Corrupt police[*]

Klaus Abbink[†] Dmitry Ryvkin[‡] Danila Serra[§]

January 30, 2020

Abstract

We employ laboratory experiments to examine the effects of corrupt law enforcement on crime within a society. We embed corruption in a social dilemma setting where citizens simultaneously choose whether to obey the law or to break the law and impose a negative externality on others. Police officers observe citizens' behavior and decide whether to impose fines on law-breakers or, in treatments with corruption, extort bribes from any citizen. In the first study, we find that the presence of police substantially reduces crime, as compared to a baseline setting without police. This is true also when police officers are corrupt. This result is driven by corrupt police officers using bribes in a targeted manner as a substitute for official fines to punish law-breakers. In the second study, we test the effectiveness of two reward mechanisms aimed at reducing police corruption, both of which are based on society-wide police performance measures and not on the observation/monitoring of individual officers. We find that both mechanisms make bribery more targeted toward law-breakers, and one of them leads to a moderate reduction in crime.

Keywords: corruption, bribery, crime, police, experiment
JEL classification codes: D73, K42, C92

[*] We thank an advisory editor and three anonymous referees for their insightful comments. This project was funded in part by the Australian Research Council Discovery Project grant DP140101900.

[†] Department of Economics, Monash University, Clayton, VIC 3800, Australia; klaus.abbink@monash.edu.

[‡] Department of Economics, Florida State University, Tallahassee, FL 32306-2180, USA; dryvkin@fsu.edu.

[§] Department of Economics, Texas A&M University, College Station, TX 77845-4228; dserra@tamu.edu.

1 Introduction

Law enforcement is crucial for the efficient functioning of the social interactions and market exchanges taking place within a society. However, rule compliance may not be achieved if law enforcement officers are corrupt, i.e., if officers bend the rules or make new ones with the sole purpose of personal gain. Corruption among law enforcement officers, and the police in particular, is, in fact, widespread. Transparency International's 2017 Global Corruption Barometer,[1] which is based on surveys of 162,136 adults in 119 countries, identifies the police as the sector perceived by the public as the most corrupt, with 36% of the respondents around the globe believing that most or all police officers are corrupt. This percentage increases to 47% when restricting the analysis to Sub-Saharan countries.

Data on self-reported bribery experiences give a similar picture. In Sub-Saharan Africa, over one-quarter of the people who come into contact with the police – and more than half in Liberia, Sierra Leone and Ghana – report having to pay a bribe (Transparency International, 2015). In Georgia and Ukraine, police corruption had been so pervasive that in 2003 and 2005, respectively, both governments took the drastic measure of firing all traffic police overnight and starting afresh. Several years later, the success of the draconian measures could not be more different in the two countries. While police corruption practically vanished in Georgia and "many observers believe that the roads were actually safer without the traffic people waving motorists over all the time" (World Bank, 2012, p.16), the new Ukrainian traffic police did not seem to be much better than the old one and scored a miserable 4.3 on the Global Corruption Barometer. In fact, it worked so badly that in 2015, the Ukrainian government took a second attempt of a ground-up reform, again including the firing of all officers.

In situations where it is common for police officers to accept bribes to leave rule violations unpunished and/or extort bribes from law-abiding citizens, could the presence of law enforcers incentivize rather than reduce the occurrence of criminal activities? It is obvious that when officers extract bribes completely independent of actual infringements, a police force is no longer a deterrence. Moreover, could such a police force even drive otherwise law-abiding citizens towards rule-breaking, as a response to the constant nuisance of the bribe demands?

While there exists a large theoretical literature on optimal law enforcement,[2] including studies of optimal incentives for corrupt law enforcers (e.g., Mookherjee and Png, 1995; Polinsky and Shavell, 2001), and a vast empirical literature on crime deterrence,[3] little is known on the impact of corrupt law enforcement on rule-breaking within a society. We address this question by employing a novel laboratory experiment that simulates the interactions between citizens facing the social dilemma of choosing whether to obey the law or break it at the expense of others, and

[1] https://www.transparency.org/_view/publication/8064

[2] See Garoupa (1997) and Kaplow and Shavell (2002) for a review of the theoretical literature on optimal law enforcement.

[3] See Levitt and Miles (2007) for a recent review.

police officers deciding whether to give fines to law-breakers or demand bribes from law-breakers and/or law-abiders. We first assess the impact of officers' corruption on citizens' rule breaking behavior, and then examine whether corruption in law enforcement may be reduced through incentive systems that do not rely on the monitoring of individual officers; rather, they rely on aggregate outcomes, such as the crime rate or the bribery rate observed in the society as a whole and possibly measurable through citizen surveys.

The empirical literature on corruption is vast and fast growing,[4] yet only a handful of studies (e.g., Olken and Barron, 2009; Foltz and Opoku-Agyemang, 2015) focus on police corruption. The crucial challenge lies in the difficulty of measuring and examining the behavior of agents that are in charge of registering and prosecuting the same rule violations that they may be responsible for perpetuating. Hence, it is difficult to access reliable data on both the bribes pocketed by police officers and the rule violations taking place under their watch.

By employing a novel laboratory experiment we are able to overcome the measurement and identification challenges that come with any attempt to study corrupt police in the field. We depart from all previous bribery experiments (see, e.g., Abbink, Irlenbusch and Renner, 2002; Abbink et al., 2014; Armantier and Boly, 2013; Banuri and Eckel, 2015; Barr and Serra, 2010; Ryvkin, Serra and Tremewan, 2017; Salmon and Serra, 2017) by embedding corruption in law enforcement into a social dilemma environment where citizens simultaneously decide whether to obey or break the law, knowing that breaking the law benefits them but generates a negative externality on fellow citizens. Police officers in charge of clusters of citizens observe rule violations and decide whether to impose fines on, or demand bribes from citizens. Fines can only be imposed on rule violators, whereas bribes can be demanded from both rule-abiding and rule-breaking citizens.

Our design allows for assessing the frequency of both collusive and extortionary corruption. The former corresponds to the demand of bribes lower than fines exclusively from rule-violators. This is the kind of corruption that benefits both briber and bribee and that is referred to as "corruption with theft" in the seminal paper by Shleifer and Vishny (1993). Extortionary corruption, or "corruption without theft," is the demand of bribes for the provision of services that citizens are entitled to. In our specific setting, it is the demand of bribes from citizens who did not commit any law violation, or the demand of bribes higher than the official fine from citizens who committed a law violation. While collusive corruption is widespread around the world and may involve large amounts of money and quid pro quos between public officials and wealthy or influential citizens, extortionary bribery is small scale yet it affects citizens in their everyday lives in countries where corruption is endemic and misbehavior of public officials goes unpunished. For example, in India, data collected by Transparency International (TI, 2013) show that about half of the population pays bribes to obtain common government services, and 80% of these bribes are paid to avoid harm rather than to get a benefit.

[4] For recent surveys of the literature, see Sequeira (2012) and Olken and Pande (2012).

In our experiment, by repeating the game and randomly rematching citizens to police officers, we are able to assess the impact of extortionary corruption on the citizens' decision to break the law in the future. We contrast the corrupt police environment to two benchmarks: One where the police cannot engage in corruption, i.e., officers can only impose fines on rule-breakers, and one where there are no police officers and, therefore, citizens can only rely on self-governance.

By focusing on illegal or unethical behavior, our baseline rule-breaking setting relates to the vast experimental literature on cheating (for a review, see Abeler, Nosenzo and Raymond, 2019) and misbehavior in other contexts, such as tax reporting (e.g., Coricelli et al., 2010; Luttmer and Singhal, 2014; Mascagni, 2018) and managerial decision-making (e.g., Butler, Serra and Spagnolo, 2019; Cason, Friesen and Gangadharan, 2016; Schmolke and Utikal, 2018). A crucial difference between this literature and our setting is that the law-breaking game we employ has a strategic element, i.e., subjects do not decide whether to act unethically in isolation. Instead, they know that their law-breaking behavior will affect others negatively and, more importantly, that others' law-breaking behavior will affect them negatively. This way, our baseline setting is closer to that of a prisoner's dilemma game with multiple players and the treatments with police relate to the literature on public good games with centralized punishment (e.g., Baldassarri and Grossman, 2011; Markussen, Putterman and Tyran, 2016; Andreoni and Gee, 2012; Zhang et al., 2014) and prisoner's dilemma games with third party punishment (e.g., Fehr and Fischbacher, 2004). Here, the novel feature of our design is that our enforcement authority – the police – can demand and pocket bribes from, rather than just impose penalties on citizens. We know of only two studies, one by Muthukrishna et al. (2017) and the other by Buffat and Senn (2017), which employ a public good game with a corruptible monitor. Both experimental investigations allow contributors to a public good to offer a bribe to their monitor to influence his or her punishment decisions. In contrast, we employ a prisoner's dilemma game where it is the enforcement authority – the police officer – that takes the initiative of demanding bribes from "defectors" or "cooperators" at his or her own discretion.

We find that law enforcement significantly reduces rule-breaking behavior by citizens, even when law enforcers are corrupt. In fact, the societal crime rate is more than 60% higher without a police force than with it. Somewhat surprisingly, corrupt police officers are no less effective than honest officers in enforcing the law, even though bribery is rampant. This is because, when available, bribes substitute fines as a law enforcement mechanism, i.e., officers turn to using bribes to punish rule-breakers. However, high bribes being transferred from citizens to officers lead to severe wealth redistribution within the society. Moreover, corrupt officers engage also in extortionary corruption by targeting law-abiding citizens, which induces the latter to become law-breakers in the future. This is very much in line with behaviors often observed in public goods experiments where peer punishment of cooperators, i.e., anti-social punishment, induces cooperators to free ride in subsequent rounds (e.g., Cinyabuguma, Page and Putterman, 2006;

Herrmann, Thöni and Gächter, 2008; Nikiforakis, 2008).

Can police corruption be curbed by using incentive systems that do not require the observation and monitoring of individual officers? In our second set of treatments, we test the effectiveness of reward mechanisms that assign (small) extra compensation to officers on the basis of either the societal crime rate or the societal bribery rate. While our reward schemes are similar in scope to more traditional pay for performance incentives (e.g., Banuri and Keefer, 2015; Barrera-Osorio and Raju, 2017; Hasnain, Manning and Pierskalla, 2012), the novelty is that rewards are not based on objective measures of individual performance, which would require assessing whether officers collected fines and/or demanded bribes from law-breaking or law-abiding citizens. Instead, rewards are based on the joint performance of the police force, as measured by the overall reduction in corruption or crime in society. This is similar to the field experiment conducted by Khan, Khwaja and Olken (2015) in Pakistan, where tax collectors in teams of three were given a bonus based on the amount of tax revenues they jointly collected above a benchmark (set on the basis of past performance). However, in our case, police performance is measured as the level of corruption or crime in the society as a whole. Such measures can be elicited through citizen victimization surveys and surveys of corruption experiences – which are commonly conducted by non-governmental organizations in high corruption countries – therefore bypassing formal, and possibly corrupt, law enforcement institutions.

In the specific setting of the game, in any given round we provide small monetary rewards to officers based on either the fraction of citizens in society who did not break the law or the fraction of those who were not asked to pay a bribe. For selfish, payoff-maximizing agents, these monetary incentives should have no impact on officers' behavior, as they are too small to make up for the loss in bribery earnings that an officer would suffer if acting honestly. In contrast, the results show that both reward systems significantly reduce the occurrence of extortionary corruption, i.e., the demand of bribes from law-abiding citizens. Moreover, the system that assigns rewards to officers on the basis of the bribery rate within the society also lowers, albeit moderately, the percentage of citizens breaking the law.

A note on the external validity of our experiment is warranted. We acknowledge that decisions made by student subjects in an experimental lab do not directly correspond to decisions involved in corrupt transactions between regular citizens and police officers in the field. Our goal, however, is not to estimate the "general level" of corruption or preferences for law-breaking in a society, which are, of course, different depending on the nature of the subject pool and parameters of the environment. Our interest, instead, is in comparisons across parallel treatments with varying contexts and incentive systems, and in the underlying decision-making mechanisms.[5] Studying our research questions in the field would be prohibitively costly or unfeasible because it is impossible to find governments willing to engage in radical experiments with the police

[5] See Kessler and Vesterlund (2011) and Camerer (2011) for a general discussion of the external validity of laboratory experiments.

force. Additionally, there is evidence that, at least in some contexts, treatment effects uncovered in stylized lab experiments on corruption are in line with those in field experiments with non-student subjects or corruption-related experiences of the general population in the same country (Armantier and Boly, 2013; Barr and Serra, 2010). There is also evidence that correlations between answers to an incentivized survey are similar for student subjects and non-student subjects from a representative sample of the US population (Snowberg and Yariv, 2018).

The rest of the paper is organized as follows. In Section 2, we describe our experimental design and procedures. Section 3 provides a set of formal equilibrium predictions based on a standard model with selfish payoff-maximizing agents as well as on a behavioral model. Experimental results and robustness checks are presented in Section 4, and Section 5 concludes.

2 Experimental design

2.1 The Corrupt Police game

In the *Corrupt Police* game we employ in this paper, there is a society consisting of 3 police officers and 9 private citizens. Private citizens are randomly divided into groups of 3, and each group is randomly assigned a police officer. All players start the game with an endowment of 100 experimental currency units (ECU). The game proceeds in two stages. In the first stage, three citizens in each group simultaneously and independently decide whether or not to break the law. If a citizen breaks the law, she earns 20 ECU while each of the other two citizens in her group loses 20 ECU. In the second stage, the police officer assigned to the group observes the law-breaking decisions of the three citizens and makes a separate decision for each. If a citizen broke the law, the officer can impose a fine of 27 ECU[6] or demand a bribe between zero and 50 ECU. If a citizen did not break the law, the officer can still demand a bribe between zero and 50 ECU.[7] If an officer imposes a fine, the citizen has to pay the fine, and the money is divided equally among the 9 citizens in the society. That is, each citizen in the society receives 3 ECU every time a fine is imposed on some citizen, including herself. As a result, the net loss from a fine for a citizen is 24 ECU.[8] We decided to impose externalities within the group, but spread the fine incomes across the whole society. This should reflect the local character of the externalities. A speeding driver, for example, endangers the people traveling or living in the immediate area, but not so much the wider society. Fine income, on the other hand, flows into the general budget, with little local effect. If an officer demands a bribe, the citizen has to pay

[6] While it is reasonable to assume that imposing a fine has some costs to the officer (for instance, the cost of time required to fill in the required documents), here we abstract from such costs both for simplicity and because in the environment of interest (high corruption societies) these costs are likely to be small, given the officers' high discretionary power.

[7] Demanding a bribe of zero effectively amounts to the officer doing nothing. This was a separate option for subjects in the experiment, to make the distinction between bribery and doing nothing more salient.

[8] We implemented the redistribution of fines to keep the game budget-neutral. The net loss from a fine, 24 ECU, is slightly above the benefit from law-breaking, 20 ECU.

the bribe, and the money is transferred to the officer. This is because we are simulating an environment where corruption is widespread and officers are never punished. Hence, citizens have zero bargaining power and no outside option.

Formally, enumerate citizens in the society by $i = 1, \ldots, 9$ and assume, without loss of generality, that group 1 includes citizens 1, 2 and 3. Similarly, enumerate officers by $j = 1, 2, 3$ and assume that officer 1 is assigned to group 1. Let v_i denote a binary variable equal 1 if citizen i breaks the law and zero otherwise; similarly, let f_i denote a binary variable equal 1 if a police officer imposes a fine on citizen i. Finally, let b_i denote the bribe demanded by an officer from citizen i, if any.

The payoff of citizen 1 if she *does not* break the law is $\pi_0^{C1} = 100 - 20(v_2 + v_3) + 3\sum_{i=2}^{9} f_i - b_1$. This is because the citizen may suffer negative externalities of 20 or 40, respectively, if one or both of the other citizens in the group break the law, may get some positive returns from fines collected in the society, and may have to pay a bribe to the matched police officer. If citizen 1 breaks the law, her payoff is $\pi_1^{C1} = 100 + 20 - 20(v_2 + v_3) - 24f_1 + 3\sum_{i=2}^{9} f_i - b_1$, where $f_1 = 1$ is only possible if $b_1 = 0$, since the officer will either impose a fine, or demand a bribe, or do nothing. Since officers receive fixed wages and can only increment their income through bribery, the payoff of officer 1 is $\pi^{O1} = 100 + \sum_{i=1}^{3} b_i$.

The game is repeated for 15 rounds. At the end of each round, citizens get feedback on the law-breaking choices of the other two citizens in their group and on the police officer's decision (do nothing, impose a fine or demand a bribe) regarding her. Citizens do not observe officers' decisions about the other two citizens in their group or other citizens in their society. At the beginning of each round, the 9 citizens that form a society are randomly re-matched in groups of 3, and are randomly re-assigned to a police officer. Police officers do not have observable identifiers, and citizens' id numbers (1,2,3) are re-shuffled in each round, to ensure that subjects cannot track others' actions over time and form long-term relationships.

Our random re-matching protocol could be seen as a hybrid stranger-partner protocol, given that all interactions happen within societies of 9 citizens and 3 officers. This is on purpose, as we aim to examine the actions of officers who are stationed in given districts or neighborhoods, and who therefore become familiar, in the long run, with the law-breaking taking place in their jurisdiction (e.g., the local roads or intersections). At the same time, we do not want the formation of long-term relationships to play a role in the behavior of officers, as we envision the jurisdiction to be large enough for the officer to not personally know every citizen/driver.

2.2 Treatments

2.2.1 Study 1: The effects of police corruption

The goal of Study 1 is to assess the effects of corrupt police on law-breaking and other outcomes within a society. In addition to the main Corrupt Police treatment described in Section 2.1, we implemented two additional control treatments: *No Police* and *Honest Police*. Under No Police, the society consists of 9 private citizens, and there are no police officers. The payoffs of the citizens are determined as above, but excluding the possibility of fines and bribes. Therefore, the payoff of citizen 1 in this case is $\pi_0^{C1} = 100 - 20(v_2 + v_3)$ if she does not break the law, and $\pi_1^{C1} = 120 - 20(v_2 + v_3)$ if she does.

In the Honest Police treatment, the structure of the society and payoffs are the same as in Corrupt Police, but here police officers can only impose fines on law-breaking citizens, and cannot demand and pocket bribes. Therefore, the payoff of citizen 1 is $\pi_0^{C1} = 100 - 20(v_2 + v_3) + 3\sum_{i=2}^{9} f_i$ if she does not break the law, and $\pi_1^{C1} = 120 - 20(v_2 + v_3) - 24f_1 + 3\sum_{i=2}^{9} f_i$ if she does.

2.2.2 Study 2: Anti-corruption reward systems

As discussed in the Introduction, using standard monitoring and punishment mechanisms to reduce corruption may not be the most efficient strategy in systemically corrupt societies. In this study, we examine the effectiveness of two alternative reward schemes in curbing police corruption – *Rewards Low Crime* and *Rewards Low Corruption*. Performance pay for police officers is fundamentally a contentious issue. Ideally, police officers should serve the public diligently and impartially, not maximize the number of arrests or the fine revenues. Specifically rewarding non-corrupt behavior seems even harder; next to the ethical issue there is the very practical problem that non-corrupt behavior is just as unobservable as corrupt behavior (apart from the cases where a corrupt officer is caught, which are frustratingly rare in societies with pervasive corruption). Therefore, reward schemes can only rely on *aggregate outcomes* that can be measured via surveys or official statistics. In addition, there are inevitable injustices: An individual officer has not much direct control over the aggregate outcome (s)he is paid for. This makes it politically difficult to introduce such rewards as part of the regular pay package for police officers. However, it would be possible for NGOs to implement such schemes, as they are private actors and not the officers' direct employer. This limits the size of the payments, as NGOs would not have the budget to pay rewards that would constitute a substantial part of an officer's wage, let alone compensate for the maximum revenues police officers can obtain from all-out corruption. Nevertheless, small rewards may still have a motivating effect, since the officer has a goal to work towards (low crime rates or low corruption), and that (s)he can contribute to by refraining from or reducing his or her own bribe demands.

In this study, the Corrupt Police game serves as the baseline, and the same structure of the game is preserved in both treatments with rewards. In the *Rewards Low Crime* treatment, each police officer is rewarded with 6 ECU for each citizen in the society who did not break the law. Formally, the payoff of officer 1 is $\pi^{O1} = 100 + \sum_{i=1}^{3} b_i + 6\sum_{i=1}^{9}(1 - v_i)$. Therefore, in addition to the fixed wage and collected bribes (if any), each officer can earn a reward ranging from 0 to 54 ECU depending on the level of law-breaking in the society (with zero reward if all 9 citizens

Treatments	Sessions	Societies	Subjects
No Police	3	6	54
Honest Police	3	6	72
Corrupt Police	5	8	96
Rewards Crime	4	8	96
Rewards Corruption	7	8	96
Total	22	36	414

Table 1: Experimental treatments.

in the society break the law and the maximum reward of 54 ECU if all 9 citizens do not break the law).

In the *Rewards Low Corruption* treatment, each police officer is rewarded with 6 ECU for each citizen in the society who did not have to pay a bribe. In this case, $\pi^{O1} = 100 + \sum_{i=1}^{3} b_i + 6\sum_{i=1}^{9} \mathbb{1}_{b_i=0}$ (where $\mathbb{1}_{b_i=0}$ is the indicator equal 1 if $b_i = 0$ and zero otherwise). Here, rewards range between 0 and 54 ECU depending on the level of corruption in the society.

In both settings, it is immediately clear that the reward of 6 ECU is small compared to the possible bribe income (recall that the officer can demand bribes of up to 50 ECU). Neither do the rewards create a social dilemma situation among the officers – even if all officers refrain from demanding bribes, they are not better off than if they were all corrupt. Thus, the rewards can only work if officers feel some moral costs when they extract bribes. An NGO trying to introduce such a scheme would need to hope that the presence of a reward exacerbates these moral costs.

2.3 Implementation and procedures

A total of 414 subjects (62% of them female) participated in 22 sessions of the experiment, with one or two non-interacting societies in each session. Table 1 provides a summary of sessions, societies and subjects by treatment. All sessions were conducted at the XS/FS laboratory of Florida State University (FSU). Subjects were recruited through the web-based platform ORSEE (Greiner, 2015) from a pool of 3000+ pre-registered FSU students. We used a between-subject design, with treatments randomly assigned at the session level, and each subject only participating in one session. The experiment was implemented in z-Tree (Fischbacher, 2007), with subjects making decisions at visually separated computer terminals.[9]

[9] Using data from a post-experimental questionnaire, we assessed the quality of randomization in the experiment via pairwise comparisons of observable subject characteristics (means, using the t-test, and proportions, using the chi-square test) across treatments. We do not find any gender difference across treatments. We also do not see any significant differences in age, with the exception of the comparison of Honest Police and Rewards Low Crime (20.61 vs. 21.29, $p = 0.043$). There are no statistically significant differences in the proportion of economics majors. We only see one significant difference in the proportions of religious participants, between Corrupt Police and Rewards Low Corruption (0.42 versus 0.35, $p = 0.078$). Overall, we conclude that the randomization was successful.

We use strongly framed instructions (cf. Appendix A). Officers are told that their mission is to enforce the law, and bribery is termed as such. For citizens, the law-breaking decision is termed as such in the form of a hypothetical scenario. Experimental instructions were read out loud, with paper copies distributed to subjects.

The experiment followed an anonymous identification protocol customary for corruption experiments. When subjects entered the lab, they were assigned id numbers. Subjects were informed during instructions that their names would not be recorded and there would be no way to associate their names with decisions they made in the experiment. At the end of the session, subjects' earnings were linked to their id numbers, but not to their names, which were only used for accounting purposes.

The main part of the experiment consisted of 15 rounds. Societies remained fixed, but the groups of three citizens were randomly re-matched after each round, and police officers (whenever present) randomly re-assigned, within each society. We did this to avoid repeated-game effects. While police officers in reality do act in a local area, the interaction between individual officers and citizens is not so regular that it would warrant a partners matching protocol. One of the 15 rounds was selected randomly for payment. After the main part, subjects completed a demographic questionnaire. Earnings were converted into US$ at the exchange rate of $1 for 10 ECU. The experiment lasted about 60 minutes, with subjects earning $19.52 on average,[10] including a $10 show-up fee.

3 Theory and predictions

We start by identifying the subgame-perfect Nash equilibrium (SPNE) of the Corrupt Police game assuming all players are monetary payoff maximizers. It is easy to see that the strictly dominant strategy for police officers in the second stage is to demand the maximum bribe of 50 from each of the citizens in their group. In turn, in the first stage it is the strictly dominant strategy for citizens to break the law. Thus, the only SPNE is where all citizens break the law and all officers demand a bribe of 50, resulting in payoffs of 30 ECU for citizens and 250 ECU for police officers. Similarly, all citizens break the law in the equilibrium of the No Police game, resulting in payoffs of 80. In the Honest Police game, the situation is a bit more complicated since officers are indifferent between imposing and not imposing a fine on a law-breaking citizen. Consider an officer who imposes a fine on a law-breaker with some probability $p \in [0, 1]$. A law-breaking citizen's expected loss is then $24p$, whereas her gain from law-breaking is 20. Thus, there is a continuum of equilibria with law-abiding citizens and $p \in (\frac{5}{6}, 1]$, and a continuum of equilibria with law-breaking citizens and $p \in [0, \frac{5}{6})$. Finally, at $p = \frac{5}{6}$ there is a continuum of equilibria with citizens breaking the law with any probability $q \in [0, 1]$.

[10] Subjects in the role of citizens earned $17.20, while subjects in the role of police officers earned $27.89, on average.

The equilibrium in the games with rewards is not different from the one in the baseline Corrupt Police game. Indeed, in Rewards Low Crime, officers' rewards are affected by the behavior of citizens for whom breaking the law is still a strictly dominant strategy. In Rewards Low Corruption, an officer only gains 6 ECU by not demanding a bribe but can gain 50 ECU by demanding the maximum bribe from a citizen.

A behavioral model Due to the presence of strong framing in the instructions, we expect that officers obtain a moral benefit from law enforcement and incur a moral cost from bribery, while citizens incur moral costs from law-breaking.

Suppose that each police officer is characterized by a privately observable moral cost parameter, c, with a commonly known distribution $F(\cdot)$. There is a critical value $c = \hat{c}$ such that, when encountering a law-abiding citizen, an officer with $c > \hat{c}$ does not demand a bribe and an officer with $c < \hat{c}$ demands the maximum bribe of 50.[11] Further, there is a $\delta > 0$ such that, when encountering a law-breaker, an officer with $c > \hat{c} + \delta$ imposes a fine and an officer with $c < \hat{c} + \delta$ demands a bribe of 50. Parameter δ characterizes the fact that it is morally less costly to demand a bribe from a law-breaker than from a law-abiding citizen. This can be justified via a simple social norm argument; by breaking the law, the citizen makes law-breaking look more socially acceptable to the officer. It can also be justified if the officer views bribery as a form of punishment and hence the moral cost of bribery is reduced by the moral benefit from law enforcement. In other words, δ characterizes *targeted extortion*: The fraction $F(\hat{c} + \delta) - F(\hat{c})$ of officers believe that it is not justifiable to demand a bribe from an honest citizen but it is morally justifiable to demand a bribe from someone who broke the law.

For citizens, let m denote a privately observable moral cost of breaking the law, with a commonly known distribution $G(\cdot)$. Modulo all irrelevant terms, the expected payoff of a law-abiding citizen of type m, therefore, is

$$\pi_0 = -50F(\hat{c}),$$

and the expected payoff of a law-breaking citizen is

$$\pi_1 = 20 - m - 50F(\hat{c} + \delta) - 24(1 - F(\hat{c} + \delta)).$$

This gives the fraction of law-breakers (i.e., those citizens for whom $\pi_1 > \pi_0$):

$$\alpha(\hat{c}, \delta) = G(50F(\hat{c}) - 26F(\hat{c} + \delta) - 4).$$

As expected, $\alpha(\hat{c}, \delta)$ is decreasing in δ. The effect of $\hat{c}$, for a given δ, is generally ambiguous

[11] We adopt a simplifying assumption that the officer's moral cost of bribery is independent of the bribe size, and hence all bribes will be either zero or 50. A more refined model could have moral cost increasing gradually with bribe size but it would not qualitatively change the results.

because, on the one hand, an increase in $\hat{c}$ leads to a reduction of π_0 (and hence the opportunity cost of law-breaking goes down) but, depending on the shape of F it may either increase or decrease the level of targeted bribery. However, when F is uniform or has a decreasing density for $c > \hat{c}$, α is increasing in $\hat{c}$. Even if F has an increasing density, it should be increasing fast enough for the second effect to dominate. We conclude that in most cases the effect of $\hat{c}$ on α is positive.

Consider now the effects of reward systems in treatments Rewards Low Crime and Rewards Low Corruption. Due to anonymity and random rematching, rewards cannot serve as a credible enforcement mechanism, and hence they do not affect equilibrium predictions. That said, rewards can introduce additional framing effects and change expectations and social norms. The reward system based on crime rate will likely lead to an increase in δ, i.e., in the fraction of officers who use targeted extortion. The reason is that the reward makes the officer's mission of fighting crime more salient, and if some officers observe that imposing a fine is not sufficiently harsh to deter crime they can resort to bribery. In contrast, the reward system based on bribery rate will likely reduce $\hat{c}$ as it explicitly rewards for honest behavior and thus makes the moral cost of corruption more salient. Both of these should reduce crime rate, according to the model above, as long as the effect of $\hat{c}$ on $\alpha(\hat{c}, \delta)$ is positive. However, in one case this reduction will occur due to an increase in bribery while in the other due to a reduction in bribery and increase in the number of fines. It is possible, therefore, that citizens will react differently: In the case of increased bribery, citizens can retaliate by increasing law-breaking, especially because it also reduces the rewards for officers; while in the case of reduced bribery citizens can respond by a reduction in law-breaking as a sign of good will.

The two reward systems also create additional elements of repeated play and interconnectedness among citizens' and officers' decisions within a society. In Rewards Low Crime, officers' rewards depend directly on citizens' law-breaking behavior; hence, officers have an incentive to be more forward-looking and apply punishment in a targeted manner. In Rewards Low Corruption, officers' bribery decisions generate a negative externality on fellow officers reducing their rewards. Targeted bribery can then emerge as a form of cooperation among police officers.

4 Results

4.1 Study 1

4.1.1 Summary statistics and basic results

Table 2 presents summary statistics by treatment for the main variables of interest, with standard errors clustered by society in parentheses. For each variable, we provide the averages over the entire duration of the experiment (All t) as well as over the second half ($t > 7$). Columns (1) and (2) show average crime rate defined as the fraction of citizens breaking the law. The highest

Treatments	Crime rate, %		% Fined		% Paid bribe		Bribe size	
	All t	$t > 7$	All t	$t > 7$	All t	$t > 7$	All t	$t > 7$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
No Police	84.7	86.3						
	(4.0)	(4.6)						
Honest Police	48.9	51.9	87.6	89.7				
	(3.8)	(3.8)	(3.9)	(3.4)				
Corrupt Police	50.0	54.1	14.6	11.2	69.6	77.4	44.7	45.8
	(4.4)	(5.1)	(4.4)	(4.1)	(4.0)	(4.2)	(1.4)	(1.6)

Table 2: Summary statistics for Study 1 (standard errors clustered by society in parentheses).

Treatments	Citizens' payoffs		Officers' payoffs		Payoff ratio	
	All t	$t > 7$	All t	$t > 7$	All t	$t > 7$
	(1)	(2)	(3)	(4)	(5)	(6)
No Police	83.1	82.7				
	(0.8)	(0.9)				
Honest Police	90.2	89.6	100	100	0.902	0.896
	(0.8)	(0.8)			(0.008)	(0.008)
Corrupt Police	58.4	53.2	193.3	206.4	0.375	0.316
	(2.9)	(2.9)	(7.3)	(7.8)	(0.034)	(0.036)

Table 3: Average payoffs and citizen-to-officer payoff ratios in Study 1 (standard errors clustered by society in parentheses).

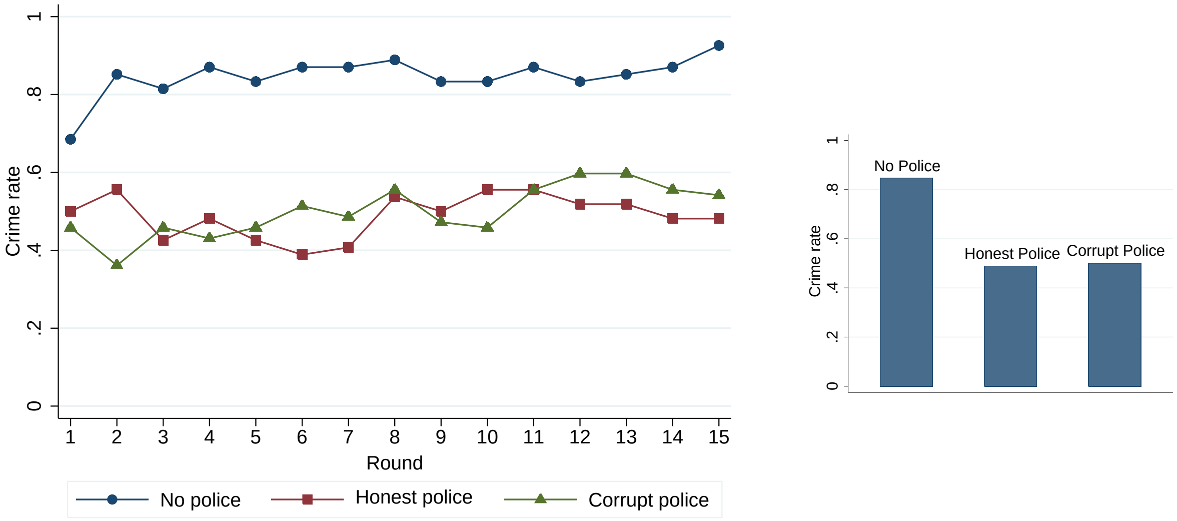

Figure 1: Average crime rates in Study 1, by treatment.

crime rate, around 85%, is observed in the No Police treatment,[12] whereas Honest Police and Corrupt Police have very similar crime rates around 50%. Figure 1 visualizes the averages and also shows how average crime rate changes over time. The separation of No Police from the other two treatments is quite evident, and there is no obvious difference in crime rate between Honest Police and Corrupt Police.

For basic statistical comparisons, we calculate averages in each society over all 15 rounds and use Wilcoxon rank-sum test treating each society as one independent observation. For crime rate, the tests produce $z = 3.10$, $p = 0.002$ (No Police vs. Corrupt Police), $z = 2.88$, $p = 0.004$ (No Police vs. Honest Police), and $z = -0.129$, $p = 0.897$ (Corrupt Police vs. Honest Police).

Columns (3) and (4) of Table 2 show the fraction of cases when a law-breaking citizen was fined. Officers impose a fine in almost 90% of cases in Honest Police, but in less than 15% of cases in Corrupt Police ($z = -3.102$, $p = 0.002$). These results indicate that police officers take their mission of law enforcement seriously but readily switch to bribery as soon as it becomes available.

For the Corrupt Police treatment, columns (5) and (6) show the fraction of citizens who were asked to pay a bribe and columns (7) and (8) show average bribe sizes. The frequency of bribery reaches 77%, and the average bribe size of 45 is at 90% of the maximum bribe of 50. These results show, consistent with the assumptions of the model in Section 3, that, for the most part, police officers either do not demand a bribe at all or demand the maximum bribe.

Table 3 shows average payoffs of citizens and police officers as well as their ratio that serves as a measure of inequality. Without police, citizens' payoffs are close to the (socially inefficient) equilibrium level of 80, reflecting the fact that the vast majority of citizens are breaking the law. In Honest Police, citizens' payoffs are higher, close to 90 ($z = -2.88$, $p = 0.004$). In Corrupt Police, the payoffs are the lowest, around 55, due to large bribe transfers ($z = 3.10$, $p = 0.002$, No Police vs. Corrupt Police; $z = 3.10$, $p = 0.002$, Honest Police vs. Corrupt Police). As expected, officers' payoffs are much larger in Corrupt Police, leading to strong inequality. However, combined payoffs of citizens and police officers are similar in Honest Police and Corrupt Police due to a similar reduction in crime ($z = -0.516$, $p = 0.606$).

We summarize the basic results as follows.

Result 1 *(a) The presence of police reduces crime: Crime rate is more than 30 percentage points higher in the No Police treatment as compared to treatments with police.*

(b) Police officers actively use fines to enforce the law in the Honest Police treatment, but rarely use fines when bribery is available.

[12] In No Police, there is a significant upward trend in crime rate over all rounds ($p = 0.03$); however, there is no time trend for $t > 7$ ($p = 0.30$), i.e., crime rate stabilizes in the long run. Due to random re-matching, No Police is effectively a three-player prisoner's dilemma, and sustaining cooperation in such settings is challenging. For comparison, Cooper et al. (1996) find long-run cooperation rates of 22% in two-player one-shot prisoner's dilemmas.

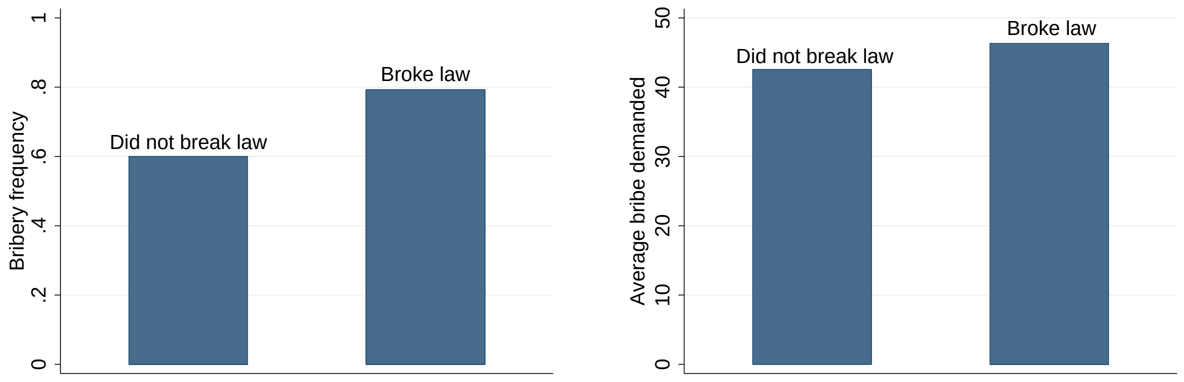

Figure 2: Average bribery rates and bribe sizes in Study 1, Corrupt Police.

(c) Bribes substitute fines as a law enforcement mechanism: There is no significant difference in crime rates between the Honest Police and Corrupt Police treatments.

(d) Citizens' payoffs are the highest in Honest Police and the lowest in Corrupt Police. Payoff inequality is strongest in Corrupt Police.

4.1.2 Officers' behavior in Corrupt Police

Figure 2 shows the average frequency of bribery (left) and bribe sizes (right) in the Corrupt Police treatment, separately for citizens who did and did not break the law. As seen from the figure, police officers engage in targeted bribery: They are about 20 percentage points more likely to demand a bribe from a citizen who broke the law ($z = 1.680$, $p = 0.093$, Wilcoxon sign-rank test). Law-breakers pay higher bribes than honest citizens, conditional on being demanded a bribe, although the size of the effect is small (3.8 ECU, $z = 2.240$, $p = 0.025$, Wilcoxon sign-rank test).[13] These results are consistent with the behavioral model in Section 3.

Result 2 *In Corrupt Police, officers engage in targeted bribery. They are more likely to demand a bribe from citizens who broke the law, and they demand higher bribes from law-breakers than from law-abiding citizens.*

Models of social preferences, with agents exhibiting various forms of utility from social efficiency and aversion to inequality (e.g., Fehr and Schmidt, 1999; Charness and Rabin, 2002), have generally been successful in explaining behavior in social dilemmas. However, it is easy to see that while the behavior we observe in Study 1 is consistent with the "moral homo oeconomicus" model described in Section 3, it cannot be explained by social preferences. First,

[13] We created paired observations for the sign-rank tests by computing, for each society, the average frequency of bribery and bribes demanded for citizens who did and did not break the law.

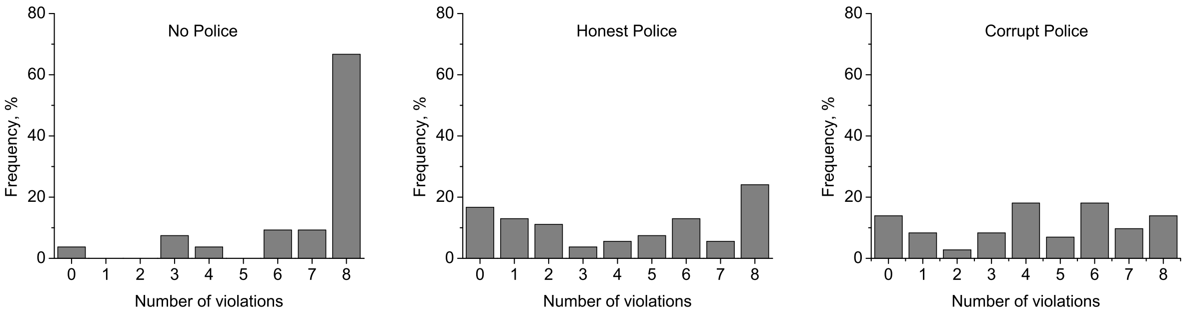

Figure 3: The histograms of the number of violations in Study 1 (rounds 8-15), by treatment.

overall efficiency plays no role as both fines and bribes are purely re-distributive, and breaking the law reduces social welfare. Second, police officers could reduce inequality by imposing fines on law-breakers; instead, they demand bribes close to 50 ECU with very imprecise targeting, thereby increasing inequality (cf. citizens' payoffs in Table 2).

4.1.3 Citizens' behavior

In this section, we analyze in more detail the dynamics of citizens' behavior. Specifically, we explore whether citizens' law-breaking is consistent across time (i.e., if there are stable "honest" and "criminal" behaviors), as well as how citizens react to the law-breaking by other citizens (conditional cooperation; see, e.g., Fischbacher, Gächter and Fehr, 2001; Thöni and Volk, 2018) and fines and/or bribe demands by police officers.

Figure 3 shows the histograms of the number of violations by citizens in the second half of the experiment (rounds 8-15), by treatment. Define a subject's behavior as honest (respectively, criminal) if she broke the law 0 or 1 (respectively, 8 or 7) times during these 8 rounds. As seen from the figure, more than 75% of citizens exhibit stable criminal behavior in the No Police treatment, and other types of behavior are barely present. In Honest Police, nearly 30% of citizens behave honestly, and similarly almost 30% show criminal behavior. Finally, in Corrupt Police, about 20% exhibit honest behavior, ∼25% are criminal, and there is a sizable number of intermediate cases breaking the law 2-6 times. Thus, in the absence of police the society nearly universally converges to the inefficient law-breaking equilibrium. In Honest Police, the landscape is dominated by relatively stable honest and criminal behaviors. In this treatment, police impose fines on law-breakers 89.7% of the time (cf. Table 2), which is only slightly above the probability $p = \frac{5}{6} \approx 0.83$ at which citizens are indifferent between breaking and not breaking the law (cf. Section 3). In Corrupt Police, the extreme behaviors are no longer dominant; citizens' behavior becomes more reactive.

Table 4 shows the results of pooled OLS regressions, by treatment, of dummy variable $Broke\ law_t$ (=1 if the citizen broke the law in round t and zero otherwise) on indicators for

	No Police (1)	Honest Police (2)	Corrupt Police (3)
Broke law$_{t-1}$	0.692**	0.770***	0.729***
	[0.11,1.32]	[0.40,1.01]	[0.51,1.01]
Others broke law$_{t-1}$	0.203	0.135**	0.137***
	[-0.20,0.52]	[0.05,0.19]	[0.11,0.16]
(Others broke law$_{t-1}$)×(Broke law$_{t-1}$)	-0.181	-0.243**	-0.149***
	[-0.40,0.19]	[-0.33,-0.12]	[-0.23,-0.07]
Fined$_{t-1}$		-0.0549	-0.195
		[-0.24,0.14]	[-0.58,0.03]
Paid bribe$_{t-1}$			0.170***
			[0.06,0.31]
(Paid bribe$_{t-1}$)×(Broke law$_{t-1}$)			-0.309***
			[-0.65,-0.07]
Intercept	0.192*	0.116**	0.0977**
	[-0.82,1.27]	[0.02,0.25]	[0.03,0.18]
Observations	756	756	1,008
Societies	6	6	8
R-squared	0.177	0.263	0.157

Table 4: OLS regressions for law-breaking by citizens. Significance levels, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, and 95% confidence intervals in brackets are based on wild cluster bootstrap.

various events in the previous round.[14] *Broke law$_{t-1}$* is an indicator for whether the citizen broke the law in the previous round; *Others broke law$_{t-1}$* is the number of subjects from the citizen's group who broke the law in the previous rounds; *Fined$_{t-1}$* is an indicator for whether the citizen was fined.

Given the relatively small number of independent clusters (societies), in all regressions in this paper we report statistical significance and 95% confidence intervals based on the wild cluster bootstrap method (Cameron, Gelbach and Miller, 2008). We also use wild cluster bootstrap for post-regression Wald tests of linear hypotheses. Following the recommendations of Cameron and Miller (2015) and Roodman et al. (2019), we use the six-point distribution Webb weights for regressions with fewer than 10 clusters, and the default two-point Rademacher weights for regressions with more than 10 clusters. The number of replications is set to 9,999 in all cases.

Coefficient estimates on variables *Broke law$_{t-1}$*, *Others broke law$_{t-1}$* and their interaction show a similar pattern across the three treatments, although the latter two are not statistically significant in No Police. Law-breaking is fairly persistent; someone who broke the law in round $t-1$ is more likely to break the law in round t. The positive coefficient on *Others broke law$_{t-1}$* and the negative coefficient of a similar magnitude on *Others broke law$_{t-1}$* × *Broke law$_{t-1}$* indicate that subjects who did not break the law at $t-1$ are more likely to break the law at t

[14] Probit regressions produce similar results. Throughout the paper, we opted for using linear probability models for the ease of interpreting coefficients. Probit marginal effects would be uninterpretable in the presence of interactions.

the more others in their group broke the law; however, the effect of $Others\ broke\ law_{t-1}$ on law-breaking is absent for those who broke the law at $t-1$. Thus, we observe *one-sided* conditional cooperation, or *contagion*, whereby citizens are converted by their peers from law-abiders into law-breakers, leading to escalation of law-breaking in the No Police treatment. This may be a consequence of shifting social norms (the moral cost of law-breaking decreases as others practice it more), aversion to inequality (by breaking the law, a citizen can equalize her payoff with those of other law-breakers) or simple retaliation (although the scope for retaliation is limited due to anonymity and random re-matching).

Interestingly, model (2) shows that being fined in Honest Police does not deter law-breaking. This is likely due to the fact that nearly all law-breakers are fined in that treatment (recall that $Fined_{t-1}$ can only be equal to 1 if $Broke\ law_{t-1} = 1$), and hence law-breaking is driven by types and responses to the behavior of other citizens. Similarly, there is no effect of being fined in Corrupt Police where citizens are fined very rarely (cf. Table 2); however, in that treatment citizens exhibit strong reactions to bribery. As seen from model (3), a citizen who was law-abiding at $t-1$ is 17 percentage points more likely to break the law if she had to pay a bribe. In contrast, a citizen who broke the law at $t-1$ and had to pay a bribe, is 14 percentage points less likely to break the law again, although the effect is not statistically significant (two-sided $p = 0.17$, the Wald test).

Result 3 *In Corrupt Police, both law-breaking by other citizens and having to pay a bribe lead to an increase in the likelihood that an honest citizen will break the law in the following round. In Honest Police, law-breaking by others has a similar effect, while in No Police the effect has the same sign but is not statistically significant.*

Why, despite Result 3, is there no escalation of crime rate over time in the treatments with police (cf. Figure 1)? In Honest Police, since fines do not reverse law-breaking, the only remaining explanation is the presence of stable citizens' behaviors. This is consistent with Figure 3. These behaviors are different from those in No Police, and hence the crime rate is different, because the presence of fines changes the payoff calculation for citizens. Given the frequency of fines, citizens in Honest Police are only slightly better off not breaking the law, on average, and approximately half of them remain honest.

In Corrupt Police, the mechanism is different. Here, fines as well as bribes reverse law-breaking, although the effects are not statistically significant. A citizen who broke the law and was fined at $t-1$ is 19.5 percentage points less likely to break the law at t; similarly, a citizen who broke the law and paid a bribe at $t-1$ is 14 percentage points less likely to break the law at t. These effects are noisy but, as a result, citizens' behaviors are less stable, and a large fraction of citizens go back and forth between breaking and not breaking the law, cf. Figure 3.

Thus, the dynamics of crime rate in Corrupt Police are determined jointly by citizens' reactions to bribery and fines and the fact that, similar to fines, bribery is targeted.

4.1.4 Robustness checks

In order to test the robustness of our findings, we conducted two additional treatments. The first is a replication of the No Police treatment with a different subject pool – students at Texas A&M University (TAMU). We implemented such replication to test whether the high baseline frequency of law-breaking in the No Police treatment (∼85% over all rounds) was an anomaly due to some unobservable characteristics of our original FSU student sample. Using the same protocol and the same sample size as in the original No Police treatment, we find no statistical difference in crime rates between the two samples. At TAMU, the rate of law-breaking is 0.79 over all 15 rounds and 0.83 for $t > 7$, compared to 0.85 and 0.86, respectively, at FSU ($z = 0.882$, $p = 0.378$ over all 15 rounds; $z = 0.642$, $p = 0.521$ for $t > 7$; Wilcoxon rank-sum test). The replication leads us to believe that the high baseline frequency of law-breaking is not an anomaly.[15]

The second additional treatment (run at FSU) is a version of Corrupt Police with the maximum bribe size capped at 27, which is equal to the official fine police officers could impose on law-breaking citizens. The goal of this treatment, CP(27), is to assess whether the high frequency of bribery observed in the original Corrupt Police treatment, where the maximum bribe was set at 50, is due to police officers using higher bribes for personal gain or to enforce the law more effectively. By setting the maximum bribe equal to the official fine we eliminate the potential for bribery to be more effective than ticketing in deterring crime.

Ticketing is virtually nonexistent in CP(27), with average ticketing rate of 0.076 (all t) and 0.055 ($t > 7$). Bribery rates are high – 0.85 (all t) and 0.90 ($t > 7$) – and nearly all bribes are at the maximum, with average bribe sizes 25.3 (all t) and 25.9 ($t > 7$). In fact, bribery rate in CP(27) is higher than in Corrupt Police ($z = 2.197$, $p = 0.028$). We still see targeted bribery, with law-breaking citizens paying bribes 22.7 percentage points more frequently than law-abiding citizens ($z = 1.782$, $p = 0.075$), which is slightly larger in magnitude than the targeted bribery observed in Corrupt Police (17.8 percentage points).

The average crime rate in CP(27) is 0.73 over all rounds and 0.80 for $t > 7$. Comparing these averages to the crime rates observed in No Police (0.85 and 0.86), we obtain $z = 2.005$, $p = 0.045$ (all t) and $z = 1.127$, $p = 0.260$ ($t > 7$); whereas comparing them to the crime rates in Corrupt Police (0.50 and 0.54), we get $z = -2.711$, $p = 0.007$ (all t) and $z = -2.711$, $p = 0.007$ ($t > 7$). We conclude that police officers primarily use bribes for personal gain. Had they tried to enforce the law in CP(27), they could have achieved much better results by using fines as in Honest Police where a well-targeted fine (by design) of 27 produces a 51.9% long-run crime rate – much lower than the 80% in CP(27). Imperfectly targeted bribes of 27 instead lead to a long-run crime rate similar to No Police.

[15] While the percentage of law-breakers is higher than what is observed in cheating games such as the die roll games or the tax compliance games, it is not unusual for games where unethical behavior has a strategic element, such as in our setting, where individuals suffer negative externalities caused by others' law-breaking.

Treatments	Crime rate, %		% Fined		% Paid bribe		Bribe size	
	All t	$t > 7$	All t	$t > 7$	All t	$t > 7$	All t	$t > 7$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Corrupt Police	50.0	54.1	14.6	11.2	69.6	77.4	44.7	45.8
	(4.4)	(5.1)	(4.4)	(4.1)	(4.0)	(4.2)	(1.4)	(1.6)
Rewards Low Crime	45.0	45.8	12.6	12.1	56.9	61.3	43.2	43.7
	(7.4)	(9.2)	(5.1)	(5.0)	(7.9)	(9.0)	(1.8)	(2.3)
Rewards Low Corruption	38.8	37.5	22.7	15.7	50.5	56.3	46.7	47.4
	(5.4)	(7.3)	(8.8)	(7.4)	(11.9)	(13.0)	(1.2)	(0.9)

Table 5: Summary statistics for Study 2 (standard errors clustered by society in parentheses).

We can draw two general conclusions from the comparison of crime rates in Corrupt Police, CP(27) and Honest Police. First, the size of the penalty matters. In Corrupt Police, where law-breakers could be forced to pay 50, although targeting is imperfect, we see much lower crime than in CP(27), where the maximum bribe was 27. Second, for any given degree of targeting, the deterrence effect of a penalty may depend on how the penalty is perceived by the law-breaker. In CP(27), a law-breaker being asked to pay a bribe of 27 is likely to react differently than if he/she were asked to pay a fine of 27 (as it is the case in Honest Police), due to the fact that the source of punishment is an unethical action (bribery) and, therefore, such punishment carries less normative signaling than a legal fine. This may have contributed to the ineffectiveness of bribes in deterring crime in CP(27).[16] While bribes carry low normative signaling also in Corrupt Police, the fact that officers can punish law-breakers with bribes that are much larger than the official fine proves (relatively) effective in deterring crime.

4.2 Study 2

4.2.1 Summary statistics and basic results

Table 5 shows the averages of the main variables of interest in the three treatments of Study 2. Average crime rate, summarized in columns (1) and (2), is also shown in Figure 4. Crime rate appears to be the lowest in Rewards Low Corruption and the highest in Corrupt Police ($z = 1.68$, $p = 0.093$ for comparison of average crime rates between Corrupt Police and Rewards Low Corruption, but no statistically significant differences for other pairwise comparisons, Wilcoxon rank-sum test). The frequency of fines starts off somewhat higher in Rewards Low Corruption as well, but nearly equalizes across the three treatments towards the end of the experiment (no statistically significant differences, Wilcoxon rank-sum test). Rewards Low Corruption also appear to have the lowest frequency of bribery; however, the differences between treatments are not statistically significant. As in Study 1, and consistent with the behavioral model in Section 3, average bribe sizes are very high; furthermore, they are similar across treatments.

[16] We thank an anonymous referee for bringing this to our attention.

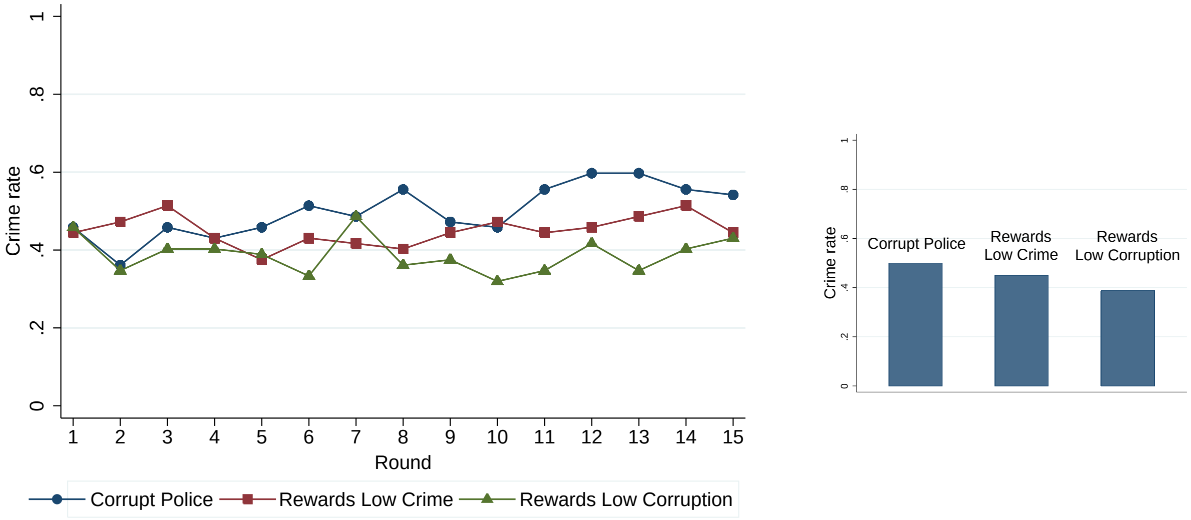

Figure 4: Average crime rates in Study 2, by treatment.

Treatments	Citizens' payoffs		Officers' payoffs		Officers' rewards		Payoff ratio	
	All t	$t > 7$	All t	$t > 7$	All t	$t > 7$	All t	$t > 7$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Corrupt Police	58.4	53.2	193.3	206.4			0.375	0.316
	(2.9)	(2.9)	(7.3)	(7.8)			(0.034)	(0.036)
Rewards Low Crime	66.4	64.0	203.5	209.6	29.7	29.3	0.376	0.353
	(4.7)	(5.6)	(7.9)	(9.5)	(3.4)	(5.0)	(0.037)	(0.044)
Rewards Low Corruption	68.7	65.9	197.4	203.6	26.8	23.6	0.396	0.370
	(6.8)	(7.6)	(11.4)	(12.1)	(6.4)	(7.0)	(0.055)	(0.062)

Table 6: Average payoffs, officers' rewards, and citizen-to-officer payoff ratios in Study 2 (standard errors clustered by society in parentheses).

Table 6 shows average payoffs of citizens and police officers. For the latter, in the Rewards treatments, the payoffs include wages, bribes and rewards. We also report the average rewards received by officers in a separate column. Finally, the last two columns of Table 6 show the citizen-to-officer payoff ratios. Citizens' payoffs are higher in the treatments with rewards as compared to Corrupt Police. Officers' payoffs that are generated by bribery are lower in the Rewards treatment. However, due to the received rewards, officers' total payoffs are essentially the same as in Corrupt Police. The resulting levels of inequality are, therefore, lower in the presence of rewards. However, the estimates are noisy and the differences are not statistically significant.

In the following section, we explore officers' behavior in more detail.

Treatments	If citizen did not break law				If citizen broke law			
	% Demanded bribe		Bribe size		% Demanded bribe		Bribe size	
	All t	$t > 7$	All t	$t > 7$	All t	$t > 7$	All t	$t > 7$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Corrupt Police	60.0	70.5	42.5	43.3	79.3	83.3	46.31	47.6
	(4.2)	(4.4)	(2.2)	(2.7)	(6.2)	(5.8)	(0.83)	(1.1)
Rewards Crime	32.0	38.8	40.3	40.1	87.4	87.9	44.47	45.6
	(5.7)	(8.0)	(4.0)	(4.4)	(5.1)	(5.0)	(2.23)	(2.5)
Rewards Corruption	34.8	40.0	47.2	46.9	75.2	83.3	46.33	47.8
	(11.3)	(12.4)	(1.6)	(1.6)	(9.7)	(7.9)	(1.17)	(1.0)

Table 7: Officers' behavior in corruption treatments (standard errors clustered by society in parentheses).

4.2.2 Officers' behavior

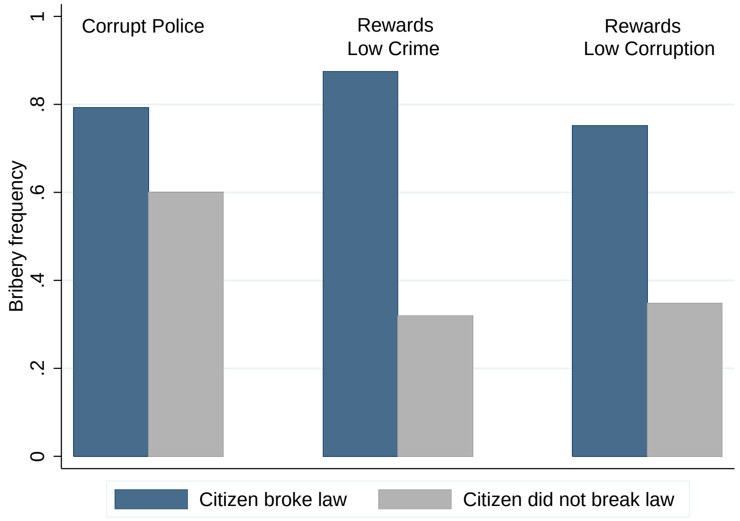

Figure 5: Average frequency of bribery for citizens who did and did not break the law, by treatment.

Recall that in the three treatments of Study 2 – Corrupt Police, Rewards Low Crime and Rewards Low Corruption – officers could demand a bribe from a citizen regardless of whether the citizen broke the law. Table 7 shows average bribery rates and bribe sizes for citizens who did and did not break the law. As before, we provide the averages over all rounds (All t) and over the second half of the experiment ($t > 7$). The averages over all rounds are also visualized in Figure 5.

In the three corruption treatments, police officers impose fines very rarely, cf. Table 5. Instead, as seen from Table 7 and Figure 5, they engage in targeted bribery. That is, police

	Demanded bribe		Bribe size	
	All t	$t > 7$	All t	$t > 7$
	(1)	(2)	(3)	(4)
Rewards Low Crime	-0.286***	-0.317***	-2.4	-3.3
	[-0.44,-0.14]	[-0.52,-0.13]	[-13.17,6.96]	[-15.69,7.17]
Rewards Low Corruption	-0.260*	-0.305*	4.5	3.5
	[-0.51,0.04]	[-0.58,0.01]	[-2.25,10.18]	[-4.09,10.67]
Citizen broke law	0.179*	0.127	3.7*	4.3
	[-0.03,0.38]	[-0.05,0.31]	[-0.03,7.54]	[-1.05,10.45]
(Citizen broke law)	0.374***	0.362***	0.68	1.3
×(Rewards Low Crime)	[0.19,0.55]	[0.18,0.53]	[-10.43,12.79]	[-9.85,13.93]
(Citizen broke law)	0.227**	0.304***	-4.4*	-3.3
×(Rewards Low Corruption)	[0.02,0.42]	[0.11,0.49]	[-8.98,0.65]	[-10.15,3.67]
Round	0.016***	0.010**	0.248*	-0.144
	[0.010,0.022]	[0.001,0.019]	[-0.03,0.55]	[-0.54,0.22]
Intercept	0.478***	0.591***	40.4***	45.0***
	[0.37,0.59]	[0.48,0.71]	[34.45,45.38]	[38.01-50.73]
Observations	3,240	1,728	1,912	1,123
Societies	24	24	24	24
R-squared	0.217	0.196	0.045	0.056

Table 8: OLS regression results for officers' actions. Significance levels, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, and 95% confidence intervals in brackets are based on wild cluster bootstrap.

officers are more likely to demand a bribe from a citizen who broke the law. This effect is present in all treatments, but is especially pronounced in the treatments with rewards ($z = 1.68$, $p = 0.093$ in Corrupt Police; $z = 2.521$, $p = 0.012$ in Rewards Low Crime; $z = 2.521$, $p = 0.012$ in Rewards Low Corruption, Wilcoxon sign-rank tests).

Table 8 presents pooled OLS regression results for officers' decision variables. Columns (1) and (2) show the results of a linear probability model for binary variable *Demanded bribe* (=1 if the officer demanded a bribe from the citizen in the given round, 0 otherwise). Columns (3) and (4) show the results for *Bribe size* (the bribe the citizen had to pay if a bribe was demanded in a given round). Corrupt Police is the base category. The explanatory variables are the two treatment dummies, an indicator for whether the citizen broke the law, its interactions with the treatments, and the time trend. In each case, we present results using data from all rounds (All t) as well as using data from the second half of the experiment ($t > 7$). The following results are based directly on the coefficient estimates in Table 8 or on the Wald tests comparing the appropriate coefficients or their linear combinations.

Result 4 *(a) Police officers are less likely to demand a bribe from law-abiding citizens in both Rewards treatments as compared to Corrupt Police.*

(b) There are no differences between treatments in the rates of bribe demands from law-breakers.

	Citizen broke law			
	All t	$t > 7$	All t	$t > 7$
	(1)	(2)	(3)	(4)
Rewards Low Crime	-0.050	-0.083	0.003	-0.001
	[-0.22,0.14]	[-0.30,0.15]	[-0.085,0.093]	[-0.082,0.080]
Rewards Low Corruption	-0.112	-0.167*	-0.029	-0.037
	[-0.26,0.04]	[-0.36,0.02]	[-0.092,0.031]	[-0.093,0.015]
Broke law$_{t-1}$			0.513***	0.618***
			[0.46,0.58]	[0.52,0.71]
# Others broke law$_{t-1}$			0.080***	0.056**
			[0.04,0.12]	[0.01,0.10]
Bribed$_{t-1}$			0.196***	0.230***
			[0.13,0.27]	[0.16,0.30]
Bribed$_{t-1}$×Broke law$_{t-1}$			-0.201***	-0.209***
			[-0.29,-0.12]	[-0.32,-0.09]
Intercept	0.500***	0.542***	0.113***	0.072**
	[0.40,0.60]	[0.42,0.67]	[0.05,0.18]	[0.01,0.14]
Observations	3,240	1,728	3,024	1,728
Societies	24	24	24	24
R-squared	0.0085	0.019	0.186	0.340

Table 9: OLS regression results for citizens' actions. Significance levels, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, and 95% confidence intervals in brackets are based on wild cluster bootstrap.

(c) Targeted bribery is stronger in the Rewards treatments.

(d) There are no long-term effects of treatments or law-breaking on average bribe sizes (conditional on demanding a bribe).

Results 4(a)-(c) are illustrated in Figure 5 showing the frequencies of police officers demanding bribes from law-abiding and law-breaking citizens in each treatment of Study 2. While targeted bribery is observed in all treatments, it is much more pronounced in the treatments with rewards, due to a reduction in the frequency of bribery for law-abiding citizens.

4.2.3 Citizens' behavior

In this section, we analyze the law-breaking decisions of citizens. Table 9 shows the results of pooled OLS regressions for a binary variable *Broke law* (=1 if the citizen broke the law in the current round, 0 otherwise). Treatment Corrupt Police serves as the baseline. Columns (1) and (2) establish the presence of a basic treatment effect whereby the long-run crime rate is lower in Rewards Low Corruption relative to Corrupt Police ($p = 0.085$ for $t > 7$, based on wild cluster bootstrap). The effect is noisy, but it is consistent with the result of the nonparametric Wilcoxon rank-sum test for all rounds, $p = 0.093$). The long-term reduction in crime rate is about 17 percentage points.[17]

Result 5 *Rewards Low Corruption produces a moderately lower long-run crime rate as compared to Corrupt Police. There is no statistically significant effect for Rewards Low Crime, and no statistically significant difference in crime rate between the two treatments with rewards.*

Columns (3) and (4) present the results of a dynamic regression where treatment differences are washed out by the presence of the lagged dependent variable $Broke\ law_{t-1}$. The latter variable is highly significant, indicating that law-breaking is persistent. A citizen who broke the law in the previous round is about 60 percentage points more likely to break the law again, in the long run, as compared to a citizen who did not. The positive and significant coefficient on $\#Others\ broke\ law_{t-1}$, the number of other citizens in the group who broke the law in the previous round, indicates the presence of some degree of conditional cooperation, which we also observed in Study 1.

The coefficient estimates on variable $Bribed_{t-1}$ (=1 if a bribe was demanded from the citizen in the previous round, 0 otherwise) and its interaction with law-breaking at $t-1$ are highly significant and almost exactly cancel each other, implying that a citizen who did not break the law but was asked to pay a bribe at $t-1$ is about 20 percentage points more likely to break the law at t (as compared to an honest citizen who did not pay a bribe), but a citizen who broke the law and paid a bribe is as likely to break the law again as a law-breaker who was not asked to pay a bribe.

4.2.4 The effects of rewards on officers' behavior

The main finding in the previous section is that while both reward systems reduce bribe demands from law-abiding citizens, only the system that gives officers rewards based on the societal bribery rate (Rewards Low Corruption) is somewhat effective at reducing crime. It is possible that the observed differences in citizens' behavior (Result 5) are driven by differences in officers' reactions to the rewards (cf. the discussion at the end of Section 3). In this section, we analyze responsiveness to rewards in more detail.

Table 10 shows the results of pooled OLS regressions for binary variable *Demanded bribe* (=1 if the officer demanded a bribe from a given citizen) separately for Rewards Low Crime and Rewards Low Corruption treatments. In columns (1) and (3), we control for the number of bribes the officer demanded in the previous round ($\#Bribes\ demanded_{t-1}$), whether the citizen broke the law, and the size of the officer's reward in the previous round ($Reward_{t-1}$). In columns (2) and (4), we additionally control for the interaction of law-breaking in the current round with the previous round reward.

[17] As a robustness check, we also compared the two crime rates for periods $t > t_c$ using different cutoff points t_c. The comparison of means using wild cluster bootstrap is significant at 10% for $t_c = 7, 9, 10, 11, 12$.

	Officer demanded bribe			
	Rewards Low Crime		Rewards Low Corruption	
	(1)	(2)	(3)	(4)
# Bribes demanded$_{t-1}$	0.175***	0.176***	0.145*	0.145*
	[0.08,0.025]	[0.08,0.24]	[-0.03,0.25]	[-0.03,0.25]
Citizen broke law	0.486***	0.537**	0.266**	0.160*
	[0.39,0.59]	[0.20,1.03]	[0.08,0.46]	[-0.003,0.637]
Reward$_{t-1}$	0.023**	0.029	-0.040	-0.049*
	[0.001,0.053]	[-0.030,0.084]	[-0.096,0.010]	[-0.107,0.005]
Citizen broke law		-0.010		0.026
×Reward$_{t-1}$		[-0.081,0.047]		[-0.035,0.093]
Round	0.005	0.005	0.001	0.002
	[-0.002,0.014]	[-0.002,0.013]	[-0.08,0.012]	[-0.007,0.013]
Intercept	-0.106	-0.136	0.370*	0.409*
	[-0.35,0.13]	[-0.54,0.21]	[-0.07,0.88]	[-0.06,0.91]
Observations	1,008	1,008	1,008	1,008
Societies	8	8	8	8
R-squared	0.440	0.440	0.489	0.495

Table 10: OLS regression results for officers' decision to demand a bribe. Significance levels, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, and 95% confidence intervals in brackets are based on wild cluster bootstrap.

As seen from Table 10, the likelihood that a police officer demands a bribe is highly persistent from one round to the next, and increases if the citizen breaks the law, as expected. However, police officers' reactions to rewards are different in the two treatments. In Rewards Low Crime, officers are more likely to demand a bribe the larger the reward they received in the previous round (column (1)). When broken down by whether or not the citizen broke the law (column (2)), the effect of rewards is not significant for law-abiding citizens, but positive and significant for law-breakers ($p = 0.040$, Wald test for the sum of coefficients on $Reward_{t-1}$ and its interaction with law-breaking, wild cluster bootstrap). In contrast, in Rewards Low Corruption officers are less likely to demand a bribe the more they have been rewarded in the previous round, when facing a law-abiding citizen. It can be argued that this is due to the fact that rewards themselves are decreasing in bribery; however, we control for the individual officer's propensity to demand bribes through variable $\#Bribes\ demanded_{t-1}$. Also, rewards are based on the frequency of bribery in the society as a whole. Column (4) further shows that the negative effect of rewards is observed for law-abiding citizens, but not for law-breakers ($p = 0.46$, Wald test, wild cluster bootstrap).

These differences are illustrated in Figure 6 showing the dependence of average frequency of bribery on the level of officers' rewards in the previous round (measured as $9 - (\#\ crimes)$ in Rewards Low Crime and $9 - (\#\ bribes)$ in Rewards Low Corruption, based on society-wide measures of crime and corruption). For law-abiding citizens (left), the frequency of bribery is

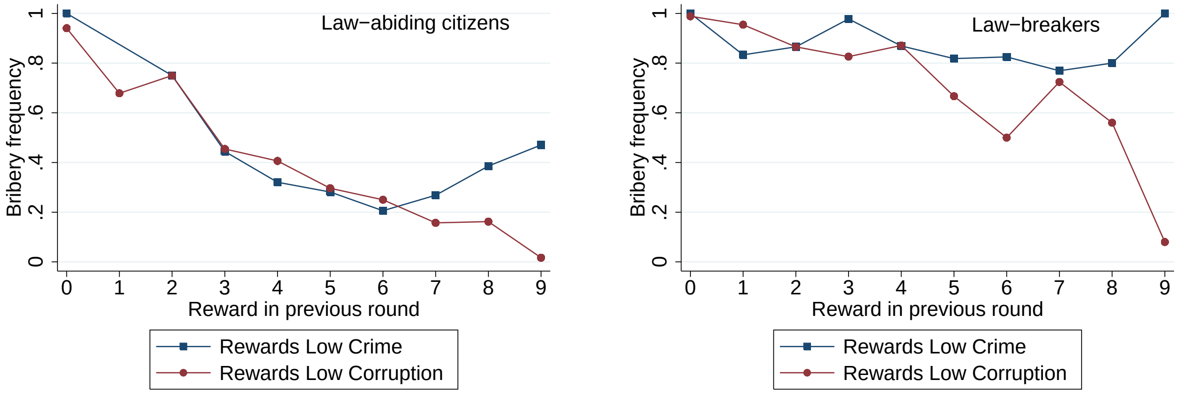

Figure 6: Average bribery rates as functions of officers' rewards in the previous round. *Left*: for law-abiding citizens. *Right*: for law-breakers.

monotonically decreasing in rewards in Rewards Low Corruption, but is U-shaped in Rewards Low Crime. For law-breakers (right), the situation is similar except the frequency of bribery is generally higher. Still, it declines with rewards in Rewards Low Corruption (which may be partly due to reverse causality), but stays at almost 100% and does not decline with rewards in Rewards Low Crime.

Result 6 *In Rewards Low Crime, police officers are more likely to demand a bribe the more they are rewarded in the previous round. In Rewards Low Corruption, rewards have the opposite effect.*

The results in this section show that the different nature of rewards – the fact that in one case rewards are based on citizens' law-breaking and in the other on officers' corruption – may explain why one of them is successful in reducing crime while the other is not. When officers are rewarded for low crime they keep using bribes as a primary law enforcement instrument even when rewards are high because they perceive these rewards as evidence of success. Citizens then have no incentive to decrease law-breaking. On the other hand, when officers are rewarded for not demanding bribes, the rewards reinforce lower corruption levels and citizens can avoid paying bribes by not breaking the law.

5 Conclusions

Corruption among police officers is widespread across the globe, yet little is known about its effects on crime. We conducted a novel laboratory experiment to examine whether the presence of corrupt officers may incentivize citizens to break the law. We found that this is the case, as law-abiding victims of extortion are significantly more likely to break the law in the future.

Nevertheless, the societal crime rate is still lower under corrupt police than under no police. This is because officers use bribes to enforce the law, i.e., they do not indiscriminately demand bribes from law-abiding and law-breaking citizens. Rather, they engage in targeted bribery, whereby law-breaking citizens are significantly more likely to face extortion of bribes than law-abiding citizens.

The bribes demanded are high – close to the maximum possible bribe in our setting – which contributes to the deterrent role of bribery, given that the highest possible bribe is much larger than the fine that an officer could impose on law-breaking citizens. If we restrict the bribes to be no larger than the fine, the societal crime level approaches that observed with no police in the long run. Still, when translating our findings to outside the lab behavior, we can confidently assume that in systemically corrupt societies officers have discretion to demand bribes that are higher than the official fines, which would lead to the crime deterrence result we observe in our primary corruption treatment.

We also tested the effectiveness of two anti-corruption mechanisms that do not rely on the observation and monitoring of individual officers' behaviors – which would be prohibitively costly and unfeasible in high corruption societies. Instead, we implemented two incentive mechanisms that assigned small rewards to all police officers in a society on the basis of either the societal crime rate or the societal bribery rate. Under these mechanisms, the rewards are higher the lower the percentage of citizens breaking the law or being asked to pay a bribe, depending on the treatment. We found both mechanisms to be highly effective in reducing the demand of bribes from law-abiding citizens – termed "extortionary corruption" – and, consequently, in increasing the average earnings of citizens. Moreover, the reward mechanism relying on the societal bribery rate also leads to a decline, albeit a moderate one, in the societal crime rate. This result is driven by officers responding to the received rewards by reducing their demands of bribes from law-abiding citizens, which in turn positively affects the likelihood that citizens obey the law in the future.

In our setting, the highest reward that an officer could get in any given time period (if all other officers were honest or if all citizens obeyed the law) was approximately equal to the maximum bribe that he or she could obtain from one citizen only. On average, the reward that an officer received was about half of a bribe he could have demanded from one citizen. Given that the kind of corruption we are studying is prevalent in developing countries, we believe that it would be feasible, i.e., not prohibitively costly, for an international organization or an NGO to use such reward systems to reduce police corruption. Translating our incentive mechanisms to a field setting would imply awarding small rewards to officers on the basis of crime rates and bribery rates, which could be measured through citizen surveys, which are commonly conducted in high corruption countries.

Overall, our study offers important and novel insights into both the unintended consequences of police corruption on citizens' willingness to break the law and on the use of bribes as law enforcement instruments in settings characterized by high rates of crime and corruption. Moreover, our analysis highlights the effectiveness of an anti-corruption mechanism consisting of assigning small financial rewards to all officers on the basis of the bribery rate observed (e.g., through surveys) in the society as a whole. An interesting extension of our study would be to make such rewards endogenous, by allowing citizens to voluntarily contribute to a "police rewards fund."

Two final remarks are warranted. First, we note that even when corrupt officers engage in targeted bribery, targeting is imperfect. In the Corrupt Police treatment, 60% of law-abiding citizens had to pay a bribe. In the treatments with rewards, the percentage of law-abiding citizens who are extorted bribes goes down to 30%, but it is still significantly different from zero. Second, although officers may successfully use corruption to (imperfectly) deter crime, the use of bribes rather than fines as law enforcement instruments generates a severe redistribution of wealth within the society, i.e., it substantially increases inequality by enriching police officers at the expense of citizens.

References

Abbink, Klaus, Bernd Irlenbusch, and Elke Renner. 2002. "An experimental bribery game." *Journal of Law, Economics, and Organization*, 18(2): 428–454.

Abbink, Klaus, Utteeyo Dasgupta, Lata Gangadharan, and Tarun Jain. 2014. "Letting the briber go free: An experiment on mitigating harassment bribes." *Journal of Public Economics*, 111: 17–28.

Abeler, Johannes, Daniele Nosenzo, and Collin Raymond. 2019. "Preferences for truth-telling." *Econometrica*, 87(4): 1115–1153.

Andreoni, James, and Laura K. Gee. 2012. "Gun for hire: delegated enforcement and peer punishment in public goods provision." *Journal of Public Economics*, 96(11): 1036–1046.

Armantier, Olivier, and Amadou Boly. 2013. "Comparing corruption in the laboratory and in the field in Burkina Faso and in Canada." *Economic Journal*, 123(573): 1168–1187.

Baldassarri, Delia, and Guy Grossman. 2011. "Centralized sanctioning and legitimate authority promote cooperation in humans." *Proceedings of the National Academy of Sciences*, 108(27): 11023–11027.

Banuri, Sheheryar, and Catherine Eckel. 2015. "Cracking down on bribery." *Social Choice and Welfare*, 45(3): 579–600.

Banuri, Sheheryar, and Philip Keefer. 2015. *Was Weber Right? The Effects of Pay for Ability and Pay for Performance on Pro-Social Motivation, Ability and Effort in the Public Sector.* The World Bank.

Barr, Abigail, and Danila Serra. 2010. “Corruption and culture: An experimental analysis.” *Journal of Public Economics*, 94(11): 862–869.

Barrera-Osorio, Felipe, and Dhushyanth Raju. 2017. “Teacher performance pay: Experimental evidence from Pakistan.” *Journal of Public Economics*, 148: 75–91.

Buffat, Justin, and Julien Senn. 2017. “Corruption and cooperation.” *Working Paper.* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3018606.

Butler, Jeffrey V., Danila Serra, and Giancarlo Spagnolo. 2019. “Motivating whistleblowers.” *Management Science, forthcoming.*

Camerer, Colin. 2011. “The promise and success of lab-field generalizability in experimental economics: A reply to Levitt and List (2007+).” *Methods of Modern Experimental Economics*, , ed. Guillaume Frechette and Andrew Schotter. Oxford University Press.

Cameron, A. Colin, and Douglas L. Miller. 2015. “A practitioner’s guide to cluster-robust inference.” *Journal of Human Resources*, 50(2): 317–372.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. “Bootstrap-based improvements for inference with clustered errors.” *The Review of Economics and Statistics*, 90(3): 414–427.

Cason, Timothy N., Lana Friesen, and Lata Gangadharan. 2016. “Regulatory performance of audit tournaments and compliance observability.” *European Economic Review*, 85: 288–306.

Charness, Gary, and Matthew Rabin. 2002. “Understanding social preferences with simple tests.” *The Quarterly Journal of Economics*, 117(3): 817–869.

Cinyabuguma, Matthias, Talbot Page, and Louis Putterman. 2006. “Can second-order punishment deter perverse punishment?” *Experimental Economics*, 9(3): 265–279.

Cooper, Russell, Douglas V. DeJong, Robert Forsythe, and Thomas W. Ross. 1996. “Cooperation without reputation: Experimental evidence from prisoner’s dilemma games.” *Games and Economic Behavior*, 12(2): 187–218.

Coricelli, Giorgio, Mateus Joffily, Claude Montmarquette, and Marie Claire Villeval. 2010. “Cheating, emotions, and rationality: an experiment on tax evasion.” *Experimental Economics*, 13(2): 226–247.

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

Fehr, Ernst, and Urs Fischbacher. 2004. "Social norms and human cooperation." *Trends in Cognitive Sciences*, 8(4): 185–190.

Fischbacher, Urs. 2007. "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental Economics*, 10(2): 171–178.

Fischbacher, Urs, Simon Gächter, and Ernst Fehr. 2001. "Are people conditionally cooperative? Evidence from a public goods experiment." *Economics letters*, 71(3): 397–404.

Foltz, Jeremy D., and Kweku A. Opoku-Agyemang. 2015. "Do higher salaries lower petty corruption? A policy experiment on West Africa's highways." *Working Paper*. http://cega.berkeley.edu/assets/miscellaneous_files/118_-_Opoku-Agyemang_Ghana_Police_Corruption_paper_revised_v3.pdf.

Garoupa, Nuno. 1997. "The theory of optimal law enforcement." *Journal of Economic Surveys*, 11(3): 267–295.

Greiner, Ben. 2015. "Subject pool recruitment procedures: organizing experiments with ORSEE." *Journal of the Economic Science Association*, 1(1): 114–125.

Hasnain, Zahid, Nick Manning, and Jan Henryk Pierskalla. 2012. *Performance-related pay in the public sector: a review of theory and evidence.* The World Bank.

Herrmann, Benedikt, Christian Thöni, and Simon Gächter. 2008. "Antisocial punishment across societies." *Science*, 319(5868): 1362–1367.

Kaplow, Louis, and Steven Shavell. 2002. "Economic analysis of law." *Handbook of Public Economics*, 3: 1661–1784.

Kessler, Judd, and Lise Vesterlund. 2011. "The external validity of laboratory experiments: Qualitative rather than quantitative effects." *Methods of Modern Experimental Economics*, , ed. Guillaume Frechette and Andrew Schotter. Oxford University Press.

Khan, Adnan Q., Asim I. Khwaja, and Benjamin A. Olken. 2015. "Tax farming redux: Experimental evidence on performance pay for tax collectors." *The Quarterly Journal of Economics*, 131(1): 219–271.

Levitt, Steven D., and Thomas J. Miles. 2007. "Empirical study of criminal punishment." *Handbook of Law and Economics*, 1: 455–495.

Luttmer, Erzo F.P., and Monica Singhal. 2014. "Tax morale." *Journal of Economic Perspectives*, 28(4): 149–68.

Markussen, Thomas, Louis Putterman, and Jean-Robert Tyran. 2016. "Judicial error and cooperation." *European Economic Review*, 89: 372–388.

Mascagni, Giulia. 2018. "From the lab to the field: A review of tax experiments." *Journal of Economic Surveys*, 32(2): 273–301.

Mookherjee, Dilip, and Ivan Paak-Liang Png. 1995. "Corruptible law enforcers: how should they be compensated?" *The Economic Journal*, 145–159.

Muthukrishna, Michael, Patrick Francois, Shayan Pourahmadi, and Joseph Henrich. 2017. "Corrupting cooperation and how anti-corruption strategies may backfire." *Nature Human Behaviour*, 1(7): 0138.

Nikiforakis, Nikos. 2008. "Punishment and counter-punishment in public good games: Can we really govern ourselves?" *Journal of Public Economics*, 92(1-2): 91–112.

Olken, Benjamin A., and Patrick Barron. 2009. "The simple economics of extortion: Evidence from trucking in Aceh." *Journal of Political Economy*, 117(3): 417–452.

Olken, Benjamin, and Rohini Pande. 2012. "Lifting the curtain on corruption in developing countries." *VOX.* https://pdfs.semanticscholar.org/5a8b/abdd443b8703f895a7780389af0ee0afb19b.pdf.

Polinsky, A. Mitchell, and Steven Shavell. 2001. "Corruption and optimal law enforcement." *Journal of Public Economics*, 81(1): 1–24.

Roodman, David, Morten Ørregaard Nielsen, James G. MacKinnon, and Matthew D. Webb. 2019. "Fast and wild: Bootstrap inference in Stata using boottest." *The Stata Journal*, 19(1): 4–60.

Ryvkin, Dmitry, Danila Serra, and James Tremewan. 2017. "I paid a bribe: An experiment on information sharing and extortionary corruption." *European Economic Review*, 94: 1–22.

Salmon, Timothy C., and Danila Serra. 2017. "Corruption, social judgment and culture: An experiment." *Journal of Economic Behavior & Organization*, 142: 64–78.

Schmolke, Klaus Ulrich, and Verena Utikal. 2018. "Whistleblowing: incentives and situational determinants." *Available at SSRN 3198104.*

Sequeira, Sandra. 2012. "Chapter 6 Advances in Measuring Corruption in the Field." In *New advances in experimental research on corruption.* 145–175. Emerald Group Publishing Limited.

Shleifer, Andrei, and Robert W. Vishny. 1993. "Corruption." *The quarterly journal of economics*, 108(3): 599–617.

Snowberg, Erik, and Leeat Yariv. 2018. "Testing the waters: Behavior across participant pools." *NBER Working Paper #24781*. https://www.nber.org/papers/w24781.

Thöni, Christian, and Stefan Volk. 2018. "Conditional cooperation: Review and refinement." *Economics Letters*, 171: 37–40.

Zhang, Boyu, Cong Li, Hannelore De Silva, Peter Bednarik, and Karl Sigmund. 2014. "The evolution of sanctioning institutions: an experimental approach to the social contract." *Experimental Economics*, 17(2): 285–303.

A Experimental instructions

The instructions below are for the main part of the *Rewards Low Crime* treatment. Instructions for other treatments are straightforward modifications and available from the authors upon request.

Instructions for Part 1

This part of the experiment consists of 15 decision rounds. At the end of the experiment, one of the 15 rounds will be randomly chosen to base your actual earnings on.

In this experiment you will be randomly assigned to one of two roles: **Private Citizen** or **Police Officer**. You are part of a society made up of 9 Private Citizens and 3 Police Officers and will only interact with Citizens and Officers within your society.

In each round, you will be randomly assigned to a group including one Police Officer and 3 Private Citizens from your society. The identities of the Officer and the Citizens will not be disclosed.

Private Citizens
Each Private Citizen in your group will start with a monetary endowment of 100 ECU and will have to decide whether or not to break the law, for instance by committing a traffic violation, or making false statements in a tax report, or claiming unearned benefits, etc.

Breaking the law generates a monetary **benefit of 20 ECU** to the law breaker, but would also generate a monetary **loss of 20 ECU** to each of the other 2 Private Citizens in the group.

If, for instance, one Citizen decides to break the law while the other 2 Citizens do not, the law breaker would earn 100+20 = 120 ECU whereas the other two Citizens would earn 100-20 = 80 ECU. If, instead, two Citizens break the law, they both earn 100+20-20 = 100 ECU, whereas the Citizen who did not break the law would earn 100-20-20 = 60 ECU. If all 3 Citizens break the law, they all earn 100+20-20-20 = 80 ECU.

The 3 Private Citizens will make the decision to break or not break the law simultaneously.

Police Officers
The Police Officer in your group is in charge of deterring law breaking by Private Citizens in your group. The Police Officer will receive a lump-sum wage of 100 ECU, and will be able to observe whether each of the Private Citizens in your group breaks the law.

After observing the law breaking by the 3 Private Citizens, if any, the Police Officer will have to make one of the following decisions for each Private Citizen in your group:

1) **Give a fine of 27 ECU (only if the Citizen broke the law)**. The Police Officer can only give a fine to Citizens who broke the law, and if the Citizen is given a fine, he or she has no choice but to pay the fine. The amount collected in fines will be equally redistributed among all

the 9 Private Citizens in your society. This means that if the Police Officer gives a fine to one Citizen, and therefore collects 27 ECU, each of the 9 Citizens in your society will subsequently receive 27/9 = 3 ECU. If the Officer gives a fine to two Citizens, and therefore collects 54 ECU, each of the 9 Citizens in your society will subsequently receive 54/9 = 6 ECU. Finally, if the Officer gives a fine to all three Citizens, and therefore collects 81 ECU, each of the 9 Citizens in your society will subsequently receive 81/9 = 9 ECU.

2) **Demand a bribe** between 1 and 50 ECU. If the Police Officer decides to demand a bribe, the Officer will have to choose the size of the bribe, in the range between 1 and 50 ECU. The Officer will be able to keep whatever he or she collects in bribes. If the Police Officer decides to demand a bribe from one or more Private Citizens, the Private Citizen(s) will have no choice but to pay the bribe.

3) Not give a fine and not demand a bribe.

In addition to their lump-sum wages, Police Officers can receive **financial rewards** depending on how successful they are in deterring law breaking by the Citizens in their society. Specifically, each Police Officer will receive **6 ECU for each Citizen in the society who does not break the law**. Recall that there are 9 Citizens in the society. If all Citizens break the law, Police Officers will receive no financial rewards in addition to their wage. If 8 Citizens break the law and 1 Citizen doesn't, each Officer will receive 6 ECU. If 7 Citizens break the law and 2 Citizens don't break the law, each Officer will receive 6*2=12 ECU. If 6 Citizens break the law and 3 Citizens don't, each Officer will receive a reward of 6*3=18 ECU, etc. Finally, if none of the Citizens breaks the law, each Officer receives a reward of 6*9=54 ECU.

The payoff of a Private Citizen in each round will depend on:

- Whether or not the Citizen and other Citizens in the group break the law.
- Whether or not the Police Officer in the group gives the Citizen a fine or demands a bribe from the Citizen. Note that the Police Officer can demand a bribe from a Citizen even if that Citizen did not break the law.
- The total amount of money that the 3 Police Officers in a society collect in fines. This money will be divided equally among all 9 Private Citizens in the society.

The payoff of the Police Officer in each round is determined as follows:

- If the Officer does not demand a bribe and does not give a fine to any of the Private Citizens in the group, the Officer will earn his or her wage of 100 ECU.
- If the Officer gives a fine to one or more Private Citizens, he or she earns the wage of 100 ECU. Whatever the Officer collects in fines will be redistributed among all 9 Private Citizens in the society.
- If the Officer demands a bribe from one or more Private Citizens, he or she earns the wage of 100 ECU and can keep the money he or she collects as bribes.
- If fewer than 9 Citizens in the society break the law, the Officer will receive a reward of 6 ECU for each Citizen who did not break the law.

After the Citizens in a group make their rule-breaking decisions, and after their matched Police Officer makes his or her decisions, each Private Citizen in the group will be told whether the

Officer gave him or her a fine, demanded a bribe, or abstained from both actions. Each Citizen will be also informed about the decisions made by the other two Citizens in the group. However, Citizens will not be informed about the Officer's decisions concerning the other Citizens in the group.

At the beginning of each round roles of Private Citizens and Police Officers will be retained, but groups will be randomly re-matched within your society. It is possible that you will be matched with the same Citizens and/or Police Officer in multiple rounds but you will not know their identities anyway.

Are there any questions?

The experiment is about to begin. Please stay quiet and do not communicate with other participants or look at their monitors. If you have a question from this point on, please raise your hand and one of us will assist you in private.