Deterrence, Peer-Effect, and Legitimacy in Anti-Corruption Policy

Boly, Amadou; Gillanders, Robert; Miettinen, Topi

Published in:
Journal of Legal Studies

DOI:
10.1086/703128

Publication date:
2019

Document Version
Peer reviewed version, als known as post-print

This is the post-print version (author's manuscript as accepted for publishing after peer review but prior to final layout and copyediting) of the article:
Amadou Boly, Robert Gillanders, and Topi Miettinen, Deterrence, Contagion, and Legitimacy in Anticorruption Policy Making: An Experimental Analysis, *The Journal of Legal Studies* 48, no. 2 (June 2019): 275-305.
https://doi.org/10.1086/703128

Deterrence, Contagion, and Legitimacy in Anti-Corruption Policymaking: An Experimental Analysis

Amadou Boly[1], Robert Gillanders[2] and Topi Miettinen[3]

February 2019

Abstract

In our framed laboratory experiment, two Public Officials, A and B, make consecutive decisions regarding embezzlement from separate funds. Official B observes Official A's decisions before making his/her own. We find a contagion effect of embezzlement in that facing a corrupt official A increases the likelihood of embezzlement by Official B. Likewise, deterrence matters in that higher detection probabilities significantly decrease the likelihood of embezzlement. Crucially, when the same deterrence policy applies to both officials, detection is more effective in curbing embezzlement if chosen by an honest public official A rather than a corrupt public official A. This legitimacy effect may help explain why anti-corruption policies can fail in countries where the government itself is believed (or known) to be corrupt.

JEL: C91, D73, K42.

Keywords: Contagion-effect; Corruption; Deterrence; Embezzlement; Laboratory Experiment; Legitimacy.

[1] African Development Bank (AfDB), Macroeconomics Policy, Forecasting and Research Department, Abidjan, Cote d'Ivoire. Corresponding author: a.boly@afdb.org.
[2] Dublin City University Business School, Dublin, Ireland, and Hanken School of Economics, Helsinki, Finland. E-mail: robert.gillanders@dcu.ie.
[3] Hanken School of Economics, Department of Economics & Helsinki GSE, P.O. Box 479, FI-00101 Helsinki, Finland. E-mail: topi.miettinen@hanken.fi.
This research project started while Amadou Boly was a Research Fellow at the UNU-WIDER. This article is published here with due acknowledgement of UNU-WIDER, Helsinki, which commissioned the original research under the project Macro-Economic Management (MEM). The views expressed here are those of the author(s) and do not necessarily represent or reflect those of the United Nations University nor the African Development Bank Group.
We thank the Busara Center for Behavioral Economics for conducting the experiments. We would also like to thank the editor, an anonymous referee, participants at the African Development Bank internal seminar series, the 2016 Canadian Economic Association Annual Conference, the 2016 PEGNet Conference, the UNU-MERIT/MGSoG seminar series, and the 2018 Public Choice Society Meetings for comments and suggestions.

1 INTRODUCTION

Corruption has been found to have undesirable effects on key economic metrics such as macroeconomic growth (Mauro 1995), firm growth (Fisman and Svensson 2007), and income equality and poverty (Gupta et al. 2002). On account of these large costs of corruption and the fact that corruption is particularly prevalent in developing and transition economies, anti-corruption laws and policies often constitute important elements in both internally and externally initiated reforms. A commonly advocated anti-corruption approach is deterrence, which is justified by appeals to models of rational criminal behaviour. These models assume that an illegal act, such as corruption, is preferred and chosen if its net expected benefit is higher than that of legal alternatives (e.g. Becker 1968; Eide et al. 2006). As a result, government authorities can increase compliance with the law by increasing the risks (probability of detection) and/or costs (severity of sanctions) associated with illegal acts.

The available experimental evidence suggests that the existence of a probability of detection and punishment can indeed curb corruption (for example: Abbink et al. 2002; Olken 2007; Hanna et al. 2011). In a typical corruption experiment, a control treatment with zero detection probability is compared to an experimental treatment with one positive level of detection probability exogenously imposed by the researcher (see e.g. Abbink et al. 2002; Serra 2012). The present paper builds on this literature by considering environments in which the level of the detection is endogenously chosen, allowing us to analyse the role of legitimacy in anti-corruption policymaking.[4] Specifically, in our framed corruption experiment conducted in Kenya, there are two Public Officials, called A and B. They both receive a salary and are entrusted with separate funds to be spent on social projects. Each public official has the opportunity to embezzle from the fund under his/her control before sending the remaining amount to a recipient. The recipient is different for each public official and is chosen randomly from a list of local NGOs and charities. Embezzlement is inefficient from a social point of view as the amount sent to each recipient is doubled while the amount embezzled is not.

There are two different treatments with a mechanism for detection, wherein detection automatically leads to punishment, which entails a loss of all earnings for the period. In the Endogenous and Discretionary (ED) treatment, Public Official A must choose a level of detection probability which applies only to Public Official B.[5] In the Endogenous and Non-Discretionary (END) treatment, Public

[4] See e.g. Fehr and Rockenbach (2003) and Falk and Kosfeld (2006) for labour market experiments with endogenous monitoring.

[5] This is analogous to a weak institutional environment, with endogenous detection and discretionary punishment institutions; for example, where the judicial and police systems act in the service of the government (as opposed to the state). As a result, opposition leaders are jailed while government supporters are shielded from prosecution.

Official A has the same power over the anti-corruption policy but detection and punishment applies to both public officials.[6]

Authorities are considered legitimate when the public views them as having both the legal and the moral authority for law enforcement (Tyler 2006). Legitimacy enhances compliance with the law even when the likelihood of sanctions is low (Tyler 2006). In contrast, a lack of legitimacy could translate into behaviour contrary to that sought, resulting in non-compliance with the law or even increased criminal behaviour (Kagan and Scholz 1984; Fehr and Rockenbach 2003). We operationalize this concept in our experimental design by allowing the 'public official' who chooses the strength of the detection probability to be corrupt. These decisions are observed by a second public official who then makes his or her own decision regarding corruption.

Our findings suggest that there is a contagion effect in corruption whereby a corrupt policy maker generates corruption in others. We also confirm the deterrent effects of monitoring and punishment in that increasing the detection probability significantly reduces the likelihood and level of embezzlement. However, this raw deterrent effect of monitoring and punishment is only present in an institutional setting in which the policy maker is exempt from the policy's provisions, namely the ED treatment. In a setting with equality before the law (END treatment), the effectiveness of monitoring and punishment on the behaviour of the second public official is found to depend on the legitimacy of the policy maker. Specifically, *ceteris paribus*, we find that in this institutional setting, monitoring and punishment have an effect on the corrupt behaviour of the other official only when the policy maker is honest. In other words, a lack of legitimacy undermines the effectiveness of deterrence as an anti-corruption mechanism. This legitimacy effect is not present in the setting with procedural asymmetry (ED treatment).

We contribute to the literature on the effects of deterrence on corruption in several ways. First, to the best of our knowledge, this is the first experiment to demonstrate that a non-monetary factor such as legitimacy can affect the effectiveness of anti-corruption monitoring and punishment in a significant way, to the extent that the same policy actions produce different outcomes. As deterrence is more effective when chosen by honest officials, legitimacy can reduce the costs of enforcement. This is in line with Levi and Sacks (2009) who argue that citizens who perceive a regime as legitimate are more likely to comply with its precepts (even when the probability that non-compliance would be detected is low). Also, our results may help explain why anti-corruption policies can fail in countries where the government itself is believed (or known) to be corrupt. The governments of many developing and transition economies clearly face such legitimacy concerns. The implications of our findings may

[6] This situation can also be described as a weak institutional environment, with endogenous detection but non-discretionary punishment institutions, for example, when the judicial and police systems work independently, but under 'manipulable' monitoring and punishment institutions.

extend to other policymaking situations where the actions of the policy maker run contrary to the objective of the promoted policy. For example, a policy maker trying to curb tax evasion while putting his/her own wealth in a tax haven may face legitimacy issues that will affect the effectiveness of the said policy. Overall, the implication of our findings should be of interest and practical value to anti-corruption advocates and policy makers.

This paper is organized as follows. Section 2 discusses the relevant literature including the existing experimental literature on the effects of deterrence on corruption. In Section 3, we describe the main features of experimental design and in Section 4, our empirical approach. Section 5 presents the results, which give rise to a theoretical treatment in the online appendix OA. We conclude in Section 6.

2 LITERATURE REVIEW AND FURTHER MOTIVATION

As mentioned above, the potential of monitoring and punishment as an anti-corruption tool has been well studied in the experimental literature. The ground-breaking contribution in this regard can be found in Abbink et al. (2002) who find that even a very small exogenously determined probability of being caught coupled with a severe punishment can significantly and meaningfully reduce bribery. The effectiveness of this type of anti-corruption policy is also evident in the complex experimental setting used in Azfar and Nelson (2007) and Barr et al. (2009). Evidence from the field is provided by Olken (2007) who finds that government audits are effective at reducing corruption in the context of Indonesian infrastructure projects.

However, some studies have reached somewhat different conclusions. Schulze and Frank (2003) conclude that monitoring and punishment damages intrinsic motivation; as they find that monitoring reduces the number of subjects that choose the highest level of bribe but the average bribe actually increases. Serra (2012) finds that while low-level monitoring does not deter corruption alone, it is effective in a mixed top-down and bottom-up accountability system. Overall, these studies suggest that monitoring and punishment can be an important element of an effective anti-corruption strategy. We find that this effect holds in our experimental framework.

We then show that the effectiveness of deterrence can be mitigated by the legitimacy (here being corrupt or honest) of the person enacting the policy. While long and widely studied outside economics (see e.g. Weber 1964; Tyler 1990; or Papachristos et al. 2012), legitimacy has received attention only more recently in economics with a few papers underscoring its relevance theoretically (see e.g. Tyran and Feld 2006; Schnellenbach 2007; Basu 2015; Akerlof 2016) and empirically (Chen 2013). Akerlof (2016) explores the constraints that the need for legitimacy imposes on organizational behaviour outcomes such as the rejection of overqualified workers or above-market-clearing wages. Tyran and Feld (2006) find that formal legal sanctions for non-contribution to a public good game are more effective when they were first legitimated through voting rather than when they are externally imposed. Using a dataset on World War I deserters, Chen (2013) finds evidence that the higher execution rate of Irish soldiers compared to British soldiers by the British, regardless of the crime, stimulated absences; particularly Irish absences. We contribute to the literature on legitimacy by presenting evidence that an anti-corruption policy maker's decision to act corruptly reduces the effectiveness of any given level of detection chosen by the policy maker. This finding could reflect a process in which he is delegitimized in the eyes of others who are subject to the provisions of his policy.

Our paper also relates, to some extent, to experimental studies that show that others' behaviour can influence an individual's own attitudes and behaviour. For example, it has been found that most people contribute more to public goods if others do so (Brandts and Schram 2001; Fischbacher and Gächter 2010) or that tax compliance depends on the behaviour of others in society (e.g. Fortin et al. 2007).[7] In particular, d'Adda et al. (2017) run an experiment that allows for behaviour that is reminiscent of corruption and show that groups with (likely) dishonest leaders are more likely to cheat. Jones and Kavanagh (1996) conclude that the ethical behaviour of employees is influenced by the ethical behaviour of their peers and managers. Similarly, Pierce and Snyder (2008) demonstrate that a firm's ethical norms can influence those of its workers. Specifically, they find that the pass rate of vehicle inspectors adjusts to conform to the norm prevailing at the facility in which they are working. Our experiment provides additional evidence by allowing for the first-moving policy maker to 'set the tone' of the organization that the subjects find themselves operating in. Indeed, Lambsdorff (2015) argues that the 'tone at the top' is '… [m]aybe the most important factor in fighting corruption…' (Lambsdorff 2015: 10).

3 EXPERIMENTAL DESIGN

The data used in this paper were generated from a framed laboratory experiment which was carried out at the Busara Center for Behavioral Economics, in Nairobi, Kenya. Our subjects are mostly university students from the University of Nairobi and they come from a variety of disciplines. In the remainder of this section, we outline our basic procedure and describe our experimental treatments in detail.[8]

3.1 Procedure

Busara staff members read aloud instructions (see OB) at the start of each session. After this, subjects were invited to ask clarification questions and their understanding of the task at hand was then tested with comprehension questions. The duration of each session was roughly one hour.

---

[7] These peer effects extend to various areas such as donations (Shang and Croson 2009; Smith et al. 2015), work effort supply (Falk and Ichino 2006; Bandiera et al. 2010), and legally deviant and norm-breaking behaviours (Bikhchandani et al. 1998; Keizer et al. 2008).

[8] The experimental design is identical to that of Boly and Gillanders (2018).

In our design, we model embezzlement which is a misuse (typically for private gains) of another party's money or property, to which the embezzler has legal access but not legal ownership; in particular, when public or external aid funds are captured by officials or politicians. For example, Svensson and Renikka (2004) show that during 1991–1995, schools in Uganda received, on average, only 13 per cent of funds disbursed by the government; while the majority received zero. Most of the funds for school grants was siphoned by local officials (and politicians). Francken, Minten and Swinnen (2009) also unveil evidence of public funds embezzlement in the education sector of Madagascar. Likewise, in Chad, Gauthier and Wane (2009) show that out of the funds allocated by the Ministry of Health to the regional administrations and primary health centres, amounting to 60 per cent of the Ministry's non-wage recurrent expenditures, only 18 per cent actually reached the regional administrations and only 1 per cent reached the health centres, which are the frontline health service providers. In Indonesia, Suryadarma and Yamauchi (2013) find that only 69 per cent of disbursements of an anti-poverty program actually reached the intended beneficiaries. In all cases above, embezzlers clearly undermine economic development by diverting public resources (including foreign aid) allocated to education, health or poverty-reduction programs.

Our framed laboratory experiment mirrors therefore a situation in which public officials have the opportunity to embezzle public funds. We model this in the following way. Our participants take one of two roles, Public Official A or Public Official B, which they keep throughout the experiment and play a sequential move game. New pairs consisting of one of each type of public official are formed randomly at the start of each round. Payoffs are expressed in terms of Experimental Currency Units (ECU) during the sessions before being converted to Kenyan Shillings at the end of the experiment at a rate of 8ECU to 1Ksh.

Public Official A and Public Official B are each paid a salary of 1,140ECU at the start of every round. They are then each allocated a fund amounting to 2,280ECU, which they are aware is intended to be spent on 'social projects'. Public Official A moves first and has to choose whether to keep 0ECU or 760ECU from the social fund.[9] If Public Official A chooses to keep 760ECU, this amount is added to his payoff for the round. The balance (2,280−Amount Kept) is multiplied by 2 and, after conversion into Kenyan Shillings, is sent to a recipient, called Recipient 1. This recipient is randomly selected at the end of the experiment from a list of local non-governmental organizations (NGOs) and local charity

[9] We restricted Official A's choice in this way so that Official B received a simple binary signal which increased comparability across treatments and individuals. The extent of A's possible embezzlement is similar to some evidence from the field which suggests that only 69% of intended anti-poverty funds in Indonesia reach the intended beneficiaries (Suryadarma and Yamauchi, 2013)

funds. Carrying out our experiment in Kenya and using real donations to local NGOs adds further ecological validity to our study.[10]

After observing the choice of Public Official A, Public Official B makes his/her decisions. Whereas Public Official A faced a binary embezzlement choice, Public Official B can opt to embezzle any whole number between 0 and 2,280 from the social funds under his control. The amount that Public Official B chooses to keep is transferred to his/her private account. As was the case with the funds passed on by Public Official A, the remainder of the fund (2,280−Amount Kept) is doubled, converted into Kenyan Shillings, and sent to a recipient. However, this recipient, called Recipient 2, is different from Recipient 1 and is also randomly selected from a list of local NGOs and local charity funds.[11]

Each session lasted for 40 independent rounds. In each round, each A was randomly matched (with equal probability) with a new B (random strangers). Public Official A was not informed about the choice made by Public Official B for the first 20 rounds but for the final 20 rounds he or she observed how much Public Official B transferred to Recipient 2. Once all 40 rounds were complete, the subjects were asked to answer a survey that included questions on demographics, socio-economic status, attitudes to and experiences of corruption. Finally, one of the rounds was randomly drawn, and the payments to the participant and the recipient organization were carried out according to the outcome of that round.

3.2 Treatments

We implemented two main experimental treatments with detection and punishment. The first treatment, which we call Endogenous and Discretionary (ED), gives Public Official A the responsibility of choosing at no cost a probability of detection and punishment which must be selected from the values 0 per cent, 5 per cent, 10 per cent, 15 per cent, 20 per cent, 25 per cent, or 30 per cent. Detection is discretionary in that this mechanism only applies to Public Official B. This setting captures a weak institutional environment in which Public Official A faces no risk when engaging in corruption while Public Official B does. In other words, the principle of equality before the law does not hold. Public Official B acts only after he or she has observed the choices made by Public Official A with respect to

---

[10] As discussed by Abbink and Serra (2012), the use of NGOs or charities as recipients of non-embezzled funds is a useful way to model the negative impact of corruption on public well-being. There are, however, two potential issues. First, there may be some loss of control regarding a subject attitude towards a particular NGO or charity. To mitigate this effect, we simply informed the subject that the local NGO would be drawn randomly from a list. Second, subjects' donation behaviour outside the experiment is unknown, if the subject had already given to a charity recently or if he/she decides to be corrupt in the experiment and give later. This may, however, be a non-optimal choice given the multiplicative factor in our experiment.

[11] We chose two different recipients, one for Public Official A and another for Public Official B, in order to make sure that A's choice does not substitute for the choice made by B or influence its marginal effect (see e.g. Francois 2000, 2003 for further theorizing as to why such issues may matter, though in a slightly different context).

the level of detection probability and embezzlement. Detection means that Public Official B loses all earnings for that round.

Treatment two, Endogenous and Non-Discretionary (END), also gives Public Official A the power to select the likelihood of detection (at no cost and from among the same values as above) but this probability applies to both public officials. That is to say that the enforcement of the law is non-discretionary. This is a stronger institutional setting in the sense that equality before the law is a feature but note that the framework is still manipulable. A public official who is detected embezzling loses his/her salary for the round and the amount embezzled in that round. Independent and separate draws are carried out for each public official meaning that in situations where both are corrupt one can be detected and punished while the other is not. Once again, Public Official B observes the choices of Public Official A before making his own decision.

An additional treatment, Exogenous and Non-Discretionary (XND), exogenously sets the probability of detection at 30 per cent and applies it to both public officials. As with END, independent and separate draws are carried out for each public official. This treatment is used to test the robustness of our results relative to legitimacy.

The monitoring mechanism functions as follows. Once the public officials have made their decisions, the computer generates a random number between 1 and 100. In treatments where both public officials are subject to the mechanism, separate and independent draws are made for each public official. Say a public official opts to keep a positive amount of the social fund for himself and the probability of detection that has been chosen (or exogenously imposed) is W per cent (W equals 0, 5; 10; 15; 20; 25; or 30). If the randomly generated number for that public official falls between 1 and W (inclusive) then the public official's decision to embezzle is detected and punished. For that specific round, the public official loses both his salary and the embezzled funds but this does not affect the payoffs in any other round. If the randomly generated number falls between (W+1) and 100 (inclusive) then the public official in question gets to keep both his salary and the amount kept. The detection and punishment mechanism operates identically in all treatments. The probability value is chosen by Public Official A in the ED and END treatments and is exogenously set at 30 per cent in the XND treatment.

3.3 Participants and payoffs

Across the three treatments, 198 subjects participated at the Busara Center for Behavioral Economics, in Nairobi, Kenya. Half took the role of Public Official A and the other half took the role of Public Official B. Sixty-four subjects served in the ED treatment 68 in the END, and 66 in XND treatment. Table 1 presents summary statistics for our sample of Public Officials B. They are roughly 21 years old on average and most of them are male. Economics majors make up large proportions of the sample in all of our treatments. The average monthly expenses are around 9500Ksh which is equivalent to around €80.[12] Thus the average earnings from the experiment as described below represent a significant sum to our participants. There are some differences across treatments in some of these characteristics. Though the subjects in ED and END, the treatments that are of particular interest for this paper, are rather similar, we include these factors in our regression analysis.

While survey data on an individual's relationship to corruption may be prone to certain biases, subjects' responses summarized in Table 1 suggest that our subjects have an experience of corruption. In addition to asking about their experiences of corruption, our survey also probed our subjects' attitudes to and understanding of corruption. In terms of perceptions of corruption, 5% of public official Bs think that a few government officials are involved in corruption, 78% think that some of them are, and the remainder think that all of them are. Most of our subjects (63%) most often hear about corruption in the context of scandals involving politicians and bureaucrats. 27% of our subjects most often hear about corruption in the context of harassment bribes levelled on ordinary people by government officials and 8% in the context of scandals involving companies and rich individuals. 89% agree that Kenyan law is such that both bribe takers and givers are acting illegally. 100% of public official Bs profess to agree with the statement "it is always wrong for a government official to take a bribe." While survey data on an individual's relationship to corruption may be prone to certain biases, these responses suggest that our subjects have an understanding of the practical and moral facets of corruption.

[**Table 1 about here**]

One period from the 40 was chosen at random to calculate the payoffs.[13] With an exchange rate of 8ECU = 1KSh, the average total earnings (i.e. salary plus embezzlement) for those in the role of Public Official A was 208KSh in the ED treatment, 194KSh in the END treatment, and 145KSh in the XND treatment. Public Officials B earned 307KSh in the ED treatment, 306KSh in the END treatment, and 185KSh in the XND treatment. In addition, each subject received a fixed payment of 400Ksh for their participation. The NGOs Green Belt Movement and Impacting Youth Trust (Mathare) served as Recipient 1 and Recipient 2 respectively after being randomly drawn from a list of local NGOs. 48,285KSh were transferred to Recipient 1 and 58,900KSh to Recipient 2 after the experiment had ended. These amounts were calculated by taking the total amount sent to Recipient 1 (Recipient 2) by those in the role of

[12] This information comes from a post experiment survey question and refers to the participant's own expenses. We do not specify what the participant should include in the calculation.
[13] Even though we only use the first 20 rounds for this analysis, all 40 were used for the purposes of calculating payoffs so that participants had an incentive to treat all rounds equally seriously.

Public Official A (Public Official B) using one randomly determined period per subject and an exchange rate of 8ECU = 1KSh.[14]

4 EMPIRICAL APPROACH

In this paper, we are interested in explaining the corrupt behaviour of Public Official B.[15] We also restrict our analysis to the first 20 rounds; given that in the final 20 rounds, Public Official A was informed about Public Official B's embezzlement decision. We analyse the data using summary statistics and statistical tests (mainly Mann-Whitney tests with individual average choices as independent units of observations), followed by regression analyses. In our regression analyses, we estimate several equations:

$$y_{it} = \alpha + \beta END + \rho Honest\_A_{it} + \delta Detection_{it} + \mu_i + \varepsilon_{it} \quad (1)$$

$$y_{it} = \alpha + \beta END + \rho Honest\_A_{it} + \delta Detection_{it} + \gamma(Honest\_A_{it} \times Detection_{it}) + \mu_i + \varepsilon_{it} \quad (2)$$

$$y_{it} = \alpha + \beta END + \rho Honest\_A_{it} + \delta Detection_{it} + \gamma(Honest\_A_{it} \times Detection_{it}) + \varphi(Honest\_A_{it} \times END) + \psi\,(Detection_{it}\ \mathrm{x}\ END) + \omega(Honest\_A_{it} \times Detection_{it}\ x\ END) + \mu_i + \varepsilon_{it} \quad (3)$$

$$y_{it} = \alpha + \rho Honest\_A_{it} + \delta Dummy\_Detection_{it} + \mu_i + \varepsilon_{it} \quad (4)$$

$$y_{it} = \alpha + +\rho Honest\_A_{it} + \delta Dummy\_Detection_{it} + \gamma(Honest\_A_{it} \times Dummy\_Detection_{it}) + \mu_i + \varepsilon_{it} \quad (5)$$

y_{it} is the dependent variable. It is a dummy variable which equals 1 when Official B embezzles (0 otherwise) though we also run the regressions to explain the amount, between 0 and 2,280ECU, embezzled by Official B. The variable END is a dummy variable which equals 1 for the END treatment and 0 otherwise. $Honest\_A_{it}$ is a dummy variable which equals 1 when Official A is honest and 0 otherwise. $Detection$ is the level of detection chosen by Official A and is treated as a continuous variable; while for $Dummy\_Detection_{it}$ a dummy is created for each level of detection. The constant term, individual effect term, and the error term are respectively α, μ_i and ε_{it}. The subscript i is for individual subjects and t denotes the round. Equations 1 to 3 apply only to the pooled data regression. Equations 1 to 3 (without the END variable) and Equations 4 to 5 apply to the individual treatment regressions.

As stated above, the main variable of interest is the likelihood that a Public Official B is corrupt; but we also briefly discuss results which examine the amount embezzled by a corrupt Public Official B. We

[14] Each participant was notified once the funds were transferred. The participants also received a text message notifying them that there were receipts and formal letters of NGO payments available for viewing and collection at Busara's offices if they so wished.
[15] The behavior of Public Official A is studied in a companion paper (see Boly and Gillanders, 2018).

use a random-effects Logit model to analyse Public Official B's decision to embezzle or not. All standard errors are clustered at the session level.

5 MAIN RESULTS

We expect three distinct effects. Firstly, we expect to see a deterrence effect. Such an effect will be evident if we find a downward-sloping relationship between the level of detection probability and corrupt behaviour. Secondly, we expect to see a contagion effect whereby a Public Official B who witnesses corrupt behaviour on the part of a Public Official A will follow suit. A difference in the average level of corrupt outcomes by the type of Public Official A will be evidence of this contagion effect. Finally, and most importantly, we are interested in the possibility that policies originating from a corrupt official are less effective than those promulgated by an honest policy maker. Such a legitimacy effect could present itself in two ways. Firstly, if the slope of the deterrence effect differs by type of Public Official A we would have evidence of a legitimacy effect. The effect of changes in detection would be different depending on the behaviour of Public Official A. Secondly, and conceptually equivalently, differences in the overall marginal effects of Public Official A's type (honest or corrupt) at each level of detection probability would be evidence of the same policy having different effects depending on the behaviour of the policy maker.

5.1 Share of corrupt decisions by Officials B

Summary statistics on the average share of corrupt decisions made by Public Official B are given in Table 2. Note that individual average choices are used as independent units of observations. The average shares of corrupt decisions in the ED, END, and XND treatments are respectively 88 per cent, 91 per cent, and 79 per cent. Compared to the XND treatment, we find that corruption is significantly greater in the END treatment (p-value = 0.04, two-sided Mann-Whitney) but not in the ED treatment. No significant difference is found between the ED and the END treatment (p-value = 0.31, Mann-Whitney).

[**Table 2 about here**]

Figure 1 plots, overall and for each treatment, the percentage of Public Officials B who made corrupt decisions at each point on our detection scale. It also includes reference points for the XND treatment where this probability was exogenously set. The share of corrupt decisions conditional on Public Official A being honest is indicated by solid line while the share of corrupt decisions conditional on A being corrupt is indicated by dashed line. Bubble size corresponds to the percentage of times a given level of detection was chosen. XND H and XND C show the honest and corrupt choices by Official A in the XND treatment. The pooled data suggest that deterrence is effective in that no matter the type of

Public Official A, we see a downward slope indicating that higher levels of detection lower the likelihood of a corrupt Public Official B. However, the type of Public Official A does seem to matter in that the line for honest Public Official As is always below the line for corrupt Public Officials A.

[**Figure 1 about here**]

In the ED treatment, a negative relationship between the probability of detection and the share of corrupt Public Officials B can be observed and the relationship appears similar for each type of Public Official A. This suggests that the decision by Public Official A to be corrupt or honest does not influence the effectiveness of the chosen level of detection in terms of deterring Public Official B from embezzlement. In contrast, in our END treatment, there is a clear difference depending on whether Public Official A is corrupt or honest. At each point on our detection scale, the share of corrupt Public Officials B is appreciably lower when the probability has been chosen by an honest Public Official A as opposed to a corrupt Public Official A. In addition, increasing the probability does not seem to dissuade Public Official B from being corrupt when that probability has been chosen by a corrupt Public Official A. When Public Official A is honest, we do see some evidence of a downward slope. These differences are indicative of a legitimacy effect.

We now proceed to a regression analysis to test the statistical significance and magnitude of these apparent effects, using a random-effects Logit model to analyse Public Official B's decision to embezzle or not. The results are in line with our graphical analysis for the most part.

5.1.1 Pooled Data

Figure 1 suggests that different relationships are at play in the ED and END treatments. To establish this formally, in Table 3 we pool the data from the ED and END treatments. Column 1 looks at the main effects of Public Official A's behaviour and the level of detection probability on the likelihood that Public Official B acts corruptly. We find a significant and negative relationship between honest behaviour by Public Official A and the likelihood of embezzlement by Public Official B (at the 1 per cent level). We also find that deterrence has a negative and statistically significant effect on the likelihood that Public Official B acts corruptly (at the 10 per cent level). This deterrence effect is in line with much of the experimental literature outlined above. Our pooled data thus suggests the presence of a contagion effect and of a deterrence effect.

[**Table 3 about here**]

The regression framework allows us to go deeper and study the interaction of Public Official A's behaviour and the level of detection that he chooses. Column 2 of Table 3 includes the main effects and the interaction effect (Equation 2). The reference group is a Public Official B in the ED treatment who is paired with a corrupt Public Official A who has selected a zero probability of detection. The coefficient for 'detection level' is negative suggesting that detection is effective when chosen by an honest Public Official A but the effect is not significant at traditional levels. The coefficient for the dummy representing an honest Public Official A is negative (and significant at the 10 per cent level) indicating that the likelihood of embezzlement by Public Official B is higher when facing a corrupt Public Official A instead of an honest Public Official A. Graphically, this would mean that the line of predicted Logit values for corrupt Public Officials A will lie above the line for their honest counterparts.

In Column 3, we run a three-way interaction between the END treatment, Public Official A's type and detection levels. The main objective is to see if there are differences between the ED and the END treatments. The results suggest that while detection has a negative and significant effect on the likelihood of embezzlement (see "Detection Level" coefficient), its impact is lower in the END treatment compared to the ED treatment (see "END x Detection Level" coefficient). There is also a significant difference in the impact of detection when chosen by an honest official A (compared to a corrupt official A) in the END treatment (see "END x A's Behaviour=1 x Detection Level" coefficient); while such a difference in not found in the ED treatment (see "A's Behaviour=1 x Detection Level" coefficient). A test of the restriction that the coefficients on the treatment effect and interactions involving END are jointly equal to zero further supports the notion that the END and ED treatments generate different behaviours (p= 0.000). To explore further the differences between the ED and END treatments, we ran our analysis by individual treatments in the sections below.

5.1.2 Individual Treatments Data

We already noted above that Figure 1 as well as the results in Column 3 of Table 3 suggest that the effects of interest are heterogeneous across treatments. Table 4 therefore analyses the ED and END treatments individually. We begin again by looking at the main effects in columns 1 and 2. In the ED treatment, deterrence seems to be at work, as the coefficient on detection level is negative and significant at the 1 per cent level (column 1). However, we do not see evidence of Official A's behaviour having an effect in this institutional setting. Things are different in the institutional setting represented by the END treatment where a contagion effect appears to be at work, with the coefficient for Public Official A's behaviour being negative and significant at the 1 per cent level, while the coefficient for detection level is not significant (column 2).

[**Table 4 about here**]

The full model presented in Column 3 of Table 4 shows that Official A's behaviour has no significant effect on Official B's likelihood to embezzle in the ED treatment but that there is a deterrence effect. Thus, we conclude that in the particular institutional framework captured by the ED treatment, monitoring and punishment are effective and the behaviour of the rule maker is largely irrelevant. These conclusions are in line with the graphical analysis in Figure 1 and the findings of the literature on exogenously given detection levels that began with Abbink et al. (2002).

In the END treatment (Table 4, Column 4), we find no evidence of a direct effect of Official A's behaviour on that of Official B and we can also see that detection levels chosen by a corrupt Public Official A have no significant effect on the likelihood of embezzlement by Public Official B. However, the negative and significant interaction term suggests that stronger detection choices made by an honest Public Officials A tend to decrease the likelihood of embezzlement by Public Official B. This is consistent with the idea of a legitimacy effect. Once again, these conclusions are fully consistent with the graphical analysis presented in Figure 1.

The choices of probability can be sharply different between the ED and END treatments, as suggested by Figure 2. As is demonstrated in Boly and Gillanders (2018), there is a tendency to choose low levels of detection in the END treatment and higher levels in the ED treatment.

[**Figure 2 about here**]

To account for these differences in probability choices by Public Official As, the last four columns of Table 4 include dummies for each level of detection. Columns 5 and 6 of Table 4 support the conclusions of the main effects models which treated detection as a continuous variable. As we have reason to believe that interactions may be important in the END treatment at least, more interesting are the final two columns of Table 4. In Column 7, we do find a contagion effect in the ED treatment but only when the detection level is zero. That is to say that when participants are treated equally, as they always are in END, there is a simple contagion effect. At all positive detection levels, monitoring and punishment is an effective deterrent to corrupt behaviour but this effect is in general not contingent on Public Official A's behaviour. The opposing sign and comparable magnitude of the interaction coefficients relative to the effect of Public Official A's behaviour show this. As an example, consider the 5% detection level, if a corrupt Public Official A chooses that detection level, the deterrence effect is captured by the coefficient 2.774. If an honest Public Official A chooses that detection level, the deterrence effect is captured by the sum of main and the interaction effects, i.e. 2.774+1.483-1.750, which is approximately equal to the deterrence effect for a corrupt Public Official A. A similar reasoning holds for most detection levels. Finally, Column 8 shows that there is legitimacy effect in the END in that the effectiveness of a given level of detection in terms of deterring corruption is conditional on the behaviour of Public Official A. On average, the anticorruption policy is only effective if the policymaker is herself honest.

Our analysis of the amount embezzled by Public Official B is presented in the online appendix OC. These results are very similar to those for the likelihood of embezzlement.[16] However, it should be noted that since for a given probability of detection embezzling 1ECU is as likely to result in punishment as embezzling 2,280ECU, it is not obvious that higher probabilities should lead to an individual embezzling a lower amount. The model presented in OA offers insights into this observed behaviour.

5.1.3 Discussion

Our results thus indicate that in the END treatment, a higher probability of detection and punishment can effectively deter corruption, but only when the policy is put in place by a policy maker who is himself 'clean'. Thus, in institutional settings in which equality before the law is observed, it is vital that those at the top set the right 'tone', if endogenously chosen anti-corruption measures of this type are to be effective. The results on the effects of institutional settings on legitimacy can be related to the perceived procedural fairness of the system. Indeed, the procedural fairness literature suggests that legitimacy springs from a shared perception between all relevant parties and outsiders about the fairness of the procedures applied—though outcomes may be unequal, at least everyone acts under a common set of rules that equally apply to all (Lind 2001; Tyler 2004). Perceived procedural fairness promotes compliance with the verdicts of the authority. Since the seminal work of Thibaut and Walker (1975), various studies have come to establish and support these views (see e.g. Lind 2001; Falk et al. 2003; Tyler 2004; Bolton et al. 2005). There are several additional channels through which the legitimacy effect could manifest itself. The first is simply fairness (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). Since in END the same behaviour by B and A implies identical outcomes, an inequity averse B prefers to behave as A does. Conformism and norms could also play a role in that B simply prefers behaving like A when the social context and thus the related norms are identical (Bernheim, 1994; Bardsley and Sausgruber, 2005; Bicchieri, 2006; Kimbrough and Vostroknutov, 2016).[17] In END where payoff implications are more similar, if B believes that A has superior information about the optimal behaviour then she may herd and choose the same action as A (Banerjee, 1992; Bikhchandani et al 1992). The END treatment results could also be explained by a simple signalling story. If a policy maker

[16] As the degree of embezzlement had to be chosen from a restricted range of values (between 0 and 2,280ECU), we employ a random-effects two-sided Tobit model for this analysis.
[17] See the online appendix OA for a theoretical model along these lines that isbased on Bicchieri (2006); Battigalli and Dufwenberg (2007), Lopez-Perez (2008) and Miettinen (2013).

(here, Public Official A) chooses an anti-corruption policy but still engages in corruption, this might signal to others (here, Public Official B) that the policy is toothless. In other words, the probability of detection is lower than what the policy maker has actually announced. While our experimental design rules out this signalling story (there is no private information about the detection probability), the possibility remains that heuristics that are valid in the real world continue to operate in lab settings where they are invalid. [18]

Recent self-image and social image models combine informational and normative explanations. These models can contribute to accounting for some puzzling features in our experiment, including the contagion effect and the fact that Public Official Bs who embezzle still transfer some of the funds to charitable organizations. These features cannot be explained by traditional models (Becker 1968). Yet, the observed patterns appear broadly consistent with a model where, in addition to the pecuniary incentives, Public Official B does not want to entertain an immoral social-image (Dufwenberg and Dufwenberg, 2018; Gneezy et al. 2018) and the incentive to do so may depend on what she thinks about the type of the audience (Ellingsen and Johanneson, 2008). These two features could explain a contagion effect (Public Official B cares less about a moral social image if the audience appears immoral) and the fact that not all funds are embezzled (see Dufwenberg and Dufwenberg, 2018; Gneezy et al. 2018). In addition, the observed legitimacy effect when rules apply to both officials and Official A is corrupt would appear if the signal value arising from an immoral audience is particularly strong, and thus the social image motivation is entirely absent. From this perspective, perhaps the explanation for our results lies in the fact that the ultimate authority that chooses the symmetric or asymmetric rules is not the first-moving official but the experimenter. Given the importance of the institutional framework in our results, future research could usefully endogenize the choice of whether the fight against corruption applies to all or whether immunity is granted only to the top level which has an influence on the anti-corruption institutions. This would allow one to study the social signalling effect on social image motivation in an even richer framework.

5.1.4 Robustness Check

Given that we find a legitimacy effect only in the END treatment, we focus our robustness check on this treatment by comparing it with the XND treatment. We only consider END data points where 30% was chosen. As a result, the only difference between these two treatments is that the detection probability is chosen by Public Official A in the END treatment, while it is exogenously set at 30 per cent in the XND treatment. The idea is to check whether Public Official A's behaviour has an impact on the effectiveness of her detection choices, focusing on the highest detection level available.

[18] We are indebted to the editor for this point.

When the detection probability is 30 per cent, the average shares of corrupt decisions are 89 per cent and 79 per cent in the END and XND treatments respectively. The results in Table 5 suggest a significantly higher likelihood of embezzlement by corrupt Officials B in the END treatment, compared to the XND treatment. In the XND treatment where detection is set exogenously, we find no significant difference in the likelihood of embezzlement by Public Official B, when facing a corrupt or an honest Public Official A (see coefficient for "A's Behaviour (Honest=1)"). In contrast, in the END treatment, we find that facing an honest Public Official A significantly decreases the likelihood of embezzlement by Public Official B at the 1 per cent level (see coefficient for "END x A's Behaviour=1" in Table 5). In other words, choosing the highest level of detection available in our setting increases (decreases) the likelihood of embezzlement when the decision-maker is corrupt (honest), compared to a situation where this level of detection is set exogenously. Such a result shows an interaction between Public Official A's behaviour and his/her detection choices further confirming the existence of a legitimacy effect.

[**Table 5 about here**]

6 CONCLUDING REMARKS

This paper draws on data obtained from a framed laboratory experiment carried out in Kenya to examine the roles of contagion effects, deterrence effects, and legitimacy effects in the fight against corruption. In our regression analysis, we labelled the main effect of Public Official A's behaviour as a contagion effect, and that of the detection level as a deterrence effect. The legitimacy effect refers to the interaction between A's behaviour and detection levels. This captured the idea that deterrence can be less effective when chosen by a corrupt Public Official A.

Crucially, we found that the importance of these effects depended on the institutional framework in which our 'public officials' are operating. When policy makers are exempt from their own laws we find evidence of a strong deterrence effect in that a greater chance of being detected and punished reduces the likelihood and the extent of corruption. This effect does not depend on the behaviour of the policy maker. In settings in which equality before the law is observed and policy makers are liable to be caught in their own net, we find that detection policies are only an effective deterrent when promulgated by honest policy makers. We also found that externally imposed rules may be superior to equally stringent rules originating from a corrupt internal policy maker. Once again, this existence of this effect was dependent on the internal policy maker being subject to the provisions of the policy.

Our findings offer several important implications in fighting corruption for policy makers and other interested parties—subject to the usual external validity caveats of experimental economics which we will address briefly below. Firstly, our results add further evidence as to the potential for detection and punishment mechanisms to play a role in curbing corruption. Our findings of a contagion effect suggest that creating a culture of honesty among the top-rank officials in systems such as the one in our experiment can have knock-on, or perhaps trickle-down, effects on others within the organization or society (Moxnes and Van det Heijden 2003; Güth et al. 2007; Levati et al. 2007; Cappelen et al. 2015). Our finding of a strong legitimacy effect adds more weight to this argument in that fostering such an honest ethic may result in the same policy being more effective. Moreover, internally generated anti-corruption detection mechanisms will only be effective in institutional settings with equality before the law when the policy maker is honest. If this condition is not met, exogenously imposed rules are preferable.

As for the applicability of our lab results to the 'real world', a few arguments can be noted. Firstly, while the magnitude of any given effect may not carry over from the lab to the field, for an experiment to be useful we need only qualitative external validity (Camerer 2014; Kessler and Vesterlund 2014). Secondly, external validity problems are not unique to experimental results (Falk and Heckman 2009; Kessler and Vesterlund 2014). Finally, the very nature of corruption makes data difficult to collect and means that there are practical and ethical challenges to evaluating anti-corruption policies and institutions in a field setting (Armantier and Boly 2012). Laboratory studies allow us to get a handle on what might work in a low-cost and ethically feasible setting (Dusek et al. 2005; Abbink 2006).

References

Abbink, Klaus. 2006. Laboratory Experiments on Corruption. In Susan Rose-Ackerman (ed.), *The Handbook of Corruption*. Cheltenham, UK, and Northampton, US: Edward Elgar Publishers.

Abbink, Klaus, Bernd Irlenbusch, and Elke Renner 2002. An Experimental Bribery Game. *Journal of Law, Economics, and Organization*, 18(2): 428–454.

Abbink, Klaus, and Danila Serra. 2012. Anticorruption Policies: Lessons from the Lab. *Research in Experimental Economics*, 15: 77–115.

Akerlof, Robert. 2016. The Importance of Legitimacy. *The World Bank Economic Review*. 30: S157-S165 doi: 10.1093/wber/lhw009.

Armantier, Olivier, and Amadou Boly. 2012. On the External Validity of Laboratory Experiments on Corruption. In Danila Serra and Leonard Wantchekon (eds), *New Advances in Experimental Research on Corruption (Research in Experimental Economics)*, Vol. 15: 117–144. Bingley, UK: Emerald Group Publishing Limited.

Azfar, Omar, and William Robert Nelson. 2007. Transparency, Wages, and the Separation of Powers: An Experimental Analysis of Corruption. *Public Choice*, 130(3–4): 471–493.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2010. Social Incentives in the Workplace. *Review of Economic Studies*, 77(2): 417–458.

Banerjee, Abhijit V. 1992. A simple model of herd behavior. *The quarterly journal of economics*, 107(3), 797-817.

Bardsley, Nicholas, and Rupert Sausgruber. 2005. Conformity and reciprocity in public good provision. Journal of Economic Psychology, 26(5), 664-681.

Barr, Abigail, Magnus Lindelow, and Pieter Serneels. 2009. Corruption in Public Service Delivery: An Experimental Analysis. *Journal of Economic Behavior & Organization*, 72(1): 225–239.

Basu, Kaushik. 2015. The Republic of Beliefs: A New Approach to Law and Economics. World Bank Policy Research Working Paper 7259. Washington, DC: World Bank.

Battigalli, Pierpaolo, and Martin Dufwenberg. 2007. Guilt in games. *American Economic Review*, 97(2), 170-176.

Beams, Joseph D, Robert M. Brown, and Larry N. Killough. 2003. An Experiment Testing the Determinants of Non-compliance with Insider Trading Laws. *Journal of Business Ethics*, 45(4): 309–323.

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

Bernheim, B. Douglas. 1994. A theory of conformity. *Journal of Political Economy*, 102(5), 841-877.

Bicchieri, Cristina. 2005. *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch. 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5), 992-1026.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch. 1998. Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades. *The Journal of Economic Perspectives*, 12(3): 151–170.

Bolton, Gary E., Jordi Brandts, and Axel Ockenfels. 2005. Fair Procedures: Evidence from Games Involving Lotteries. *Economic Journal*, 115(506): 1054–1076.

Bolton, Gary E., and Axel Ockenfels. 2000. ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, 90(1), 166-193.

Boly, Amadou, and Robert Gillanders. 2018. Anti-corruption Policy-making, Discretionary Power, and Institutional Quality. *Journal of Economic Behaviour and Organisation*, 152, 314-327

Brambor, Thomas, William Roberts Clark, and Matt Golder. 2006. Understanding Interaction Models: Improving Empirical Analyses. *Political Analysis*, 14: 63–82.

Brandts, Jordi, and Arthur Schram. 2001. Cooperation and Noise in Public Goods Experiments: Applying the Contribution Function Approach. *Journal of Public Economics*, 79(2): 399–427.

Camerer, Colin. 2014. The Promise and Success of Lab-Field Generalizability in Experimental Economics: A Reply to Levitt and List. In Guillaume Fréchette and Andrew Schotter (eds), *The Methods of Modern Experimental Economics*. Oxford: Oxford University Press.

Cappelen, Alexander W., Bjørn-Atle Reme, Erik Ø. Sørensen, and Bertil Tungodden. 2015. Leadership and Incentives. *Management Science*, 62(7), 1944-1953.

Chen, Daniel L. 2013. The Deterrent Effect of the Death Penalty? Evidence from British Commutations during World War I. NBER Working Paper, 2013. Cambridge, MA: National Bureau of Economic Research.

d'Adda, Giovanna, Donja Darai, Nicola Pavanini, and Roberto A. Weber. 2017. Do Leaders Affect Ethical Conduct?. *Journal of the European Economic Association*, 15(6), pp.1177-1213.

Dufwenberg, Martin, and Martin A. Dufwenberg. 2018. Lies in disguise–A theoretical analysis of cheating. *Journal of Economic Theory*, *175*, 248-264.

Dusek, Libor, Andreas Ortmann, and Lubomir Lizal. 2005. Understanding Corruption and Corruptibility through Experiments: A Primer. *Prague Economic Papers*, 14(2), 147–162.

Eide, Erling, Paul H. Rubin, and Joanna M. Shepherd. 2006. *Economics of Crime*. Boston, MA: Now Publishers Inc.

Ellingsen, Tore, and Magnus Johannesson. 2008. Pride and prejudice: The human side of incentive theory. *American Economic Review*, 98(3), 990-1008.

Falk, Armin, Ernst Fehr, and Urs Fischbacher. 2003. On the Nature of Fair Behavior. *Economic Inquiry*, 41(1): 20–26.

Falk, Armin, and James J. Heckman. 2009. Lab Experiments Are a Major Source of Knowledge in the Social Sciences. *Science*, 326: 535–538.

Falk, Armin, and Andrea Ichino. 2006. Clean Evidence on Peer Effects. *Journal of Labor Economics*, 24(1): 39–57.

Falk, Armin, and Michael Kosfeld. 2006. The Hidden Costs of Control. *American Economic Review*, 96(5): 1611–1630.

Fehr, Ernst, and Bettina Rockenbach. 2003. Detrimental Effects of Sanctions on Human Altruism. *Nature*, 422(6928): 137–140.

Fehr, Ernst, and Klaus M. Schmidt. 1999. A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3), 817-868.

Fisman, Raymond, and Jakob Svensson. 2007. Are Corruption and Taxation Really Harmful to Growth? Firm Level Evidence. *Journal of Development Economics*, 83(1): 63–75.

Fischbacher, Urs, and Simon Gachter. 2010. Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments. *American Economic Review*, 100(1): 541–556.

Fortin, Bernard, Guy Lacroix, and Marie-Claire Villeval. 2007. Tax Evasion and Social Interactions. *Journal of Public Economics*, 91(11): 2089–2112.

Francois, Patrick. 2000. "Public Service Motivation" as an Argument for Government Provision. *Journal of Public Economics*, 78(3): 275–299.

Francois, Patrick. 2003. Not-For-Profit Provision of Public Services. *Economic Journal*, 113(486): C53-C61.

Francken, Nathalie, Bart Minten, and Johan FM Swinnen. 2009. Media, Monitoring, and Capture of Public Funds: Evidence from Madagascar. *World Development*, *37*(1), 242-255.

Gauthier, Bernard, and Waly Wane. 2009. Leakage of Public Resources in the Health Sector: An Empirical Investigation of Chad. *Journal of African Economies*, *18*(1), 52-83.

Gneezy, Uri, Agne Kajackaite, and Joel Sobel. 2018. Lying Aversion and the Size of the Lie. *American Economic Review*, 108(2), 419-53.

Gupta, Sanjeev, Hamid Davoodi, and Rosa Alonso-Terme. 2002. Does Corruption Affect Income Inequality and Poverty?. *Economics of Governance*, 3(1): 23–45.

Güth, Werner, M. Vittoria Levati, Matthias Sutter, and Eline Van Der Heijden. 2007. Leading by Example with and without Exclusion Power in Voluntary Contribution Experiments. *Journal of Public Economics,* 91(5): 1023–1042.

Hanna, Rema, Sarah Bishop, Sara Nadel, Gabe Scheffler, and Katherine Durlacher. 2011. The Effectiveness of Anti-corruption Policy. EPPI Centre Report (No. 1909). London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Jones, Gwen E., and Michael J. Kavanagh. 1996. An Experimental Examination of the Effects of Individual and Situational Factors on Unethical Behavioral Intentions in the Workplace. *Journal of Business Ethics*, 15(5): 511–523.

Kagan, Robert A., and John T. Scholz. 1984. The Criminology of the Corporation and Regulatory Enforcement Strategies. In Keith Hawkins and John M. Thomas (eds), *Enforcing Regulation*. Boston: Kluwer-Nijhoff.

Kessler, Judd B., and Lise Vesterlund. 2014. The External Validity of Laboratory Experiments: Qualitative rather than Quantitative Effects. In Guillaume Fréchette and Andrew Schotter (eds), *The Methods of Modern Experimental Economics*: Oxford: Oxford University Press.

Keizer, Kees, Siegwart Lindenberg, and Linda Steg. 2008. The Spreading of Disorder. *Science*, 322(5908): 1681–1685.

Kimbrough, Erik O., and Alexander Vostroknutov. 2016. Norms make preferences social. *Journal of the European Economic Association*, 14(3), 608-638.

Graf Lambsdorff, Johann. 2015. Preventing Corruption by Promoting Trust: Insights from Behavioral Science. Vol. 69, No. 15 Passauer Diskussionspapiere: Volkswirtschaftliche Reihe. Passau: Universität Passau.

Lahno, Amrei M., and Marta Serra-Garcia. 2015. Peer Effects in Risk Taking: Envy or Conformity?. *Journal of Risk and Uncertainty*, 50(1): 73–95.

Levati, M. Vittoria, Matthias Sutter, and Eline Van der Heijden. 2007. Leading by Example in a Public Goods Experiment with Heterogeneity and Incomplete Information. *Journal of Conflict Resolution*, 51(5): 793–818.

Levi, Margaret, and Audrey Sacks. 2009. Legitimating Beliefs: Sources and Indicators. *Regulation & Governance*, 3(4): 311–333.

Lind, E. Allan. 2001. Fairness Heuristic Theory: Justice Judgments as Pivotal Cognitions in Organizational Relations. *Advances in Organizational Justice*, 56:88.

López-Pérez, Raúl. 2008. Aversion to norm-breaking: A model. *Games and Economic Behavior*, 64(1), 237-267.

Mauro, Paolo. 1995. Corruption and Growth. *Quarterly Journal of Economics*, 681–712.

Miettinen, Topi. 2013. Promises and conventions–an approach to pre-play agreements. *Games and Economic Behavior*, 80, 68-84.

Moxnes, Erling, and Eline Van der Heijden. 2003. The Effect of Leadership in a Public Bad Experiment. *Journal of Conflict Resolution*, 47(6): 773–795.

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

Papachritos, Adrew V., Tracey L. Meares, and Jeffrey Fagan. 2012. Why Do Criminals Obey the Law? The Influence of Legitimacy and Social Networks on Active Gun Offenders. *The Journal of Criminal Law and Criminology*, 397–440.

Pierce, Lamar, and Jason Snyder. 2008. Ethical Spillovers in Firms: Evidence from Vehicle Emissions Testing. *Management Science*, 54(11): 1891–1903.

Reinikka, Ritva, and Jakob Svensson. 2004. Local Capture: Evidence from a Central Government Transfer Program in Uganda. *The Quarterly Journal of Economics*, *119*(2), 679-705.

Schnellenbach, Jan. 2007. Tax Morale and the *Legitimacy* of Economic Policy. *Homo Oeconomicus*, 24(1): 21–45.

Schulze, Günther G., and Björn Frank. 2003. Deterrence versus Intrinsic Motivation: Experimental Evidence on the Determinants of Corruptibility. *Economics of Governance*, 4(2): 143–160.

Serra, Danila. 2012. Combining Top-down and Bottom-up Accountability: Evidence from a Bribery Experiment. *Journal of Law, Economics, and Organization*, 28(3): 569–587.

Shang, Jen, and Rachel Croson. 2009. A Field Experiment in Charitable Contribution: The Impact of Social Information on the Voluntary Provision of Public Goods. *Economic Journal*, 119(540): 1422–1439.

Smith, Sarah, Frank Windmeijer, and Edmund Wright. 2015. Peer Effects in Charitable Giving: Evidence from the (Running) Field. *Economic Journal*, 125(585): 1053–1071.

Suryadarma, Daniel, and Chikako Yamauchi. 2013. Missing Public Funds and Targeting Performance: Evidence from an Anti-Poverty Transfer Program in Indonesia. *Journal of Development Economics*, *103*, 62-76.

Thibaut, John W., and Laurens Walker. 1975. *Procedural Justice: A Psychological Analysis*. New York, NY: L. Erlbaum Associates.

Tyler, Tom R. 1990. *Why People Obey the Law*. Princeton and Oxford: Princeton University Press.

Tyler, Tom R. 2004. Enhancing Police Legitimacy. *The Annals of the American Academy of Political and Social Science*, 593(1): 84–99.

Tyler, Tom R. 2006. Psychological Perspectives on Legitimacy and Legitimation. *Annual Review of Psychology*, 57: 375–400.

Tyran, Jean-Robert, and Lars P. Feld. 2006. Achieving Compliance when Legal Sanctions are Non-deterrent. *The Scandinavian Journal of Economics*, 108(1), 135-156.

Weber, Max. 1964. *The Theory of Social and Economic Organization.* Talcott Parsons (ed.). New York, NY: Free Press.

Yechiam, Eldad, Meir Druyan, and Eyal Ert. 2008. Observing Others' Behavior and Risk Taking in Decisions from Experience. *Judgment and Decision Making*, 3(7): 493.

Tables

Table 1: Summary Statistics - Public Official B Characteristics

	ED	END	XND	Mean Differences		
	Mean (SD)	Mean (SD)	Mean (SD)	ED-END	ED-XND	END-XND
Age	21.06	21.76	21.45	-0.702	-0.392	0.310
	(2.44)	(2.15)	(2.28)			
Gender (1 if Male)	0.63	0.65	0.85	-0.022	-0.223*	-0.201⁺
	(0.50)	(0.49)	(0.36)			
Monthly Expenses (Log)	8.80	8.66	8.67	0.131	0.129	-0.002
	(0.94)	(1.31)	(1.73)			
Economics Major	0.56	0.47	0.76	0.092	-0.195⁺	-0.287*
	(0.50)	(0.51)	(0.44)			
Observations	**32**	**34**	**33**			

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

Table 2: Average Choices of Public Official B by Treatment

	(1)	(2)	(3)
	ED	**END**	**XND**
	Mean (SD)	**Mean (SD)**	**Mean (SD)**
B's Behaviour (Corrupt=1)	0.88	0.91	0.79
	(0.19)	(0.16)	(0.30)
Amount Kept by B	1487.65	1493.75	1198.49
	(512.89)	(590.52)	(689.88)
Subjects	**32**	**34**	**33**

Table 3: Likelihood of Embezzlement by Public Official B – Pooled Data (Logit)

	(1) Main Continuous	(2) Full Continuous	(3) 3 Way Interaction
A's Behaviour (honest=1)	-1.146**	-0.750 +	-0.354
	[0.423]	[0.428]	[0.895]
Detection Level	-0.029 +	-0.020	-0.045**
	[0.016]	[0.024]	[0.016]
END	0.788	0.801	0.287
	[0.680]	[0.689]	[1.255]
A's Behaviour (honest=1) x END			-0.106
			[1.162]
END x Detection Level			0.094 +
			[0.055]
A's Behaviour (honest=1) x Detection Level		-0.026	-0.006
		[0.031]	[0.034]
A's Behaviour (honest=1) x END x Detection Level			-0.114 +
			[0.063]
Age	-0.196 +	-0.201 +	-0.208 +
	[0.118]	[0.117]	[0.112]
Gender	0.286	0.297	0.303
	[0.693]	[0.703]	[0.761]
Log Monthly Expenses	0.155	0.154	0.154
	[0.327]	[0.330]	[0.329]
Economics Major	0.524	0.540	0.560
	[0.741]	[0.754]	[0.783]
Constant	7.251*	7.213*	7.775*
	[3.529]	[3.581]	[3.296]
Controls	YES	YES	YES
Observations	1320	1320	1320
Subjects	66	66	66

Standard errors are clustered at the session level and reported in brackets $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. All models include controls for age, gender, monthly expenses, and having economics as a major.

Table 4: Likelihood of Embezzlement by Public Official B - By Treatment (Logit)

	Main Continuous		Full Continuous		Main Dummy		Full Dummy	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	ED	END	ED	END	ED	END	ED	END
A's Behaviour (honest=1)	-0.338	-1.823**	-0.172	-0.264	-0.441	-1.816**	-1.750**	0.240
	[0.300]	[0.499]	[0.940]	[0.773]	[0.409]	[0.502]	[0.442]	[0.750]
Detection Level	-0.043**	-0.019	-0.040*	0.040				
	[0.012]	[0.034]	[0.016]	[0.053]				
A's Behaviour (honest=1) x Detection Level			-0.009	-0.131**				
			[0.038]	[0.049]				
Detection Level=5					-1.963**	0.579**	-2.774**	1.377^{+}
					[0.599]	[0.141]	[0.533]	[0.792]
Detection Level=10					-0.787^{+}	-0.410	-1.895**	-0.277
					[0.437]	[0.249]	[0.541]	[0.401]
Detection Level=15					-1.452	-0.476	-2.583*	0.973**
					[1.115]	[0.337]	[1.076]	[0.345]
Detection Level=20					-1.832**	-0.585	-2.522**	0.743
					[0.517]	[0.768]	[0.543]	[1.432]
Detection Level=25					-1.589**	0.727	-2.564**	2.565^{+}
					[0.562]	[1.250]	[0.583]	[1.471]
Detection Level=30					-1.996**	-0.749	-2.550**	0.730
					[0.683]	[0.773]	[0.586]	[1.238]
A's Behaviour (honest=1)x Detection Level=5							1.483**	-2.323**
							[0.173]	[0.734]
A's Behaviour (honest=1)x Detection Level=10								-0.799
								[0.713]
A's Behaviour (honest=1) x Detection Level=15							2.859**	-2.933**
							[0.851]	[0.659]
A's Behaviour (honest=1) x Detection Level=20							1.001*	-3.060^{+}
							[0.435]	[1.574]
A's Behaviour (honest=1)x							1.857**	-4.388**

Detection Level=25								
							[0.324]	[1.268]
A's Behaviour (honest=1)x Detection Level=30							0.385	-3.909**
							[1.035]	[1.252]
Age	-0.243	-0.011	-0.244	-0.028	-0.260+	0.005	-0.285*	-0.022
	[0.153]	[0.058]	[0.151]	[0.058]	[0.143]	[0.048]	[0.138]	[0.069]
Gender	0.795	-0.302	0.796	-0.248	0.706	-0.330	0.710	-0.263
	[0.887]	[1.042]	[0.889]	[1.311]	[0.898]	[1.092]	[0.908]	[1.324]
Log Monthly Expenses	0.428	0.040	0.428	0.020	0.440	0.049	0.449	0.065
	[0.761]	[0.243]	[0.761]	[0.260]	[0.794]	[0.248]	[0.795]	[0.277]
Economics Major	-0.450	1.488	-0.454	1.763	-0.418	1.470	-0.434	1.761
	[0.734]	[1.242]	[0.749]	[1.302]	[0.739]	[1.257]	[0.815]	[1.380]
Constant	5.294	4.192	5.286	4.224	6.264	3.770	7.545	3.497
	[7.833]	[3.232]	[7.846]	[3.108]	[8.132]	[3.307]	[8.154]	[3.306]
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	608	680	608	680	608	680	600	680
Subjects	32	34	32	34	32	34	32	34

Standard errors are clustered at the session level and reported in brackets. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. All models include controls for age, gender, monthly expenses, and having economics as a major.

Table 5: Likelihood of Embezzlement by Public Official B (Logit)

	Detection Level 30 **Baseline Group: Corrupt Official B in XND**
END	2.402**
	[0.658]
A's Behaviour (Honest=1)	0.063
	[0.470]
END x A's Behaviour (Honest=1)	-4.856**
	[0.738]
Constant	7.667
	[5.052]
Controls	YES
Observations	**724**
Subjects	**62**

Standard errors are clustered at the session level and reported in brackets. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.The model includes controls for age, gender, monthly expenses, and having economics as a major.

Figures

Figure 1: Share of Corrupt Public Official B's - by Detection Level and Public Official A's Type

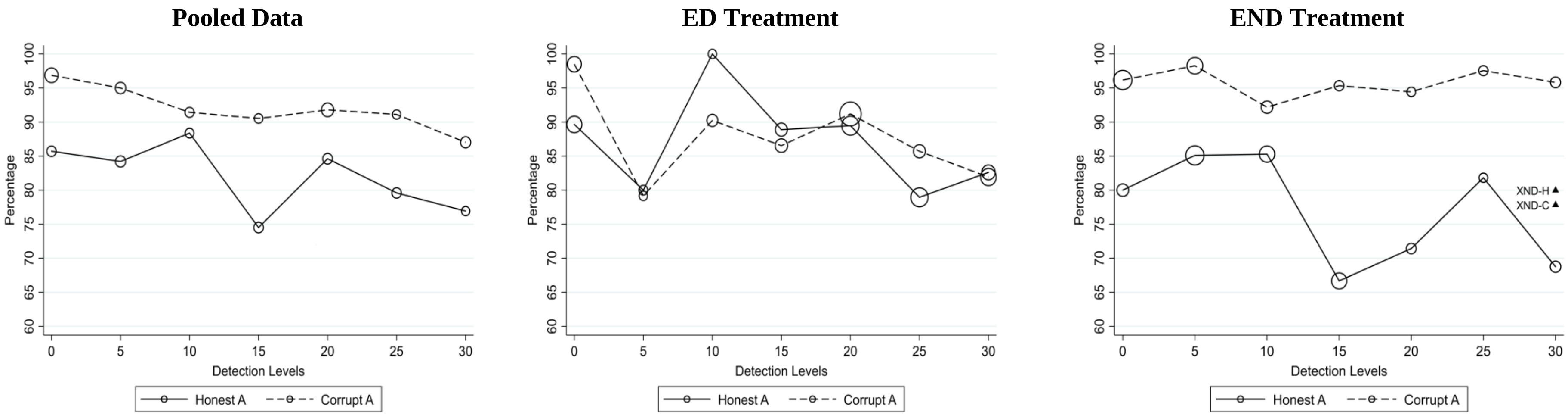

Figure 2: Probability choices by Public Official A