Do bystanders react to bribery?

Do bystanders react to bribery?*

Alice Guerra† Tatyana Zhuravleva‡

Abstract

Do individuals consider bribery as an acceptable behavior? We use a newly-designed game to study if—and under which conditions—bystanders are willing to express disapproval for bribing behavior through costly punishment. We manipulate two key dimensions: the benefits accrued by corrupt actors and the externality imposed on idle victims. We show that on average bystanders were unresponsive nearly half of the time they witnessed bribery. We also find that context specificity matters, as bystanders were more willing to punish when bribing caused them a disadvantageous inequity with respect to corrupt actors, even if bribing enhanced overall welfare. In an additional experiment testing whether social norms play any role in punishment decisions, we find that norms did not align with the observed bystanders' behavior. This further supports our main result that bystanders did not react to bribery due to a concern for the social norm, but rather for their own comparative disadvantage relative to corrupt actors.

Keywords: Bribery; Third-party punishment; Social norms; Inequity aversion; Experimental economics

JEL Classification: C92; D62; D73; D91; K42

1 Introduction

Bystanders are generally willing to incur personal costs to express disapproval for others' social misconduct through punishment, in the absence of any pecuniary benefits for themselves (Fehr and Schmidt, 1999; Fehr and Gächter, 2002; Fehr and Fischbacher, 2004; Henrich et al., 2006; Charness et al., 2008; Chaudhuri, 2011; Nikiforakis and Mitchell, 2014). They do so for different reasons, including fairness, equity, altruistic motivation to help others or to contribute to the greater good (Fehr and Schmidt, 1999; Fehr and Gächter, 2000, 2002; Charness and Rabin, 2002; Gintis et al., 2003; Masclet et al., 2003). This behavior—referred to as third-party punishment, or the responsive bystander phenomenon (Fischer et al., 2011)—has been argued to play an important role in enforcing social norms and stabilizing human cooperation by deterring unfairness or selfishness

*We thank Maria Bigoni, Marco Casari, Emanuela Carbonara, Pablo Branas Garza, Lorenz Götte, Jan Hausfeld, Alexander Nesterov, Werner Raub, Arthur Schram, and seminar participants at the HSE University, University of Bologna, Copenhagen Business School, European University Institute, for helpful comments.

†Corresponding author. Department of Economics, University of Bologna, Rimini Campus, Via Angherà 22, 47900 Rimini, Italy. ORCID iD: 0000-0003-1956-0270. E-mail: alice.guerra3@unibo.it. Tel.: +39 0541 434 266.

‡Department of Economics, HSE University, Kantemirovskaya Ulitsa, 3A, 194100 St Petersburg, Russia. E-mail: tzhuravleva@hse.ru.

(Fehr and Fischbacher, 2004; Falk et al., 2005; Charness et al., 2008; Balafoutas et al., 2014). The evidence in support of this explanation mostly comes from lab experiments using economic games, mainly conducted in the context of group cooperation and unfair economic interactions (Fehr and Fischbacher, 2002; Nikiforakis, 2010; Balafoutas and Nikiforakis, 2012; Jordan et al., 2016; Acemoglu and Jackson, 2017; Brütt et al., 2020). No evidence has yet been advanced on bystanders' responsiveness in the context of corruption, nor any other illegal behavior. Therefore, an important unresolved question is: *Do individuals consider bribery as an acceptable behavior?* Understanding how bystanders treat wrongdoers can have crucial legal and social implications (Zyglidopoulos and Fleming, 2008; Ottone et al., 2015; Krueger and Hoffman, 2016). The logic is simple yet compelling: individuals who consider corruption as an acceptable behavior are also likely to carry it out or justify it, both to themselves and others (Kubbe and Engelbert, 2017; Abbink et al., 2018; Köbis et al., 2018). This leads to corruption proliferating (Cameron et al., 2009), as it becomes a self-sustaining social norm rather than an exception (Andvig and Moene, 1990; Barr and Serra, 2010; Balafoutas and Nikiforakis, 2012; Gächter and Schulz, 2016; Kubbe and Engelbert, 2017).

In this paper, we analyze if—and under which conditions—bystanders are willing to express disapproval of corrupt behavior through third-party punishment. For this purpose, we employ a bribery experiment à la Barr and Serra (2009) where we introduce—for the first time—the role of bystanders as third-party punishers à la Fehr and Fischbacher (2004). Our (third-party) source of punishment is the methodological key to measure individuals' judgment on corruption (Chen et al., 2020). Unlike second-party punishment bribery games (for a review, see Abbink, 2006)—in which punishers are also the immediate victims—in our novel third-party punishment bribery game—in which punishers are neutral, unaffected observers—we are able to cleanly measure judgment on corruption and disentangle it from self-serving motives such as negative reciprocity or desire for revenge (Nikiforakis et al., 2012; Bone and Raihani, 2015).

The basic idea of our experiment is that each participant is randomly assigned to one of the following roles: private citizens, public officials, other members of society, or monitors. As the simplest bribery situation, each citizen can offer a bribe to a public official in exchange for a corrupt service (e.g. avoiding a fine, exemption from a regulation, tax reduction, ascending a waiting list), which the official can either accept or reject. Accepting the bribe implies providing the corrupt service to the citizen while imposing negative externalities to all other members of society.

The key novel feature of our research is represented by the role of monitors, who act as third-party punishers. They are neither engaged in nor affected by bribery, but they can express disapproval towards corruption by sacrificing their own resources to impose a *larger* monetary cost on the corrupt actors, in the absence of any benefits for themselves. For example, the punishment cost can represent the opportunity cost of reporting corrupt actors to a higher instance, calling the police, appearing in court as a witness, etc. (Cameron et al., 2009; Chaudhuri et al., 2016; Jordan et al., 2016). In most cases, those costs are much lower than the amount of punishment actually imposed on the parties. Hence, as is common in third-party punishment games (Alatas et al., 2009; Jordan et al., 2016), we assume that if a monitor chooses a punishment amount of p, the corrupt

actors suffer a payoff reduction of $2p$ each.

Monitoring corruption by humans has been analyzed by Azfar and Nelson (2007) and Barr et al. (2009) in the context of public service delivery. A key difference is that the monitors in these two contributions are part of the institutional environment (attorney generals in Azfar and Nelson, 2007; public servants in Barr et al., 2009). Instead, in our paper monitors are third parties, in principal "external" to the corruption environment. Our monitors are *randomly selected* among participants—rather than being appointed or elected (cf: Azfar and Nelson, 2007)—and they are provided with *a possibility of reporting* observed corruption, rather than *a duty of monitoring*. Crucially, while those contributions were interested in analyzing the impact of institutional design on monitors' behavior (e.g. whether being appointed *vs* elected affects their being more vigilant), we rather look at monitors' behavior as a measure for individuals' disapproval towards corruption. In other terms, our setting follows the standard third-party punishment paradigm (Fehr and Gächter, 2000, 2002; Charness and Rabin, 2002), in that the monitor's role is meant to measure the way in which individuals without any institutional role perceive corruption, as well as the extent to which those individuals would subsequently be willing to take some action to condemn corruption, if given the choice.[1]

Our experiment is designed to mimic the real-world scenario of impartial bystanders who have witnessed a case of corruption, or know someone who has taken and/or accepted bribes. Those third parties can incur some personal cost (e.g. time, effort) to impose a *larger* cost on the corrupt actors. As our monitors do not have any institutional power to directly punish corrupt activities, the punishment cost imposed upon wrongdoers could represent the cost incurred from being reported to a higher instance, arrested after a bystander's call to the police, or sentenced with an eyewitness testimony. Whether it is direct or indirect, the important feature here is that corrupt actors suffer a cost due to a bystander's reaction. Studying bystanders' intervention is essential, as they might be the only party witnessing the case and being able to report it, thus being able to (even indirectly) impose a punishment.

An illustrative translation of our setting into a real-world scenario is the attempt to bribe a traffic warden by personally offering a small bribe amount to avoid a larger official fine. Traffic warden bribery is a relatively common bribery scenario globally, and it is also a public situation in which there are likely to be bystander witnesses to the event. One might also consider the example in Barr et al. (2009) of a hospital patient who requires a change of linen and offers a bribe to a nurse in exchange for having it done immediately, at the expense of other patients who will have to wait for longer. Imagine that an external person witnesses this situation: Would s/he stay passive or rather react, e.g. by calling the head of department, or the police? Our paper precisely analyzes this kind of situations.

We study the conditions under which bystanders react by manipulating two main components of bribery: the private benefits accrued by the corrupt actors, and the negative externality imposed on idle victims. In designing the two treatments, our goal is to analyze how individuals trade

[1] Our monitors cannot be equated to whistleblowers (Choo et al., 2019; Butler et al., 2020; Mechtenberg et al., 2020), who are directly affected by corrupt transactions and fear retaliation, public scrutiny, or ostracism, which does not apply to our monitors.

off the benefits *vs* externality of corruption, in *both* bribing behavior and third-party punishment. Greater benefits make bystanders worse off with respect to corrupt actors; instead, greater externality increases inequity between the corrupt actors and their idle victims, but makes bystanders better off with respect to the latter. We are particularly interested in studying whether bystanders always condemn corruption, or if their behavior varies with treatments, i.e. with the different payoff inequities generated by corrupt acts.

We find that on average bystanders were *only* willing to incur personal costs to punish corrupt behavior in nearly half of the cases. Our treatment manipulations show that bribery conditions matter: bystanders punished more often as the benefits accrued by corrupt actors increased, but not when negative externality increased. In terms of payoff differential, bystanders used punishment as a response to their *own* greater payoff disadvantage relative to corrupt actors, but not to the greater inequity between corrupt actors and their idle victims, which made them better off with respect to the latter. This suggests that bystanders' behavior was motivated by a *self-focused* inequity aversion rather than an anti-bribery norm: they only cared about their *own* payoff disadvantages relative to corrupt actors, and not also about the negative externality of corruption on others with no personal interest at stake, as prior contributions on third-party punishment would have rather predicted (Fehr and Schmidt, 1999; Fehr and Fischbacher, 2004; Jordan et al., 2016).

To better interpret bystanders' behavior, we conducted an additional experiment—with a new set of subjects—to directly test whether social norms towards bribery play any role in punishment decisions. We elicited social norms by adapting Krupka and Weber's (2013) procedure to our bribery experiment. Here, we find that norms did not align with the observed bystanders' behavioral patterns across treatments. Excluding a social norm-driven behavior provides further support to our main result: third parties did not react for the greater good, but rather to reduce their payoff disadvantage with respect to corrupt actors.

The remainder of the paper is organized as follows. In Section 2, we describe our bribery experiment and its implementation, before we subsequently report the results in Section 3. In Section 4, we describe the additional norm-elicitation experiment, its implementation, and the results. In Section 5, we discuss our findings and conclude with directions for future research.

2 Bribery experiment

2.1 Design

At the beginning of the experiment, participants are randomly matched in eight groups of four players. In each group, the following roles are randomly assigned and retained throughout the end of the experiment: a citizen, a public official, another member of society, and a monitor. All players receive an initial endowment of 50 tokens (1 token=0.20 Euro). The game is sequential. The first player—who acts as a citizen—can offer a bribe of three tokens to the official in exchange for a corrupt service, the value of which to him is 3α, with $\alpha > 1$. If the citizen does not offer a bribe, his final payoff remains 50. If the citizen does offer a bribe, his final payoff depends upon the official's and the monitor's decisions.

The second player—who acts as a public official—is informed about the citizen's decision, and if a bribe is offered, he can either accept or reject it. If the official accepts the bribe, he automatically has to supply the corrupt service to the citizen at the expense of the other members of society. In this case, the citizen's and the official's payoffs each increase by the value of the corrupt service, namely 3α, with $\alpha > 1$. The rationale for this design choice is to ensure mutual gains from corruption for both the citizen and the official, while avoiding excessively unequal payoffs.[2] If the official rejects the bribe, he gives back the three tokens to the citizen and does not supply any corrupt service.

The third players—who act as other members of society—are idle victims: they cannot take any actions, but *all* of them incur a cost 3γ, with $\gamma \geq 1$, for *every* bribe exchanged among all citizen-official pairs. Each other member of society's final payoff is $50 - 3n\gamma$, where $n \in \{0, \ldots, 8\}$ is the number of citizen-official pairs who exchanged bribes.[3]

The fourth player—who acts as a monitor—is a third-party punisher. This role represents the key original feature of our design. The monitor receives information about the choices of the citizen and the official of his group and he can punish—at his own personal cost—the citizen, the official, or both. Specifically, if a bribe has been offered, the monitor can choose either not to take any action, or to forego—out of his endowment—an amount $p_C \in [0, 10]$ to punish the citizen, and $p_O \in [0, 10]$ to punish the official. The punishment is costly to the monitor, whose payoff is reduced by the amount that he has chosen to punish, and it imposes a *larger* monetary cost on the citizen and official by reducing their payoff by twice the punishment amount, $2p_C$ and $2p_O$, respectively.

We use a 2×2 between-subject design where we vary the size of benefits accrued by the corrupt actors, α—setting it either low, $\alpha_L = 3$, or high, $\alpha_H = 6$—and the size of externalities suffered by the other members of society, γ, setting it either low, $\gamma_L = 1$, or high, $\gamma_H = 2$. These manipulations serve to analyze whether individuals' behavior vary with different payoff inequities. Holding negative externality constant, greater benefits make bystanders *worse off* with respect to corrupt actors (9 tokens *fewer* with respect to each corrupt actor's payoff), while their relative payoff with respect to the idle victims does not change. On the other hand, holding benefits constant, greater externality does not change bystanders' relative payoff with respect to corrupt actors, but makes them *better off* with respect to the idle victims (from a minimum of 3 tokens *more* with respect to *each* victim's payoff if one bribe overall was exchanged in a round, to a maximum of 24 tokens *more* if eight bribes overall were exchanged). In terms of overall welfare, in the presence of low

[2] In Abbink et al. (2002), Abbink (2004), and Abbink and Hennig-Schmidt (2006), the bribe is tripled before passing on to the official. The rationale for this multiplication is that the marginal utility for any given bribe is likely to be greater for an official than for a business person or a large firm, given that the latter generally have higher income than the former. This rationale may also be applied to our bribery context, as evidence shows that public server providers have lower income than service recipients (Barr et al., 2009; Abbink and Serra, 2012).

[3] This design choice is similar to Barr and Serra (2009), who highlight that the presence of a group of passive victims could introduce interdependence between individuals' choices within sessions. Some individuals may—or may not—engage in corruption conditional on their expectations concerning the number of other participants choosing to act corruptly. Cameron et al. (2009) avoided this interdependence issue by letting each act of bribery affect only one of the other members of society. Here, we follow Barr and Serra (2009), since we aim to mimic real-world cases where bribery harms many individuals (who are often unidentifiable), and not only one person, as in Cameron et al. (2009).

externality and high benefits, bribery is welfare-enhancing: the total benefits (equal to 36, as the citizen and official both gain 18) exceed the total externality (equal to 24, as the other members suffer a loss of 3 each). In the other cases, bribery is welfare-reducing.[4]

Fig. 1 shows the structure of the game and the final payoffs, expressed in tokens and computed as follows. The citizen's final payoff equals 50 if he does not offer a bribe, $47+3\alpha$ if he offers a bribe that is accepted, minus $2p_C$ if he is punished, and 50 if he offers a bribe that is rejected, minus $2p_C$ if he is punished. The official's final payoff is 50 if he is not offered a bribe, 50 if he is offered a bribe but rejects it, minus $2p_O$ if he is punished, and $53+3\alpha$ if he accepts a bribe, minus $2p_O$ if he is punished. Let $n \in \{0,\ldots,8\}$ be the total number of bribes exchanged in a round among all eight citizen-official pairs. The other member of society's final payoff equals $50-3n\gamma$, ranging between 50 (no bribes exchanged at all) and $50-24\gamma$ (bribes exchanged in all eight citizen-official pairs), and it can reach a minimum of 2 if $\gamma = 2$. The monitor's final payoff equals 50 if he does not punish, minus $p_C \in [0,10]$ and $p_O \in [0,10]$ if he does punish, and it ranges between 30 and 50.[5] In equilibrium, with standard preferences where all players behave as payoff maximizers, the game has the following Nash equilibrium: the citizen offers the bribe, the official accepts it, the monitor does not take any action. As an illustrative example, Fig. B1 in Appendix B shows the game structure in the 'High Externality High Benefits' treatment if the monitor foregoes 10 tokens out of his endowment to punish each corrupt actor, and if bribes have been exchanged in the other seven citizen-official pairs.

To gain a better understanding of the game and treatment variations, consider an example in which a citizen offers a bribe, an official accepts it, and a bystander decides not to punish (his payoff remains 50). Under either low or high externality, the citizen's and official's payoffs are 56 and 62, respectively, under low benefits, and 65 and 71 under high benefits. The payoff differential between the bystander and *each* corrupt actor is -9 when moving from low to high benefits (while holding externality constant), and 0 when moving from low to high externality (while holding benefits constant). Assume that one bribe overall has been exchanged in the society. Under either low or high benefits, each idle victim's payoff is 47 under low externality, and 44 under high externality. The payoff differential between the bystander and *each* idle victim is 0 when moving from low to high benefits (while holding externality constant), and $+3$ when moving from low to high externality (while holding benefits constant).

—-INSERT FIGURE 1 HERE—

The game comprises ten identical and independent rounds, in which the equilibrium does not vary with the number of rounds that the game is played. The roles remain fixed throughout

[4]Only a few experiments thus far have analyzed the distinction between welfare-reducing *vs* welfare-enhancing corruption, albeit not in a third-party punishment framework (Abbink et al., 2002; Barr and Serra, 2009; Cameron et al., 2009). The results are mixed. For example, in Abbink et al. (2002), individuals' engagement in corruption was not affected by welfare considerations, whereas in Cameron et al. (2009) greater social loss discouraged second-party punishment and enhanced corrupt actions, albeit only in Australia, yet not in India, Indonesia, and Singapore.

[5]For reasons of experimental design symmetry, we deliberately chose to provide monitors with the ability to punish officials even if they reject a bribe. Similar to Barr and Serra (2009) and Cameron et al. (2009), we used loaded terms in the instructions, such as "bribe" and "punishment." See also Abbink and Hennig-Schmidt (2006) and Serra (2011) on framing effects in bribery games.

the experiment, whereas in each round participants are randomly rematched to different groups. Accordingly, within one session, groups are reshuffled in every round. To cancel out any learning effects from punishment, after each round neither the citizen, the official, nor the other member of society are informed about the monitors' choices. Moreover, to rule out any peer effect and any bias from knowing the results of the previous round, monitors receive no feedback about the other monitors' choices.[6]

Finally, subjects were presented with debriefing questionnaires to collect their socio-demographics and risk attitudes. We obtained a measure risk aversion as an individual self-assessment on a ten-point scale, where one means "not at all prepared to take risk," and ten "very much prepared to take risk" (Dohmen et al., 2011). We use this survey data as control variables in the econometric analysis (Section 3).

2.2 Procedure

We conducted the experiment in May 2019, with eight sessions (two for each treatment) at the Bologna Laboratory for Experiments in Social Science (BLESS) at the University of Bologna in Italy. Each session lasted approximately 70 minutes in total, with average earnings of 14.56 euros per subject, including a 5 euro show-up fee. The treatment order was randomized to control for session effects (Fréchette, 2012).

Participants were recruited using ORSEE (Greiner, 2015) among graduate and undergraduate students who met the following criteria: being an Italian citizen, and born in the North of Italy. These recruitment filters were applied to rule out any cultural and socio-economic differences in corrupt behaviors that previous contributions have proven to exist between Italy and other countries (Treisman, 2000), as well as within Italy (Del Monte and Papagni, 2007). The experiment was computerized using oTree (Chen et al., 2016), and participants performed all tasks via computer. At the beginning of the experiment, written instructions were handed out and read out loud by the experimenter (sample instructions are provided in Appendix A.1). In order to prevent any bias, the same experimenter and laboratory staff conducted all of the sessions. To ensure that all subjects understood the instructions, a computer-based quiz with eight comprehension questions was conducted before starting the experiment, with direct feedback and explanations in case of an incorrect answer.

A total of 256 subjects participated (64 per treatment): 48.4% females, with a mean age of 24.02, and 51.2% had a university degree (Table B1). A series of balance tests confirmed that the sample is balanced across treatments and roles in terms of personal characteristics (gender, age, education level) and risk aversion.

[6]The repetition of independent rounds, coupled with the random matching ("strangers") protocol and the absence of any feedback to participants allows us to cleanly isolate bystanders' reactions to different bribery situations from strategic motives for punishing, conditional cooperation, negative reciprocity, or reputation formation. For a discussion on this design choice, see Abbink and Serra (2012). See also Duffy and Ochs (2009) for a comparison between fixed matching ("partners") design and random matching ("strangers") design in a sequence of indefinitely-repeated two-player prisoner's dilemma games.

3 Results

We report the experimental results on bribing behavior (Section 3.1) and bystanders' behavior (Section 3.2), whereby this latter is the main focus of our paper.

In our non-parametric analyses, following Barr et al. (2009), we conducted Somers' *D*-tests (Newson, 2002), in which we account for the non-independence of observations within each session by clustering. Similarly, in our parametric analyses, we clustered standard errors at the session level to take account of the within-session interdependency. To correct for the small number of clusters (eight experimental sessions), the *p*-values are determined using the score bootstrap procedure for probit regressions (Kline and Santos, 2012), and the wild-cluster bootstrap procedure for linear regressions (Cameron et al., 2008; Cameron and Miller, 2015), with 10,000 replications and Webb's (2014) weights. Throughout our regression analyses, we use data from all ten rounds, and include controls for gender, age, education level (1 for university degree, 0 for high-school diploma), and risk aversion (ranging from 1 "highly risk taking" to 10 "highly risk averse").[7] Our results are robust to controlling for round fixed effects, start-game and end-game effects (see Appendix C).

3.1 Bribing behavior

Overall, pooling across treatments, citizens offered a bribe 52.03% of the time (333 times out of 640), and officials accepted one if it was offered 70.27% of the time (234 times out of 333). Individuals' bribing behavior was not significantly sensitive to treatments, as revealed by summary statistics and Somers' *D*-tests (Table 1), as well as probit regressions (Table 2).

—-INSERT TABLE 1 HERE—

—INSERT TABLE 2 HERE—

Result 1 *Individuals' bribing behavior was not significantly responsive to greater benefits nor greater externality.*

3.2 Bystanders' behavior

We now turn to the analysis of third-party punishment by treatments. Following Carpenter and Matthews (2009) and Chaudhuri et al. (2016), we analyze bystanders' behavior as stemming from two sequential, interrelated decisions. The first decision is binary, i.e. to either punish corrupt actors or not. The second decision concerns how much to relinquish for punishment, conditional on punishing. This allows us to account for the high percentage of bystanders who chose zero punishment (Fig. 2), and analyze whether bribery conditions affect the two decisions differently. Accordingly, we define the following two variables. *Frequency of Punishment* is a dummy variable equal to 1 if a bystander foregoes a positive amount out of his endowment for punishment. *Size*

[7] In a related paper, we analyze gender differences in bribery behavior and third-party punishment. See Guerra and Zhuravleva (2020).

of Punishment is a discrete variable ranging from 1 to 10, measuring the number of tokens that a bystander foregoes to punish either the citizen or the official, conditional on punishing.

—-INSERT FIGURE 2 HERE—

Table 3 reports summary statistics of bystanders' behavior and Somers' *D*-tests. Overall, pooling across treatments, bystanders decided to punish corrupt citizens (those who offered a bribe; 333 out of 640) and corrupt officials (those who accepted a bribe; 234 out of 640) slightly more than half of the time, i.e. 53.45% and 53.85%, respectively (Panel A, Table 3). More disaggregated information reveals that bystanders decided to punish *only* corrupt citizens 16.52% of the time, *only* corrupt officials 2.14% of the time, and *both* corrupt actors in a pair 51.71% of the time. These percentages are substantially lower than those found in cooperation games (e.g. Fehr and Fischbacher, 2004), in which the frequency of third-party punishment upon free riders goes up to 80% or more. Our data further reveals that those bystanders who punished forego on average 5.089 tokens out of their endowment to punish citizens and 5.746 tokens to punish officials (Panel B, Table 3). Overall, this corresponds to approximately 50% of the maximum amount that they could have foregone to punish each player. As a final remark here, it is interesting to note that citizens' and officials' beliefs about monitors' decisions—as measured in the post-experiment survey—*did not* correspond to actual punishment behavior (cf: Jordan et al., 2016, p.754). Indeed, our data reveals that the majority of corrupt actors (65%) believed that monitors would not have punished them *at all*, while in fact punishment occurred 53.5% of the time towards citizens, and 54.7% of the time towards officials.

Result 2 *Bystanders punished corrupt actors nearly half of the time when they witnessed bribery, and relinquished approximately 50% of the maximum amount that they could have foregone out of their endowment to punish each corrupt actor.*

—-INSERT TABLE 3 HERE—

We next turn to investigating the effect of our manipulations on third-party punishment. Somers' *D*-tests reveal that bystanders' behavior was sensitive to bribery conditions. Greater benefits—under either low or high externality—prompted bystanders to punish corrupt actors more often (Panel A, Table 3), both citizens (42.53% in LB_LE *vs* 66.02% in HB_LE, p=0.038; 39.39% in LB_HE *vs* 61.04% in HB_HE; p=0.044), and officials (48.15% in LB_LE *vs* 65.85% in HB_LE, p=0.071; 32.56% in LB_HE *vs* 58.18% in HB_HE; p=0.000). The size of punishment also increased, although this effect was not statistically significant (Panel B, Table 3). Greater externalities had no significant impact on bystanders' punishment.

Regression analyses confirm those results, as shown in Table 4. Panel A reports probit estimates of the frequency of punishment, and Panel B reports the OLS estimates of the size of punishment. All specifications include treatment dummies, with the LB_LE treatment as the baseline. To test all treatment effects, Table 4 also reports the p-values of Wald tests of hypotheses relating to the coefficients of HB_LE *vs* HB_HE, and LB_HE *vs* HB_HE. Consistent with the non-parametric

tests, the regression estimates show that greater benefits—under either low or high externality—significantly increased the likelihood of punishment towards both citizens and officials (Panel A), and the size of punishment only towards citizens at $p < 0.10$ (Panel B). The regression estimates also reveal that greater externality under low benefits significantly *reduced* the likelihood of punishment towards officials (Panel A). In terms of payoff differentials, third parties were more likely to punish when bribery yielded them a disadvantageous inequity vis-à-vis corrupt actors (which occurred under greater benefits), and not when it produced greater inequity between corrupt actors and their idle victims while yielding them an advantageous inequity (which occurred under greater externality). In terms of welfare, these results reveal that bystanders were more willing to punish welfare-enhancing bribes than welfare-reducing bribes, as the former yield them a comparative payoff disadvantage relative to corrupt actors.

—INSERT TABLE 4 HERE—

Result 3 *Greater benefits—under either low or high externality—significantly increased the likelihood of punishment towards corrupt actors, as well as the size of punishment towards citizens. Greater externality had no effect except under low benefits, in which the likelihood of punishment towards officials significantly decreased.*

Finally, we analyze whether bribery conditions led to any change in bystanders' behavior in terms of punishment differentiation between citizens and officials, i.e. whether they tended to punish one more than the other. OLS regressions show no significant differences (Table B3).

4 Norm-elicitation experiment

We explore whether and to what extent the behavior observed in the bribery experiment can be explained by social norms. We are particularly interested in people's perceptions of the appropriateness of engaging in corrupt behavior and punishing bribery. For this purpose, following Krupka and Weber's (2013) procedure, we conducted an additional norm-elicitation experiment with a new set of subjects, which we describe in the following.

4.1 Design

We elicited social norms towards bribery by adapting Krupka and Weber's (2013) procedure to our bribery experiment. Specifically, we recruited a new set of subjects and asked them to evaluate each of the five decisions that players can take in our bribery game in the four treatments described in Section 2. Subjects were presented four scenarios, corresponding to the four treatments of the bribery experiment (see instructions in Appendix A.2). Each scenario involved a possible corrupt act and punishment decision. For each scenario, we elicited judgments concerning the appropriateness of five different actions: (i) a citizen's decision to offer a bribe to a public official (*OfferBribe*); (ii) the official's decision to accept the bribe (*AcceptBribe*); (iii) a bystander's decision to punish the citizen for offering the bribe (*PunishOffer*); and the bystander's decision

to punish the official for (iv) accepting (*PunishAccept*) or (v) rejecting the bribe (*PunishReject*). In each scenario, we asked subjects to assess the social appropriateness of each action on a four-point scale taking the following values: 'socially very unacceptable,' 'socially quite unacceptable,' 'socially quite acceptable,' and 'socially very acceptable.' Applying a within-subject design, each participant was presented all four scenarios (i.e. four treatments), each with the five actions to assess. To mitigate the order effects, we randomly varied the order of the four scenarios.

For each action, subjects were incentivized to choose the social appropriateness rating that other participants were most likely to assign to that action: they received 2 euros for participation, and an additional 7 euros if their assessment matched the modal response from all participants' assessments in one randomly-determined action. As discussed in Krupka and Weber (2013) and related contributions (Erkut et al., 2015; Banerjee, 2016; Huber and Huber, 2020; Schmidt et al., 2020), this gives subjects an incentive to reveal what they perceive to be the *socially* recognized perceptions of the appropriateness of the described action, rather than their own personal perception of appropriateness (on personal *vs* social norms, see e.g. Wenzel, 2004a,b; Bicchieri, 2005, 2010; Burks and Krupka, 2012).

4.2 Procedure

We recruited another 42 students (47.6% females; mean age of 23.523; 64.2% with a university degree; see Table B1) with the same procedure as in the bribery experiment described in Section 2.2. Accordingly, subjects were recruited via ORSEE (Greiner, 2015) at the BLESS at the University of Bologna in Italy, with being an Italian citizen born in the North of Italy representing the participation criteria. We further follow Krupka and Weber (2013) and the related literature (e.g. Banerjee, 2016; Huber and Huber, 2020) in recruiting a *separate*—yet comparable (Table B1)—set of subjects for the norm elicitation experiment: as no subject in this norm-elicitation experiment participated in the bribery experiment, this should lead social appropriateness ratings elicited here that are independent from subjects' actual behavior in the bribery game.

We conducted the experiment in December 2020 by setting up an *ad-hoc virtual* lab, given that physical labs were closed due to the Covid-19 pandemic. Specifically, we programmed the experiment in Qualtrics, and invited participants to a Zoom meeting where the instructions were read aloud by the experimenter. Participants were recommended in advance to have a stable internet connection and avoid small screens (cell phones, tablets). Details on subjects' device characteristics are reported in Table B2. In our regression analyses, we also control for screen resolution, in addition to subjects' demographics. Participants were allowed to ask questions by typing them in the Zoom chat, and/or unmuting their microphone. To ensure anonymity, participants were asked to replace their name in the Zoom meeting with an ID code, which we randomly assigned them and privately communicated via email through Qualtrics the day before the experiment. Screenshots of the Zoom meeting and more details of the experimental procedure are provided in Appendix A.2. The session lasted about 30 minutes, and subjects were paid via PayPal.

4.3 Results

Following Krupka and Weber (2013), we converted participants' social norm ratings into numerical scores. A rating of 'socially very unacceptable' received a score of -1, 'socially quite unacceptable' a score of $-1/3$, 'socially quite acceptable' a score of $+1/3$, and 'socially very acceptable' a score of $+1$.

Table 5 presents participants' appropriateness ratings for each of the five possible actions per treatment, including the full distribution of responses as well as the mean and standard deviation. The last column reports the pooled results across treatments. The modal responses (shaded and bold) are remarkably similar across treatments: the large majority of subjects thought that offering and accepting bribes is socially very inappropriate (in line with d'Adda et al., 2016), punishing corrupt behaviors—either offering or accepting a bribe—is socially very appropriate, whereas punishing a public official who rejects a bribe is socially very inappropriate.

Some differences emerge in the average ratings between treatments, which we formally analyze by using Somers' D-tests and linear regressions with standard errors clustered at the subject level. We start by using Somers' D-tests to compare the ratings elicited between pairwise treatments, for each of the five actions. For the actions OfferBribe, AcceptBribe, and PunishReject, we detect statistically significant differences between ratings elicited in the LB_LE *vs* LB_HE treatments (all $ps < 0.10$). This means that greater externality under low benefits led individuals to rate those actions as more inappropriate. Moreover, greater externality under high benefits led individuals to rate OfferBribe and PunishReject as more inappropriate (HB_LE *vs* HB_HE; p=0.029 for OfferBribe; p=0.015 for PunishReject). Regarding benefit effects, we find that greater benefits under high externality led individuals to rate AcceptBribe as *less* inappropriate (LB_HE *vs* HB_HE; p=0.027), and greater benefits under low externality led them to rate PunishOffer as *less* appropriate (LB_LE vs HB_LE; p=0.102). These results are confirmed by regression analyses where we clustered standard errors at the subject level (Table 6), and robust to a set of controls (Table B4).

Relating social appropriateness ratings in the norm-elicitation task to actual behavior in the bribery game, we observe that treatment variations do not generally produce comparable effects on social norms as choices. Greater externality—especially under low benefits—led individuals to rate corrupt acts as more inappropriate, and the corresponding punishment decisions as more appropriate. Instead, in the bribery game, greater externality had no effects on bribery engagement (Result 1) nor bystanders' punishment decisions (Result 3). An exception—which stands opposed to social norms—is that greater externality under low benefits *reduced* the likelihood of punishment towards public officials. Moreover, social norms were not (or only weakly) significantly responsive to greater benefits. Instead, greater benefits had a significant impact on bystanders' behavior, yielding higher punishment rates. An exception—which again has social norms as opposed to behavior—is that greater benefits under low externality led individuals to rate punishment towards officials as *less* appropriate, whereas it led bystanders to actually punish them *more* often. Taken together, these results suggest that social norms elicited by a separate group of subjects are not in line with—and may even contradict—our observed bystanders' behavioral patterns.

Result 4 *Greater externality led individuals to rate corrupt acts as more socially inappropriate, and the corresponding punishment decisions as more appropriate. Social norms did not significantly change with greater benefits. Overall, elicited norms did not align with—and may even contradict—the observed bystanders' behavioral patterns across treatments.*

5 Conclusion

A substantial body of experimental literature reveals that a large proportion (80% or more) of bystanders are willing to administer costly punishment as unfair behavior increases (e.g. Fehr and Schmidt, 1999; Fehr and Gächter, 2000, 2002; Charness and Rabin, 2002; Gintis et al., 2003; Masclet et al., 2003; Fehr and Fischbacher, 2004; Henrich et al., 2006; Charness et al., 2008; Chaudhuri, 2011; Nikiforakis and Mitchell, 2014). This evidence mostly derives from cooperation games (e.g. public good games), in which free riding represents the norm violation. Our results reveal that this large percentage of responsive bystanders substantially drops in a bribery context, where bystanders were *only* willing to incur personal costs to punish corrupt behavior in nearly half of the cases. Nonetheless, bribery conditions matter. Bystanders were only responsive to bribery under greater benefits, which yielded them a disadvantageous inequity relative to corrupt actors. Instead, they were unresponsive to bribery under greater externality, which increased inequity between corrupt actors and their idle victims, but yielded them an advantageous inequity relative to the latter.

Third-party punishment appears to demonstrate bystanders' *self-focused* inequity aversion, rather than a genuine displeasure towards corruption. This interpretation is further supported by the results from our additional norm-elicitation experiment à la Krupka and Weber (2013): people's assessment of a 'anti-bribery norm' violation did not align with—and may even contrast—bystanders' punishment decisions. This casts doubts on the idea that third parties are willing to punish at their own costs if there is no personal interest at stake, as prior contributions would rather predict (Carpenter et al., 2004; Jordan et al., 2016; Chaudhuri, 2011). Instead, our analysis reveals a dark side of altruistic punishment (Leibbrandt and López-Pérez, 2011; Guala, 2012; Clavien et al., 2012), whereby bystanders were only willing to punish if successful bribery caused them a disadvantageous payoff with respect to corrupt actors, even if that violation enhanced the overall welfare. Instead, they did not care about the greater inequity between corrupt actors and their idle victims, even if that bribery act reduced the overall welfare, since they were relatively better off relative to corruption victims.

Our findings call for further investigations into the motivations driving altruistic punishment. Among others, one open question for future research follows from our result that greater benefits only significantly increased the size of punishment towards citizens. Why did officials escape the bystanders' wrath at the *intensive* margin? Along similar lines, greater externality significantly reduced the likelihood of punishment towards officials, but not towards citizens. A plausible explanation—which warrants further investigation—is because the citizen was the first mover, being responsible for initiating a corrupt act. Our results further suggest that bystanders' responsive-

ness may depend upon the moral frame (cf: Banerjee, 2016), as analyzing third-party punishment in the context of group cooperation may differ from doing so in illegal frames. Future research could investigate bystanders' reactions to other illegal actions such as tax evasion, money extortion, and vote-buying. Finally, it is our future research agenda to explore whether and under which conditions third-party punishment is associated with emotional reactions such as anger (Jordan et al., 2016), and conduct a similar experiment in different countries to investigate *where*—i.e. in which societies—corruption is punished due to concerns for the social norm *or* inequity aversion.

References

Abbink, K. (2004). Staff rotation as an anti-corruption policy: An experimental study. *European Journal of Political Economy 20*(4), 887–906.

Abbink, K. (2006). *Laboratory Experiments on Corruption*. Cheltenham, UK: Edward Elgar Publishing.

Abbink, K., E. Freidin, L. Gangadharan, and R. Moro (2018). The effect of social norms on bribe offers. *The Journal of Law, Economics, & Organization 34*(3), 457–474.

Abbink, K. and H. Hennig-Schmidt (2006). Neutral versus loaded instructions in a bribery experiment. *Experimental Economics 9*(2), 103–121.

Abbink, K., B. Irlenbusch, and E. Renner (2002). An experimental bribery game. *The Journal of Law, Economics, & Organization 18*(2), 428–454.

Abbink, K. and D. Serra (2012). Anticorruption policies: Lessons from the lab. In *New Advances in Experimental Research on Corruption*, pp. 77–115. Emerald Group Publishing Limited.

Acemoglu, D. and M. O. Jackson (2017). Social norms and the enforcement of laws. *Journal of the European Economic Association 15*(2), 245–295.

Alatas, V., L. Cameron, A. Chaudhuri, N. Erkal, and L. Gangadharan (2009). Gender, culture, and corruption: Insights from an experimental analysis. *Southern Economic Journal*, 663–680.

Andvig, J. C. and K. O. Moene (1990). How corruption may corrupt. *Journal of Economic Behavior & Organization 13*(1), 63–76.

Azfar, O. and W. R. Nelson (2007). Transparency, wages, and the separation of powers: An experimental analysis of corruption. *Public Choice 130*(3-4), 471–493.

Balafoutas, L., K. Grechenig, and N. Nikiforakis (2014). Third-party punishment and counter-punishment in one-shot interactions. *Economics Letters 122*(2), 308–310.

Balafoutas, L. and N. Nikiforakis (2012). Norm enforcement in the city: A natural field experiment. *European Economic Review 56*(8), 1773–1785.

Banerjee, R. (2016). On the interpretation of bribery in a laboratory corruption game: moral frames and social norms. *Experimental Economics 19*(1), 240–267.

Barr, A., M. Lindelow, and P. Serneels (2009). Corruption in public service delivery: An experimental analysis. *Journal of Economic Behavior & Organization 72*(1), 225–239.

Barr, A. and D. Serra (2009). The effects of externalities and framing on bribery in a petty corruption experiment. *Experimental Economics 12*(4), 488–503.

Barr, A. and D. Serra (2010). Corruption and culture: An experimental analysis. *Journal of Public Economics 94*(11-12), 862–869.

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

Bicchieri, C. (2010). Norms, preferences, and conditional behavior. *Politics, Philosophy & Economics 9*(3), 297–313.

Bone, J. E. and N. J. Raihani (2015). Human punishment is motivated by both a desire for revenge and a desire for equality. *Evolution and Human Behavior 36*(4), 323–330.

Brütt, K., A. Schram, and J. Sonnemans (2020). Endogenous group formation and responsibility diffusion: An experimental study. *Games and Economic Behavior 121*, 1–31.

Burks, S. V. and E. L. Krupka (2012). A multimethod approach to identifying norms and normative expectations within a corporate hierarchy: Evidence from the financial services industry. *Management Science 58*(1), 203–217.

Butler, J. V., D. Serra, and G. Spagnolo (2020). Motivating whistleblowers. *Management Science 66*(2), 605–621.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics 90*(3), 414–427.

Cameron, A. C. and D. L. Miller (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources 50*(2), 317–372.

Cameron, L., A. Chaudhuri, E. Nisvan, and L.Gangadharan (2009). Propensities to Engage in and Punish Corrupt Behavior: Experimental Evidence from Australia, India, Indonesia and Singapore. *Journal of Public Economics 93*(8), 843–851.

Carpenter, J. and P. H. Matthews (2009). What norms trigger punishment? *Experimental Economics 12*(3), 272–288.

Carpenter, J. P., P. H. Matthews, et al. (2004). Why punish? Social reciprocity and the enforcement of prosocial norms. *Journal of Evolutionary Economics 14*(4), 407–429.

Charness, G., R. Cobo-Reyes, and N. Jiménez (2008). An investment game with third-party intervention. *Journal of Economic Behavior & Organization 68*(1), 18–28.

Charness, G. and M. Rabin (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics 117*(3), 817–869.

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics 14*(1), 47–83.

Chaudhuri, A., T. Paichayontvijit, and E. Sbai (2016). The role of framing, inequity and history in a corruption game: Some experimental evidence. *Games 7*(2), 13.

Chen, D. L., M. Schonger, and C. Wickens (2016). oTree - An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance 9*, 88–97.

Chen, H., Z. Zeng, and J. Ma (2020). The source of punishment matters: Third-party punishment restrains observers from selfish behaviors better than does second-party punishment by shaping norm perceptions. *PLOS One 15*(3), e0229510.

Choo, L., V. Grimm, G. Horváth, and K. Nitta (2019). Whistleblowing and diffusion of responsibility: An experiment. *European Economic Review 119*, 287–301.

Clavien, C., C. J. Tanner, F. Clément, and M. Chapuisat (2012). Choosy moral punishers. *PLOS One 7*(6), e39002.

d'Adda, G., M. Drouvelis, and D. Nosenzo (2016). Norm elicitation in within-subject designs: Testing for order effects. *Journal of Behavioral and Experimental Economics 62*, 1–7.

Del Monte, A. and E. Papagni (2007). The determinants of corruption in Italy: Regional panel data analysis. *European Journal of Political Economy 23*(2), 379–396.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association 9*(3), 522–550.

Duffy, J. and J. Ochs (2009). Cooperative behavior and the frequency of social interaction. *Games and Economic Behavior 66*(2), 785–812.

Erkut, H., D. Nosenzo, and M. Sefton (2015). Identifying social norms using coordination games: Spectators vs. stakeholders. *Economics Letters 130*, 28–31.

Falk, A., E. Fehr, and U. Fischbacher (2005). Driving forces behind informal sanctions. *Econometrica 73*(6), 2017–2030.

Fehr, E. and U. Fischbacher (2002). Why social preferences matter–the impact of non-selfish motives on competition, cooperation and incentives. *The Economic Journal 112*(478), C1–C33.

Fehr, E. and U. Fischbacher (2004). Third-party punishment and social norms. *Evolution and Human Behavior 25(2)*, 63–88.

Fehr, E. and S. Gächter (2000). Cooperation and punishment in public goods experiments. *American Economic Review 90*(4), 980–994.

Fehr, E. and S. Gächter (2002). Altruistic punishment in humans. *Nature 415*(6868), 137.

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

Fischer, P., J. I. Krueger, T. Greitemeyer, C. Vogrincic, A. Kastenmüller, D. Frey, M. Heene, M. Wicher, and M. Kainbacher (2011). The bystander-effect: a meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies. *Psychological Bulletin 137*(4), 517.

Fréchette, G. R. (2012). Session-effects in the laboratory. *Experimental Economics 15*(3), 485–498.

Gächter, S. and J. F. Schulz (2016, Mar). Intrinsic honesty and the prevalence of rule violations across societies. *Nature 531*(7595), 496–499.

Gintis, H., S. Bowles, R. Boyd, and E. Fehr (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior 24*(3), 153–172.

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

Guala, F. (2012). Reciprocity: Weak or strong? What punishment experiments do (and do not) demonstrate. *Behavioral and Brain Sciences 35*(1), 1–15.

Guerra, A. and T. Zhuravleva (2020). Do women always behave as corruption cleaners? *Available at SSRN: https://ssrn.com/abstract=3601696*. Date accessed: January 28, 2021.

Henrich, J., R. McElreath, A. Barr, J. Ensminger, C. Barrett, A. Bolyanatz, J. C. Cardenas, M. Gurven, E. Gwako, N. Henrich, et al. (2006). Costly punishment across human societies. *Science 312*(5781), 1767–1770.

Huber, C. and J. Huber (2020). Bad bankers no more? Truth-telling and (dis)honesty in the finance industry. *Journal of Economic Behavior & Organization 180*, 472 – 493.

Jordan, J., K. McAuliffe, and D. Rand (2016). The effects of endowment size and strategy method on third party punishment. *Experimental Economics 19*(4), 741–763.

Kline, P. and A. Santos (2012). A score based approach to wild bootstrap inference. *Journal of Econometric Methods 1*(1), 23–41.

Köbis, N. C., D. Iragorri-Carter, and C. Starke (2018). A social psychological view on the social norms of corruption. In *Corruption and Norms*, pp. 31–52. Springer.

Krueger, F. and M. Hoffman (2016). The emerging neuroscience of third-party punishment. *Trends in Neurosciences 39*(8), 499–501.

Krupka, E. L. and R. Weber (2013). Identifying social norms using coordination games: why does dictator game sharing vary? *Journal of the European Economic Association 11(3)*, 495–524.

Kubbe, I. and A. Engelbert (2017). *Corruption and Norms: Why Informal Rules Matter*. Springer.

Leibbrandt, A. and R. López-Pérez (2011). The dark side of altruistic third-party punishment. *Journal of Conflict Resolution 55*(5), 761–784.

Masclet, D., C. Noussair, S. Tucker, and M.-C. Villeval (2003). Monetary and nonmonetary punishment in the voluntary contributions mechanism. *American Economic Review 93*(1), 366–380.

Mechtenberg, L., G. Muehlheusser, and A. Roider (2020). Whistleblower protection: Theory and experimental evidence. *European Economic Review*, 103447.

Newson, R. (2002). Parameters behind nonparametric statistics: Kendall's tau, somers' d and median differences. *The Stata Journal 2*(1), 45–64.

Nikiforakis, N. (2010). Feedback, punishment and cooperation in public good experiments. *Games and Economic Behavior 68*(2), 689–702.

Nikiforakis, N. and H. Mitchell (2014). Mixing the carrots with the sticks: Third party punishment and reward. *Experimental Economics 17*(1), 1–23.

Nikiforakis, N., C. N. Noussair, and T. Wilkening (2012). Normative conflict and feuds: The limits of self-enforcement. *Journal of Public Economics 96*(9-10), 797–807.

Ottone, S., F. Ponzano, and L. Zarri (2015). Power to the People? An experimental analysis of bottom-up accountability of third-party institutions. *The Journal of Law, Economics, & Organization 31*(2), 347–382.

Schmidt, R., C. Schwieren, and A. N. Sproten (2020). Norms in the lab: Inexperienced versus experienced participants. *Journal of Economic Behavior & Organization 173*, 239–255.

Serra, D. (2011). Combining top-down and bottom-up accountability: Evidence from a bribery experiment. *The Journal of Law, Economics, & Organization* (28(3)), 569–587.

Treisman, D. (2000). The causes of corruption: a cross-national study. *Journal of Public Economics 76*(3), 399–457.

Webb, M. D. (2014). Reworking wild bootstrap based inference for clustered errors. *Queen's Economics Department Working Paper No. 1315*.

Wenzel, M. (2004a). An analysis of norm processes in tax compliance. *Journal of Economic Psychology 25*(2), 213–228.

Wenzel, M. (2004b). The social side of sanctions: personal and social norms as moderators of deterrence. *Law and Human Behavior 28*(5), 547.

Zyglidopoulos, S. C. and P. J. Fleming (2008). Ethical distance in corrupt firms: How do innocent bystanders become guilty perpetrators? *Journal of Business Ethics 78*(1-2), 265–274.

Tables and figures to be included in the main text

Table (1) Summary statistics of bribing behavior, by treatment

Treatment	Bribe Offers			Bribe Acceptance		
	Freq.	Pct.	N	Freq.	Pct.	N
LB_LE	87	54.37%	160	54	62.07%	87
HB_LE	103	64.38%	160	82	79.61%	103
LB_HE	66	41.25%	160	43	65.15%	66
HB_HE	77	48.13%	160	55	71.43%	77
Total	333	52.03%	640	234	70.27%	333
H_0	p-value for Somers'D-test			p-value for Somers'D-test		
LB_LE=HB_LE	0.720			0.258		
LB_HE=HB_HE	0.700			0.704		
LB_LE=LB_HE	0.568			0.873		
HB_LE=HB_HE	0.483			0.477		

Notes. The variable *Bribe Offers* is a dummy variable equal to 1 if citizens offered a bribe. The variable *Bribe Acceptance* is a dummy variable equal to 1 if officials accepted a bribe. p-values for Somers' D-tests are clustered at the session level.

Table (2) Marginal effects probit regression of bribing behavior

	Bribe Offers		Bribe Acceptance	
	(1)	(2)	(3)	(4)
HB_LE	0.101	0.060	0.176*	0.193
	(0.175)	(0.167)	(0.094)	(0.127)
LB_HE	-0.129	-0.153	0.028	0.041
	(0.142)	(0.145)	(0.105)	(0.124)
HB_HE	-0.061	-0.055	0.088	0.098
	(0.135)	(0.133)	(0.120)	(0.177)
Female		-0.043		0.064
		(0.074)		(0.183)
Age		-0.021**		-0.005
		(0.009)		(0.018)
University degree		0.083		0.038
		(0.063)		(0.115)
Risk Aversion		-0.007		-0.008
		(0.022)		(0.037)
$H_0^{\dagger}$				
$\beta_{HB\_LE} = \beta_{HB\_HE}$	[0.267]	[0.445]	[0.290]	[0.335]
$\beta_{LB\_HE} = \beta_{HB\_HE}$	[0.535]	[0.416]	[0.538]	[0.572]
Observations	640	640	333	333

Notes. The variable *Bribe Offers* is a dummy variable equal to 1 if citizens offered a bribe. The variable *Bribe Acceptance* is a dummy variable equal to 1 if officials accepted a bribe. Standard errors (in parentheses) are adjusted for clustering at the session level. Significance levels (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$) are determined by the score bootstrap procedure to correct for the small number of clusters (Kline and Santos, 2012).

† Wald tests of hypotheses on the parameters of the regression model; p-values reported in squared parentheses.

Table (3) Summary statistics of bystanders' behavior, by treatment

Panel A: Frequency of Punishment

Treatment	Punish citizens			Punish Officials		
	Freq.	Pct.	N	Freq.	Pct.	N
LB_LE	37	42.53%	87	26	48.15%	54
HB_LE	68	66.02%	103	54	65.85%	82
LB_HE	26	39.39%	66	14	32.56%	43
HB_HE	47	61.04%	77	32	58.18%	55
Total	178	53.5%	333	126	53.85%	234
H_0	p-value for Somers'D-test			p-value for Somers'D-test		
LB_LE=HB_LE	0.038			0.071		
LB_HE=HB_HE	0.044			0.000		
LB_LE=LB_HE	0.661			0.105		
HB_LE=HB_HE	0.738			0.276		

Panel B: Size of Punishment

Treatment	Punish citizens			Punish Officials		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N
LB_LE	3.973	2.061	37	4.961	2.918	26
HB_LE	5.441	3.533	68	6.037	3.403	54
LB_HE	4.346	2.134	26	5.428	2.593	14
HB_HE	5.872	2.763	47	6.031	3.431	32
Total	5.089	2.961	178	5.746	3.229	126
H_0	p-value for Somers'D-test			p-value for Somers'D-test		
LB_LE=HB_LE	0.513			0.524		
LB_HE=HB_HE	0.231			0.812		
LB_LE=LB_HE	0.681			0.844		
HB_LE=HB_HE	0.613			0.952		

Notes. *Frequency of Punishment* is a dummy variable equal to 1 if a bystander relinquished a positive amount out of his endowment to punish corrupt citizens (those who offered a bribe) and/or corrupt officials (those who accepted a bribe). *Size of Punishment* is a discrete variable ranging from 1 to 10, which measures the amount of tokens a bystander relinquished to punish, conditional on punishing. p-values for Somers' D-tests are clustered at the session level.

Table (4) Regression analysis of bystanders' punishment behavior

	Punish Citizens		Punish Officials	
	(1)	(2)	(3)	(4)
Panel A: Frequency of Punishment (marginal effects)				
HB_LE	0.229***	0.240**	0.172**	0.180**
	(0.086)	(0.110)	(0.067)	(0.077)
LB_HE	-0.031	-0.025	-0.154***	-0.184***
	(0.040)	(0.058)	(0.055)	(0.017)
HB_HE	0.178***	0.222***	0.096*	0.097
	(0.042)	(0.067)	(0.056)	(0.061)
Female		0.114		0.077
		(0.109)		(0.143)
Age		0.001		-0.010
		(0.023)		(0.019)
University degree		-0.052		-0.050
		(0.186)		(0.192)
Risk Aversion		-0.041		-0.041
		(0.027)		(0.031)
$H_0^\dagger$				
$\beta_{HB\_LE} = \beta_{HB\_HE}$	[0.600]	[0.824]	[0.081]	[0.166]
$\beta_{LB\_HE} = \beta_{HB\_HE}$	[0.000]	[0.000]	[0.000]	[0.000]
Observations	333	333	234	234
Panel B: Size of Punishment (OLS)				
HB_LE	1.468*	1.625*	1.075	1.735
	(0.650)	(0.775)	(0.937)	(1.196)
LB_HE	0.373	0.419	0.467	0.852
	(0.606)	(0.724)	(1.208)	(1.713)
HB_HE	1.899***	2.055**	1.070	1.419
	(0.322)	(0.865)	(1.021)	(1.145)
Female		-2.263**		-1.277
		(0.938)		(1.637)
Age		-0.022		-0.001
		(0.157)		(0.302)
University degree		-0.014		-0.328
		(0.928)		(1.093)
Risk Aversion		0.118		-0.469
		(0.184)		(0.310)
Constant	3.973***	5.026	4.962***	8.124
	(0.316)	(4.354)	(0.743)	(8.389)
$H_0^\dagger$				
$\beta_{HB\_LE} = \beta_{HB\_HE}$	[0.475]	[0.558]	[0.995]	[0.668]
$\beta_{LB\_HE} = \beta_{HB\_HE}$	[0.022]	[0.019]	[0.626]	[0.568]
Observations	178	178	126	126

Notes: Panel A reports the marginal effects of probit estimates of *Frequency of Punishment*—a dummy variable equal to 1 if a bystander relinquished a positive amount out of his endowment to punish corrupt citizens (those who offered a bribe) and/or corrupt officials (those who accepted a bribe). Panel B reports the OLS estimates of *Size of Punishment*—a discrete variable ranging from 1 to 10 and measuring the amount of tokens a bystander relinquished to punish, conditional on punishing. Standard errors (in parentheses) are adjusted for clustering at the session level. Significance levels (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$) are determined by the score bootstrap procedure (Kline and Santos, 2012) in Panel A, and by the wild cluster bootstrap procedure (Cameron et al., 2008) in Panel B, to correct for the small number of clusters.

† Wald tests of hypotheses on the parameters of the regression model; p-values reported in squared parentheses.

Table (5) Appropriateness ratings across treatments

	Treatment				
	LB_LE	HB_LE	LB_HE	HB_HE	Total
OfferBribe					
- - (%)	42.86	**42.86**	**57.14**	**52.38**	**48.81**
- (%)	**45.24**	40.48	38.10	40.48	41.07
+ (%)	7.14	14.29	2.38	4.76	7.14
+ + (%)	4.76	2.38	2.38	2.38	2.98
Mean	-0.508	-0.492	-0.667	-0.619	-0.571
Std. Dev.	0.532	0.527	0.448	0.469	0.496
AcceptBribe					
- - (%)	**73.81**	**69.05**	**83.33**	**73.81**	**75.00**
- (%)	14.29	16.67	11.90	16.67	14.88
+ (%)	7.14	11.90	4.76	4.76	7.14
+ + (%)	4.76	2.38	0.00	4.76	2.98
Mean	-0.714	-0.683	-0.857	-0.730	-0.746
Std. Dev.	0.554	0.536	0.346	0.532	0.499
PunishOffer					
- - (%)	2.38	2.38	7.14	4.76	4.17
- (%)	11.90	23.81	11.90	14.29	15.48
+ (%)	38.10	30.95	30.95	**40.48**	35.12
+ + (%)	**47.62**	**42.86**	**50.00**	**40.48**	**45.24**
Mean	0.540	0.429	0.492	0.444	0.476
Std. Dev.	0.520	0.581	0.621	0.569	0.570
PunishReject					
- - (%)	**61.90**	**73.81**	**76.19**	**83.33**	**73.81**
- (%)	28.57	16.67	16.67	16.67	19.64
+ (%)	2.38	4.76	2.38	0.00	2.38
+ + (%)	7.14	4.76	4.76	0.00	4.17
Mean	-0.635	-0.730	-0.762	-0.889	-0.754
Std. Dev.	0.574	0.532	0.506	0.251	0.487
PunishAccept					
- - (%)	7.14	4.76	9.52	7.14	7.14
- (%)	7.14	11.90	4.76	9.52	8.33
+ (%)	23.81	19.05	21.43	23.81	22.02
+ + (%)	**61.90**	**64.29**	**64.29**	**59.52**	**62.50**
Mean	0.603	0.619	0.603	0.571	0.599
Std. Dev.	0.608	0.592	0.643	0.622	0.611

Notes: Responses are 'Socially Very Unacceptable' (- -) $=-1$; 'Socially Quite Unacceptable' (-) $=-1/3$; 'Socially Quite Acceptable' (+) $=+1/3$; 'Socially Very Acceptable' (+ +) $=+1$. Modal responses are shaded and bold.

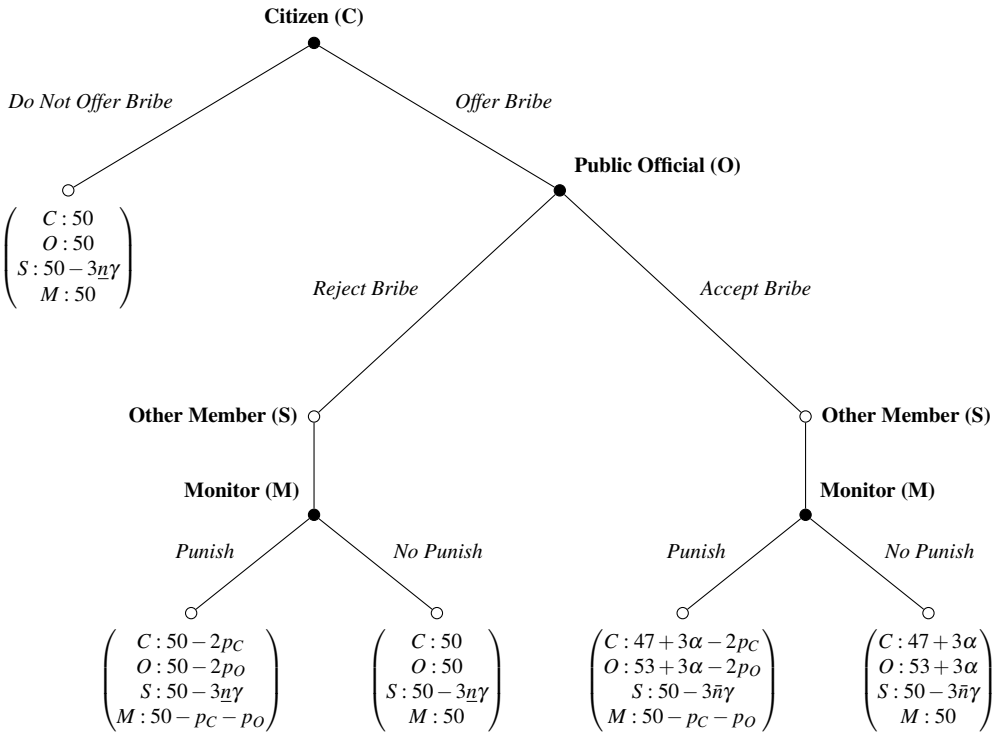
Table (6) OLS regression of appropriateness ratings

	(1) OfferBribe	(2) AcceptBribe	(3) PunishOffer	(4) PunishReject	(5) PunishAccept
HB_LE	0.016	0.032	-0.111*	-0.095	0.016
	(0.074)	(0.083)	(0.060)	(0.075)	(0.071)
LB_HE	-0.159**	-0.143*	-0.048	-0.127*	-0.000
	(0.064)	(0.071)	(0.084)	(0.070)	(0.073)
HB_HE	-0.111*	-0.016	-0.095	-0.254***	-0.032
	(0.065)	(0.078)	(0.071)	(0.086)	(0.076)
Constant	-0.508***	-0.714***	0.540***	-0.635***	0.603***
	(0.083)	(0.086)	(0.081)	(0.089)	(0.095)
$H_0^\dagger$					
$\beta_{HB\_LE} = \beta_{HB\_HE}$	[0.032]	[0.541]	[0.831]	[0.024]	[0.541]
$\beta_{LB\_HE} = \beta_{HB\_HE}$	[0.266]	[0.032]	[0.558]	[0.121]	[0.690]
Observations	168	168	168	168	168

Notes: Standard errors clustered at the subject-level are in parentheses. Results are robust to controls (see Table B4). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

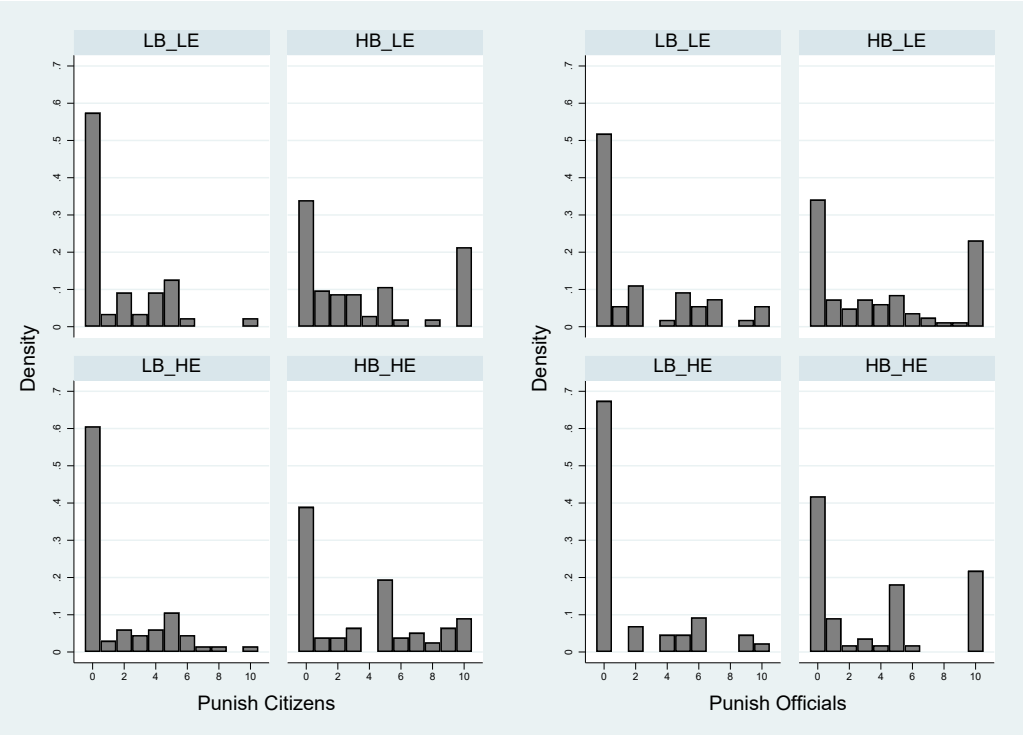
† Wald tests of hypotheses on the parameters of the regression model; p-values reported in squared parentheses.

Figure (1) *The bribery game structure*



Notes: 50 is the initial endowment; 3 is the bribe amount; p_C and p_O are the amounts chosen by the monitor to punish the citizen and the public official, respectively; $n \in [0,\ldots,8]$ is the total number of bribes exchanged in a round, with $\underline{n} \in [0,\ldots,7]$ and $\bar{n} \in [1,\ldots,8]$; $\alpha \in \{3,6\}1$ and $\gamma \in \{1,2\}$ represent the treatments (benefits and externality, respectively).

Figure (2) The distribution of bystanders' punishment expenditures, by treatment



Appendix A Experimental instructions

A.1 Bribery experiment (treatment HB_HE)

Welcome! And thanks for participating. This is an experiment to evaluate how people make economic decisions. If you pay attention, the instructions will help you make these decisions and earn a sum of money. This sum depends both on yours decisions during the experiment , and on the decisions of the other participants.

The earnings will be calculated in tokens, converted to Euro at the end of the experiment and paid in cash at the end of today's session. For every 5 tokens, you will receive 1 Euro. In addition, you will receive 5 Euros for your participation.

Your answers will remain confidential and final payments will be made in a sealed envelope. You are not obliged to communicate your earnings to anyone. It is important not to communicate in any way with other people in the room until the end of the experiment. The use of cell phones is prohibited. Please take a moment to turn off your cell phones and put away all outside materials. You can ask questions at any time. If you have a question, please raise your hand and we will respond in private.

Anonymity. When you entered the room , you took a piece of paper with a number. This is your identification code and will be used to guarantee your anonymity. The decisions you make during the experiment will be matched to your identification code, never to your name. No one (i.e. neither the people who will analyze the data, nor other participants, nor any other person) will ever be able to match your name to the decisions you will take during the experiment.

Ten Rounds. This experiment consists of **10 rounds**. Each round is completely independent of the others: the decisions taken in a round will not influence neither your gain nor the procedures in the following rounds. At the end of the experiment, the computer will randomly choose one round for your payment.

Roles and Groups. At the beginning of today's experiment, you will be randomly assigned to one of four roles: a citizen, a public official, another member of society, a monitor.

At the beginning of each round, the computer will randomly form **groups of four participants**, each with a different role. The composition of the groups will randomly change from round to round. You will never know the identity of the other members of the group, nor will the other members of the group know yours.

Tasks. At the beginning of each round every player receives **50 tokens**. Each round consists of **four stages**.

First stage: the citizen's task. If you are a citizen, you can decide to offer a bribe to the public official in your group.

- If you do not offer a bribe to the public official, your payoff remains equal to 50 tokens. The round ends and the next one starts.
- If you offer a bribe to the public official, you offer 3 tokens to the public official. These 3 tokens will be returned to you only if the public official rejects the bribe. Please see the second stage.

Second stage: the public official's task. If you are a public official, you will be notified of the decision made by the citizen in your group. If the citizen in your group has offered you a bribe, you can decide whether to accept it or reject it. Accepting the bribe implies accepting the 3 tokens offered by the citizen and granting a favour to the citizen, which causes a loss to the other members of society. Specifically:

If you accept the bribe:

- You receive 18 additional tokens (in addition to the 3 tokens offered by the citizen);
- The citizen receives 18 additional tokens;
- Each other member of society in this room loses 6 tokens, and may lose more tokens if bribes are offered and accepted in the other groups in this room;
- You and the citizen may receive a punishment by the monitor (see the fourth stage).

If you reject the bribe:

- You receive 0 additional tokens (and return the 3 tokens to the citizen);
- The citizen receives 0 additional tokens;
- Each other member of society in this room may lose tokens if bribes have been offered and accepted in the other groups in this room;
- You and the citizen may receive a punishment by the monitor (see the fourth stage).

Third stage: the role of the other members of society. If you are another member of society, your role is idle: you cannot take any decision. *Per each bribe* offered and accepted in this room, you lose 6 tokens.

While waiting for the other participants' decisions, we will ask you to answer some questions, with no effects on your earnings.

Fourth stage: the monitor's task. If you are a monitor, you will receive information about the decisions made by the citizen and the public official in your group.

You can choose whether — and *how much* — to punish the citizen and/or the public official, at your own expenses. In particular, you can choose an amount between 0 and 10 tokens to punish the citizen, and an amount between 0 and 10 tokens to punish the public official. Your payoff will be reduced by the punishment amount you have chosen.

The punishment amount that you choose to punish the citizen will be multiplied by 2, and the payoff of the citizen will be reduced by this doubled amount. The same holds for the public official.

For example, if you spend 5 tokens to punish the citizen and 2 tokens to punish the public official: your earnings are reduced by 7 tokens ($= 5 + 2$); the citizen's earnings are reduced by 10 tokens ($= 5 \times 2$); the earnings of the public official are reduced by 4 coins ($= 2 \times 2$).

A.2 Norm-elicitation experiment

A.2.1 Instructions

Welcome! And thanks for participating to this study, which will last approximately half an hour.

Warning: It is preferable not to use a mobile device, as this study may not be displayed correctly on small screens. Also, please note that you will not be able to go back to previous pages during the entire study.

With your participation, you will contribute to the research and earn money. You will receive 2 euros as participation fee, and you can earn more money based on your decisions and those made by the other participants. You will be paid via PayPal within 14 days. Participants who leave the study at any stage prior to its completion are not eligible for payment.

Your identity will remain anonymous, and your choices will remain confidential, i.e. not disclosed to any other participant and not linked to your identity. The anonymized data will be used for scientific purposes only.

Read the instructions carefully and do not communicate with others during the entire study. If you have a question, click on "raise your hand," and the experimenter will help you.

[NEW PAGE]

In this study, your task is to evaluate possible choices of various individuals, in different situations. We will ask you to indicate, for each choice, which one you think most people find inappropriate or appropriate.

Let's make an example. There is a group of people consisting of a citizen, a public official, a

monitor, and eight other members of society. At the beginning, each of them has 50 tokens.

The citizen can decide whether or not to offer a bribe to the public official in exchange for a favor. If the citizen offers the bribe, the public official can decide whether to accept or reject it. For every bribe offered and accepted, eight other members of society lose some tokens. Specifically:

If the citizen offers the bribe and the public official accepts it:

- The citizen and the public official earn 9 extra tokens each (therefore, total gain = +18);
- The other eight members of society lose 3 tokens each (thus, total loss = -24).

If the citizen offers the bribe and the public official rejects it:

- The citizen and the public official earn no extra tokens;
- The eight members of society lose nothing.

If the citizen does not offer the bribe to the public official:

- As above (no gain, no loss).

The monitor observes the decisions. If the citizen offers the bribe, the monitor can punish the citizen and/or the public official, at his own expenses.

If the monitor decides to punish:

- His earnings will be reduced by the amount of the punishment;
- The earnings of the citizen and the public official will each be reduced by double the amount of the punishment.

What is your task?

In the following pages, you will read situations similar to the one described above, and you will be asked to evaluate each of the possible choices of the various individuals (e.g. offer a bribe; reject a bribe; punish the citizen; etc), and decide whether this choice is socially very inappropriate, socially quite inappropriate, or socially quite appropriate, socially very appropriate.

By socially appropriate, we mean the behavior that most people agree on its being the right or ethical thing to do.

Remember: we are not asking what you personally think should be done, but what you think most people find appropriate or inappropriate.

How much can you earn?

Your earnings for this experiment will be computed as follow. You will be asked to assess 20 different actions (divided in four different pages), similar to the one described in the example.

At the end of the study, we will randomly select one of these 20 actions. For that randomly selected action, we will check which response was most frequently given by the participants in this study (this is called the modal response).

If you guess the modal response, you receive 7 euros; otherwise, you only receive the participation fee of 2 euros. In other terms, if you give the answer more frequently given by the participants in this study, then you receive another 7 euros.

[NEXT PAGE]

Are there any questions? If everything is clear, then click on the arrow to start.

[NEXT PAGE]

If the citizen offers the bribe and the public official accepts it:

- Citizen and public officer earn *[9 LB / 18 HB]* extra tokens each (thus, total earnings = + *[18 LB / 36 HB]*)
- The other eight members of society lose *[3 LE / 6 HE]* tokens each (thus, total loss = - *[24 LE / 48 HE]*).

If the citizen offers the bribe and the public official rejects it:

- Citizens and public officials do not earn anything extra
- The eight members of society lose nothing

If the citizen does not offer the bribe to the public official: as above (no gain, no loss).

If the monitor decides to punish:

- His earnings will be reduced by the amount of the punishment;
- The earnings of the citizen and the public official will each be reduced by double the amount of the punishment.

In the table below, we show all the possible choices. For each of them, indicate whether it is socially very inappropriate, socially quite inappropriate, socially quite appropriate or socially very appropriate.

Remember:

- We are not asking what you personally think should be done, but what you think that most people find socially inappropriate or appropriate.
- If one of these choices is the one randomly selected for payment, you receive 7 euros if your answer is the one given most frequently by the other participants in this study.

A.2.2 Information email to participants

The following email (translated from Italian to English) was privately sent to each participants via Qualtrics the day before the experiment. It contains the personal anonymous ID code, instructions for the Zoom meeting, and the Qualtrics link for the experiment.

Welcome!

Thank you for agreeing to participate in our study, which will be conducted online on Zoom on December 22nd at 10 am.

In the following you can find the instructions to participate: read them carefully!

For this online study, you need to have a stable internet connection, and it is strongly recommended to use a computer (avoid tablets or cell phones) as the study may not display properly on small screens.

From 9:40 am on Tuesday 22nd, you can access the virtual lab on Zoom by clicking on the following link:

[ZOOM LINK HERE]

Alternatively, you can log in using the following Meeting ID and passcode:
Meeting ID: 949 6367 3994
Passcode: JqzK8K

IMPORTANT: To ensure your anonymity, we have randomly assigned you an ID CODE, which you need to use on Zoom instead of your name.

Your ID CODE: $m : //ExternalDataReference$

Changing the name on Zoom—and replacing it with your ID CODE—is simple! You need to follow these short instructions: *[LINK TO A PDF FILE ON GOOGLE DRIVE HERE]*

When all the participants will be present, we will begin the real study, which you can access

via the following link (the password will be communicated to you on Zoom):

$l : //SurveyLink?d = TaketheSurvey$

Or copy and paste this URL into your internet browser:

$l : //SurveyURL$

For any problem, do not hesitate to contact me via email: alice.guerra3@unibo.it

We are looking forward to our Zoom meeting!

Best

Alice Guerra

A.2.3 Screenshots

Figure (A1) Exemplary screenshot of the decision screen (in Italian)

	socialmente molto inappropriata	socialmente alquanto inappropriata	socialmente alquanto appropriata	socialmente molto appropriata
Il cittadino offre la tangente	○	○	○	○
Il pubblico ufficiale accetta la tangente offerta	○	○	○	○
Lo spettatore punisce il cittadino che offre una tangente	○	○	○	○
Lo spettatore punisce il pubblico ufficiale che rifiuta la tangente offerta	○	○	○	○
Lo spettatore punisce il pubblico ufficiale che accetta la tangente offerta	○	○	○	○

Figure (A2) Exemplary screenshot of the Zoom meeting with anonymous ID codes

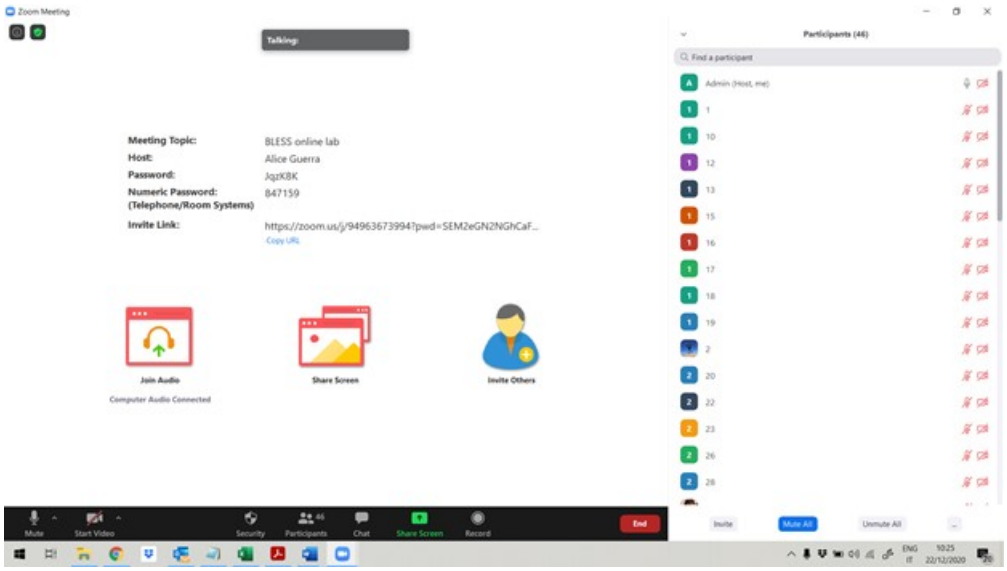
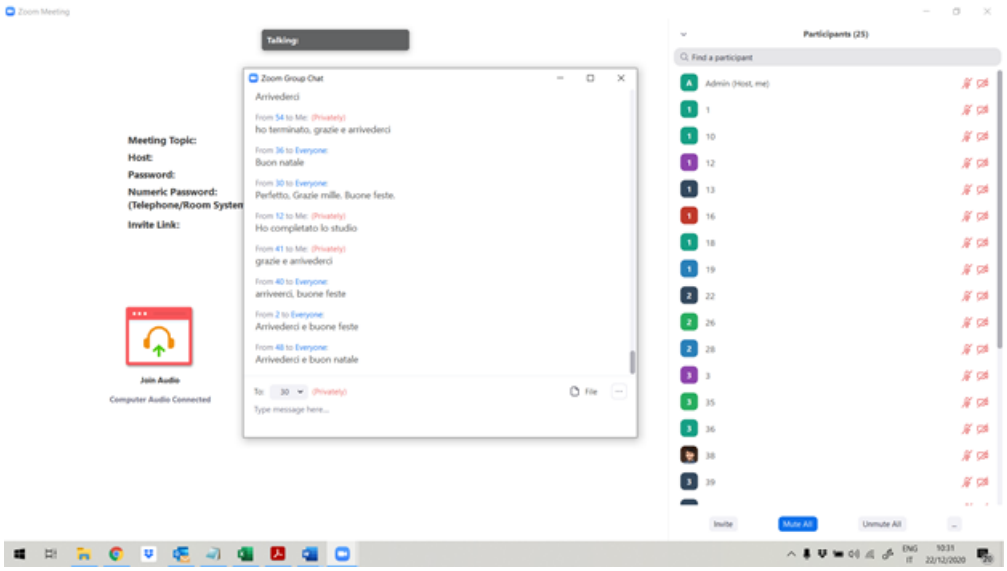


Figure (A3) Exemplary screenshot of the chat over the Zoom meeting



Appendix B Additional Figures and Tables

Table (B1) Subjects' demographics

Variable	Bribery experiment			Norm-elicitation exp.		
	mean	s.d	median	mean	s.d	median
Female (proportion)	0.484	0.501	0	0.476	0.505	0
Age	24.02	4.068	23	23.524	3.285	23
University degree (proportion)	0.512	0.501	1	0.643	0.485	1
Risk aversion	5.949	1.955	6			
	N=256			*N=42*		

Notes: The table reports demographic information of participants for both bribery experiment and the additional norm-elicitation experiment.

Table (B2) Subjects' device characteristics in the norm-elicitation experiment

Variable	Freq.	Pct.	Cum.Pct.
Screen resolution:			
1280x720	6	14.286	14.286
1280x800	1	2.381	16.667
1366x768	17	40.476	57.143
1368x912	1	2.381	59.524
1440x900	5	11.905	71.429
1536x864	7	16.667	88.095
1600x900	1	2.381	90.476
1920x1080	3	7.143	97.619
2048x1152	1	2.381	100
Browser:			
Chrome	32	76.19	76.19
Edge	3	7.143	83.333
Firefox	4	9.524	92.857
Safari	3	7.143	100
Operating System:			
Linux x86_64	2	4.762	4.762
Macintosh	6	14.286	19.048
Ubuntu	1	2.381	21.429
Windows NT 10.0	32	76.19	97.619
Windows NT 6.1	1	2.381	100
	N=42		

Notes: The table reports subjects' device characteristics in the norm-elicitation experiment.

Table (B3) OLS regression of bystanders' punishment towards citizens *vs* officials

DV: Punish Citizens − Punish Officials	(1)	(2)
HB_LE	0.870	0.726
	(0.567)	(0.662)
LB_HE	1.056	1.186
	(0.764)	(0.785)
HB_HE	1.089	0.982
	(0.891)	(1.006)
Female		-0.333
		(0.771)
Age		0.0361
		(0.113)
University degree		0.499
		(0.562)
Risk Aversion		0.226
		(0.150)
Constant	-0.870*	-3.090
	(0.444)	(3.174)
$H_0^\dagger$		
$\beta_{HB\_LE} = \beta_{HB\_HE}$	[0.805]	[0.804]
$\beta_{LB\_HE} = \beta_{HB\_HE}$	[0.975]	[0.844]
Observations	234	234

Notes: The table reports the OLS regression estimates, where the Dependent Variable (DV) is the difference in the size of punishment towards corrupt citizens (those who offered a bribe) and/or corrupt officials (those who accepted a bribe). Standard errors (in parentheses) are adjusted for clustering at the session level. Significance levels (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$) are determined by the wild cluster bootstrap procedure (Cameron et al., 2008) to correct for the small number of clusters.

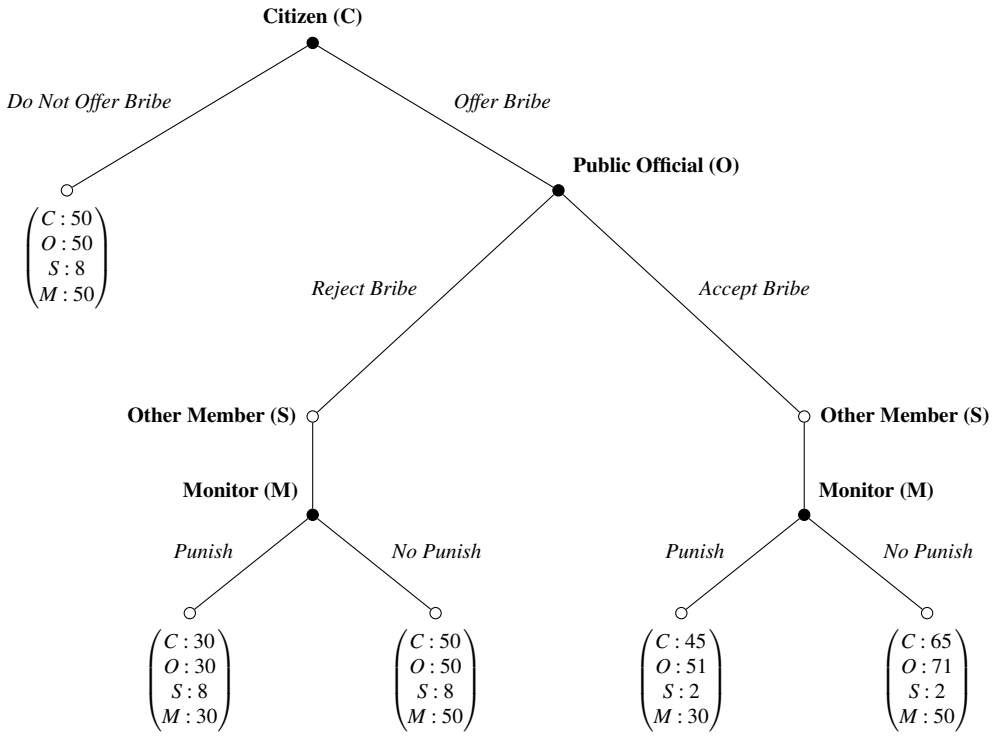
† Wald tests of hypotheses on the parameters of the regression model; *p*-values reported in squared parentheses.

Table (B4) OLS regression of appropriateness ratings (with controls)

	(1) OfferBribe	(2) AcceptBribe	(3) PunishOffer	(4) PunishReject	(5) PunishAccept
HB_LE	0.016	0.032	-0.111*	-0.095	0.016
	(0.075)	(0.084)	(0.061)	(0.075)	(0.071)
LB_HE	-0.159**	-0.143*	-0.048	-0.127*	-0.000
	(0.065)	(0.072)	(0.085)	(0.071)	(0.073)
HB_HE	-0.111*	-0.016	-0.095	-0.254***	-0.032
	(0.065)	(0.079)	(0.072)	(0.087)	(0.077)
Female	0.342***	0.345***	-0.561***	0.208	-0.634***
	(0.113)	(0.105)	(0.145)	(0.125)	(0.151)
Age	0.052*	0.066***	-0.029	-0.022	-0.012
	(0.030)	(0.024)	(0.021)	(0.013)	(0.021)
University degree	-0.056	0.035	0.129	0.323**	-0.139
	(0.139)	(0.102)	(0.137)	(0.128)	(0.132)
Screen Resolution	0.056**	0.041	-0.034	-0.006	0.014
	(0.026)	(0.026)	(0.027)	(0.027)	(0.030)
Constant	-2.087***	-2.618***	1.536***	-0.401	1.218**
	(0.693)	(0.564)	(0.500)	(0.312)	(0.525)
$H_0^{\dagger}$					
$\beta_{HB\_LE} = \beta_{HB\_HE}$	[0.034]	[0.546]	[0.834]	[0.026]	[0.547]
$\beta_{LB\_HE} = \beta_{HB\_HE}$	[0.272]	[0.034]	[0.563]	[0.125]	[0.694]
Observations	168	168	168	168	168

Notes: Standard errors clustered at the subject-level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
† Wald tests of hypotheses on the parameters of the regression model; p-values reported in squared parentheses.

Figure (B1) *The bribery game structure in the HB_HE treatment: An illustrative example*



Notes: This is an illustrative example of the structure of the game in the HE ($\gamma_H = 2$) HB ($\alpha_H = 6$) treatment, with $p_C = 10$ and $p_O = 10$ being the amount chosen by the monitor to punish the citizen and the official, respectively; $\underline{n} = 7$ the total number of bribes exchanged in the other seven citizen-official pairs; 50 the initial endowment; 3 the amount of the bribe.

Appendix C Supplementary Material

Supplementary material—which contains robustness checks for round fixed effects, start-game and end-game effects—can be found in the online version at *[INSERT LINK HERE]*.