Career incentives and corruption: A lab experiment

César Martinelli[*]   Naila C. Sebastián Esandi[†]

September 30, 2022

Abstract

We propose a model of political career incentives and corruption, and take it to the lab. As predicted by the model, moral incentives and the desire for reelection interact to refrain politicians from taking bribes in the early stage of their careers. Treatments with weaker reelection incentives do worse in terms of inducing good initial behavior of politicians, but may do better in terms of inducing good behavior at a later stage. The probability of voters' mistakes and, possibly, the distribution of moral motivations seem to vary with the treatment, with strategic behavior being apparently more common in environments with perfect information about politicians' actions.

1 Introduction

At least since the *Federalist Papers*, the expectation of running for election in the future has been considered a mean to discipline politicians and align their behavior with the interests of their constituents, among other possible functions. This insight has motivated a burgeoning literature in political economy dedicated to understanding the incentives that elections provide to politicians in environments in which there are possible conflicts of interests between politicians and their constituents, and in which actions of politicians are not perfectly well observed. Barro [1973], Ferejohn [1986], Banks and Sundaram [1993] and Fearon [1999] provided the initial contributions.

Most of the political economy literature on elections as a disciplining devise has a common game theoretic structure [Besley, 2007, Ashworth, 2012, Duggan and Martinelli, 2017]. This common structure can be illustrated succinctly by the *accountability game* described in Figure 1. Politicians have some private information about their objectives, which may differ from those of voters, as well as opportunities to privately profiteer at the expense of the common good. If they hope to run for office in the future, however, they may be deterred from acting against the interests of their voters because incriminating evidence may be leaked to voters, who may decide to punish the worst behaving politicians.

In fact, it may be rational from the point of view of voters to refuse reelection to politicians who have behaved badly, if past behavior is an indication of future behavior. This is the case if politicians have idiosyncratic, possibly privately known, predispositions, opportunities and moral costs associated to deviating from the common good, which in game theoretic terms can be modeled as the politician's *type*. However, politicians may be inclined to dissemble their type by behaving well before the election, in the expectation of reaping

possibly larger rewards in the future.

Equilibrium behavior in the accountability game is most easily described if we assume that there is a continuum of politicians' types, measuring their predisposition to act in favor of the common good. As in Martinelli [2022], suppose the opportunity to deviate from the common good is a bribe. Then politicians' types in equilibrium will sort into those who refuse the bribe before the elections and would do so in the future as well, those who would take the bribe both before and after the election, and those who refuse the bribe before the election in the expectation of reaping larger rewards (licit or illicit) in the future. We can refer to these three sets of types as *honest*, *corrupt*, and *opportunist* politicians, respectively. When voters have some information about past behavior, and use it at the voting booth, elections weed out corrupt politicians, and provide partial incentives for opportunist politicians to serve the common good.

Equilibrium behavior in the accountability game, however, is a delicate construction. It requires voters to pay attention to politicians' past behavior, and use it prospectively in order to determine whom to reelect (or promote to a higher position). From the point of view of each voter, selecting good politicians is a public good, which provides incentives to free ride on the attention paid by other voters. Moreover, the probability that a single vote affects the outcome is small or nil, which may greatly diminish the sense of responsibility of citizens when casting a vote, a point forcefully made by Schumpeter [1942] (see e.g. Martinelli [2006] and Matějka and Tabellini [2021] for game theoretic formulations). Incentives for politicians are more clear, insofar as they are forward looking and expect voters to reward good behavior.

In this paper, we take the accountability game to the lab. We consider a simple version of the accountability game, with two politicians and three

citizens—the minimal numbers to have political competition among politicians and a collective choice problem among citizens. Politicians initially have the opportunity to accept or reject a bribe, with the latter action having negative consequences for citizens. Citizens, in turn, can choose among the two politicians, after observing, with some probability, the past action of each politician. Last, the politician who is elected has, with some probability, the opportunity to accept or reject a bribe.

In each session, participants alternate in the roles of citizens and politicians, with randomly formed groups in each of ten rounds. We consider four treatments, varying the probability of observing information about politicians' behavior and the probability of being offered bribes after the election—in both cases the probability can be 1/2 or 1. Each session is dedicated to a single treatment, so the analysis of treatment effects is inter subjects.

By analogy to Smith [1982], who argues that markets in the lab are microeconomic systems, we think of our experimental setup as a political economy system. Accepting bribes and being elected give monetary rewards to politicians, and having politicians who accept bribes provides a monetary loss for citizens. The bribe's social cost exceed the individual reward to the politician. In addition to the monetary rewards implemented in the lab, the experiment presupposes that participants in their role as politicians have, to varying degrees, an intrinsic moral predisposition against taking bribes—be it because of altruism, utilitarian considerations, gratitude to voters, or fairness considerations. This unobservable moral predisposition corresponds to the notion of types in the game-theoretic model.

Summarizing the results, participants in their role as politicians generally behave according to theory: most of their observed behavior can be rationalized as the result of honest, corrupt, or opportunist strategies. Voters generally,

4

but not perfectly, react to available information. Treatments in which voters have perfect information about politicians' past action do better in terms of inducing good initial behavior of politicians, but may do worse in terms of inducing good final behavior of politicians. That is, there is a trade off between incentives and selection. The probability of voters' mistakes and possibly the distribution of moral motivations seem to vary with the treatment. This is not completely surprising: mistakes have different costs depending on the environment, and moral motivations are social in nature.

There is a growing body of experimental literature on corruption, including Barr and Serra [2009], Serra [2012], and the volume edited by Serra and Wantchekon [2012]; there is comparatively less on the accountability game and its connection to corruption in the lab. The experimental political economy literature is summarized by Palfrey [2016]. Cason and Mui [2003] provide a seminal contribution to studying experimentally environments that involve voting and explicit political processes. There is also a growing empirical literature on accountability and corruption, including, inter alia, Ferraz and Finan [2008], Ferraz and Finan [2011] and Costas-Pérez et al. [2012].

The remainder of this paper is organized as follows. The accountability game is described in section 2. Notions of equilibrium with and without mistakes are discussed in sections 3 and 4. The experiment design and implementation is discussed in sections 5 and 6. Experimental results are described in section 7. Conclusions are gathered in section 8.

2   The electoral accountability game

In this section we develop a simple model of accountability that forms the basis of the lab experiment. In particular, we adapt the model of Martinelli

[2022] to a lab environment by considering several voters rather than a representative one, correlated rather than independent observation of signals about politicians, and (in the equilibrium definitions) the possibility of politicians and voters making mistakes.

We consider two politicians, $i = A, B$, initially occupying lower level office, and n citizens, $j = 1, \ldots, n$, where n is odd. Each politician, while at the lower level office, has to make a public policy decision $x_i \in \{0, 1\}$ (the first period policy choice). Citizens obtain a payoff $g > 0$ if a lower rank politician adopt the policy $x_i = 1$ rather than $x_i = 0$. Each lower rank politician, however, can receive a bribe $b > 0$ if adopting the policy $x_i = 0$ rather than $x_i = 1$, where the bribe is offered with probability $p_1 > 0$.

Politicians' policy choices are not directly observed by citizens prior to casting their votes; instead, with probability q, citizens receive a perfectly informative signal of both politicians' policy choices $((s_A, s_B) = (x_A, x_B))$, and with probability $1 - q$ they receive an uninformative signal $((s_A, s_B) = (\emptyset, \emptyset))$. We can think of q as the quality of public information, and the signal $s_i \neq \emptyset$ as conclusive evidence of a wrongful $(x_i = 0)$ or rightful $(x_i = 1)$ policy choice from the viewpoint of citizens.

After signals are received, citizens vote to promote one of the two politicians to higher office. The promoted politician has to to make a public policy decision $x_2 \in \{0, 1\}$ (the second period policy choice). Citizens obtain a payoff g if the promoted politician adopt the policy $x_2 = 1$ rather than $x_2 = 0$. The promoted politician, however, can receive a bribe b if adopting the policy $x_2 = 0$ rather than $x_2 = 1$, where the bribe is offered with probability $p_2 > 0$. The promoted politician also receives a reward of $r > 0$ associated with higher office.

We assume that politicians differ in their warm-glow feeling after providing

Nature	Politicians	Nature	Citizens	Nature	Elected politician
politician types θ_i and bribe offers χ_i	policy choices x_i	signals s_i	election decision	bribe offer χ_2	policy choice x_2

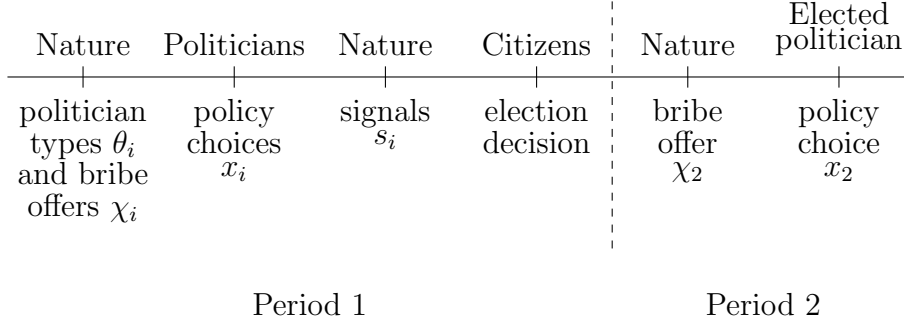Period 1   Period 2

Figure 1: The electoral accountability game.

a policy convenient to citizens. In particular, we assume that each politician has a (privately observed) parameter θ_i representing the psychic payoff obtained from providing a good policy, where θ_i is obtained from some continuous distribution density f (the same distribution for both politicians) with support $[0, \bar{\theta}]$, where $\bar{\theta} > b$. We also assume that all players discount the second period payoffs according to $\delta \in (0, 1]$.

Summarizing, the payoffs of voters $j = 1, \ldots, n$ are given by

$$(x_A + x_B + \delta x_2)g,$$

and the payoffs of politicians $i = A, B$ are given by

$$x_i\theta_i + (1 - x_i)\chi_i b + \delta\phi_i(r + x_2\theta_i + (1 - x_2)\chi_2 b),$$

where $\chi_i = 1$ if politician i is offered a bribe in period one and $\chi_i = 0$ otherwise, $\phi_i = 1$ if the politician is promoted and $\phi_i = 0$ otherwise, and $\chi_2 = 1$ if the promoted politician is offered a bribe and $\chi_2 = 0$ otherwise.

Payoffs are realized by the end of the game. Figure 1 illustrates the timing of events.

3 Responsive equilibrium

In this section we define the strategies and equilibrium for the accountability game described in the previous section.

A strategy for citizen $j = 1, \ldots, n$ is a mapping

$$\nu_A^j : \{0,1\}^2 \to [0,1],$$

specifying the probability $\nu_A^j(s_A, s_B)$ with which citizen j votes for politician A, and the complementary probability $\nu_B^j(s_A, s_B) = 1 - \nu_A^j(s_A, s_B)$ with which the citizen votes for politician B, given the observed signals.

A strategy $\sigma_i = (\sigma_{i1}, \sigma_{i2})$ for politician $i = A, B$ is a pair of mappings

$$\sigma_{i1} : [0, \overline{\theta}] \times \{0,1\} \to [0,1] \quad \text{and} \quad \sigma_{i2} : [0, \overline{\theta}] \times \{0,1\} \to [0,1]$$

where $\sigma_{i1}(\theta_i, \chi_i)$ and $\sigma_{i2}(\theta_i, \chi_2)$ specify the probability with which the politician chooses $x_i = 1$ and $x_2 = 1$, respectively, conditional on the politician's type and whether the politician is offered a bribe or not, and in the second period, conditional as well on the politician's being promoted.[1]

A belief system for citizens is a pair of mappings $\beta_A : \{0,1\} \to \mathcal{F}$ and $\beta_B : \{0,1\} \to \mathcal{F}$, where $\mathcal{F}$ is the set of probability density functions with support $[\underline{\theta}, \overline{\theta}]$, so that $\beta_j(s_j)[\cdot]$ represents the updated (common) beliefs of citizens about politician j's type after observing the signal s_j.

A perfect Bayesian equilibrium in the accountability game is a profile of strategies for politicians and voters and a belief system for citizens such that

---

[1]In principle, the politician's decision in case of being promoted could be influenced as well by the public information revealed by the end of the first period, including the public signals about both politicians and the vote tally. This additional information is irrelevant for politicians: given our assumptions about politicians' motivations, behavior in the second period depends trivially on the politician's type.

politicians play best responses to voters' strategies, citizens play best responses to the strategies of politicians and other citizens, given their updated beliefs, and the belief system is consistent with the politicians' strategies.

A perfect Bayesian equilibrium is consistent with sincere voting if citizens vote to promote with probability one the politician that has the largest probability of choosing the best policy choice in the second period, given the updated beliefs and politicians' strategies. Sincere voting is a refinement of perfect Bayesian equilibria: it precludes equilibria in which, for instance, all citizens vote for the same politician no matter what, so that no citizen is ever decisive.

We refer to the following strategy for citizens as the *responsive* strategy:

$$
\nu_A^j(s_A, s_B) = \begin{cases} 1 & \text{if } s_A = 1 \text{ and } s_B = 0 \\ 1/2 & \text{if } s_A = s_B \\ 0 & \text{if } s_A = 0 \text{ and } s_B = 1 \end{cases}.
$$

A *responsive equilibrium* is a perfect Bayesian equilibrium in which citizens play responsive strategies. Intuitively, if a responsive equilibrium is consistent with sincere voting, it requires that citizens interpret bad signals as evidence that the politician is less likely to provide good policies in the future, and vote according to those beliefs.

As it turns out, if voters play responsive strategies, the best response of politicians has a simple form. We say that politicians play *cutoff* strategies if for each politician there is a cutoff $\theta_i^* \in [\bar{\theta}, \underline{\theta}]$ such that

$$
\sigma_{i1}(\theta_i, \chi_i) = \begin{cases} 0 & \text{if } \theta_i < \theta_i^* \text{ and } \chi_i = 1 \\ 1 & \text{if } \theta_i > \theta_i^* \text{ or } \chi_i = 0 \end{cases},
$$

and

$$\sigma_{i2}(\theta_i, \chi_i) = \begin{cases} 0 & \text{if } \theta_i < b \text{ and } \chi_2 = 1 \\ 1 & \text{if } \theta_i > b \text{ or } \chi_2 = 0 \end{cases}.$$

We have:

Theorem 1. *There is a unique responsive equilibrium strategy profile. In this equilibrium, politicians play cutoff strategies, with cutoffs*

$$\theta_A^* = \theta_B^* = \theta^* \equiv \max\left\{0, \frac{(1 - \frac{1}{2}\delta q p_2)b - \frac{1}{2}\delta q r}{1 + \frac{1}{2}\delta q(1 - p_2)}\right\}.$$

Moreover, this strategy profile is consistent with sincere voting.

Proof. If politician i is promoted and offered a bribe, the politician optimally takes the bribe if $\theta_i < b$ and rejects the bribe if $\theta_i > b$ (as described in the definition of cutoff strategies). Hence, the discounted payoff gain for the politician in case of being promoted is $\delta(r + (1 - p_2)\theta_i + p_2 \max\{\theta_i, b\})$.

Now consider the problem of the politician if offered a bribe in period 1. If the politician rejects the bribe rather than accepting it, the politician gets a payoff loss in period 1 of $b - \theta_i$.

If voters follow responsive strategies, rejecting the bribe rather than accepting it increases the probability of promotion in $\frac{1}{2}q$, regardless of the strategy followed by the other politician. To see this, note that if signals are uninformative, rejecting the bribe does not change the probability of promotion. If signals are informative, which happens with probability q, and the other politician gets a good signal, politician i gets promoted with probability $\frac{1}{2}$ if getting a good signal and with probability 0 otherwise. Similarly, if the other politician gets a bad signal, politician i gets promoted with probability 1 if

getting a good signal and with probability $\frac{1}{2}$ otherwise. Thus, the increase in the probability of promotion if choosing a good policy rather than a bad policy is $\frac{1}{2}q$ regardless of the action of the other politician.

From previous arguments, the politician should reject the bribe in period 1 if

$$\tfrac{1}{2}q\delta(r + (1 - p_2)\theta_i + p_2 \max\{\theta_i, b\}) > b - \theta_i,$$

and should accept the bribe if the reverse inequality holds. The left-hand side of the previous inequality is positive and increasing in θ_i, while the right-hand side is decreasing in θ_i, and negative if $\theta_i > b$. Hence, either the politician prefers to reject the bribe for all values of θ_i—equivalent to a cutoff $\theta^* = 0$— or there is a unique $\theta^* \in (0, b)$ such that the politician prefers to reject the bribe if $\theta_i > \theta^*$ and to accept it if $\theta_i < \theta^*$. Manipulating the inequality above, we get the value of θ^* reported in the statement of the theorem.

It remains to show that there is a belief system for citizens that is consistent with the politicians' strategies and such that citizens vote sincerely given their beliefs. If $\theta^* \in (0, b)$, posterior beliefs of citizens can be derived from Bayesian updating. Since politicians obtaining the same signal induce the same posterior beliefs, and politicians obtaining different signals are such that given the beliefs of citizens the politician with the better signal has a larger probability of getting a type $\theta_i > b$, responsive strategies in this case are consistent with sincere voting.

If $\theta^* = 0$, observing $s_i = 0$ is off the equilibrium path. We can set $\beta_i(1)[\theta_i] = f(\theta_i)$ (as determined by Bayesian updating) and $\beta_i(0)[\theta_i] = f(\theta_i)$; given these beliefs, responsive strategies are consistent with sincere voting. $\square$

Note that the responsive equilibrium is unique, and not only the responsive equilibrium strategy profile, except for the possible (and inconsequential)

multiplicity of off-equilibrium beliefs if $\theta^* = 0$.

Figure 2 illustrates the behavior of politicians as a function of their private information. Politicians with types above b are "honest" and never take bribes. Politicians with types between θ^* and b are "opportunist" and take bribes only if promoted. Politicians with types between 0 and θ^* are "corrupt" and take bribes in every opportunity. For simplicity, we take $p_1 = p_2 = p$ in the figure.
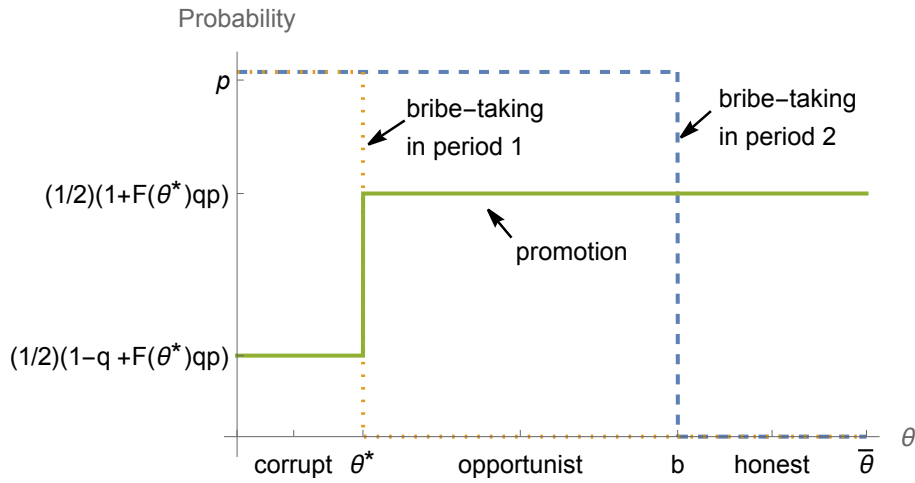

Figure 2: Politicians' equilibrium behavior and promotion probabilities.

Comparative statics are the same as in the quantal responsive equilibrium described in the next section, since the responsive equilibrium is a limit case.

There may be other perfect Bayesian equilibrium. For instance, consider the *counter responsive* strategy for citizens, in which citizens randomize between the two politicians when observing the same signal for both, and favor the politician who obtains the signal 0 when observing different signals. We say that a perfect Bayesian equilibrium is counter responsive if citizens play counter responsive strategies. Intuitively, a counter responsive equilibrium requires that citizens expect every politician to take bribes whenever possible

in the first period. Thus, observed signals do not change citizens' prior beliefs about politicians, and since citizens are indifferent, they can promote a politician who obtains a bad signal.

If voters play counter responsive strategies, accepting the bribe rather than rejecting it increases the probability of promotion in $\frac{1}{2}q$, regardless of the strategy followed by the other politician. A counter responsive equilibrium exists if and only if even the most "public spirited" politician is willing to take bribes in period 1 given the (perverse) electoral incentive, that is

$$\overline{\theta} - b \leq \tfrac{1}{2}\delta q(r + \overline{\theta}),$$

or equivalently

$$\overline{\theta}(1 - \tfrac{1}{2}\delta qr) \leq b + \tfrac{1}{2}\delta qr.$$

4   Quantal responsive equilibrium

In this section we consider the possibility that players in the game may deviate occasionally from best response equilibrium strategies, as in a quantal response equilibrium [McKelvey and Palfrey, 1995]. In particular, an agent quantal response equilibrium of the accountability game is a perfect Bayesian equilibrium in which each player in each opportunity to play chooses a best response action with probability $\lambda \in (0, 1)$ and otherwise plays each available action with equal probability;[2] the limit case when λ approaches one is the responsive equilibrium described in the previous section.

A *quantal responsive equilibrium* is an agent quantal response equilibrium in which best responding citizens play responsive strategies. Note that a quan-

---

[2]We adapt the definition of McKelvey and Palfrey [1996, 1998] to our game.

tal responsive equilibrium necessarily induces sincere voting from best responding citizens since, given that other citizens may make mistakes, the probability of a citizen being decisive is positive.

In a quantal responsive equilibrium the probability that a voter votes for a politician that obtained a good signal when the other politician got a bad signal is $\lambda + \frac{1}{2}(1 - \lambda) = \frac{1}{2}(1 + \lambda)$. Thus, the probability that a politician that obtained a good signal gets promoted if the other politician got a bad signal is $\frac{1}{2}(1 + \Lambda) < 1$, where

$$\Lambda = \left(\frac{1}{2}\right)^{n-1} \sum_{k=\frac{n+1}{2}}^{n} \binom{n}{k} \left((1 + \lambda)^k (1 - \lambda)^{n-k} - 1\right) \in (0, 1).$$

Intuitively, Λ indicates how close is the electorate as a whole to play a responsive strategy.

We have:

Theorem 2. *For any given $\lambda \in (0, 1)$, there is a unique quantal responsive equilibrium. Moreover, in this equilibrium, politicians play cutoff strategies, with cutoffs*

$$\theta_A^\lambda = \theta_B^\lambda = \theta^\lambda \equiv \max\left\{0, \frac{(1 - \frac{1}{4}\delta q \Lambda p_2(1 + \lambda))b - \frac{1}{2}\delta q \Lambda r}{1 + \frac{1}{2}\delta q \Lambda(1 - \frac{1}{2}p_2(1 + \lambda))}\right\}.$$

We omit the proof since it just follows the steps of the proof of Theorem 1. In particular, rejecting the bribe rather than accepting it increases the probability of promotion in $\frac{1}{2}\Lambda q$, regardless of the strategy followed by the other politician, and the discounted payoff gain for the politician in case of being promoted, in case $\theta_i < b$, is

$$\delta(r + (1 - \frac{1}{2}p_2(1 + \lambda)\theta_i + \frac{1}{2}p_2(1 + \lambda)b).$$

Thus, if $\theta^\lambda > 0$, it must satisfy

$$b - \theta^\lambda = \tfrac{1}{2}\Lambda q\delta(r + (1 - \tfrac{1}{2}p_2(1 + \lambda)\theta^\lambda + \tfrac{1}{2}p_2(1 + \lambda)b),$$

which yields the expression in the statement of Theorem 2. Note that there are no off-equilibrium beliefs in a quantal responsive equilibrium, because a bad signal in case $\theta^\lambda = 0$ can be interpreted as the result of a politician who is not playing a best response.

We provide now some results that are useful for interpreting the experiment. We have

Corollary 1. *The willingness to take bribes in period 1 is $\frac{1-\lambda}{2} + \lambda F(\theta^\lambda)$ and in period 2 is $\frac{1+\lambda}{2} - \lambda(1 - F(b))(1 + \Lambda qF(\theta^\lambda))$. In particular, the willingness to take bribes in period 1 is larger than in period 2 if and only if*

$$q < \frac{F(b) - F(\theta^\lambda)}{(1 - F(b))F(b)}.$$

Finally, consider a change in parameters from q and p_2 to q' and p_2', holding the other parameters constant (including λ and the distribution F), and let the initial and final cutoff be θ^λ and θ'^λ, respectively. Then the willingness to take bribes in period 1 increases if $\theta^\lambda < \theta'^\lambda$, and the willingness to take bribes in period 2 decreases if $qF(\theta^\lambda) < q'F(\theta'^\lambda)$.

Proof. The probability that a politician is willing to take a bribe in the first period is the probability that the politician chooses randomly multiplied by $1/2$, plus the probability that the politician is corrupt and plays a best response, that is $\frac{1-\lambda}{2} + \lambda F(\theta^\lambda)$.

The probability that a politician is willing to take a bribe in the second period is the probability that the promoted politician chooses randomly mul-

15

tiplied by $1/2$, plus the probability that the promoted politician plays a best response and is not honest. If the electorate chooses randomly, which happens with probability $1 - \Lambda$, or if the electorate chooses a best response but has uninformative signals, which happens with probability $\Lambda(1 - q)$, adding up to $1 - q\Lambda$, the probability of promoting an honest politician is simply $1 - F(b)$. If instead the electorate chooses a best response and has informative signals, which happens with probability Λq, the probability of promoting an honest politician is

$$(1 - F(b))^2 + 2(1 - F(b))F(\theta^\lambda) + 2 \times \tfrac{1}{2}(1 - F(b))(F(b) - F(\theta^\lambda))$$
$$= (1 - F(b))(1 + F(\theta^\lambda)).$$

Thus, the willingness to take bribes in the second period is

$$\tfrac{1-\lambda}{2} + \lambda \left(1 - ((1 - F(b))(1 - q\Lambda) + \Lambda q(1 - F(b))\left(1 + F(\theta^\lambda)\right)\right)$$
$$= \tfrac{1+\lambda}{2} - \lambda(1 - F(b))(1 + \Lambda q F(\theta^\lambda)).$$

The difference between the willingness to take bribes in the second and the first period is then

$$\lambda \left[F(b) - F(\theta^\lambda) - \lambda\Lambda q F(\theta^\lambda)(1 - F(b)) \right].$$

The other clauses in the statement of the corollary follow. $\qquad\square$

5 Experimental design and hypotheses

We test the predictions of the model in a laboratory experiment. Our experimental design is built over groups of five subjects; two of them chosen at random are assigned the role of *Public Officials* and the other three are assigned the role of *Regular Citizens*.[3] In each session, subjects play ten rounds, with groups and possibly roles changing in each round.

In each round, as in the model, each of the public officials is offered a bribe; this is equivalent to taking $p_1 = 1$ in the model. After public officials have taken or rejected the bribe, regular citizens observe a signal about the action chosen by the public official, with a probability that depends on the treatment. After observing the signals, regular citizens vote on whom of the two politicians to promote, with the winner chosen by simple majority. The winner takes the role of *Elected Public Official* and is offered a bribe with a probability that depends on the treatment. Finally, subjects are informed of their payoffs in the round.

For comparison purposes, we deviate slightly from the model, and have the elected public official deciding whether to accept or reject the bribe before knowing whether the bribe will be offered. Also, for simplicity, we have the public official doing the good public decision whenever not taking a bribe. Neither change has an effect on equilibrium.[4]

Payoffs accrued in the lab experiment are described in Table 1. Payoffs in the first period depend on the actions of the public officials, with $x_i = 0$ denoting taking the bribe and $x_i = 1$ denoting not taking the bribe. Payoffs in

---

[3]With respect to the framing entailed by assigning roles, note that Barr and Serra [2009] find no framing effects in a previous bribery experiment.

[4]The experiment design precludes choosing the bad policy by mistake when there is no bribe offering, which we assumed on the definition of quantal responsive equilibrium.

First period payoff	$x_A = 1$ $x_B = 1$	$x_A = 0$ $x_B = 1$	$x_A = 1$ $x_B = 0$	$x_A = 0$ $x_B = 0$
Public Official A	25	45	25	45
Public Official B	25	25	45	45
Regular Citizen	25	15	15	5

Second period payoff	$x_2 = 1$ $\chi_2 = 0$	$x_2 = 1$ $\chi_2 = 1$	$x_2 = 0$ $\chi_2 = 0$	$x_2 = 0$ $\chi_2 = 1$
Elected Public Official	25	25	25	45
Non-elected Public Official	5	5	5	5
Regular Citizen	25	25	25	15

Table 1: Payoffs in points.

the second period depend on whom of the officials is elected, which is decided by regular citizens, on the action of the elected public official, with $x_2 = 0$ denoting willingness to take the bribe and $x_2 = 1$ denoting unwillingness to take the bribe, and on whether the bribe is offered, with $\chi_2 = 0$ denoting a bribe was not offered and $\chi_2 = 1$ denoting a bribe was offered.

In terms of the model, the payoffs accrued in the lab in each round correspond to setting the gain for each citizen of an official taking the good policy to $g = 25 - 15 = 10$, the bribe to $b = 45 - 25 = 20$, the reward for promotion to $r = 25 - 5 = 20$, the discount $\delta = 1$, and the probability of bribes in the first period $p_1 = 1$. The psychic payoff associated with taking a policy the benefits regular citizens, θ_i is not observable; we expect heterogeneity in behavior of subjects in the role of public officials arising from different values of this payoff. Note that the gain of a good policy for the three citizens together is 30 and is larger than the cost for the official of relinquishing a bribe.

We consider four treatments, depending on the probability q that damaging information is revealed to regular citizens if a public official takes a bribe, and

	$q = 1$	$q = \frac{1}{2}$
$p_2 = 1$	PI100	II100
$p_2 = \frac{1}{2}$	PI50	II50

Table 2: Treatments.

	PI100	PI50	II100	II50
θ^*/b	0	$1/5$	$1/2$	$5/9$
θ^λ/b	$\dfrac{1-\frac{\Lambda}{4}(1+\lambda)-\frac{\Lambda}{2}}{1+\frac{\Lambda}{4}(1-\lambda)}$	$\dfrac{1-\frac{\Lambda}{8}(1+\lambda)-\frac{\Lambda}{2}}{1+\frac{\Lambda}{8}(3-\lambda)}$	$\dfrac{1-\frac{\Lambda}{8}(1+\lambda)-\frac{\Lambda}{4}}{1+\frac{\Lambda}{8}(1-\lambda)}$	$\dfrac{1-\frac{\Lambda}{16}(1+\lambda)-\frac{\Lambda}{4}}{1+\frac{\Lambda}{16}(3-\lambda)}$

Table 3: Cutoffs.

on the probability p_2 that a bribe is offered in the second period. In particular, we consider the cases of perfect ($q = 1$) and imperfect information ($q = \frac{1}{2}$) about bribe-taking, and high ($p_2 = 1$) and low ($p_2 = \frac{1}{2}$) probabilities of bribery for the elected official. The treatments and their abbreviations are summarized in Table 2. Each group of five subjects is assigned to one of the treatments for the ten rounds and is not informed about the existence of other treatments.

Using the parameter values implied by the lab experiment, the cutoffs corresponding to the four treatments are provided in Table 3. Cutoffs are presented as a fraction of the value of the bribe, $b = 20$. The first line provides the responsive equilibrium cutoffs, and the second line the cutoffs corresponding to quantal responsive equilibria. The latter depend on the value of λ, but in our lab implementation the ranking is the same as in the responsive equilibrium for any strictly positive value of λ.

Based on the quantal responsive equilibrium and the calculated cutoffs, we have the following predictions.

Hypothesis 1. *In each treatment, after observing a bad signal about only one of the public officials, voters are more likely to vote for the other public official.*

Hypothesis 1 is a consequence of best-responding citizens playing respon-

sive strategies.

Hypothesis 2. *In each treatment, if a public official takes a bribe before promotion, then the official is likely to take a bribe if promoted.*

Hypothesis 2 is a consequence of best-responding politicians playing cutoff strategies.

The next hypothesis is a conditional statement. Corollary 4 establishes a very permissive condition under which the observed willingness to take bribes of politicians in period 2 is larger than the observed in period 1, which depends on $F(b) - F(\theta^\lambda)$ and $1 - F(b)$. Though we cannot observe these numbers, they correspond to the proportion of opportunists and honest, which we can estimate for each treatment.

Hypothesis 3. *In each treatment, if the estimated fractions of honest and opportunist politicians, respectively $\hat{h}$ and $\hat{o}$, satisfy $q\hat{h}(1 - \hat{h}) < \hat{o}$, then public officials are more likely to be willing to take bribes in the second period than in the first period.*

Intuitively, the condition in the statement is that there are enough opportunists; it is satisfied for all q if $\hat{o} > 3 - 2\sqrt{2} \approx 0.17$.

Hypothesis 4. *In terms of the willingness to take bribes in the first period, the four treatments are ranked as follows:*

$$PI100 < PI50 < II100 < II50.$$

Hypothesis 4 follows from Corollary 1 and the calculated cutoffs.

Our final hypothesis ranks the willingness to bribe in the second period across treatments. It follows from Corollary 4 and it is partly a conditional

statement. Intuitively, one of the comparisons requires that the effect of changes in the quality of information is not obliterated by changes in the proportion of corrupt politicians.

Hypothesis 5. *In terms of the willingness to take bribes in the second period, the four treatments are ranked as follows:*

$$PI100 > PI50 \quad and \quad II100 > II50.$$

Moreover, if the estimated fraction of corrupt politicians under the treatments PI50 and II100, respectively $\hat{c}$ and $\hat{c}'$, satisfy $\hat{c} < \hat{c}'/2$, then

$$PI50 > II100.$$

Intuitively, the condition for a complete ranking is that the proportion of corrupt politicians is not overly sensitive to the quality of information.

6 Experimental setup

All the sessions of this experiment were run in the Interdisciplinary Center for Economic Science (ICES) laboratory for experimental economics at George Mason University (GMU) and all the subjects were members of the Mason community. The funds needed to carry out these experiments were also provided by ICES and GMU. Every session lasted under 60 minutes and involved a number of subjects that was a multiple of 5 and varied from 10 to 20, that is two to four groups per session.

Subjects received a show-up fee of $10. Their additional earnings depended on their decisions during the game, and could vary from $5 to $30. These

additional earnings were introduced in points currency throughout the game and were converted to US dollars at the end of the session with a rate of $5\,\text{points} = \$1$. At the beginning of the session, subjects were given an anonymous label that we used to record their behavior. Once they had their card, they were assigned a seat and asked to watch a 5-minute video with the instructions for the experiment. This video is an animated presentation with a voice over that explains how to play the game and how rewards will be calculated.

There are four versions of the instructions corresponding to the four treatments. After watching the instructions, participants were required to take a short quiz about them, to make sure that they understood properly the instructions. Once the quizzes were checked and they have resolved their questions about the procedure of the experiment, subjects were taken to an `otree` app [Chen et al., 2016]. To access the app, subjects were required to type in their participant label. Then, they were taken to a welcome screen that summarizes the instructions they saw in the video.

There were two sections in each session. In the first section, subjects play ten rounds of the game corresponding to the treatment for the session. In the second section, subjects answered a risk assessment question together with a short survey.

At the beginning of each round, participants were assigned a role for the round. Over the ten rounds, each participant played four times as a *Public Official* and six times as a *Regular Citizen*. Although role assignment followed a certain pattern, group assignment was random each round. There was no communication between the subjects during each session. Screenshots for the game played in each round are provided in the Appendix.

After the ten rounds, one of them was randomly selected to determine the

	PI100	PI50	II100	II50	Total
Women	26	28	30	22	106
Men	31	26	18	31	106
Non binary	1	0	0	2	3
Undeclared	2	1	2	0	5
Undergrad	36	33	32	37	138
Graduate	23	22	16	16	77
Undeclared	1	0	2	2	5
Total	60	55	50	55	220

Table 4: Distribution of subjects.

payoff of the participants. All the rounds had the same probability of being selected. Players were then taken to a screen informing about which round was selected and what was their reward in US dollars.

After being informed of the reward of the game, participants were taken to a screen with the instructions for a last bonus question that consisted of a list of random lottery pairs, as described in Harrison and Rutström [2008], to assess their risk aversion, and a short survey. The survey ask their gender, their level of studies, and their major. At the end, subjects got a screen with instructions to wait in their seats until they are taken to the payment room. Once the session was over, subjects were paid in US dollars when they gave back the card with their label. The mean payment per subject for this one-hour session was $19.45.

Table 4 breaks down the participants in the experiment by treatment, gender and study level. As indicated, there were between ten and twelve groups of five students in each round per treatment.

	responsive vote (%)	responsive electorate (%)	one sided p–value	estimated λ	estimated Λ
PI100	79.01	88.89	< 0.0001	0.58	0.78
PI50	74.36	88.46	< 0.0001	0.49	0.77
II100	76.67	80.00	0.0026	0.53	0.60
II50	72.73	81.82	0.0001	0.45	0.64
Total	76.08	86.96	< 0.0001	0.52	0.74

Table 5: Responsive votes in the lab.

7 Experimental results

7.1 Responsive strategies

We consider first the behavior of citizens, and in particular whether they played or not responsive strategies. In Table 5, we consider elections in which citizens had observed one politician who took a bribe and one who did not. The first column reports on the percentage of times that citizens cast votes for the politician who did not take the bribe, as prescribed by the responsive strategy, for each of the treatments. The second column reports on the percentage of times that the politician who did not take the bribe won the election.

In the third column of Table 5, we report the p–values for a binomial test in which the null hypothesis is that the probability of a citizen casting a responsive vote is 50% and the alternative is that the probability is larger than 50%. In every case we can reject the null at the 1% significance level; p–values are higher for II100 and II50 since samples are smaller (30 and 66, respectively). The evidence provides strong support for Hypothesis 1, as expected.

We use the first and second column to estimate λ and Λ, that is the frequency of voting according to the responsive strategy rather than randomly for individual citizens and for groups. These values are reported in the third and fourth column of Table 5. In particular, letting z be the frequency of casting

a responsive vote, the estimated λ satisfies $\hat{\lambda} + \frac{1}{2}(1 - \hat{\lambda}) = z$ or equivalently $\hat{\lambda} = 2z - 1$, and similarly for the estimated Λ. Depending on the treatment, between 45% and 60% of citizens seem to be playing responsive strategies. Due to composition effects, groups do better, with a frequency between 60% and 80% of playing responsively.

7.2   Cutoff strategies

We now consider the behavior of politicians, and in particular whether they played or not cutoff strategies. In Table 6, we report the percentage of promoted politicians that took each possible sequence of public policy decisions. The first three rows correspond to the policy sequences that would be adopted by honest, opportunist, and corrupt politicians, respectively, as described in the partition in Figure 2. The fourth row corresponds to the policy sequence that is inconsistent with cutoff strategies.

In the fifth row of Table 6, we report the p–values for a binomial test in which the null hypothesis is that the probability that a politician that has taken a bribe in period 1 and has been promoted will take a bribe again is 50%, and the alternative is that the probability that the politician will take a bribe again is larger than 50%. In every case, we can reject the null at the 1% significance level. The evidence provides strong support for Hypothesis 2, as expected.

While, according to the model, the inconsistent sequence can only happen by mistake, other policy sequences can also be reached by mistake, so an estimate of $1 - \lambda$ is a fortiori larger than the frequency of the inconsistent sequences. Similarly, the observed sequences of policy choices of promoted politicians do not give us directly the proportion of politicians who follow

sequence (x_i, x_2)	PI100	PI50	II100	II50	Total
honest $(1, 1)$	15.00	10.91	10.00	24.55	15.23
opportunist $(1, 0)$	50.83	50.91	19.00	22.73	36.59
corrupt $(0, 0)$	30.83	36.36	65.00	43.64	43.18
inconsistent $(0, 1)$	3.33	1.82	6.00	9.09	5.00
one sided p–values	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001

Table 6: Policy sequences in the lab.

each strategy, since (i) politicians make mistakes, and (ii) the probability of promotion after $x_i = 0$ is different from the probability of promotion after $x_i = 1$.

To estimate the probabilities of each possible best response as well as the probability of playing a best response, consider a given treatment and let the probabilities of a best response being opportunist, honest and corrupt be respectively o, h, and c, with $h + c = 1 - o$. The sequence $(0, 1)$ can be reached by a corrupt or honest politician if making one mistaken and one correct choice, which happens with probability $\frac{1 - \lambda^2}{4}$, or by an opportunist politician if making two mistaken choices, which happens with probability $\frac{(1 - \lambda)^2}{4}$, multiplied by the probability of being reelected after $x_i = 0$. Thus, the expected number of $(0, 1)$ sequences is the total number of politicians (promoted or not), multiplied by $(1 - o)\frac{1 - \lambda^2}{4} + o\frac{(1 - \lambda)^2}{4}$, multiplied by the probability of reelection after $x_i = 0$. The probability of reelection after $x_i = 0$ can be estimated by the number of times that a sequence $(0, 0)$ or $(0, 1)$ is observed, divided by the number of times a politician adopts $x_i = 0$.

Let v be the given by the number of times the sequence $(0, 1)$ is observed divided by the total number of $(0, 0)$ and $(0, 1)$ sequences, and let w be the number of times a politician adopts $x_i = 0$ divided by the total number of politicians. (These fractions can be calculated using Tables 6 and 8.) Then

equating the number of times the sequence $(0, 1)$ is observed to the expected number, we get the equation

$$vw = (1-o)\frac{1-\lambda^2}{4} + o\frac{(1-\lambda)^2}{4}. \tag{1}$$

Similarly, let v' be the given by the number of times the sequence $(1, 0)$ is observed divided by the total number of $(1, 1)$ and $(1, 0)$ sequences, and let w' be the number of times a politician adopts $x_i = 1$ divided by the total number of politicians. Then equating the number of times the sequence $(1, 0)$ is observed to the expected number, we get the equation

$$v'w' = (1-o)\frac{1-\lambda^2}{4} + o\frac{(1+\lambda)^2}{4}. \tag{2}$$

Solving the system of equations 1 and 2, we get[5] the following estimated values of λ and o:

$$\hat{\lambda} = v'w' - vw + \sqrt{(v'w' - vw)^2 - 2(v'w' + vw) + 1}.$$

and

$$\hat{o} = (v'w' - vw)/\hat{\lambda}.$$

Similarly, equating the number of times the sequences $(1, 1)$ and $(0, 0)$ are observed to the expected number, we get the equations

$$(1 - v')w' = h\frac{1+\lambda^2}{4} + o\frac{1-\lambda^2}{4} + c\frac{(1-\lambda)^2}{4} \tag{3}$$

---

[5]We take the larger root of the quadratic equation defining λ to guarantee that the estimated o is smaller than one.

	PI100	PI50	II100	II50
honest	9.91	6.66	9.94	22.99
opportunist	38.90	36.38	14.03	10.17
corrupt	52.01	56.96	76.54	66.85
$\hat{\lambda}$	0.83	0.91	0.85	0.73

Table 7: Estimated strategies of politicians (%) and prob. of best response.

and

$$(1 - v)w = h\frac{(1-\lambda)^2}{4} + o\frac{1-\lambda^2}{4} + c\frac{(1+\lambda)^2}{4}. \tag{4}$$

Combining these two equations and using the expression for $\hat{o}$ we get

$$\hat{h} = \frac{1}{2}\left[1 + \frac{w'(1-2v')}{\hat{\lambda}} - \frac{w(1-2v)}{\hat{\lambda}}\right]$$

and

$$\hat{c} = \frac{1}{2}\left[1 + \frac{w}{\hat{\lambda}} - \frac{w'}{\hat{\lambda}}\right].$$

In Table 7, we report estimates of the fraction of politicians playing opportunist, honest and corrupt strategies, as well as the estimated probability of playing a best response λ, for each treatment. Except for the case of the II50 treatment, estimated values of λ for politicians are much larger than those for individual citizens and electorates.

Given the observed behavior of politicians, it was a best response for citizens to play a responsive strategy, under the assumption of maximizing monetary payoffs. Conversely, given the observed behavior of citizens, it was a best response for politicians to restrict themselves to cutoff strategies, under the assumption that moral motivations and monetary payoffs worked as modeled. Thus, participants played best responses much more accurately in their role as politicians than in their role as citizens.

	first period			second period		
	men	women	total	men	women	total
PI100	47.58	52.88	51.67	76.05	89.13	81.67
PI50	57.69	53.57	56.36	88.13	86.00	87.27
II100	65.28	75.00	72.50	83.78	82.76	84.00
II50	66.13	59.09	62.27	68.75	61.9	66.36
Total	58.49	60.61	60.23	78.35	80.61	79.77

Table 8: Willingness to accept bribes per treatment and period, in %.

	second − first period		men − women (1st period)	
	98% confidence interval	one sided p–value	98% confidence interval	two sided p–value
PI100	(0.365, 0.585)	< 0.0001	(-0.208, 0.102)	0.4253
PI50	(0.380, 0.602)	< 0.0001	(-0.116, 0.199)	0.5426
II100	(0.015, 0.245)	0.0133	(-0.257, 0.063)	0.1492
II50	(0.007, 0.265)	0.0192	(-0.087, 0.228)	0.2949

Table 9: Differences in probability of willingness to accept bribes by period and by gender.

7.3   Bribery in the first versus second period

Note that the condition of Hypothesis 3, $q\hat{h}(1 - \hat{h}) < \hat{o}$, is satisfied in every treatment. In Table 8, we report the percentage of politicians who take bribes in the first period, and the percentage who are willing to take bribes in the second period, disaggregated by treatment and gender.[6] Consistent with Hypothesis 3, bribery in the first period is less frequent than the willingness to take bribes in the second period in every treatment.

In Table 9, we report on z–tests for the difference between the probability of willingness to accept bribery in the second period and the probability of accepting bribery in the first period. We provide 98% confidence intervals as well as p–values where the null hypothesis is that the difference is equal to zero

[6]The total columns consider men, women, and undeclared and nonbinary subjects (see Table 4).

and the alternative hypothesis is that the difference is larger than zero. We can reject the null hypothesis with varying degrees of confidence, most clearly for the perfect information treatments.

7.4 Treatment effects in the first period

We turn now to comparative statics predictions. Table 8 is moderately supportive of Hypothesis 4, with treatments effects going in the expected direction in the first period, except in the comparison between II100 and II50. The II50 treatment may have been cognitively the hardest, given uncertainty with respect to the availability of information before the election and bribe opportunities in the second period. The estimated λ for voters is lower in this treatment than in the others, although due to composition effects, the estimated Λ for electorates is similar to the II100 treatment.

A subtler reason for the reverse ranking of II100 and II50 is that in the latter observed bribes in the second period are less frequent simply because bribe opportunities happen half the time. If the distribution of moral costs is not constant across treatments because it is influenced by the observed frequency of bribe taking, bribe taking in the first period in the II50 may have been more stigmatic than in the II100 treatment simply because less bribe taking is observed in the second period, even if actions are anonymous in a lab setting.

Table 10a reports on z–tests for the difference between the probability of willingness to accept bribery in the first period for different pairs of treatments. We provide p–values where the null hypothesis is that the difference is equal to zero and the alternative hypothesis is that the difference goes in the direction predicted by the theory. We can reject the null hypothesis with

	PI50	II100	II50
PI100	0.1566	< 0.0001	0.0107
PI50		0.0002	0.1037
II100			†

(a) First period

	PI50	II100	II50
PI100	†	†	0.0036
PI50		0.1216	0.0001
II100			0.0014

(b) Second period

Table 10: Differences in probability of willingness to accept bribes by period and by treatment, one sided p–values. †: Reversal with respect to theory prediction.

varying degrees of confidence when comparing perfect information with imperfect information treatments, but not when comparing the two incomplete information treatments.

7.5 Treatment effects in the second period

Note that the condition of Hypothesis 5 is satisfied. The idea behind Hypothesis 5 is that, due to worse selection, a reduction of bribery in the first period is associated with an increase in the willingness to take bribes in the second period. This prediction fails in the comparison between the two PI treatments and in the comparison between the two II treatments, although it holds in the comparison between the PI treatments and II50—a large increase in bribery in the first period may have improved selection of promoted politicians. Overall, the evidence suggests different λ or different distribution of θ over treatments due to social considerations.

Table 10b reports on z–tests for the difference between the probability of willingness to accept bribery in the first period for different pairs of treatments. We provide p–values where the null hypothesis is that the difference is equal to zero and the alternative hypothesis is that the difference goes in the direction predicted by the theory. We can reject the null hypothesis with varying degrees

of confidence when comparing the treatments with strongest incentives (PI100 and PI50) with the treatment with the weakest promotion incentives (II50).

7.6 Gender effects

The relationship between gender and attitudes respect to corruption has been the subject of an animated debate. For instance, in a cross country experimental comparison, Alatas et al. [2009] find that women are less tolerant of corruption in some countries while in others there are no behavioral differences between men and women. The survey by Chaudhuri [2012] has a similar message; across a variety of experiments, it is either the case that women behave in a less corrupt manner or there are no significant gender differences. We present the frequency of willingness to take bribes in the first period disaggregated by gender in Table 8. Our results seem to indicate that gender and bribe-taking are independent; generally speaking, men and women take a bribe in the first period with similar frequency.

In Table 9, we report on z–tests for the difference between the probability of men and the probability of women accepting bribes in the first period. We provide 98% confidence intervals as well as p–values where the null hypothesis is that the difference is equal to zero and the alternative hypothesis is that the difference is different from zero. We cannot reject the null hypothesis, and generally we cannot discard the hypothesis that the probabilities of bribe taking for men and women are equal.

7.7 Learning

Last, we consider whether there are significant changes in the willingness to take bribes over the ten rounds in each treatment. In Figure 3, we report on
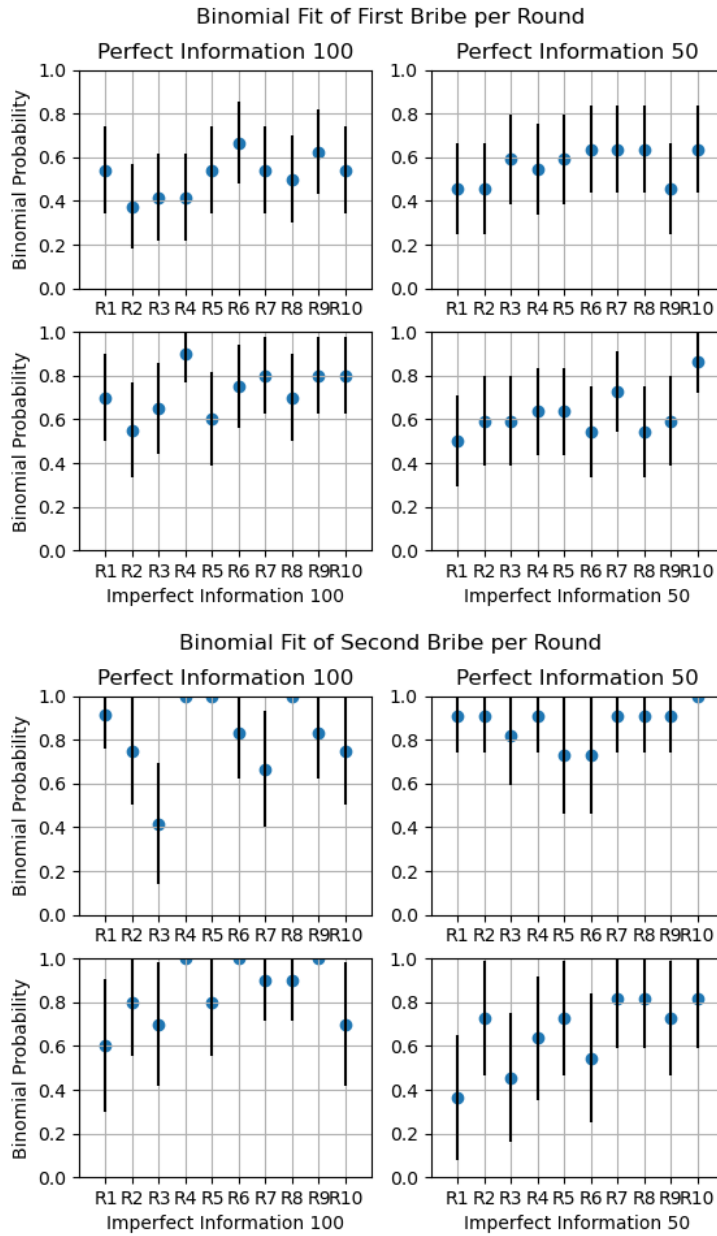
Figure 3: Binomial fit of willingness to take bribes in the first (top) and second (bottom) periods, with 95% confidence intervals.

estimates of the probability of willingness to take bribes per round in each period in each treatment. Confidence intervals overlap for the most, and it is hard to see any trend, except in the case of the second period in the II50 treatment, in which there is a greater willingness to accept bribes in the last rounds than in the first round.

The II50 is the treatment in which there is less direct information about the willingness to accept bribes of other politicians, since the probability of directly observing bribe taking is 50% by the end of the first period, and the probability that there is an opportunity for bribe-taking is only 50% in the second, so the evidence is consistent with a reduction in the moral cost associated with accepting bribes in the second period.

8 Conclusions

In this paper, we take a model of accountability and corruption to the lab, varying the quality of information available to citizens as well as the bribery opportunities for politicians at the final stage of their careers. Within each treatment, citizens and politicians behave according to theory predictions—citizens attempt to hold politicians responsible for past actions, and politicians in turn react strategically, accepting bribes less frequently in the initial stage of their careers in the hope of winning elections and possibly reaping bribes in the final stage.

Comparative statics predictions are generally borne when comparing the two treatments with perfect information (PI100 and PI50) with the treatments with imperfect information (II100 and II50). In every one-to-one comparison, there is more cheating in the first period in the imperfect information treatments, and there is more cheating in the second period in the perfect informa-

tion treatments (except in the comparison between PI100 and II100). That is, consistent with the theory, treatments with perfect information induce more opportunist behavior from politicians, who take fewer bribes in the initial stage but are more willing to take bribes in the final stage.

Comparative statics predictions are not borne, however, when comparing treatments with the same information quality, but different probability of getting a bribe in the second period (PI 100 versus PI50, and II100 versus II50). In particular, there is clearly more cheating in the II100 treatment than in the II50 treatment in the first period, even though incentives for strategic behavior are stronger in the former.

Looking more closely into the behavior in the II50 treatment, it apparent that voters and politicians are less likely to play best responses and, moreover, the estimated fraction of honest politicians is much larger than on other treatments. This is not entirely surprising. Since information is less frequently available in this treatment, and the behavior of politicians in the second period is less important, voters have diminished incentives to pay attention. More subtly, since fewer bribes are observed—both because there is less direct observation of bribes in the first period and fewer opportunities in the second—it may be more morally troublesome to take bribes.

Our conjecture about the observed differences in behavior between II50 and other treatments is consistent, for instance, with the "self respect" motivation for pro social behavior discussed by Bénabou and Tirole [2006]. Previous experimental evidence indicates that pro social behavior, such as refusing a bribe in our setting, is influenced by social history; see e.g. Berg et al. [1995]. Social history is bound to be specially relevant for explaining behavior in political economy experiments. In turn, this is a feature of experiments that is reminiscent of the empirical literature on corruption. As noted by della Porta

and Vanucci [1999], among others, the diffusion of corruption reduces its moral costs by creating a parallel, implicit normative system.

The results of the experiment indicate that institutional changes in the lab environment have consequences for the behavior of politicians and voters not only due to directly changing strategic incentives, but also due to changes in the the incentives to pay attention and in the incentives for moral behavior. These are channels of influence which the theoretical literature on accountability has not put much stress on, and which seems worth incorporating explicitly in the analysis.

In the experiment, politicians are chosen randomly from the set of citizens. It would be useful to investigate more realistic selection procedures, involving decisions of the potential candidates themselves. Self selection into politics is likely to be affected by moral considerations, including the possible stigma of being a politician in an environment where they are known to take bribes.

Finally, we have constrained ourselves to a finite horizon version of the accountability game. Infinite horizon versions—representing, say, long-lived political parties—have been object of discussion in the literature as well. Anesi and Buisseret [2022], for instance, show that if voters do not discount the future much, there is no trade off between selection and incentives and voters can approximate arbitrarily their optimal payoff. Equilibrium behavior may require involved retrospective behavior rules. Taking such models to the lab is of course challenging. This is an area in which much work remains to be done.

Appendix: Screenshots

At the beginning of each round, after roles were assigned, *Public Officials* were taken to a screen where they were offered a bribe, and were asked whether they wish to take it or not (see Figure A1 for this and other screenshots). The rest of the players saw a waiting screen. These waiting screens were used every time subjects need to wait for others to take action.

The next screen was shown to all in the group. It contained or not information about the bribery, depending on the probability assigned to the treatment. The screen also informed each player of their reward (in points) for the first period. *Regular Citizens* can infer the value of $x_A + x_B$ from seeing the period payoff. Despite this, if subjects did not get explicit information about who engaged in corrupt activities, they were not able to tell apart the actions of different *Public Officials*. Therefore, learning the reward did not bias their voting decision.

The next button took *Regular Citizens* to a screen where they need to vote for one of the *Public Officials* to be elected for the role in higher office. The options were *Public Official A* and *Public Official B*.[7] After voting, everyone got a message with the results. *Public Officials* were told whether they were elected or not, and *Regular Citizens* were informed about whether A or B won the election.

In the second period, the *Elected Public Official* was taken to a screen with the question "If you are offered a bribe, will you want to take it?" for the treatments with $p_2 = 0.5$ and with the question "You have been offered

---

[7]We expected voters to be indifferent when there was no information displayed or when $s_A = s_B$. In these 302 cases, *Public Official A* was voted 51.66%. The p–value of the statistical t-test of equal frequency is 0.3192, so we do not have enough evidence to reject the hypothesis of equal frequency. All subjects played *Public Official A* twice and *Public Official B* twice.

a bribe, do you want to take a bribe?" for the treatments with $p_2 = 1$. The *Elected Public Official* is required to choose while the other subjects wait.

Once the *Elected Public Official* made a decision, the bribe was offered with probability p_2. The last screen of the period included a message informing about whether the *Elected Public Official* had received (accepted and been offered) the bribe or not. Similarly to corresponding screen in the first period, there was also a line with the second period reward. Finally, everyone got a message with their total reward for the round. After they clicked on the Next button, they were taken to the role page again to begin a new round.

A bribe of 45 points is being offered to you by a third party. You are required to decide whether to take it or not.

Please, choose one of the options:
○ I want to take the bribe.
○ I do not want to take the bribe.

First period bribe-taking decision.

• Public Official A has taken the bribe: True. • Public Official B has taken the bribe: False. Your reward for the first period is: 15 points.	There is no information to be displayed about the bribery. Your reward for the first period is: 15 points.

First period bribery results. Left: information is displayed. Right: information is not displayed.

As a citizen, you are asked to vote for one of the public officials to be elected for the next round.

Please, choose one of the following options.
○ Public official A.
○ Public official B.

Citizens' election screen.

A bribe of 45 points will be offered to you by a third party with a probability of 50%. You are required to make a decision on whether to take it or not.

Please, choose one of the options:
○ If I am offered a bribe, I want to take it.
○ I do not want to take a bribe.

A bribe of 45 points is being offered to you by a third party. You are required to decide whether to take it or not.

Please, choose one of the options:
○ I want to take the bribe.
○ I do not want to take the bribe.

Second period bribe-taking decision. Top: $p_2 = 50\%$. Bottom: $p_2 = 100\%$.

The elected public official has received the bribe: False.

Your reward for this period is: 25 points.

Second-period bribery results.

Figure A1: Screenshots

References

Vivi Alatas, Lisa Cameron, Ananish Chaudhuri, Nisvan Erkal, and Lata Gangadharan. Gender, culture, and corruption: Insights from an experimental analysis. *Southern Economic Journal*, 75(3):663–680, 2009.

Vincent Anesi and Peter Buisseret. Making elections work: Accountability with selection and control. *American Economic Journal: Microeconomics* (forthcoming), 2022.

Scott Ashworth. Electoral accountability: Recent theoretical and empirical work. *Annual Review of Political Science*, 15:183–201, 2012.

Jeffrey Banks and Rangarajan Sundaram. Moral hazard and adverse selection in a model of repeated elections. In William Barnett, Melvin Hinich, and Norman Schofield, editors, *Political Economy: Institutions, Information Competition, and Representation*. Cambridge University Press, New York, 1993.

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

Robert Barro. The control of politicians: An economic model. *Public Choice*, 14:19–42, 1973.

Roland Bénabou and Jean Tirole. Incentives and prosocial behavior. *American Economic Review*, 96(5):1652–1678, 2006.

Joyce Berg, John Dickhaut, and Kevin McCabe. Trust, reciprocity, and social history. *Games and Economic Behavior*, 10:122–142, 1995.

Timothy Besley. *Principled Agents? The Political Economy of Good Government.* Oxford University Press, New York, 2007.

Timothy N. Cason and Vai-Lam Mui. Testing political economy models of reform in the laboratory. *American Economic Review*, 93(2):208–212, 2003.

Ananish Chaudhuri. Gender and corruption: A survey of the experimental evidence. In Danila Serra and Leonard Wantchekon, editors, *New Advances in Experimental Research on Corruption*, chapter 2. Emerald Group Publishing Limited, 2012.

Daniel L. Chen, Martin Schonger, and Chris Wickens. otree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97, 2016.

Elena Costas-Pérez, Albert Solé-Ollé, and Pilar Sorribas-Navarro. Corruption scandals, voter information, and accountability. *European Journal of Political Economy*, 28(4):469–484, 2012.

Antonella della Porta and Alberto Vanucci. *Corrupt Exchanges: Actors, Resources, and Mechanisms of Political Corruption.* Aldine de Gruyter, New York, 1999.

John Duggan and César Martinelli. The political economy of dynamic elections: Accountability, commitment, and responsiveness. *Journal of Economic Literature*, 55(3):916–984, 2017.

James Fearon. Electoral accountability and the control of politicians: Selecting good types versus sanctioning poor performance. In Adam Przeworski, Susan C. Stokes, and Bernard Manin, editors, *Democracy, Accountability and Representation*, pages 55–97. Cambridge University Press, 1999.

John Ferejohn. Incumbent performance and electoral control. *Public Choice*, 50:5–25, 1986.

Claudio Ferraz and Frederico Finan. Exposing corrupt politicians: the effects of Brazil's publicly released audits on electoral outcomes. *Quarterly Journal of Economics*, 123(2):703–745, 2008.

Claudio Ferraz and Frederico Finan. Electoral accountability and corruption: Evidence from the audits of local governments. *American Economic Review*, 101(4):1274–1311, 2011.

Glenn W. Harrison and E. Elisabet Rutström. Risk aversion in the laboratory. In James C. Cox and Glenn W. Harrison, editors, *Risk Aversion in Experiments*. Emerald Group Publishing Limited, 2008.

César Martinelli. Would rational voters acquire costly information? *Journal of Economic Theory*, 129(1):225–251, 2006.

César Martinelli. Accountability and grand corruption. *American Economic Journal: Microeconomics* (forthcoming), 2022.

Filip Matějka and Guido Tabellini. Electoral competition with rationally inattentive voters. *Journal of the European Economic Association*, 19(3):1899–1935, 2021.

Richard D. McKelvey and Thomas R. Palfrey. Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10(1):6–38, 1995.

Richard D. McKelvey and Thomas R. Palfrey. A statistical theory of equilibrium in games. *Japanese Economic Review*, 47(2):186–209, 1996.

Richard D. McKelvey and Thomas R. Palfrey. Quantal response equilibria for extensive form games. *Experimental Economics*, 1(1):9–41, 1998.

Thomas R. Palfrey. Experiments in political economy. In John Kagel and Alvin Roth, editors, *Handbook of Experimental Economics*, volume 2, pages 45–58. Princeton University Press, 2016.

Joseph Schumpeter. *Capitalism, Socialism and Democracy*. Harper & Brothers, 1942.

Danila Serra. Combining top-down and bottom-up accountability: evidence from a bribery experiment. *Journal of Law, Economics, & Organization*, 28 (3):569–587, 2012.

Danila Serra and Leonard Wantchekon, editors. *New Advances in Experimental Research on Corruption*, volume 15 of *Research in Experimental Economics*. Emerald Group Publishing Limited, 2012.

Vernon L. Smith. Microeconomic systems as an experimental science. *American Economic Review*, 72(5):923–955, 1982.