated by shaded cells in Table 3, differ significantly in magnitude across the three environments. These results are remarkable as they suggest that some micro determinants of corrupt behaviour are identical in a developed and in a developing country, both the laboratory and in the field.

Regarding treatment effects, Table 3 confirms that subjects in all three environments are significantly less likely to accept the bribe when they receive a higher wage. In addition, we find the magnitude of this treatment effect to be statistically indistinguishable across the three environments.[23] More specifically, we find that a 40% increase in the grader's wage leads to a 29% to 35% reduction in the probability that the bribe will be accepted.[24] Finding a treatment effect with the same direction and magnitude in all three environments is noteworthy, as it suggests that, at least in some dimensions, corruption experiments conducted in the laboratory in developed countries can be consistent with behaviour observed both in the laboratory and in the field in developing countries.

Table 3 also indicates that graders reject the bribe slightly more frequently when monitored, although the effect is not statistically significant in any of the three environments. This absence of treatment effect is not necessarily surprising as the type of monitoring we implemented was not aimed at catching the subjects who accept the bribe. Instead, it was designed to catch corrupt graders when they reciprocate by reporting fewer mistakes for the briber's paper.

Finally, the results in Table 3 confirm the existence of a significant difference between the laboratory experiments in Montreal and Ouagadougou. Indeed, we find that doubling bribe increases its probability of acceptance in the laboratory in Ouagadougou by 38%, while it appears to have no effect in the laboratory in Montreal. Although this result may indicate that subjects have different price elasticities in each country, we conjecture that it may reflect a pure level effect. Indeed, although the bribe is doubled in the two countries, the absolute increase in purchasing power is nearly four times larger in Ouagadougou. Expressed differently, the bribe increase of C$4.15 (1,000 FCFA) in the high bribe treatment corresponds to 1/2 hour (4 hours) of work at the mandatory minimum wage in Canada (Burkina Faso).

[23] Throughout the article, the findings that some treatment effects have the same direction in different environments should be considered stronger than the findings that the magnitudes of these treatment effects are statistically indistinguishable across environments. Indeed, although our sample size is sufficient to identify treatment effects in each of the three environments, it may not be large enough to identify significant differences between the magnitudes of these treatment effects.

[24] Although the magnitudes are not directly comparable, the direction of these treatment effects are consistent with several laboratory experiments (Jacquemet, 2005; Barr *et al.*, 2009) and empirical analyses (van Rijckeghem and Weder, 2001; Alt and Lassen, 2003). Our results are also in line with the views of numerous practitioners and international institutions that often recommend paying civil servants above their private sector alternative as a means of deterring corruption. Singapore and Hong Kong are often presented as successful examples of such a policy. These countries are typically ranked among the least corrupt, and they are known to pay high salaries to their civil servants.

2.2. *The Number of Mistakes Reported for the Bribe Paper*

We now try to identify whether, despite having no financial incentive to reciprocate, the graders who accept the bribe respond to the briber's request for reporting fewer mistakes. A comparison of columns 2 and 3 in Table 2 reveals that in all but one of the 12 treatments conducted (the monitoring treatment conducted in the laboratory in Ouagadougou), the subjects who accept the bribe report fewer mistakes for the bribe paper (paper 11) than the subjects who reject the bribe. In addition, an average rejecter reports a failing grade (more than 15 mistakes) for the briber in all 12 treatments (Table 2, column 3), while the average number of mistakes reported by accepters is often below 15 (Table 2, column 2).

These results, however, are not necessarily evidence of reciprocation toward the briber. Instead, the differences in the average number of mistakes reported for paper 11 could simply reflect the finding in the previous Section that bad graders are more likely to accept the bribe. In an effort to control non-parametrically for grading quality, we designed the experiment so that another examination paper (paper 7) has the same number of mistakes as the bribe paper (i.e. 20 mistakes). All else equal, the probability that a grader finds fewer mistakes in the bribe paper than in paper 7 should be the same in the absence of reciprocation regardless of the grader's quality and regardless of his decision to accept or reject the bribe. In contrast, we find clear evidence that accepters report fewer mistakes for the briber's examination. Indeed, a comparison of columns 4 and 5 in Table 2 indicates that, compared with a rejecter, accepters in each of the 12 treatments are more likely to report fewer mistakes for the bribe paper than for paper 7. Furthermore, regardless of the environment, the effect is significantly more pronounced in the high wage treatment than in the control treatment. Indeed, the proportion of accepters who report fewer mistakes in paper 11 than in paper 7 roughly doubles when subjects receive a higher wage (Table 2, column 4). The treatment effect is even stronger for the high bribe treatments conducted in the laboratory and in the field in Ouagadougou. Indeed, an accepter in Ouagadougou is nearly three times more likely than in the control treatment to report fewer mistakes for the bribe paper than for paper 7 (Table 2, column 4). Finally, in all three environments, monitoring and punishment do not seem to affect the relative number of mistakes reported for paper 7 and 11 in any predictable manner. As documented later on, however, the subjects' response to monitoring and punishment is highly gender specific.

The comparisons just presented, although strongly suggestive of reciprocation, cannot be considered conclusive evidence. Indeed, they do not control for other factors that may explain why accepters report fewer mistakes than rejecters for the bribe paper. In particular, we found in the previous section that subjects whose grading quality deteriorates over time are more likely to accept the bribe. To isolate possible reciprocation effects better we estimate regression models in which we control for the subjects characteristics. The endogenous variable is the number of mistakes reported for the bribe paper. To highlight gender differences, we report the estimates of two models per environment (Models 1 and 2) in Table 4.[25] Not surprisingly, the

[25] We report the outcomes of several alternative specifications in the online Appendix Tables D2–D4. The results produced by these models yield the same conclusions as those presented here.

Table 4

Number of Mistakes Reported for Paper 11

	Environment					
	Field Ouagadougou		Lab Ouagadougou		Lab Montreal	
Variable	1	2	3	4	5	6
Constant	16.457***	16.434***	17.372***	17.328***	17.026***	16.987***
	(0.442)	(0.447)	(0.262)	(0.276)	(0.647)	(0.587)
Female	−1.182**	−0.029	−0.894***	−0.511	−2.466***	−0.787
	(0.506)	(0.574)	(0.301)	(0.397)	(0.564)	(0.553)
Age	−0.094	−0.111	0.039	0.044	0.053	0.043
	(0.079)	(0.081)	(0.041)	(0.041)	(0.036)	(0.033)
Time	−0.006	−0.012	−0.003	−0.002	−0.015	−0.015
	(0.008)	(0.008)	(0.009)	(0.009)	(0.016)	(0.014)
Religiousness	−0.205	−0.166	0.074	0.029	−0.238	−0.181
	(0.152)	(0.151)	(0.101)	(0.106)	(0.212)	(0.195)
Precision	1.093***	1.029***	0.635**	0.637***	1.510***	1.462***
	(0.223)	(0.229)	(0.238)	(0.221)	(0.230)	(0.202)
Improvement	0.569***	0.522***	0.410***	0.416***	1.137***	1.076***
	(0.149)	(0.148)	(0.089)	(0.090)	(0.241)	(0.210)
Accept	−1.746**	−1.701**	−1.317**	−1.296***	−2.239**	−2.180**
	(0.756)	(0.744)	(0.499)	(0.503)	(0.954)	(0.903)
Female × Accept		−4.139***		−2.494***		−3.785***
	—	(0.695)	—	(0.581)	—	(0.916)
High bribe treatment	−1.136	−0.968	−0.711	−0.659	−0.198	−0.018
	(0.810)	(0.790)	(0.745)	(0.743)	(0.773)	(0.796)
High wage treatment	−0.716	−0.654	0.218	0.164	−0.132	−0.064
	(0.623)	(0.637)	(0.446)	(0.460)	(0.721)	(0.674)
Monitoring treatment	0.084	0.051	−0.488	−0.415	1.116	1.216
	(0.553)	(0.549)	(0.523)	(0.531)	(0.991)	(0.890)
Accept × High bribe treatment	−1.596**	−1.642**	−1.348**	−1.316**	−0.659	−0.615
	(0.729)	(0.681)	(0.482)	(0.488)	(0.468)	(0.411)
Accept × High wage treatment	−1.960**	−2.053***	−1.636***	−1.397**	−1.889**	−1.973**
	(0.794)	(0.775)	(0.482)	(0.523)	(0.849)	(0.797)
Accept × Monitoring treatment	−0.685	−0.469	0.667	0.441	0.067	−0.957
	(0.846)	(0.979)	(0.620)	(0.629)	(0.912)	(0.841)
Female × Accept × Monitoring treatment	—	4.924***	—	5.101***	—	6.396***
		(1.254)		(1.666)		(1.786)
Adjusted R-squared	0.480	0.512	0.508	0.545	0.542	0.629
N	164		133		125	

Notes. In each environment, the subject of reference is a man who rejects the bribe in the Control treatment. The standard deviations are robust and clustered at the session level. Significance: *10%, **5%, ***1%. A cell shaded in dark (light) grey indicates that the estimated parameter is significantly different at the 5% level (10% level) than in the laboratory in Ouagadougou according to a Wald test.

parameters associated with the variables 'precision' and 'improvement' are greater than zero and highly significant in each environment, thereby indicating that subjects with better grading ability find more mistakes in paper 11.

Once we control for the subject's ability and other individual characteristics, we find the parameter associated with 'accept' in Model 1 to be significantly lower than zero in all three environments. This therefore confirms that subjects who accept the bribe reciprocate by reporting fewer mistakes for paper 11. The magnitude of the reciprocation effect, however, is gender specific. Indeed, the estimation of Model 2

with the data collected in each of the three environments (see Table 4) reveals that, while male accepters report between 1.3 and 2.2 fewer mistakes for paper 11 than male rejecters, female accepters omit an additional 2.5–4.1 mistakes (3.0–4.5 mistakes) compared with female rejecters (male rejecters). This result is consistent with several experiments showing that women tend to reciprocate more than men (see Croson and Gneezy, 2009 for a survey).

Table 4 also confirms the presence of several treatment effects. In particular, accepters in all three environments report fewer mistakes when provided with a higher wage. The total effect of a wage increase on corruption therefore appears to be ambiguous in our experiment: while it lowers the probability of accepting the bribe, it also seems to promote reciprocation towards the briber. The results from Model 2 in Table 4 also suggest that the impact of monitoring and punishment on the number of mistakes reported is gender specific in all three environments. While no effect is found for men who accept the bribe, we find that female accepters do not reciprocate as much in the monitoring treatments as they report between 4.9 and 6.4 more mistakes than their counterparts in the control treatments. 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.

Observe that, up to this point, both the direction and the magnitude of the wage and the monitoring treatment effects are statistically indistinguishable in all three environments. The only significant difference we can identify is related to the impact of a higher bribe. Indeed, we can see in Table 4 that, although the parameters corresponding to the high bribe treatment are negative in all three environments when a subject takes the bribe, they are significantly different from zero only for the laboratory and the field experiments conducted in Ouagadougou. In other words, compared with the control treatment, laboratory and field accepters in Ouagadougou report fewer mistakes for the briber, while increasing the bribe does not influence significantly how accepters grade the bribe paper in the laboratory in Montreal.

3. Discussion

As argued by several international institutions, corruption is one of the most detrimental factors currently afflicting the economies of developing countries. Because naturally occurring data on corruption are scarce and do not vary along certain desired dimensions (e.g. under different wage or monitoring structures), applied economists have had limited success in their effort to understand and combat corruption. Recently, the microdeterminants 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, then laboratory experiments could become increasingly popular in development economics as they enable the analyst to overcome the unobservability of corrupt behaviour in a controlled and cost-effective environment. First, however, it is essential to determine whether or not the insights gained in laboratory experiments on corruption can be extrapolated to where it arguably matter the most, the field in developing countries.

In an attempt to address this issue, we conducted a corruption experiment in three different environments to compare the behaviour of

(*i*) laboratory subjects in a developed and in a developing country, and .
(*ii*) laboratory and field subjects in a developing country.

In short, we hired subjects to grade examination papers, one of which came with a bribe offer and a demand for a better grade. We then recorded the frequency with which the bribe was accepted and the number of mistakes the subjects reported for the bribe paper. The experiment consisted of four treatments where we varied successively

(*i*) the amount of the bribe,
(*ii*) the wage paid to graders and
(*iii*) the level of monitoring and punishment.

The experiment was conducted in three environments: the laboratory in Montreal (Canada), the laboratory in Ouagadougou (Burkina Faso) and the field in Ouagadougou. The key difference between the laboratory and the controlled field experiment is that subjects in the field acted without knowing that they were participating in an experiment.

The results obtained in the laboratory and in the field in Ouagadougou are remarkably similar. In particular, the probability that a subject accepts the bribe in any given treatment is virtually identical in both environments. The laboratory experiments conducted in Ouagadougou and in Montreal produced slightly different results. However, once we control for observable differences between the subjects in Montreal and Ouagadougou (e.g. gender, age, ability), we find that the direction and the magnitude of several treatment effects are statistically indistinguishable across the two countries. In particular, we find that paying higher wages has an ambiguous effect on corruption: it reduces the probability that a bribe will be accepted but it promotes reciprocation towards the briber among those who accept. Likewise, regardless of the environment, we find that women who accept the bribe respond to monitoring and punishment by reporting more mistakes for the briber. The outcomes of the laboratory experiments, however, differ in one important dimension between the two locations. Indeed, doubling the amount of the bribe proposed to the grader has no effect in Montreal, while it promotes bribe taking and reciprocation in Ouagadougou. As argued in subsection 2.1, this result may reflect a stake effect, as the magnitude of the bribe increase was smaller in real terms in Montreal than in Ouagadougou.

Our results provide some elements to answer the two questions raised in the introduction. First, despite differences in the levels of corruption, we find that, once we control for the subjects' observable characteristics, the laboratory experiments conducted in Ouagadougou and Montreal produced several treatment effects that were statistically indistinguishable. In other words, our results provide some support to the hypothesis that laboratory experiments on corruption yield the same conclusions when conducted in a developed and in a developing country. Second, the laboratory and the field experiments conducted in Ouagadougou produced almost identical results. In other words, we find no evidence to support the hypothesis that attitudes towards corruption are fundamentally affected when subjects know their behaviour is being scrutinised.

Our study therefore provides some evidence to support to the external validity of corruption experiments. However, it should not be considered a definitive answer to

the question for at least two reasons. First, the robustness of our results needs to be confirmed. In particular, when moving from the laboratory to the field, we had to adjust several aspects of the experimental design to reflect the inherent differences between the two environments. Additional laboratory–field comparisons will be necessary to establish whether some of these design adjustments can explain our results. Second, the generalisability of our results needs to be established. In particular, the external validity question should be examined for other forms of corruption (e.g. embezzlement, nepotism) and more realistic corruption scenarios should be considered. Indeed, the very specific task, stakes and subject pool in our experiment may not be representative of actual corruption. Nevertheless, our results should be considered encouraging as they suggest that, at least in the specific context we considered, laboratory experiments, and in particular laboratory experiments conducted in developed countries, can be informative to understand and combat corruption in the field in developing countries.

Appendix A. The Candidates (Montreal, Canada)

Subjects, called 'candidates', were recruited to type a *dictée*. At the beginning of the typing session, each candidate was assigned to an isolated computer. Instructions were then read aloud, followed by questions. We explained 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 various subjects called 'graders'. Finally, we explained that a candidate's payment would depend in part on the average number of mistakes the grader(s) 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 examination paper'. We explained to the candidates that if they accepted to offer a bribe, 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 agreed to offer a bribe, we would not necessarily send the message and the bribe to the graders.[26]

To keep as much control as possible over the experiment, we deliberately left the candidates' instructions partly ambiguous.[27] 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. The candidates were told they would received several payments: C$20 payable immediately after the conclusion of the typing session and three additional amounts paid after the completion of the grading sessions in each of the three environments. The candidates knew that each of the additional amounts could vary between C$20 and C$60, depending on the average number of mistakes reported by the graders, and, when relevant, on the number of graders who accepted and rejected the bribe offer.

We conducted two typing sessions in Montreal at CIRANO's Bell Laboratory for Experimental Economics. Each session lasted roughly an hour and included respectively 11 and 12 subjects. All

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

[27] According with standard laboratory practices, however, we did not explicit mislead the candidates. We simply withheld some information.

23 subjects agreed to send a bribe to the graders. On average, the candidates received a total payment of C$96.48, with a maximum of C$171.41 and a minimum of C$80.00.

Federal Reserve Bank of New York, Cirano Cireq
United Nations Industrial Development Organization

Submitted: 4 July 2011
Accepted: 10 October 2012

Additional Supporting Information may be found in the online version of this article:

Appendix B: Basic Instructions Read to the Graders (Field)
Appendix C: Instructions for Graders (Laboratory)
Appendix D: Additional Tables and Results

References

Abbink, K. (2004). 'Staff rotation as an anti-corruption policy: an experimental study', *European Journal of Political Economy*, vol. 20(4), pp. 887–906.
Abbink, K., Irlenbusch, B. and Renner, E. (2002). 'An experimental bribery game', *Journal of Law, Economics, and Organization*, vol. 18(2), pp. 428–54.
Alatas, V., Cameron, L., Chaudhuri, A., Erkal, N. and Gangadharan, L. (2009*a*). 'Gender and corruption: insights from an experimental analysis', *Southern Economic Journal*, vol. 75(3), pp. 663–80.
Alatas, V., Cameron, L., Chaudhuri, A., Erkal, N. and Gangadharan, L. (2009*b*). 'Subject pool effects in a corruption experiment: a comparison of indonesian public servants and indonesian students', *Experimental Economics*, vol. 12(1), pp. 113–32.
Alt, J.E. and Lassen, D.D. (2003). 'The political economy of institutions and corruption in American states', *Journal of Theoretical Politics*, vol. 15(3), pp. 341–65.
Armantier, O. and Boly, A. (2011). 'A controlled field experiment on corruption', *European Economic Review*, vol. 55(8), pp. 1072–82.
Barr, A. and Serra, D. (2009). 'The effects of externalities and framing on bribery in a petty corruption experiment', *Experimental Economics*, vol. 12(4), pp. 488–503.
Barr, A., Lindelow, M. and Serneels, P. (2009). 'Corruption in public service delivery: an experimental analysis', *Journal of Economic Behavior and Organization*, vol. 72(1), pp. 225–39.
Becker, G.S. and Stigler, G.J. (1974). 'Law enforcement, malfeasance, and compensation of enforcers', *Journal of Legal Studies*, vol. 3(1), pp. 1–18.
Bertrand, M., Djankov, S., Hanna, R. and Mullainathan, S. (2006). 'Obtaining a driving license in India: an experimental approach to studying corruption', *Quarterly Journal of Economics*, vol. 122(4), pp. 1639–76.
Camerer, C.F. (2011). 'The promise and success of lab-field generalizability in experimental economics: a reply to Levitt and List (2007+)', Working Paper, California Institute of Technology.
Cameron, L., Chaudhuri, A., Erkal, N. and Gangadharan, L. (2009). 'Propensities to engage in and punish corrupt behavior: experimental evidence from Australia, India, Indonesia and Singapore', *Journal of Public Economics*, vol. 93(7–8), pp. 843–51.
Castillo, M., Petrie, R., Torero, M. and Viceisza, A. (2011). 'Lost in the mail: a field experiment on crime', Working Paper, George Mason University.
Croson, R. and Gneezy, U. (2009). 'Gender differences in preferences', *Journal of Economic Literature*, vol. 47 (2), pp. 448–74.
Falk, A. and Heckman, J. (2009). 'Lab experiments are a major source of knowledge in the social sciences', *Science*, vol. 326(5952), pp. 535–8.
Fisman, R. and Gatti, R. (2002). 'Decentralization and corruption: evidence across countries', *Journal of Public Economics*, vol. 83(3), pp. 325–45.
Frank, B. and Schulze, G.G. (2000). 'Does economics make citizens corrupt?' *Journal of Economic Behavior & Organization*, vol. 43(1), pp. 101–13.
Golden, M.A. and Picci, L. (2005). 'Proposal for a new measure of corruption, illustrated with Italian data', *Economics and Politics*, vol. 17(1), pp. 37–75.
Harrison, G.W. and List, J.A. (2004). 'Field experiments', *Journal of Economic Literature*, vol. 42(4), pp. 1009–55.

Hey, J.D. (1998). 'Experimental economics and deception: a comment', *Journal of Economic Psychology*, vol. 19 (3), pp. 397–401.
Hoffman, E., McCabe, K., Shachat, K. and Smith, V. (1994). 'Preference, property rights and anonymity in bargaining games', *Games and Economic Behavior*, vol. 7(3), pp. 346–80.
Hoffman, E., McCabe, K. and Smith, V. (1996). 'Social distance and other-regarding behavior in dictator games', *American Economic Review*, vol. 86(5), pp. 653–60.
Jacquemet, N. (2005). 'Corruption as betrayal: experimental evidence on corruption under delegation', Working Paper 0506, Groupe d'Analyse et de Theorie Economique (GATE).
Kaufmann, D., Kraay, A. and Mastruzzi, M. (2006). 'Measuring corruption: myths and realities', Working Paper, World Bank.
Kessler, J. and Vesterlund, L. (2011). 'The external validity of laboratory experiments: qualitative rather than quantitative effects', Working Paper, University of Pittsburgh.
Kunicova, J. and Rose-Ackerman, S. (2005). 'Electoral rules and constitutional structures as constraints on corruption', *British Journal of Political Science*, vol. 35(4), pp. 573–606.
La Porta, R., Lopez-de Silanes, F., Shleifer, A. and Vishny, R. (1999). 'The quality of government', *Journal of Law, Economics, and Organization*, vol. 15(1), pp. 222–79.
Levitt, S. and List, J.A. (2007). 'What do laboratory experiments measuring social preferences tell us about the real world', *Journal of Economic Perspectives*, vol. 21(2), pp. 153–74.
List, J.A. (2006). 'Field experiments: a bridge between laboratory and naturally occurring data', *Advances in Economic Analysis & Policy*, vol. 6(2), Article 8, pp. 1–47.
Mookherjee, D. and Png, I.P.L. (1995). 'Corruptible law enforcers: how should they be compensated?' Economic Journal, vol. 105(428), pp. 145–59.
Olken, B.A. (2007). 'Monitoring corruption: evidence from a field experiment in Indonesia', *Journal of Political Economy*, vol. 115(2), pp. 200–49.
van Rijckeghem, C. and Weder, B. (2001). 'Bureaucratic corruption and the rate of temptation: do wages in the civil service affect corruption, and by how much?' *Journal of Development Economics*, vol. 65(2), pp. 307–31.
Rivas, F. (2008). 'An experiment on corruption and gender', Discussion paper, Working Paper 0806, Universitat Autonoma de Barcelona.
Rose-Ackerman, S. (1975). 'The economics of corruption', *Journal of Public Economics*, vol. 4(2), pp. 187–203.
Rose-Ackerman, S. (1978). *Corruption: A Study in Political Economy*, New York: Academic Press.
Schulze, G.G. and Frank, B. (2003). 'Deterrence versus intrinsic motivation: experimental evidence on the determinants of corruptibility', *Economic Governance*, vol. 4(2), pp. 143–60.
Sosa, L.A. (2004). 'Wages and other determinants of corruption', *Review of Development Economics*, vol. 8(4), pp. 597–605.