Journal of Economic Behavior and Organization 227 (2024) 106693

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Research paper

A corruption dilemma

James Ahloy [a], Rebecca Gilland [b], John R. Hamman [c,*]

[a] Allstate Insurance, Sacramento, CA, USA
[b] Operations, Information, and Decisions Department, The Wharton School, University of Pennsylvania, Philadelphia, PA, USA
[c] xs/fs, Research Cluster and Department of Economics, Florida State University, Tallahassee, FL, USA

ARTICLE INFO

JEL classification:
C91
D62
D73
D83

Keywords:
Corruption
Social dilemma
Beliefs

ABSTRACT

We experimentally investigate the effects of uncertain negative externalities on corruption engagement and social beliefs. We report two experiments in which corruption is modeled as a common-pool resource. In our first experiment, participants face a decision to bribe a public official for a service where accepted bribes impose probabilistic external costs on the briber and other participants. We find that the decision to bribe is positively associated with the belief that others will do the same. We also find that participants overestimate their ability to avoid external costs. Experiment 2 explores endogeneity and ambiguity in types using a contextualized version of the corruption dilemma. Consistent with experiment 1, choosing to offer a bribe is positively associated with the expectation of similar behavior. Curiously, we find little evidence that beliefs about one's own type affect the decision to bribe.

1. Introduction

Corruption is widely recognized as an important social problem. Over the past two decades, awareness of corruption as an impediment to economic growth and development has swelled to the point that cracking down on it is among the primary objectives of organizations such as the World Bank, the African Development Bank, and the IMF. Despite widespread knowledge of corruption's ills and worldwide efforts to reduce its incidence, corruption and bribery continue to affect millions of lives. The World Bank places estimates of the amount paid in bribes by people each year at around US $1 Trillion. Many previous studies have investigated the roots and remedies for corrupt behavior. We contribute to this discussion by addressing situations where corruption generates external costs in expectation rather than with certainty.

One robust feature of corruption is that it is socially harmful. Although early research argued that corruption improved efficiency by "greasing the wheels" of bureaucracy (Leff, 1964; Huntington, 1968), empirical evidence has consistently found that corruption can also have profound adverse effects on social and economic life.[1]

In some instances where corruption is directly linked to negative outcomes, these outcomes are not realized with certainty. Consider a building owner who bribes an inspector to pass a safety inspection. A poorly constructed building whose problems are ignored due to corruption may not collapse and injure its occupants with certainty, but it may do so in the face of a natural disaster that it could have withstood had it been built to code.[2]

* Corresponding author.
E-mail addresses: jahloy@fsu.edu (J. Ahloy), bgilland@wharton.upenn.edu (R. Gilland), jhamman@fsu.edu (J.R. Hamman).

[1] See, for instance, Fisman and Svensson (2007), Méon and Sekkat (2005), and Knack and Keefer (1995) on the negative effects of corruption on economic growth.

[2] For example, corruption may have been partly to blame for the collapse of over 7000 schoolrooms during an earthquake in 2008 in the Chinese province of Sichuan. The disproportionate number of school collapses led to allegations from parents and engineers that the buildings had not been built to civil planning standards due to corrupt practices on the behalf of the Education Ministry and contractors.

https://doi.org/10.1016/j.jebo.2024.106693
Received 12 June 2023; Received in revised form 17 June 2024; Accepted 1 August 2024
Available online 4 September 2024

We experimentally test the extent to which beliefs influence the decision to behave corruptly when social harm occurs in expectation. It is possible that individuals form beliefs that make engaging in corruption more appealing than it otherwise should be. Although previous studies have examined the impact of negative externalities on the decision to bribe, external costs in those studies were imposed with certainty. To the best of our knowledge, this is the first corruption experiment to address situations with probabilistic negative externalities.

In this paper, corruption takes the form of collusive corruption (the exchange of bribes for a service) instead of extortionary corruption (a demand for bribes for a service to which the briber is entitled). We consider two sets of beliefs in the decision to bribe with expected external costs: beliefs about the actions of others and beliefs about the benefits of corruption. For example, an individual may choose to bribe because they believe that others are also bribing. This choice may arise from the known effect of empirical expectations (what individuals expect others to do) on decision-making. Bicchieri and Xiao (2009) found that when empirical expectations and normative expectations (what individuals think others think they should do) are in conflict, empirical expectations predict behavior in dictator games. Along these lines, Köbis et al. (2015) and Fišar et al. (2016) showed that, in laboratory corruption games, perceptions about the prevalence of corruption were positively correlated with corrupt behavior.

Beliefs about the benefits of corruption are salient when a briber may be harmed in an event that also harms others. Consider the act of bribing for a driver's license. Regardless of the manner in which they obtained their licenses, drivers are always at risk of harming themselves and others through vehicular accidents. However, improperly licensed drivers may be more likely to cause accidents than properly licensed ones, and so bribers increase the risk for themselves and others of an accident. Therefore, the perceived benefits of bribery are contingent on the probability with which a driver expects an accident under the decision to bribe. Drivers who believe that the probability of being involved in an accident is low when they bribe will expect a higher net benefit to bribery than those who believe the probability of an accident is high.

In our experimental environment, corruption can be modeled as a common-pool resource (CPR). In the CPR dilemma, a non-rival but scarce good is shared amongst a population. Citizens are individually motivated to over-extract from the resource; however, each individual acting selfishly results in a negative outcome for the group. Once the resource reaches a critical level of extraction, it is unable to replenish at a reasonable rate and the group loses supply. Common examples of common-pool resources are fisheries, water reserves, or forests. In the context of corruption, each individual is motivated to bribe in exchange for a service, as the risk factor is unlikely to be significantly affected by a single bribe. However, when all citizens choose to bribe, the risk factor surpasses a critical level where the group is harmed. For instance, recall the case of driver licensing. When one or a few citizens choose to bribe for a license, the risk of an accident does not increase significantly since most other drivers have proven their competency. Thus, citizens are individually motivated to engage in bribery in exchange for their license at a lower cost. However, as bribing becomes the norm, more drivers will be improperly licensed, and the risk of an accident will increase markedly for the entire population. Our experiments examine this tension between individual and group preferences with probabilistic outcomes.

Experiment 1 builds upon a game introduced by Barr and Serra (2009), which was designed to simulate petty corruption exchanges that impose costs on passive third parties with certainty. We employ a similar framework but modify their game to capture some key features of the class of exchanges we are interested in. Their design intentionally eliminated interaction between potential bribers as a way to prevent externalities from being reciprocal. However, in some exchanges with probabilistic externalities, it is possible for bribers to impose externalities on other bribers. Building owners who have paid off inspectors may find themselves in buildings owned by others who have done the same. Similarly, drivers who have bribed for their licenses may be on the road with other illegally certified drivers. As such, citizens in our game, who are always at risk of experiencing losses, can choose to pay an official for a service and then interact with other citizens. Externalities caused by the decision to bribe take the form of an increased risk of experiencing losses.

While subjects are not provided with any specific context for the game, instructions are presented using a corruption frame — that is, subjects are told they will play a game in which they can bribe for a service. We find that the decision to offer a bribe is strongly associated with the belief that others will choose to bribe. This finding is consistent with the literature on empirical expectations and behavior. We also find that subjects underweight their beliefs about the likelihood of incurring losses, and as such, the decision to bribe is weakly associated with these beliefs.

We then develop a modification of our corruption dilemma game to account for heterogeneity in populations. Consider the driver's license example mentioned above. It is entirely possible that a citizen possesses all of the skills and characteristics necessary for safe driving and know the local traffic laws from a friend or family member. Such a person might pay a bribe nonetheless to expedite the process, whereupon their corrupt action has not necessarily changed the possible risk they pose to themselves or others. However, an uneducated or inattentive individual who bribes to obtain a license may cost themselves an opportunity to learn the rules of the road or the importance of defensive driving. This individual, by bribing for a license, is causing an elevated risk to drivers when they get behind the wheel.

Experiment 2 explores this exact (contextualized) scenario of bribing to obtain a driving license. Experimental economists have traditionally shied away from using context when describing to subjects the environment under study, instead favoring abstract, neutrally worded descriptions. However, as Alekseev et al. (2017) discuss, the use of abstract structure could eliminate valuable social considerations that are innate to many naturally-occurring situations. This may be especially true with corruption, where such exchanges can have significant social consequences. The game implemented in Experiment 1 featured instructions with a corruption frame, making it more applicable to an entire class of corrupt exchanges. However, the potential exists for a deeper understanding of these exchanges through a stronger, more specific context. Additionally, the driving licensing context allows us to cleanly induce heterogeneity among potential bribers to capture the possibility that they differ in the amount of social harm they generate through corruption.

In Experiment 2, we again find that subjects are more willing to bribe when they believe that others in their group are going to offer bribes. In contrast with Experiment 1, beliefs about the probability of incurring losses from accidents do not predict bribing behavior. Participants hold beliefs about their types that are inconsistent with rational beliefs. However, these beliefs are not associated with the decision to offer a bribe when controlling for beliefs about the behavior of others and beliefs about the likelihood of incurring losses from accidents. When types are endogenously determined we find that participants hold diminished views of their ability to avoid accidents, providing an example of underconfidence.

2. Related literature

Our work contributes to a growing literature on experimental studies of corruption. These experiments have deepened our understanding of the effect on corruption of top-down and bottom-up monitoring and punishment (Abbink et al., 2002; Serra, 2011; Armantier and Boly, 2013; Banuri and Eckel, 2015; Salmon and Serra, 2017; Ryvkin et al., 2017; Butler et al., 2017), asymmetric reporting and leniency systems (Abbink et al., 2014; Engel et al., 2016), social norms and culture (Cameron et al., 2009; Barr and Serra, 2010; Abbink et al., 2016; Salmon and Serra, 2017), gender (Alatas et al., 2009; Rivas, 2013), and competition among public officials (Ryvkin and Serra, 2017). In particular, this experiment lies at the intersection of corruption experiments that examine the effect of negative externalities and those that elicit the beliefs of participants.

2.1. Negative externalities in corruption experiments

Abbink et al. (2002) (AIR, hereafter) present one of the earliest experiments on negative externalities generated by corruption. They model corruption as a repeated, sequential, two-player game in which the first player can offer a bribe to the second without knowing whether the second will grant a higher payoff in return. When a negative externality is imposed on all other participants as a result of a corrupt exchange, the authors find that it has no significant effect on behavior. However, punishment in the form of low-probability exclusion from the game, if bribery occurs, reduces exchanges between players. Framing effects do not change behavior in the AIR game (Abbink and Hennig-Schmidt, 2006), nor does the size of the externality (Cameron et al., 2009).

Abbink et al. (2020) study corruption in a social dilemma environment where Private Citizens can choose to obey or break the law, knowing that breaking the law imposes a negative externality on other citizens. Police officers in the environment can demand bribes from citizens or fine them for rule violations. The authors find that rule-breaking is less frequent when law enforcement is present, even when Officers accept bribes.

Barr and Serra (2009) introduce the game that serves as the foundation for that used in the experiments presented in this paper. Their experiment simulates a one-shot, petty corruption exchange between a Public Official and a Private Citizen and is used to investigate the effects of negative externalities and framing on the decision to bribe. They find that the combination of a higher externality and a loaded frame reduces the number of Private Citizens who offer bribes, but that the application of only one of the treatments results in Private Citizens increasing their bribe offers rather than abstaining.

2.2. Corruption and beliefs

There are four prominent corruption experiments that directly elicit the beliefs of participants. In a post-experiment questionnaire, AIR elicited beliefs about the probability of exclusion in their disqualification treatment using an incentivized procedure. They found that subjects tended to underestimate the probability of being excluded when asked to consider the likelihood of exclusion over multiple periods of interaction. Berninghaus et al. (2013) model corruption as a coordination game, where the larger the number of players who accept bribes, the lower the probability of detection and punishment. No external cost is present when bribery occurs, and the game is described to subjects in neutral terms. The authors use a quadratic scoring rule to indirectly elicit subjects' beliefs about others' behavior by asking for their beliefs about the detection probability.

Rivas (2013) uses a close variant of the AIR game to investigate gender differences in corrupt behavior. Subjects who offer bribes are asked for their beliefs about the likelihood of the bribe being accepted, but this elicitation is not incentivized. Fišar et al. (2016) also explore gender differences in behavior but with an experiment based on Cameron et al. (2009). In addition to asking subjects for their beliefs about the likelihood of a bribe being accepted, the authors ask subjects for their beliefs about the willingness of others to bribe or sanction bribery. All elicitation in their experiment is incentivized using a quadratic scoring rule.

The experiments reported here bear relation to the strand of the experimental corruption literature that is concerned with negative externalities (Abbink et al., 2002; Cameron et al., 2009; Barr and Serra, 2009) and the strand that is interested in the role of beliefs in the decision to bribe (Berninghaus et al., 2013; Rivas, 2013; Fišar et al., 2016). Our study is also directly related to corruption experiments that assess the effect of framing on the decision to engage in corruption.

2.3. Framing in corruption experiments

Several studies have sought to compare the outcomes of framed and neutral versions of corruption games. Abbink and Hennig-Schmidt (2006) provide one of the earliest of these studies with their investigation of the effect of framing on the Abbink et al. (2002) game. The authors vary the presentation of the game to subjects, using either a neutral frame or one that portrays the game as a corrupt transaction between a firm and a public official. Despite the negative connotations of the loaded frame, they fail to find any significant difference in behavior between the two treatments.

Barr and Serra (2009) considered the role of framing in a petty corruption game that simulated exchanges between private citizens and public officials. These exchanges generated external costs that were imposed on passive third parties. In addition to varying the magnitude of the externality experienced by the third parties, the authors manipulated the frame used to describe the exchange to subjects. In one treatment, the game was presented to subjects in neutral terms, while in another, the game was presented with loaded instructions. The loaded instructions specifically described the roles as "private citizen" and "public official", and the transfer between the two was referred to as a bribe. The authors find that the combination of a higher externality and a loaded frame reduces the number of Private Citizens who offer bribes but that the application of only one of the treatments results in Private Citizens increasing their bribe offers rather than abstaining.

Banerjee (2016) examines the effects of framing and changes in reference point in a harassment bribery game. The author compares a harassment bribery game – where Private Citizens perform a real effort task to earn an award, but can then be denied the award by a Public Official unless a bribe is paid – with loaded language to the same game with neutral language, and to a strategically similar ultimatum game. He finds that retained shares of the award were higher in the ultimatum game than in the other games, and shares were statistically indistinguishable between the loaded and neutral bribery game. The social appropriateness of the possible actions was elicited in a separate experiment, and the results indicated that actions in the bribery games were viewed as more socially inappropriate. Overall, his results suggest that a sense of entitlement to earnings by the individual being asked to bribe is necessary to impose a moral frame, and that loaded language may be neither necessary nor sufficient to create an environment of immorality. However, these results are established for an extortionary bribery game. It is easy to understand how a sense of entitlement may be important when individuals are being coerced to pay a bribe, but it is unclear how entitlement comes into play for framing a corrupt act that is collusive, as is the case with our environment.

3. Experiment 1

Experiment 1 uses mild framing to study how externality and information quality impact both the beliefs and decisions of individuals.

3.1. Corruption dilemma game

The Corruption Dilemma game involves groups of four players comprised of one Public Official and three Private Citizens. Roles are randomly assigned and fixed throughout the experiment. The stage game involved two phases: a belief elicitation phase, where subjects' beliefs about several aspects of the choice environment were elicited, followed by a decision phase, where subjects actually made decisions. Though belief elicitation was performed before subjects made decisions, the decision phase will be described first.[3]

3.1.1. Decision phase

In the decision phase, all players are endowed with 50 ECU (experimental currency units), with an exchange rate of 6 ECU to $1. Each Private Citizen must obtain a service conferred by the Public Official. Private Citizens could obtain the service properly or offer a bribe, $b \in \{1, 2, 3, \ldots, 15\}$, from their initial endowment to the matched Public Official. Bribing for the service gave Private Citizens a benefit valued at 12 ECU (obtaining without a bribe involved no further costs or benefits). While Private Citizens made their decisions, Public Officials stated their willingness to accept each of the possible bribe offers that a Private Citizen could make. Once all decisions had been made, any bribe offers from Private Citizens were automatically accepted or rejected according to the matched Public Official's decisions.

All players in the group faced a per-period risk $r_0 = 0.02$ of incurring a loss of $l = 16$ ECU. Unless noted otherwise, any time a player incurs a loss, another player from their group is chosen at random to share in the loss. Specifically, any loss results in two group members incurring the same loss of 16 ECU. For any successful bribe, the bribing Private Citizen faces an increased risk of $r_1 = 0.12$ of experiencing a loss. Thus, any Private Citizen who successfully bribes increases the risk faced by all others in their group.[4]

[3] In eliciting beliefs, a choice has to be made whether to elicit beliefs before or after subjects make their decisions. Research on the timing of belief elicitation is mixed (Schlag et al., 2015) but Schotter and Trevino (2014) point to belief elicitation increasing the rate of convergence of best-response behavior, so we chose to elicit beliefs before decisions were made. Additionally, Fišar et al. (2016) failed to find any significant differences between eliciting beliefs before or after decision-making in their corruption experiment.

[4] Though a subject could be randomly selected multiple times in a period to experience a loss, to limit the possibility of negative or highly unequal earnings, total losses per period were capped at 16 ECU and subjects were informed of this.

3.1.2. Belief elicitation phase

Subjects' beliefs were elicited in accordance with a Quadratic Scoring Rule.[5] Private Citizens and Public Officials were asked a similar set of questions that were modified slightly depending on their role. Private Citizens were first asked how many of the other members of their group of four they expected to offer a bribe and how many of those offers they expected to be accepted, while Public Officials were asked how many offers they expected to receive and accept. Subjects' earnings for these beliefs were determined according to the following equation:

$$\pi(r, x) = 6 - \frac{1}{2}(r - x)^2 \tag{1}$$

where *Known Type/Random* is the subject's reported belief and x is the actual outcome. Private Citizens were then asked to state what they believed was the likelihood of experiencing at least one loss during the block, both if they were to obtain the corrupt service and not obtain it, while Public Officials were asked what they believed was the likelihood of experiencing at least one loss during the block based on the number of offers they expected to receive. Earnings for these beliefs were determined by the following equation[6]:

$$\pi(r, we) = 6 - 6[we(1 - r)^2 + (1 - we)r^2] \tag{2}$$

where *Known Type/Random* is the reported belief and we is an indicator variable indicating that a loss was incurred during the block.

Both equations were presented to subjects and explained with examples. Subjects were also told that payoffs under the first equation were increasing in the accuracy of the report, while for the second equation, expected earnings were highest when actual beliefs were reported. At the end of the experiment, subjects were paid their earnings for a randomly chosen belief from the same block that was selected for payment for the decision phase.

3.2. Session timeline and procedures

Every session began with the elicitation of subjects' risk preferences using the Eckel and Grossman (2008) method. Earnings for this task were not shown to subjects until the end of the experiment. Following the elicitation of risk preferences, subjects were randomly divided into matching groups of eight and randomly assigned fixed roles such that each matching group had 2 subjects in the role of Public Official and 6 in the role of Private Citizen.[7] These matching groups remained the same throughout the session, thus enabling each to serve as a single independent observation.

The main portion of the session consisted of 25 periods that were divided into five blocks of 5 periods each. At the beginning of each block, two groups of four were formed within each matching group, with one Public Official and three Private Citizens to a group. These groups were fixed during a block and were randomly re-formed in the following block.

In period 1 of each block, participants played the stage game described above, first completing the belief elicitation phase followed by the decision phase. Once all decisions were made, participants saw the full results of the round based on their role.[8] In periods 2 through 5 of each block, participants saw a new results screen showing the same feedback as they saw after period 1.

3.3. Treatments

Three treatments were implemented using a between-subjects design to vary the presence of externalities and the information available to participants. By "externality", we refer to a participant's loss causing another participant to also suffer the loss. The information we vary is whether participants know the specific probabilities of loss.

In the baseline treatment, *Externality/Full Info*, the game is played exactly as described above. Any successful bribe increased the probability that the briber experiences a loss, and thus impacts the risk faced by all other members in their group. All participants know the exact probabilities they face if they successfully bribe or not.[9] Outcomes for each period were again independently determined and did not depend on the outcomes from previous periods. This interaction between Private Citizens and the other members of their group introduces an externality, since the decision to bribe lowers the expected earnings of the other group members, conditional on their decisions.

---

[5] Quadratic Scoring Rules (QSRs) are known to be incentive compatible for risk-neutral subjects. The Binarized Scoring Rule (Hossain and Okui, 2013), which is robust to other risk preferences was considered, but a QSR was chosen for simplicity. Methods for correcting QSRs for risk preferences exist (see Schlag et al. (2015) and Schotter and Trevino (2014) for detailed discussions) but Fišar et al. (2016) failed to find any significant differences between corrected and uncorrected beliefs. Again, for simplicity, we chose to use uncorrected beliefs and control for risk preferences using the Eckel and Grossman (2008) procedure.

[6] Subjects were not presented with the equation containing the indicator variable. Instead, they were show two simpler equations that were equivalent to the equation containing the indicator variable.

[7] See the Appendix for sample instructions.

[8] Public Officials saw the number and amounts of bribe offers, their acceptance threshold, the number of accepted offers, the number of group members involved in accidents, and earnings for the period. Private Citizens saw their bribe offer and whether it was accepted (if any), the total number of offers made and accepted in their group, whether they were involved in an accident, the number of accidents in their group, and their earnings for the period. In all periods after the first, all players also saw their cumulative earnings for the block.

[9] These probabilities are strictly the probabilities of causing a simultaneous loss. The introduction of interaction between group members in this treatment turns experiencing a loss in each period into a compound event. Thus, the per-period probabilities of suffering a loss are higher than the corresponding probabilities of causing a simultaneous loss.

Table 1
Per-period probabilities of a loss and expected group benefit.

Number of accepted offers	Role	Per-period prob. of loss	Expected net group benefit
Zero	Private Citizen	0.03	190.49
	Public Official	0.02	
One	Private Citizen (briber)	0.13	
	Private Citizen (non-briber)	0.07	186.77
	Public Official	0.05	
Two	Private Citizen (briber)	0.16	
	Private Citizen (non-briber)	0.10	183.76
	Public Official	0.08	
Three	Private Citizen	0.19	181.42
	Public Official	0.1	

In *No Externality/Full Info*, no externalities were present, and subjects were fully informed about their probabilities of experiencing losses whether they were to successfully bribe or not. In this treatment, one participant successfully offering a bribe did not increase the risk faced by anyone else in the group. We can compare behavior in this simplified setting to our baseline treatment to see how environmental complexity affects behavior.

The third and final treatment, *Externality/Partial Info*, identical to *Externality/Full Info* except that subjects were not informed of the probabilities of causing a simultaneous loss. Rather, they were told that the probability was drawn from the range [5, 20] for a Private Citizen who made a successful bribe offer and [0, 5] for all other Private Citizens, with values not necessarily equally likely to be drawn. This treatment was motivated by the possibility that individuals in the field may have some sense of the probabilities of causing losses under different bribery decisions but do not know the exact likelihood.

In all treatments, the parameters chosen ensure that Private Citizens have an expected financial incentive of approximately 4 ECU to offer a bribe. For example, in *Externality/Full Info*, if no Private Citizens in a group choose to bribe in a particular block, the per-period probability that each incurs a loss is about 3.3%. Thus, the expected loss for a Private Citizen during the block is 2.4 ECU. If, instead, one Private Citizen had chosen to bribe, that Private Citizen would have faced a per-period probability of incurring a loss of about 13%. Thus, the expected loss would have been 10.4 ECU, but the Private Citizen would have gained 12 ECU as the value of the corrupt service, for an expected net gain of 1.6 ECU. The movement from an expected loss of 2.4 ECU to an expected net gain of 1.6 ECU results in an overall improvement of 4 ECU. Similar calculations can be applied to *No Externality/Full Info* and *Externality/Partial Info*. While Private Citizens have an individual incentive to make a successful bribe offer, group welfare is lowest when all Private Citizens bribe and highest when all Private Citizens abstain from bribing, capturing the social dilemma aspect of our environment. Per-period probabilities of experiencing a loss and corresponding group benefits are summarized in Table 1.[10]

A Public Official's earnings for the decision phase in each block consisted of the initial endowment and any accepted bribe offers less any losses incurred during the block. For a Private Citizen who successfully bribed, earnings for the decision phase in each block were the initial endowment and the value of the corrupt service less the amount of the accepted bribe offer and any losses incurred during the block. Earnings for all other Private Citizens were simply the initial endowment less any losses incurred during the block. At the end of a session, subjects were paid their earnings for the decision phase for one randomly selected block.

3.3.1. Procedures

Subjects were recruited using the online system ORSEE (Greiner, 2015) from a population of undergraduate students who had pre-registered as potential experimental participants. Either 16 or 24 subjects participated in each session, with every lasting around 60 min. Each subject was paid a $7 show-up fee plus any additional earnings from the experiment. On average, subjects earned $17.51 per session. The experiment was implemented using z-Tree (Fischbacher, 2007) and conducted at the xs/fs lab at Florida State University.

3.4. Predictions

Predictions for the experiment rest primarily on the psychological cost framework used in Barr and Serra (2009) and Drugov et al. (2014). In this framework, individuals experience a psychological cost when engaging in activities that are harmful to others or are perceived as immoral, with the cost increasing in the harm done to others and the degree to which the act is seen as immoral.

[10] For an example derivation of the per-period loss probabilities, consider the simplest case of no bribe offers in a block, and index the Private Citizens in a decision group by 1 through 3. Let $\overline{A_i}$ be the event that Private Citizen i does not experience a loss in a given period, $\overline{B_{ij}}$ be the event that Private Citizen i does not experience a loss caused by Private Citizen j in the same period, and B_{ii} be the event that Private Citizen i causes a simultaneous loss. Then, for Private Citizen 1, $\overline{A_1} = \bigcap_{j=1}^{3} \overline{B_{1j}}$, and $\Pr(\overline{A_1}) = \Pr(\bigcap_{j=1}^{3} \overline{B_{1j}})$. $\Pr(\overline{B_{11}}) = 0.98$ and $\Pr(\overline{B_{1j}}) = \Pr(\overline{B_{jj}}) + (\Pr(B_{jj}) * (2/3)) = 0.98 + (0.02 * (2/3)) \approx 0.9933$, for $j \neq 1$. Since whether a Private Citizen causes a loss is independent of whether another Private Citizen causes a loss, $\Pr(\overline{A_1}) = \Pr(\overline{B_{11}}) * \Pr(\overline{B_{12}}) * \Pr(\overline{B_{13}}) = 0.98 * (0.9933)^2 \approx .9669$. Thus, $\Pr(A_1)$, the event that Private Citizen 1 experiences a loss in a given period, is $1 - .9669 \approx 0.0331$.

Table 2
Experimental treatments.

Treatments	Sessions	Indep. Obs	Subjects
Externality/Full info	3	7	56
No Externality/Full info	2	6	48
Externality/Partial info	2	6	48
Total	7	18	152

As such, individuals may refrain from engaging in an activity even when it is in their financial interest to do so, due to the opposing psychological cost.

In *No Externality/Full Info*, the absence of externalities prevents a Private Citizen's decision to engage in bribery from harming others, but the language used in describing the experiment has negative moral connotations and so may carry some psychological cost. In *Externality/Full Info*, the combination of the social harm caused by the externality and the language used should result in a higher psychological cost than in *No Externality/Full Info* and consequently fewer bribe offers.

Prediction 1. There will be a higher proportion of bribe offers in *No Externality/Full Info* than *Externality/Full Info*. Though Private Citizens have a known expected financial incentive of 4 ECU to offer a bribe, assuming accurate beliefs, less than 100% will choose to offer a bribe in both treatments.

In *Externality/Partial Info*, subjects may form beliefs about the per-period probability of causing a simultaneous loss, within limits. Subjects can form these beliefs so as to reduce the psychological costs associated with choosing to bribe, either by believing that the probability of causing a simultaneous loss is low should they choose to bribe, or that the probabilities associated with each decision are close in magnitude and there is little additional harm done by choosing to bribe. Consequently, more bribe offers would be seen than in *Externality/Full Info*. Given that social efficiency is decreasing in the number of accepted bribe offers, the higher proportion of offers in *Externality/Partial Info* could potentially lead to lower efficiency levels.

Prediction 2. The proportion of bribe offers in *Externality/Partial Info* will be higher than in *Externality/Full Info*.

Decades of behavioral research have shown that individuals consistently encounter difficulty with probabilistic reasoning, with the exception of certain scenarios in which probabilities are expressed as frequencies.[11] Given the relative difficulty of the probability calculations for this experiment's environment and the manner in which probabilities are presented to subjects, Private Citizens should fail to form probability beliefs that are accurate.

Prediction 3. Private Citizens generally form inaccurate beliefs regarding the likelihood of incurring losses, conditional on the expected number of accepted bribe offers in their group, in both *No Externality/Full Info* and *Externality/Full Info*.

The findings of Bicchieri and Xiao (2009), Köbis et al. (2015), and Fišar et al. (2016) suggest that Private Citizens' empirical expectations should influence their decision-making in a manner that makes them more likely to behave as they expect others to behave. Private Citizens who believe that the decision to bribe is commonplace should thus be more willing to offer a bribe themselves.

Prediction 4. The willingness of Private Citizens to offer a bribe in all treatments will be positively correlated with the number of other Private Citizens in their group they expect to offer a bribe.

3.5. Results

We collected seven sessions comprising a total of 152 subjects, with each treatment involving at least 48 subjects. Table 2 reports our complete sample size by treatment.

We begin by examining bribing behavior by treatment. Afterward, we examine how beliefs play a role in driving behavior in the game. The following results concentrate on the beliefs and behavior of Private Citizens.

3.5.1. Bribing behavior

Fig. 1 displays the proportion of bribe offers over all five blocks. Consistent with Prediction 1, the proportion of Private Citizens who offered a bribe was always lower in *Externality/Full Info* than in *No Externality/Full Info*. Participants were able to appreciate the added risk associated with the externality in determining their bribe behavior on the extensive margin.

The greatest share of bribe offers always occurs in *Externality/Partial Info*, consistent with Prediction 2. Uncertainty over the probability of loss in *Externality/Partial Info* allows subjects to form beliefs which reduce the psychological costs associated with choosing to bribe. However, *No Externality/Full Info* does not impose social costs to bribing. This should result in higher instances of bribing than in treatments which impose a social cost. The higher instance of bribery in *Externality/Partial Info* relative to *Externality/Full Info* suggests that subjects may overlook the costs associated with social harm when they are able to form beliefs that significantly lower the perceived probability of such losses, allowing them to bribe at a higher rate. In other words, subjects

[11] See Kahneman (2003) and DellaVigna (2007) for overviews, and Liberali et al. (2012) for a discussion of how individual differences in numeracy affect probability judgments.

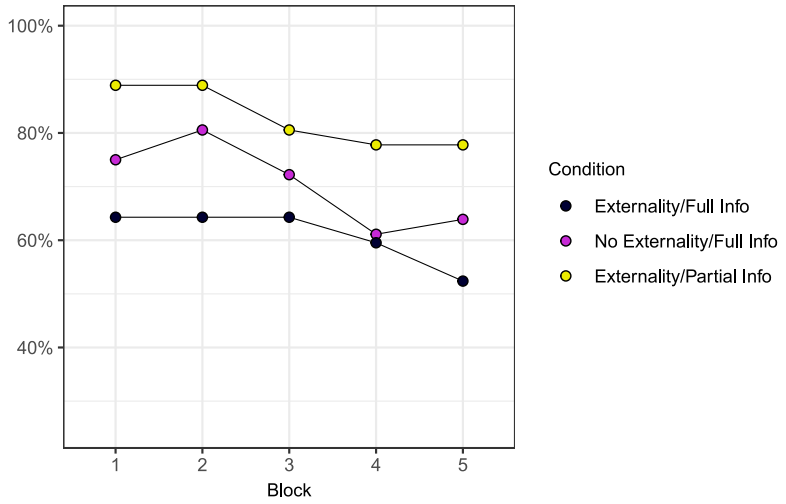

Fig. 1. Bribe offers by treatment. *Notes.* Proportion of bribes made by Private Citizens over all five blocks.

Table 3
Treatment differences, Experiment 1.

Difference	Bribe offered = 1
No Externality/Full info –Externality/Full info	11.73
	(8.11)
Externality/Partial info – Externality/Full info	26.54***
	(7.24)
Subjects	114
Blocks	5
N	570

Notes. Average partial effects of treatment expressed in percentage points. Effects estimated using random effects logistic regression with varying intercepts for subject and block to account for within-subject and within-block correlation. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

may be estimating the probability of loss to be low enough that any social costs are inconsequential because they perceive loss itself to be very unlikely.

Table 3 displays the average partial effects of treatment on the probability of offering a bribe. The directions of the effects are in line with Predictions 1 and 2, though the difference between *Externality/Full Info* and *No Externality/Full Info* falls just short of marginal significance.

Notably, while the relationships between treatments are preserved throughout all blocks, the proportion of bribe offers in all treatments decreases slightly over time. As most subjects decided to bribe in early blocks, the likelihood of experiencing a loss would be high and would have resulted in more simultaneous losses. As discussed in the next section on beliefs, subjects generally underestimated the probability of loss, though they more accurately anticipate the number of other participants who offer a bribe, which more likely drives this decline in bribe behavior over time.

3.5.2. Beliefs

We now examine subjects' beliefs towards the bribing behavior of their group members. Fig. 2 reports Private Citizens' beliefs across all five blocks regarding the number of others in their group they expect to make a bribe offer and the number of offers they expect to be accepted.

In general, Private Citizens accurately anticipated differences in bribing behavior across treatments. Anticipated offers in *Externality/Partial Info* were higher than in *Externality/Full Info*. Similarly, more offers were anticipated in *No Externality/Full Info* than in *Externality/Full Info*.

Table 4 reports the relationships between beliefs and decisions to bribe. We see that the decision to bribe is strongly associated with the belief that others in the group will also bribe, supporting Prediction 4. Unsurprisingly, Private Citizens are less likely to bribe when they believe the probability of loss to be higher if they do so.

For Private Citizens in *No Externality/Full Info*, the decision to bribe does not generate social harm, so subjects can reasonably expect one another to bribe due to limited psychological costs. The decision to bribe in both *Externality/Full Info* and *Externality/Partial Info* causes social harm and should carry higher psychological costs than in *No Externality/Full Info*. As discussed above, the ambiguity in the probabilities of causing losses in *Externality/Partial Info* permits subjects to form beliefs that reduce their psychological costs and make it easier to bribe than in *Externality/Full Info*. If subjects are aware of this, they could expect the members of their groups

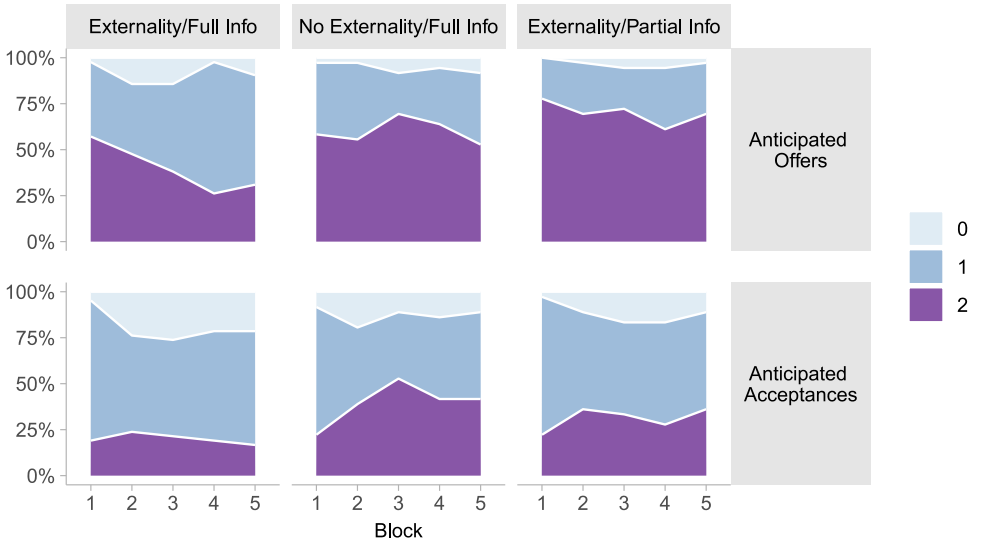

Fig. 2. Beliefs: Offers and acceptances. *Notes.* Proportion of private citizens predicting both how many others in their group will bribe (top row), and how many of these bribes will be accepted (bottom row).

Table 4
Regression analysis, Experiment 1.

	Bribe offered = 1	
Exp. offers	12.56***	12.58***
	(2.74)	(2.74)
Exp. acceptances	3.78	3.15
	(3.49)	(3.53)
Loss belief difference		−0.17*
		(0.1)
Subjects	114	114
Blocks	5	5
N	570	570

Notes. Within-subject average partial effects of beliefs, expressed in percentage points. Effects estimated using random effects logistic regression with a varying intercept for subject and mean-centered covariates, as outlined in Bell and Jones (2015). Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

to be more willing to offer bribes in *Externality/Partial Info* than if they actually knew the probabilities of causing losses, so the number of bribes in this treatment would be larger than in *Externality/Full Info*.

We now examine subjects' beliefs towards the probabilities of incurring losses during a block. Fig. 3 plots subjects' beliefs of incurring at least one loss in block 1, conditional on having a bribe accepted or not. A substantial majority of responses fall below the 45 degree line, showing that most subjects correctly understand the added risk that comes with bribing. That said, around 30% expressed the errant belief that their probability of at least one loss during the first block was greater when no bribe was exchanged than when a bribe was exchanged. A plurality of these errors come from the *No Externality/Full Info* treatment.

In *No Externality/Full Info*, Private Citizens had approximately a 47% chance of experiencing at least one loss during a block if they had successfully bribed; however, they report beliefs that are up to 17 percentage points lower on average. This bears some resemblance to AIR, where subjects consistently underestimated the probability of exclusion in their disqualification treatment.

Similarly, Private Citizens had around a 10% chance of experiencing at least one loss during a block if no bribe was exchanged, but on average they reported beliefs that are up to 16 pp higher. Generally, subjects' estimates of the probabilities of experiencing at least one loss during a block were inaccurate. For instance, in *No Externality/Full Info*, 83% of the stated beliefs for the probability of experiencing at least one loss after a successful bribe were not within 5 percentage points of the true probability. The same is true in *Externality/Full Info* – 98% of reports for the probability of experiencing at least one loss after a successful bribe were not within 5 percentage points of the true probability, given the stated expected number of accepted bribe offers. Subjects' probability beliefs are biased in directions that increase the perceived benefit of making a successful bribe, which can potentially undermine Prediction 1 by making subjects more willing to bribe. These inaccurate beliefs are, however, consistent with Prediction 3.

Relative to *No Externality/Full Info*, subject interactions in the two Externality treatments result in a higher probability of experiencing at least one loss in a block for any given bribery decision. If one assumes that Private Citizens condition their probability beliefs on the expected number of accepted offers, the absence of significant differences in the expected number of accepted offers may imply that subjects perceive the probabilities of causing a loss to be about the same across treatments.

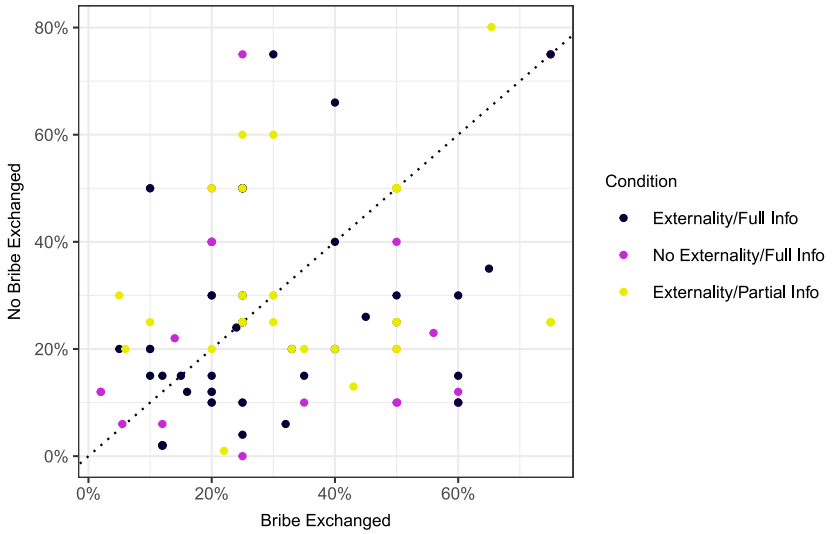

Fig. 3. Beliefs: Probability of loss. *Notes.* Elicited beliefs of incurring at least one loss, conditional on whether a bribe is exchanged.

The high incidence of inaccuracies in Private Citizens' beliefs about the probability of experiencing at least one loss during the first block points to difficulties among subjects in computing probabilities. However, the possibility exists that subjects are misreporting their beliefs, for an unknown reason, and are making payoff-maximizing decisions consistent with accurate beliefs when they choose to bribe. The median bribe amounts offered by subjects suggest that this may not be the case. As noted in the description of the decision phase, Private Citizens always have a financial incentive, in an expected sense, of approximately 4 ECU to engage in bribery, and so should not make an offer that exceeds 4 ECU. But the median bribe amount offered by subjects exceeds 4 in a majority of blocks in every treatment.

Regardless, subjects in Experiment 1 generally behave as expected. Introducing an externality appears to result in subjects expecting fewer bribe offers, estimating higher probabilities of incurring at least one loss irrespective of the bribe decision, and translating those beliefs somehow into a reduced probability of offering a bribe.

Likewise, limiting the information provided to subjects (which allows them to believe the probability of causing a loss might be lower than it is) results in an expectation of more bribe offers and even higher estimates of the probability of incurring at least one loss. These differences in beliefs translate into a higher probability of offering a bribe, possibly in response to a perceived norm.

4. Experiment 2

We know from prior work that bribing for driver certification can occur with some frequency and that many who choose to bribe are unable to operate a vehicle at a basic level (Bertrand et al., 2007). Given the potentially large losses that can accompany such a decision, this raises the question of what motivates people to bribe in spite of their incompetence. The driving environment allows a variety of beliefs to factor into the decision to bribe. If someone with an opportunity to bribe for a license believes that others in a similar situation would choose to bribe, that belief could tip the decision in favor of bribing. Beliefs about the private benefits to bribery can also be important. Consider potential bribers who differ in their inherent ability to handle a vehicle. Those who believe they have a natural aptitude for driving, whether or not this belief is aligned with reality, could expect their likelihood of being involved in an accident to be the same with or without legitimate certification. As such, those who hold this belief may view corruption as an opportunity to improve their welfare, since they are able to avoid the costs associated with legal certification without increasing their chances of suffering losses through accidents.

The driving environment is complex in that any single driver's likelihood of being involved in an accident is in part influenced by the quality and actions of myriad other drivers. A driver who is attempting to determine the likelihood of an accident according to their bribery decision must thus form beliefs over the bribery decisions of other drivers. Inaccuracies in these beliefs may lead to an incorrect determination of the likelihood of an accident and, consequently, an incorrect assessment of the benefits of bribing. Cognitive limitations in the ability to calculate probabilities may also come into play so that drivers who form accurate beliefs about the bribery decisions of others can still fail to have accurate beliefs about the likelihood of an accident and the benefits of bribing.

We explore these issues in Experiment 2, which is procedurally similar to Experiment 1, with some modifications for context and heterogeneity among bribers.[12] Subjects play in one of two roles: Licensing Official or Driver. Licensing Officials are responsible

[12] Per Alekseev et al. (2017), the use of a more natural context in presenting experimental instructions to subjects could prevent the elimination of important social considerations that occur in the field. To this end, framing the game as an exercise in bribing for a driver's license might allow for a deeper understanding of behavior in this environment. Since drivers in the field likely differ in terms of their natural ability to handle a vehicle, their potential for causing harm probably varies as well.

for administering driving exams to Drivers, who have the opportunity to bribe in order to skip the exam. Drivers come in two types, intended to reflect drivers who differ in their natural abilities to drive. One type, which represents a driver of inherently high quality, can increase social efficiency by successfully bribing, while the other type, which represents a driver of inherently low quality, reduces efficiency by successfully bribing. Regardless of their types and bribery decisions, Drivers interact with one another over several periods and can be involved in accidents, which hurt all parties involved. Prior to any decisions being made, we elicit subjects' beliefs about aspects of the driving environment.

Three treatments were conducted that vary both the method for inducing differences among potential bribers (Random assignment or based on performance on a driving quiz) and the information available to subjects about those differences (types are either Known or Unknown). In the baseline, *Known Type/Random*, two types of potential bribers exist. These types are randomly assigned and each subject is fully informed of their type. In the second treatment, *Unknown Type/Random*, there are still two randomly assigned types, but subjects do not have full information about their types. In the final treatment, *Unknown Type/Quiz*, the two types are assigned on the basis of performance on a quiz and subjects are not informed of their types.

4.1. Design

We preserve many aspects of Experiment 1, including recruitment, risk elicitation, and eight-subject matching groups. Roles are again fixed, now referred to as Driver and Licensing Official. Drivers were further assigned one of two types, Type 1 or Type 2, in a ratio of 1:2. Therefore, each group of eight had two Type 1 Drivers and four Type 2 Drivers. Type 1 Drivers represented "high quality" drivers, whereas Type 2 Drivers represented "low quality" drivers

Subjects again interacted with one another for a total of 25 periods, which were partitioned into five blocks of 5 periods each. At the beginning of a block, two groups of four were formed within each group of eight, with one Licensing Official and three Drivers to a group. These groups were unchanged during a block, but were randomly re-formed in the following block.

The first period of every block had the same two phases as demonstrated in Experiment 1: a belief elicitation phase followed by a decision phase.

4.1.1. Decision phase

At the start of the decision phase, Licensing Officials were endowed with 40 ECU and Drivers were endowed with 50 ECU. Subjects were told that Licensing Officials were in charge of granting driver's licenses to Drivers, who could obtain a license in one of two ways: either by taking a driving exam or by bribing a Licensing Official to avoid taking the exam. Drivers were then allowed to offer a bribe, $b \in \{1, 2, 3, \ldots, 15\}$, from their initial endowment to the matched Licensing Official in exchange for a license. Bypassing the exam was valued at 12 ECU for a Driver.

While Drivers made their decisions, Licensing Officials provided their willingness to accept each of the possible bribe offers that could be made. After subjects had made their decisions, any bribe offers from Drivers were automatically accepted or rejected according to the matched Licensing Official's decisions and subjects were informed of their respective outcomes. This exchange between Licensing Officials and Drivers occurred in all three treatments.

Also common to all of the treatments was a negative externality associated with a Type 2 Driver's decision to bribe. All Drivers had a 2% chance of causing an accident, which resulted in a loss of 16 ECU to both the Driver and one other randomly selected Driver from the group of four. Unlike in Experiment 1, Licensing Officials cannot experience a loss.[13] If a Type 2 Driver successfully bribed for a driver's license, the probability of causing an accident increased to 12%. For a Type 1 Driver who successfully bribed for a license, however, the probability of causing an accident remained at 2%. In this way, Type 1 Drivers behave as though they are more naturally inclined to properly handle a vehicle than Type 2 Drivers.

In the *Known Type/Random* treatment, types were randomly assigned and Drivers were made aware of their types before they made decisions. This treatment is designed to see how Drivers use the knowledge of their types to inform their bribery decision.

In the *Unknown Type/Random* treatment, types were again randomly assigned, but Drivers were not informed of their type. Instead, they received a signal about their type, whose probability of being observed varied with type. Subjects were told that they would see a signal at the start of each block in the form of either the letter A or B. The letter A had a 70% chance of being revealed to a Type 1 Driver while B had a 30% chance, with the probabilities reversed for a Type 2 Driver. Drivers were informed of their type once the experiment had ended. This treatment is intended to capture the possibility that drivers in the field do not have perfect awareness of their ability, but instead base their assessment of it on various signals they have received, with certain signals more likely to be observed by more capable drivers than less capable ones.

In *Unknown Type/Quiz*, Drivers were assigned their type based on performance on a quiz, with the top 1/3 of Drivers being assigned Type 1 and the rest Type 2. All subjects took the quiz before they were assigned roles and placed into groups of eight, in order to ensure that the groups had the correct composition. Drivers were informed of their type once the experiment had ended. The quiz contained ten questions of varying difficulty pertaining to Florida's road rules, with all questions coming from a practice exam for the road rules test administered by the Florida Department of Motor Vehicles.[14] The treatment capitalizes on actual differences

[13] Even though subjects were randomly re-matched at the start of every block and decisions were made anonymously, the small group sizes created the possibility that subjects quickly learned they were matched with Officials who never accepted bribes and so did not bother to offer one. To reduce the likelihood of this happening, given that we are mainly interested in how individuals decide to make bribe offers, we eliminated the possibility of Officials experiencing a loss to help remove any unnecessary hindrances to their accepting offers.

[14] Question difficulty was calibrated via an online survey that was issued to students in a large introductory macroeconomics course at Florida State University.

Table 5
Per-period probabilities of an accident and expected group benefit.

Number of accepted offers	Role	Per-period prob. of loss	Expected net group benefit
Zero	Driver	0.04	180.52
One			
Type 1 briber	Driver (Type 1)	0.04	192.52
	Driver (Type 2)	0.04	
Type 2 briber	Driver (Type 1)	0.09	176.92
	Driver (Type 2, non-briber)	0.09	
	Driver (Type 2)	0.14	
Two			
Type 1 and Type 2 bribers	Driver (Type 1)	0.09	188.92
	Driver (Type 2, non-briber)	0.09	
	Driver (Type 2)	0.14	
Only Type 2 bribers	Driver (Type 1)	0.13	174.30
	Driver (Type 2)	0.18	
Three	Driver (Type 1)	0.13	186.30
	Driver (Type 2)	0.18	

in driving knowledge to assign types that were designed to mimic differences in driving ability. This creates the opportunity to bring relevant elements of the field into the laboratory, such as an individual's imperfect awareness of their ability and any accompanying overconfidence.

For all treatments, the chosen parameters imply an expected financial incentive of 12 ECU for Type 1 Drivers to offer a bribe, while Type 2 Drivers have an incentive of no more than about 4.5 ECU. Though all Drivers have an incentive to bribe, group welfare is at its highest when only the Type 1 Driver bribes and is at its lowest when the Type 1 Driver does not bribe but both Type 2 Drivers bribe. Per-period probabilities of experiencing a loss and corresponding group benefits are summarized in Table 5.

A Licensing Official's earnings for the decision phase in each block consisted of the initial endowment and any accepted bribe offers. For a Driver who successfully bribed, earnings for the decision phase in each block were the initial endowment and the value of bypassing the exam less the amount of the accepted bribe offer and any losses incurred during the block. Earnings for all other Drivers were simply the initial endowment less any losses incurred during the block. At the end of a session, subjects were paid their earnings for one randomly selected block.

4.1.2. Belief elicitation phase

Subjects' beliefs were again elicited using a Quadratic Scoring Rule. Drivers were first asked how many of the other members of their group of four they expected to offer a bribe and how many of those offers they expected to be accepted, while Licensing Officials were asked how many offers they expected to receive and accept. Drivers were asked to state what they believed was the probability of experiencing at least one accident during the block both if they were to bribe for a license and if they did not, as well as how likely they thought it was that they were assigned Type 1. At the end of the experiment, subjects were paid their earnings for a randomly chosen belief from the same block that was selected for the decision phase payment.

4.2. Predictions

Predictions rely on the psychological cost framework of Barr and Serra (2009), as they did in Experiment 1, but they also incorporate predicted effects of overconfidence on behavior.

Since Type 1 Drivers can bribe without changing their probability of causing an accident, no additional social harm can be generated by their decision to do so. Thus, Type 1 Drivers who are aware of their type should be more willing to offer a bribe than their Type 2 counterparts because of the lower psychological costs associated with the decision. The stronger a Driver's belief that they have been assigned Type 1, the more willing they should be to offer a bribe. Thus, if the quiz in *Unknown Type/Quiz* allows more subjects to believe that they are Type 1 than in *Unknown Type/Random*, and the noisy Type information in *Unknown Type/Random* allows more subjects to believe they are Type 1 than in *Known Type/Random*, we should see a higher proportion of bribe offers in *Unknown Type/Quiz* when compared to *Unknown Type/Random*, and in *Unknown Type/Random* when compared to *Known Type/Random*.

Prediction 1. The proportion of bribe offers in *Unknown Type/Quiz* will be higher than in *Unknown Type/Random*, and higher in *Unknown Type/Random* than in *Known Type/Random*.

Drivers in *Unknown Type/Quiz* do not know their types, but they can form beliefs about their types based on intuition about quiz performance. The overconfidence literature discussed earlier suggests that individuals tend to overplace themselves on relatively straightforward tasks. The road rules quiz that was used to assign Type in *Unknown Type/Quiz* can arguably be classified as a straightforward task, especially when one considers the possibility that many subjects had driver's licenses. As such, overplacement on this quiz is likely, and more Drivers should believe they are Type 1 than if types had been randomly assigned, as they were in *Unknown Type/Random*.

Table 6
Summary of treatments.

Treatments	Sessions	Independent observations	Subjects
Known Type/Random	3	6	48
Unknown Type/Random	2	6	48
Unknown Type/Quiz	2	6	48
Total	7	19	144

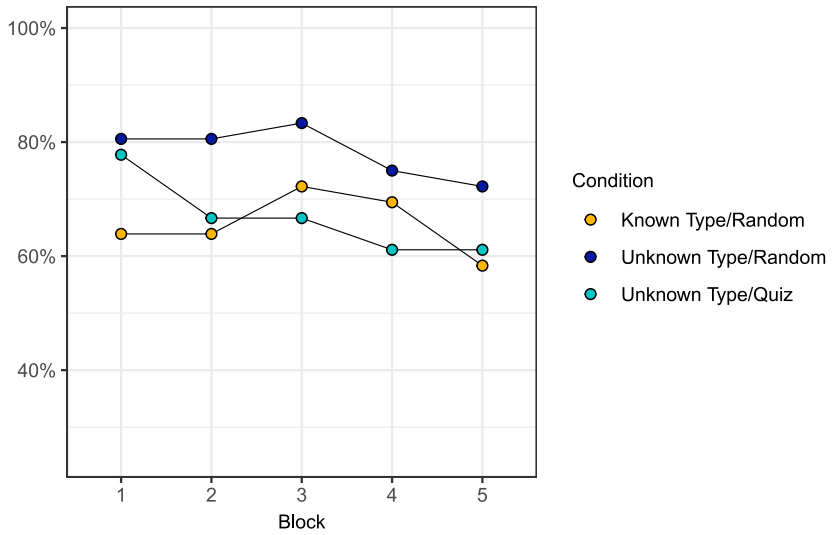

Fig. 4. Bribes by treatment. *Notes.* Proportion of bribe amounts by treatment.

Prediction 2. The proportion of subjects believing they are Type 1 will be higher in the *Unknown Type/Quiz* treatment than in the *Unknown Type/Random* treatment.

We use the same framework as Experiment 1 to predict that a Driver's decision to bribe will be greatly influenced by the belief that other Drivers will do the same.

Prediction 3. The willingness of Drivers to offer a bribe in all treatments will be positively correlated with the number of other Drivers in their group they expect to offer a bribe.

4.3. Results

Our collected sample is summarized in Table 6. Like Experiment 1, we aimed to collect data from 48 subjects per treatment.

As in Experiment 1, we begin by estimating the total effect of treatment on bribing behavior before exploring beliefs. Figures have the same units and interpretations as their counterparts in Experiment 1.

4.3.1. Bribing behavior

Differences in bribing behavior between Experiments 1 and 2 are apparent when comparing Fig. 4 to the corresponding Fig. 1. We see an overall reduction in the proportion of subjects who choose to offer a bribe in Experiment 2. While interesting, it is unclear whether this is attributable to the richer context or our introduction of heterogeneous types. However, since the new type added in Experiment 2 *reduces* the likelihood of losses, we are inclined to favor the former explanation.

Overall, we see a higher incidence of bribes in the *Unknown Type/Random* treatment than in *Known Type/Random* or the *Unknown Type/Quiz* treatment. In Experiment 1, we saw treatment differences preserved throughout all 5 blocks. However, we see fewer clear distinctions in Experiment 2. While Drivers in *Unknown Type/Random* consistently bribed more than the other two treatments, bribing behavior in *Known Type/Random* and *Unknown Type/Quiz* are indistinguishable.

Table 7 confirms these pairwise differences between treatments. We find that Drivers in *Unknown Type/Random* bribed significantly more than in *Known Type/Random*, which is consistent with Prediction 1. However, the differences observed between *Unknown Type/Random* and *Unknown Type/Quiz* are not statistically significant, which runs counter to Prediction 1. This result suggests that either the signals received from taking the quiz in *Unknown Type/Quiz* were equally as noisy as the signal in *Unknown Type/Random*, or that Drivers' beliefs that they were Type 1 were not significant predictors of bribing behavior. The latter is a key assumption of Prediction 1 and will be revisited in the next section.

Table 7
Treatment differences, Experiment 2.

Difference	Bribe offered = 1
Unknown Type/Random − Known Type/Random	15.08*
	(8.31)
Unknown Type/Quiz − Unknown Type/Random	−12.63
	(8.18)
Subjects	108
Blocks	5
N	540

Notes. Average partial effects of treatment expressed in percentage points. Effects estimated using random effects logistic regression with varying intercepts for subject and block to account for within-subject and within-block correlation. Standard errors in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 8
Regression analysis, Experiment 2.

	Bribe offered = 1	
Type 1 beliefs	0.17	0.22
	(0.2)	(0.17)
Exp. offers		7.73**
		(3.7)
Exp. acceptances		2.67
		(4.08)
Loss belief difference		−0.04
		(0.11)
Subjects	72	72
Blocks	5	5
N	360	360

Notes. Within-subject average partial effects of beliefs, expressed in percentage points. Effects estimated using random effects logistic regression with a varying intercept for subject and mean-centered covariates. Standard errors in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01.

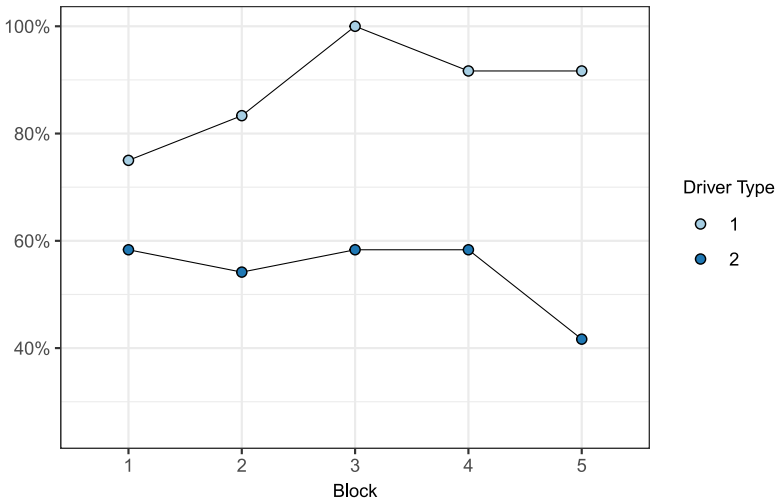

Fig. 5. Bribes by type in *Known Type/Random*.

Fig. 5 displays the proportion of Drivers of each type who offered bribes in *Known Type/Random*. Unsurprisingly, Drivers who knew they were Type 1 bribed frequently, as they did not increase their risk of loss by doing so. Adjusting for subject heterogeneity and block, the results of a logistic regression indicate that Type 1 Drivers were 36 pp more likely to offer a bribe ($p < 0.01$). It is worth noting that the number of bribe offers for Type 1 drivers increases throughout the experiment, while the number of bribe offers decreases for Type 2 drivers. This could suggest that Drivers use the outcomes of early blocks to update their probabilities of losses in the later blocks.

4.3.2. Beliefs

We now estimate the total effect of treatment on beliefs, focusing again on subjects in the role of Driver.

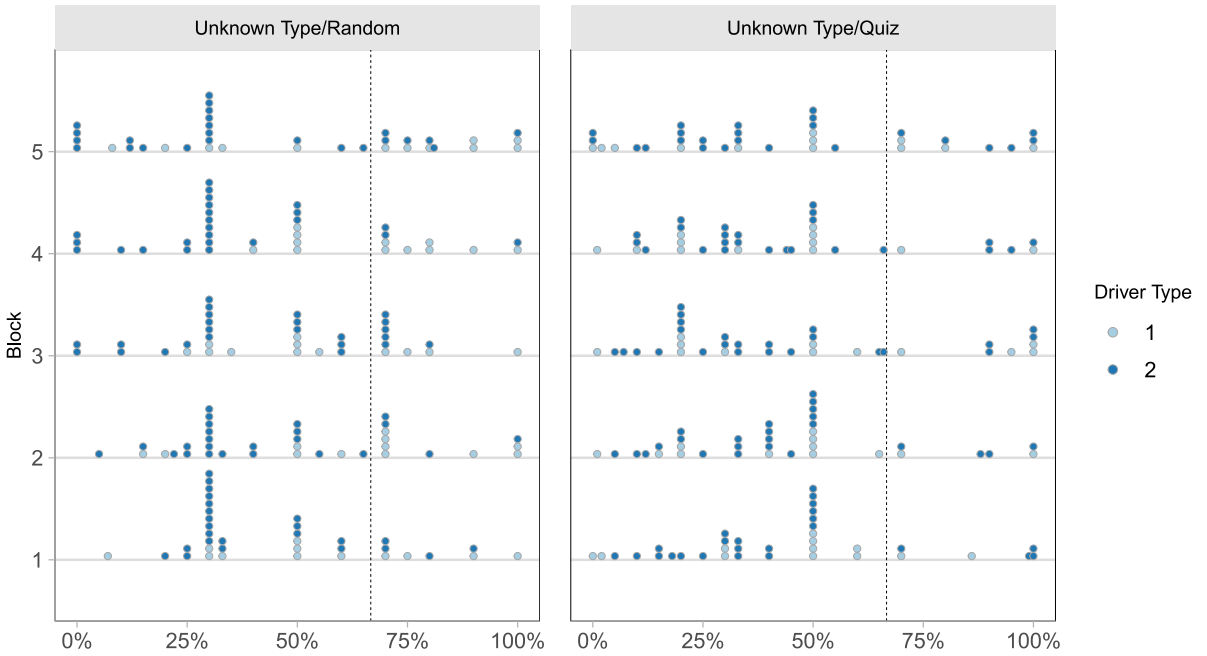

Fig. 6. Beliefs: Driver type. *Notes.* Beliefs of driver type, by treatment. Each dot represents the reported likelihood from each subject that they are Type 1. Lighter dots are actual Type 1 Drivers and darker dots are Type 2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Contrary to Prediction 2, we see no evidence that beliefs of one's type are affected by the nature of type assignment. Fig. 6 shows the distribution of Drivers' beliefs that they are Type 1, by actual type and treatment. The vertical line at 66.7% denotes the expected cutoff based on actual types. A typical overconfidence result would show more drivers to the right of this line in the right-hand panel, which we do not find. Drivers in *Unknown Type/Quiz* are actually 12 pp *less* likely to believe they are Type 1 than Drivers in *Unknown Type/Random* ($p < 0.10$ using regression analysis that controls for block and subject; results omitted for space considerations but available on request). This mild underconfidence is akin to what Healy and Moore (2008) classify as "underplacement" and, while less common than overconfidence, is well-documented in prior studies.

Drivers' beliefs about the number of other Drivers in their group who will offer a bribe and the number of these offers that will be accepted are displayed in Fig. 7. Similar to Experiment 1, Drivers generally expect the other Drivers in their groups to offer bribes to the Licensing Official. As for the expected number of acceptances, Drivers typically believe that Licensing Officials will accept one of the bribe offers from the other Drivers in their group.

Our regression analysis of the effect of beliefs on bribing behavior is reported in Table 8. We see that Drivers' type beliefs do not significantly influence their decision to bribe. This contrasts with the results of *Known Type/Random*, which showed an increase in bribing for players who knew themselves to be Type 1. It appears that Type 1 beliefs affect bribing decisions when Drivers know their type with certainty. However, when provided with noisy signals, it no longer significantly affects bribing decisions.[15] This supports our finding in the previous section, which showed no difference in bribing behavior between the two unknown treatments, *Unknown Type/Random* and *Unknown Type/Quiz*. As seen in Experiment 1, the most significant predictor of bribing behavior is the belief that others in the group will also choose to bribe, supporting Prediction 3.

In contrast to the results from Experiment 1, beliefs on the probabilities of losses incurred by the decision to bribe were not a significant predictor of bribing behavior. This result could be driving the lack of influence that type beliefs had on bribing behavior. Since loss differences were not influential on bribing decisions, the increase in loss probability between types would not have been a significant factor in the decision to bribe. Thus Drivers would make offers regardless of their believed type.

Similar to Experiment 1, we see results largely in line with predictions, though less separation by treatment. Subjects anticipate most others in their group to offer a bribe, and expect most bribes to be accepted. We again see over 50% of Drivers in all treatments offer a bribe, though this is less than in Experiment 1 and is also much less sensitive to treatments.

Curiously, we see no evidence of overconfident Drivers when types are determined endogenously by skill, which is inconsistent with the vast literature on overconfidence. This may be due to the fact that our subjects had taken their written driving test years ago.

[15] This result for the known type treatment is one reason we think a lack of variation in type belief, both within and between subjects, is why it appears as though the belief does not matter. It is unclear why there were not more subjects in the unknown type treatments with stronger beliefs, but if there had been, we would likely have seen some effect of the belief on bribing behavior.

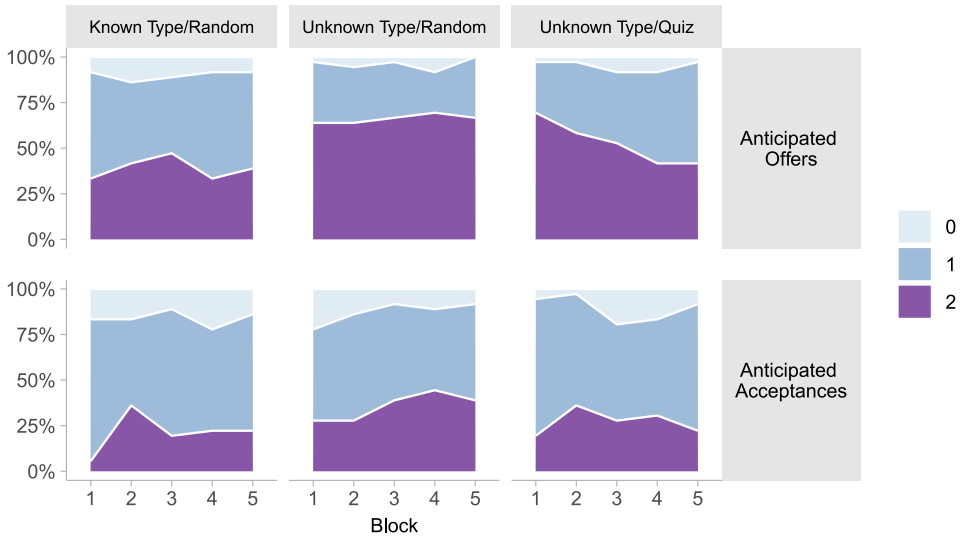

Fig. 7. Beliefs: Offers and acceptances. *Notes.* Proportion of private citizens predicting both how many others in their group will bribe (top row), and how many of these bribes will be accepted (bottom row).

While they may feel they are above-average drivers by ability, they may concurrently feel less certain that they have remembered many of the technical aspects of the rules of the road. This warrants further study, as policy aimed at improving compliance with driving tests may face an uphill struggle if too many drivers mistakenly hold inflated views of their ability.

5. Conclusion

We introduce a game designed to simulate petty corruption as a social dilemma in which the external costs that are generated by an exchange are not incurred with certainty. At all times, subjects faced some probability of causing themselves and another to experience a large loss, which was worsened substantially – both for the individual and other members of society – through bribes. We then conduct two experiments that explore behavior in this environment. The first experiment varied the presence of externality and accuracy of subject information regarding the risk of loss. In experiment 2, we introduce heterogeneity in subject types to examine whether endogenous assignment to type impacts beliefs and bribing behavior when facing probabilistic negative externalities.

Consistent with prior experimental corruption literature, we find that a substantial proportion of potential bribers – upwards of 60% in both experiments – choose to offer a bribe. In experiment 1, we find that adding externalities reduces bribing behavior somewhat. In the presence of externality, increasing information accuracy on the likelihood of loss greatly reduces bribing behavior. Experiment 2 shows that uncertainty over type reduces bribing behavior. In fact, socially inefficient outcomes in our experiments may well offer a conservative estimate of such behavior in the field due to random assignment to the role of public official. If dishonest individuals self-select into public service, as discussed by Hanna and Wang (2016), then greater social harm seems likely.

Also in line with corruption experiments that elicited the beliefs of participants, such as Köbis et al. (2015) and Fišar et al. (2016), we find in both experiments that the decision to offer a bribe is positively associated with the belief that others in the group are going to offer bribes. Interestingly, we fail to observe any impact of beliefs about one's own type on bribing behavior in Experiment 2, when we control for beliefs about other bribe offers. It seems that the social dilemma aspect of the environment dominates any individual-level motives.

Furthermore, we see no evidence of overconfidence in our endogenous treatment in Experiment 2. Our conjecture is that our subjects had taken their driving exam several years ago. Though they may feel that they are above-average drivers in practice, the quiz we used contained questions about driving rules, which they may feel less confident in having remembered.

The general structure of our game allows for a variety of potential extensions. For instance, removing the possibility of bribers harming themselves and the possibility of repeated selection to experience losses would closely fit environments in which bribery is used to bypass safety inspections or compliance checks in engineering projects. In exchanges of this nature, it is unlikely that bribers put themselves in a position to experience losses, but they certainly put others at risk. Connecting to our second experiment, it is reasonable to think that some may bribe to avoid the hassle of an inspection even if the service they provide is legally compliant, while others may bribe to avoid having to bring their service up to code.

Another way in which the experiment can be extended is through the inclusion of an intermediary who makes bribery decisions on behalf of participants. Drugov et al. (2014) examined the effect of such an intermediary in a corruption laboratory experiment and found that intermediation has the effect of lowering moral costs and increasing bribery. Bertrand et al. (2007) also found that

the participants in their Delhi, India field experiment who bribed for their driver's licenses typically did so through the use of an agent. Incorporating an intermediary into our design could thus allow the environment to more closely mimic the field.

Our findings support the argument that beliefs are a determinant of corrupt behavior, and they suggest that any policy aimed at reducing the incidence of similar petty corruption dilemmas should be designed with social enforcement mechanisms in mind. Though social enforcement mechanisms may need to be carefully tailored to the culture in which they are being implemented in order to be effective (Salmon and Serra, 2017), such mechanisms may be instrumental in mitigating the inefficiencies created by improperly obtained credentials.

Declaration of competing interest

The authors have no relevant or material financial interests that relate to the research described in this paper.

Funding for the described research was provided through the Department of Economics and the Hilton Center at Florida State University. This study involved human subjects and was reviewed and approved by the Florida State University Human Subjects Review Board before all experimental sessions were conducted at Florida State University. Documentation of funding sources and/or HSRB approval can be provided upon request.

Data availability

Data will be made available on request.

Appendix A. Supplementary materials

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jebo.2024.106693.

References

Abbink, K., Dasgupta, U., Gangadharan, L., Jain, T., 2014. Letting the briber go free: an experiment on mitigating harassment bribes. J. Public Econom. 111, 17–28.
Abbink, K., Freidin, E., Gangadharan, L., Moro, R., 2016. The effect of social norms on bribe offers.
Abbink, K., Hennig-Schmidt, H., 2006. Neutral versus loaded instructions in a bribery experiment. Exp. Econom. 9 (2), 103–121.
Abbink, K., Irlenbusch, B., Renner, E., 2002. An experimental bribery game. J. Law Econom. Organ. 18 (2), 428–454.
Abbink, K., Ryvkin, D., Serra, D., 2020. Corrupt police. Games Econom. Behav. 123, 101–119.
Alatas, V., Cameron, L., Chaudhuri, A., Erkal, N., Gangadharan, L., 2009. Gender, culture, and corruption: Insights from an experimental analysis. South. Econ. J. 663–680.
Alekseev, A., Charness, G., Gneezy, U., 2017. Experimental methods: When and why contextual instructions are important. J. Econ. Behav. Organ. 134, 48–59.
Armantier, O., Boly, A., 2013. Comparing corruption in the laboratory and in the field in Burkina Faso and in Canada. Econ. J. 123 (573), 1168–1187.
Banerjee, R., 2016. On the interpretation of bribery in a laboratory corruption game: moral frames and social norms. Exp. Econom. 19 (1), 240–267.
Banuri, S., Eckel, C., 2015. Cracking down on bribery. Soc. Choice Welf. 45 (3), 579–600.
Barr, A., Serra, D., 2009. The effects of externalities and framing on bribery in a petty corruption experiment. Exp. Econom. 12 (4), 488–503.
Barr, A., Serra, D., 2010. Corruption and culture: An experimental analysis. J. Public Econom. 94 (11), 862–869.
Bell, A., Jones, K., 2015. Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data. Polit. Sci. Res. Methods 3 (1), 133–153.
Berninghaus, S.K., Haller, S., Krüger, T., Neumann, T., Schosser, S., Vogt, B., 2013. Risk attitude, beliefs, and information in a corruption game–an experimental analysis. J. Econ. Psychol. 34, 46–60.
Bertrand, M., Djankov, S., Hanna, R., Mullainathan, S., 2007. Obtaining a driver's license in India: an experimental approach to studying corruption. Q. J. Econ. 122 (4), 1639–1676.
Bicchieri, C., Xiao, E., 2009. Do the right thing: but only if others do so. J. Behav. Decis. Mak. 22 (2), 191–208.
Butler, J.V., Serra, D., Spagnolo, G., 2017. Motivating whistleblowers.
Cameron, L., Chaudhuri, A., Erkal, N., Gangadharan, L., 2009. Propensities to engage in and punish corrupt behavior: Experimental evidence from Australia, India, Indonesia and Singapore. J. Public Econom. 93 (7), 843–851.
DellaVigna, S., 2007. Psychology and Economics: Evidence from the Field. Tech. Rep., National Bureau of Economic Research.
Drugov, M., Hamman, J., Serra, D., 2014. Intermediaries in corruption: an experiment. Exp. Econom. 17 (1), 78–99.
Eckel, C.C., Grossman, P.J., 2008. Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. J. Econ. Behav. Organ. 68 (1), 1–17.
Engel, C., Goerg, S.J., Yu, G., 2016. Symmetric vs. Asymmetric punishment regimes for collusive bribery. Am. Law Econom. Rev. 18 (2), 506–556.
Fišar, M., Kubák, M., Špalek, J., Tremewan, J., 2016. Gender differences in beliefs and actions in a framed corruption experiment. J. Behav. Exp. Econom. 63, 69–82.
Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Exp. Econom. 10 (2), 171–178.
Fisman, R., Svensson, J., 2007. Are corruption and taxation really harmful to growth? Firm level evidence. J. Dev. Econ. 83 (1), 63–75.
Greiner, B., 2015. Subject pool recruitment procedures: organizing experiments with ORSEE. J. Econ. Sci. Assoc. 1 (1), 114–125.
Hanna, R., Wang, S.-Y., 2016. Dishonesty and Selection into Public Service. Tech. Rep., National Bureau of Economic Research.
Healy, P.J., Moore, D.A., 2008. The trouble with overconfidence. Psychol. Rev. 115 (2), 502–517.
Hossain, T., Okui, R., 2013. The binarized scoring rule. Rev. Econ. Stud. 80 (3), 984–1001.
Huntington, S.P., 1968. Political Order in Changing Societies. Yale University Press.
Kahneman, D., 2003. Maps of bounded rationality: Psychology for behavioral economics. Am. Econ. Rev. 93 (5), 1449–1475.
Knack, S., Keefer, P., 1995. Institutions and economic performance: cross-country tests using alternative institutional measures. Econom. Polit. 7 (3), 207–227.
Köbis, N.C., van Prooijen, J.-W., Righetti, F., Van Lange, P.A., 2015. "Who doesn't?"—The impact of descriptive norms on corruption. PLoS One 10 (6), e0131830.
Leff, N.H., 1964. Economic development through bureaucratic corruption. Amer. Behav. Sci. 8 (3), 8–14.
Liberali, J.M., Reyna, V.F., Furlan, S., Stein, L.M., Pardo, S.T., 2012. Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment. J. Behav. Decis. Mak. 25 (4), 361–381.

Méon, P.-G., Sekkat, K., 2005. Does corruption grease or sand the wheels of growth? Public Choice 122 (1), 69–97.
Rivas, M.F., 2013. An experiment on corruption and gender. Bull. Econ. Res. 65 (1), 10–42.
Ryvkin, D., Serra, D., 2017. The industrial organization of corruption: Monopoly, competition and collusion.
Ryvkin, D., Serra, D., Tremewan, J., 2017. I paid a bribe: An experiment on information sharing and extortionary corruption. Eur. Econ. Rev. 94, 1–22.
Salmon, T.C., Serra, D., 2017. Corruption, social judgment and culture: An experiment. J. Econ. Behav. Organ..
Schlag, K.H., Tremewan, J., Van der Weele, J.J., 2015. A penny for your thoughts: A survey of methods for eliciting beliefs. Exp. Econom. 18 (3), 457–490.
Schotter, A., Trevino, I., 2014. Belief elicitation in the laboratory. Annu. Rev. Econ. 6 (1), 103–128.
Serra, D., 2011. Combining top-down and bottom-up accountability: evidence from a bribery experiment. J. Law Econom. Organ. 28 (3), 569–587.