AN EXPERIMENT ON ACCOUNTABILITY AND GRAND CORRUPTION

BACHELOR'S DEGREE THESIS

Author: Naila C. Sebastián Esandi

Supervisor: César A. Martinelli
UPC Tutor: Josep Freixas Bosch

May 2022

Interdisciplinary Center for Economic Science
Centre de Formació Interdisciplinària Superior

Bachelor's degrees: Mathematics & Engineering Physics

Abstract

In this thesis, we test in the lab a model about electoral accountability and corruption in political careers. Starting from a game-theory model by Martinelli [2022], we study it using the Bayesian equilibrium and Agent Quantal Response equilibrium theories to make predictions about the actions of both politicians and voters. We find that the regret for taking bribes explains the behavior of politicians and that players do not always pay attention to the game.

En esta tesis, ponemos a prueba en el laboratorio un modelo sobre contabilidad electoral y corrupción en carreras políticas. Partiendo del modelo de teoría de juegos de Martinelli [2022], lo estudiamos utilizando las teorías de equilibrio bayesiano y de equilibrio *quantal* de agentes para predecir las acciones tanto de los políticos como de los votantes. Encontramos que el remordimiento por la toma de sobornos explica el comportamiento de los políticos y que los jugadores no siempre están atentos al juego.

En aquesta tesi, posem a prova en el laboratori un model sobre comptabilitat electoral i corrupció en carreres polítiques. Partint del model de teoria de jocs de Martinelli [2022], l'estudiem utilitzant les teories d'equilibri bayesià i d'equilibri *quantal* d'agents per predir les accions tant dels polítics com els votants. Trobem que el remordiment per agafar un suborn explica el comportament dels polítics i que els jugadors no sempre estan atents al joc.

Keywords— game theory, microeconomics, corruption, bribery, Bayesian equilibrium, Quantal Response equilibrium, politics;

teoría de juegos, microeconomía, corrupción, sobornos, equilibrio bayesiano, equilibrio de respuesta quantal, política;

teoria de jocs, microeconomia, corrupció, suborns, equilibri bayesià, equilibri de resposta quantal, política.

American Mathematical Society code— 91B14

Acknowledgements

First of all, I need to say thanks to everyone at CFIS for the trust that you deposited in me 5 years ago. Thank you for all the opportunities you brought to me by accepting me in this program.

This Bachelor's thesis would not have been possible without Professor César Martinelli. Thanks for welcoming me with open arms since day one and for being a perfect mentor and professor.

I need to thank the ICES students and faculty too. Especially Edgar Castro for your help, guidance, and discussions about math and economics; Alex Psurek for being a great friend, mentor, colleague, and support this year (also for your voice in the instructions videos); Yi Tian for guiding me through the administrative process necessary to run experiments in economics. Also to Nick Brown, Jinpeng Shi, and Hugo Díaz, for making my experience in the US so enriching and fun.

Although I developed this thesis in the United States, there were two persons that constituted an essential link with Spain. The first one is Oriol Riera, whom I wish to thank for trusting me and making this opportunity possible for me. The second one is Josep Freixas Bosch, my local UPC supervisor, whom I thank for being my academic link to UPC during my research.

On the other hand, I want to thank my friends in Barcelona for making the most of these last 5 years and my friends at GMU for this exchange experience. Mireia and Isabella, I could not imagine making it without you. Thank you.

Finally, but most important of all, I want to thank my mum, dad, and sister for being always there for me, supporting me unconditionally, and all my family in general, for bringing so much love, life, and happiness to everything I accomplish in life.

Contents

1 Introduction

The goal of this thesis is to study the accuracy of a theoretical model predicting human behavior to the results from a lab experiment concerning real people. It is an interdisciplinary project involving mathematics, economics, software development, and data analysis.

This first section includes a review of some literature connected to the project and a list of definitions and terms that are used throughout the project. These definitions are used in game theory applied to microeconomics. The second section contains a description of the game theoretical model that we take to the laboratory and predictions of the behavior of the subjects. The third section has a description of the setup in the experimental economics laboratory. The findings of the experiment are presented in section four. Section six has a summary of how our model worked and the results we got. Finally, section six has suggestions of extensions of this project.

1.1 Previous Work

The basis of this thesis is *Accountability and Grand Corruption* by Martinelli [2022]. In that paper, he analyzes a problem that is particularly common in newly established democracies: bribe-taking by high ranking politicians. He considers environment where multiple contestants that are public officials at a certain rank fight for being selected to a higher political office. In the course of this process, the presidents, prime ministers, or authorities in democratic governments are tempted to take bribes. Taking the bribes leads them to adopting policies that are not the most beneficial for the majority of the citizens, instead they damage the general welfare.

This model is a game that combines discrete actions (taking a bribe or not, or voting for a certain public official to be elected) with continuous types (quality of the politicians). This allows a neat characterization of the equilibrium.

Our model is similar to Martinelli's (2022), but has several simplifications so that it is feasible to take it to the economics laboratory. We have 2 lower rank officials that are offered a bribe in exchange for adopting policies that are not optimal for the regular citizens. These citizens receive an information signal about the bribery with a certain probability. After getting (or not) that signal, they vote to elect one of the officials for the higher rank position. Once promoted, the elected official is given a reward for being at the higher rank and is again given the option to want to take a bribe. In this second period, the politician is only offered a bribe with a certain probability. Thus, it is expected that voters promote the official who is

more likely not to be corrupt. Since the politicians know this when they are in the lower rank position, they might be tempted not to take a bribe when they are in the lower rank to improve their chances of being elected.

We also use a cutoff structure for the equilibrium based on the types of the politicians. The cutoff type is the one that values equally the present net gain of taking the bribe in the lower rank level and the future increase of payoff that comes with the increase of probability of election related to rejecting the first bribe.

We use two equilibrium theories: Bayesian equilibrium (BE) and Quantal Response equilibrium (QRE). The first one assumes that all players are rational and will act consequently and the second one allows for mistakes on the players choice of actions.

Before us, Ferraz and Finan [2008] and Ferraz and Finan [2011] gathered empirical evidence from Brazil. Their analysis showed that, when there is a possibility of reelection, corruption is lower in the first round and in those scenarios where public information is better.

Costas-Pérez et al. [2012] study how corruption scandals affect electoral outcomes in Spain during 1996 and 2009. They find in empirical evidence that there is a significant reaction from the voters when the press coverage of the scandal is extensive. They also find no vote loss at all in cases dismissed or with reports to the courts that did not lead to a further judicial intervention.

One of our findings regarding voters behavior is that they sometimes fail to pay attention to who they vote for. This matter was studied with more depth in Matějka and Tabellini [2021] where they address voters information and attention creating a model and proving it with empirical data from general elections in the United States of America. Moreover, if we consider *paying attention* as a cost in the utility function of the voters, we can link this to Martinelli and Palfrey [2020], where they study costly voting in elections with 2 alternatives; and to Cason and Mui [2003], where they address the issue of rational voters being interested in acquiring information about the election options.

Our addition to the picture is that we intend to study these two behaviors entangled in a single experiment.

When it comes to experimental economics, Serra [2012] and Cason and Mui [2003] argue that a laboratory experiment is the right framework to study policy reforms and corruption incentives. They affirm that it allows us to measure corruption directly and that its cost of observing people's response to different policies and corruption incentives is low.

Barr and Serra [2009] find that framing in experiments about bribery does not have a significant effect in the results. Thus, in our study, we feel free to include

some framing such as the names of the roles (*Public Official* and *Regular Citizen*), the term "*bribe*", and the terms "*voting*" and "*election*" assuming that it will not affect our findings.

Finally, Serra and Wantchekon [2012] finds some evidence indicating that women are equally or less likely to engage in corruption that men, but our findings cannot corroborate this with certainty either.

1.2 Useful Definitions

In microeconomics, social situations are modeled as formal games. In other words, they become optimization problems that can be analyzed mathematically. The following definitions (Osborne and Rubinstein [1994]) are necessary in order to understand the models of this paper.

Definition 1 *A (finite)* ***extensive-form game*** *is a tuple* $< \mathcal{N}, \mathcal{H}, P, f_c, (\mathcal{I}_i)_{i \in N} >$ *where:*

- $\mathcal{N} := \{i \in 1 \div n\} \cup \{c\}$ *is the (final)* ***set of players****.* $i \in N$ $(N := \mathcal{N} \setminus \{c\})$ *represent the* n *human players in the game and player* c *is called "chance".*

- $\mathcal{H}$ *is the (finite)* ***set of histories****. It satisfies the following properties:*

  - $\emptyset \in \mathcal{H}$.
  - $(a^k)_{k=1,...,K} \in \mathcal{H} \Rightarrow (a^k)_{k=1,...,L} \in \mathcal{H}$, $\forall K, L \in \mathbf{N}$, $K > L$.

  Each element in $\mathcal{H}$, $h := (a^k)_{k=1,...,K} \in \mathcal{H}$, *is a* ***history****.*[1]
  We define the ***set of final histories*** *as* $\mathcal{Z} := \{(a^k)_{k=1,...,K} \in \mathcal{H} \mid \nexists\ a^{K+1}\ s.t.\ (a^k)_{k=1,...,K+1} \in \mathcal{H}\} \in \mathcal{H}$.
  The ***set of actions*** *available after a non-terminal history* $h \in \mathcal{H} \setminus \mathcal{Z}$ *is defined as* $A(h) = \{a | (h, a) \in \mathcal{H}\}$.

- $P : \mathcal{H} \setminus \mathcal{Z} \longrightarrow \mathcal{N}$ *is the* ***player function****. It indicates which player or players play at each step of the game.*

- $f_c : \{h \in \mathcal{H} \mid P(h) = c\} \longrightarrow \mathcal{F}_c$ *where* $\mathcal{F}_c$ *is the family of density functions that take as an input both the history* h *after which* c *plays and an element in* $A(h)$[2]*.*
  At the beginning of the game, c *determines the* ***types*** *of the players* $\theta_i \in \Theta_i$

[1] The general definition includes the possibility of K being infinite, but we do not need it for this thesis.

[2] Given $h \in \mathcal{H}$ where $P(h) = c$, $f_c(\cdot|h) \in \mathcal{F}_c$ is a function $f_c(\cdot|h) : A(h) \longrightarrow \mathbf{R}^+ \cup \{0\}$ that is Lebesgue-measurable and satisfies $\int_{A(h)} f_c(\cdot|h) = 1$.

$\forall i \in N$. *Then, each player keeps her type private and uses observations of previous moves to update her belief system about other players types.*

- *For each player $i \in N$, $\mathcal{I}_i$ is the* ***information partition*** *of player i and a partition of $H_i := \{h \in \mathcal{H} \mid P(h) = i\}$. For each $h \in H_i$, the set $I_i(h) = \{h' \in H_i \mid A(h) = A(h')\}$ is an* ***information set*** *of player i and $I_i \in \mathcal{I}_i$.*

Definition 2 *A* ***utility function*** *(or payoff function) is a function $u_i : \mathcal{Z} \longrightarrow \mathbf{R}$ defined for each player $i \in N$ that represents her preferences over the different terminal histories $h \in \mathcal{Z}$. It contains both monetary and non-monetary terms, corresponds to the objective function of the optimization problem, and is assumed to be concave.*

Notation: An extensive game $< \mathcal{N}, \mathcal{H}, P, f_c, (\mathcal{I}_i)_{i \in N} >$ can also be written as $< \mathcal{N}, \mathcal{H}, P, f_c, (\mathcal{I}_i)_{i \in N}, (u_i)_{i \in N} >$ to include the preference system of the players over the different outcomes of the game.

Definition 3 *In an extensive-form game $< \mathcal{N}, \mathcal{H}, P, f_c, (\mathcal{I}_i)_{i \in N} >$, a* ***pure strategy*** *of player $i \in N$ is a function $\sigma_i : H_i \longrightarrow A(h)$ that assigns an action of $A(h)$ for each $h \in H_i$, such that $h' \in I_i(h) \Rightarrow \sigma_i(h) = \sigma_i(h')$. Equivalently, $\sigma_i : \mathcal{I}_i \longrightarrow A(I_i)$ assigns an action in $A(I_i)$ to each $I_i \in \mathcal{I}_i$.*

Definition 4 *In an extensive-form game $< \mathcal{N}, \mathcal{H}, P, f_c, (\mathcal{I}_i)_{i \in N}, (u_i)_{i \in N} >$, a* ***behavioral strategy*** *of player $i \in N$ is a collection of independent distributions $S_i := (\sigma_i(I_i))_{I_i \in \mathcal{I}_i}$ where each $\sigma_i(I_i)$ is a density function with support $A(I_i)$. We denote a behavioral strategy $\sigma_i : \mathcal{I}_i \longrightarrow \{\sigma_i(I_i) \mid I_i \in \mathcal{I}_i\}$ and consider a pure strategy a particular case of the former. The profiles of behavioral strategies are n-tuples and elements of $S := \Pi_{i \in N} S_i$.*

Definition 5 *A* ***belief system*** *in an extensive game $< \mathcal{N}, \mathcal{H}, P, f_c, (\mathcal{I}_i)_{i \in N} >$ is $\beta : \mathcal{I} \times \mathcal{H} \longrightarrow [0, 1]$, a function that assigns to every pair of information set and history a probability of that history being reached when the corresponding information set is reached. Notice that $h \notin I \Rightarrow \beta(I)(h) = 0$. Moreover, beliefs do not depend on payoffs or equilibrium strategies and are assumed to be common to all participants with the same information. This belief system constitutes the* ***prior*** *when applying Bayes' rule.*

Definition 6 *If we define as $O : S \longrightarrow \mathcal{Z}$ the outcome of a profile of strategies $\sigma \in S$, a* ***Nash equilibrium*** *is a profile of strategies $\sigma^* \in S$ such that $\forall i \in N$, $u_i(O(\sigma_i^*, \sigma_{-i}^*)) \geq u_i(O(\sigma_i, \sigma_{-i}^*))$ for every strategy σ_i of player i.*

Definition 7 *A **perfect Bayesian equilibrium** of an extensive-form game* $< \mathcal{N}, \mathcal{H}, P, f_c, (\mathcal{I}_i)_{i\in N}, (u_i)_{i\in N} >$ *is a profile of strategies and belief systems* $\sigma, \beta = (\sigma_i(\theta_i))_{i\in N}, (\beta_i(h))_{i\in N}$ *that assigns to each player a strategy and beliefs about the types of other players for every observable history, such that:*

- *Given the beliefs systems at every* $h \in \mathcal{H} \setminus \mathcal{Z}$*, strategies are optimal for each type (sequential rationality).*
- *Initial beliefs are correct.*
- *Beliefs are only determined by actions.*
- *Beliefs are updated following Bayes' rule*[3]*.*

The next concept, *Quantal Response Equilibrium*, was introduced by McKelvey and Palfrey [1998]. It is necessary to introduce or adapt some concepts before getting to the definition of such equilibrium.

Adjustment 1: The player function is now redefined to $P : \mathcal{H} \setminus \mathcal{Z} \longrightarrow \mathcal{N} \times \mathbf{Z}^+$ and is injective.

Adjustment 2: Each $h \in \mathcal{H}$ can be rewritten as h_i^j, where $(i, j) = P(h)$ (using the new redefinition of P).

Definition 8 *The **realization probability** of a history given a profile of behavior strategies* $\rho : \mathcal{H} \times S \longrightarrow \mathbf{R}^+$ *is a strictly positive value* $\rho(h|\sigma)$ *that is attributed to a history of the game assuming that the players' choice of behavioral strategies is the n-tuple* σ*. In addition, we define the **conditional realization probability** of* h' *(with* $h' \in \mathcal{H}$ *and* $h \in h'$*) conditioned to* $h \in \mathcal{H}$ *being reached and* $\sigma \in S$ *being adopted as* $\rho(h'|h, \sigma)$*. This conditional realization probability satisfies* $\rho(h'|h, \sigma)\rho(h|\sigma) = \rho(h'|\sigma)$*.*

Adjustment 3: We extend the definition of **utility function** for player i so that it can take a strategy as an input. Therefore, we define $u_i : S^\circ \longrightarrow \mathbf{R}$, with S° the interior of S, as

$$u_i(\sigma) := \sum_{h\in\mathcal{Z}} \rho(h|\sigma)u_i(h). \tag{1}$$

[3]Bayes' rule updates the beliefs that used to be only based on priors with the new observations, i.e. $\beta_i(h, a)(\theta_i') = \frac{\sigma_i(\theta_i')(h)(a_i)\cdot\beta_i(h)(\theta_i')}{\sum_{\theta_i\in\Theta_i}\sigma_i(\theta_i)(h)(a_i)\cdot\beta_i(h)(\theta_i)}$.

Adjustment 4: In the same line, we can define the **conditional utility function** $u_i : S^\circ \longrightarrow \mathbf{R}$ for any $(h, i, \sigma) \in \mathcal{H} \times N \times S$ as

$$u_i(\sigma|h) := \sum_{\{h' \in \mathcal{Z}: h \in h'\}} \rho(h'|h, \sigma) u_i(h'). \tag{2}$$

Definition 9 *Let K be $K := \sum_{i \in N} \sum_{h \in H_i} |A(h)|$ the number of possible expected payoffs and $X := \mathbf{R}^K$ the space of possible expected payoffs. Given a profile of strategies $\sigma \in S$, we define the **profile of expected payoffs** as $\bar{u} : S^\circ \longrightarrow X$ by*

$$\bar{u}(\sigma) := (\bar{u}_1(\sigma), \ldots, \bar{u}_n(\sigma)) \tag{3}$$

where

$$\bar{u}_{i,j,a}(\sigma) \equiv u_i(a, \sigma|h_i^j) \equiv u_i(\sigma|(h_i^j, a)). \tag{4}$$

In this model, instead of maximizing the profile of expected payoffs, the players choose their strategies seeking to optimize a function

$$\hat{u}_{i,j,a}(\sigma) = \bar{u}_{i,j,a}(\sigma) + \varepsilon_{i,j,a} \tag{5}$$

where $\varepsilon_{i,j,a}$ represents the **payoff disturbance** for player i at the history $h_i^j \in H_i$ if she chooses $a \in A(h_i^j)$. On the one hand, ε is presumed to be private information to each player. On the other hand, this payoff disturbances $\varepsilon_{i,j,a}$ are assumed to be statistically independent and to exist $\forall i \in N,\ h_i^j \in H_i,\ a \in A(h_i^j)$.

Definition 10 *We define the **set of improving disturbances** for player i at history $h_i^j \in H_i$ for an action $a \in A(h_i^j)$ and a profile of expected payoffs $\bar{u}$ as*

$$R_{i,j,a}(\bar{u}) := \{\varepsilon \mid \bar{u}_{i,j,a} + \varepsilon_{i,j,a} \geq \bar{u}_{i,j,\hat{a}} + \varepsilon_{i,j,\hat{a}}\ \forall \hat{a} \in A(h_i^j)\}. \tag{6}$$

Definition 11 *The **probability of improvement** of $\hat{u}_{i,j,a}$ for player i when adopting action $a \in A(h_i^j)$ at history h_i^j is defined as*

$$\sigma_{i,j,a}(\bar{u}) = \int_{r_{i,j,a}(\bar{u})} f(\varepsilon) d\varepsilon \tag{7}$$

*where f is the density function of the distribution of ε. This density function is required to be **admissible**, i.e. to satisfy that:*

- *ε is an absolutely continuous random vector with respect to Lebesgue measure, with $f(\varepsilon)$ the density function of its joint distribution.*
- *$\varepsilon_{i,j}$ are statistically independent.*
- *the expected value of $\varepsilon_{i,j,a}$ exists for all $i, j, a \in A(h_i^j)$.*

In this model, each agent ij of a player i chooses the maximum of $\hat{u}_{i,j,a}$ at each information set h_i^j and acts independently of the other agents of the same player. Thus, this type of quantal response equilibrium that we use is called **Agent Quantal Response Equilibrium** (**AQRE**).

Definition 12 *For any extensive form game $G =< \mathcal{N}, \mathcal{H}, P, f_c, (\mathcal{I}_i)_{i\in N}, (u_i)_{i\in N} >$ and an error structure $f(\varepsilon)$, a behavioral strategy $\sigma^* \in S$ is an **AQRE** if it is a fixed point of $\sigma \circ \bar{u}$. It is a vector $\sigma^* \in S^\circ$ such that $\forall i \in N$, $1 \leq j \leq J_i$, $a \in A(h_i^j)$, $\sigma^*_{i,j,a} = \sigma_{i,j,a}(\bar{u}(\sigma^*))$.*

In the original *Quantal Response Equilibria for Extensive Form Games* paper, McKelvey and Palfrey [1998] prove that an AQRE exists for any admissible f.

2 Experimental Model

The theoretical model in Martinelli [2022] cannot be brought to the lab as it is. On the one hand, it is not possible to study that many variables simultaneously. On the other hand, the subjects that will take part in the experiment are ideally a random sample of the society, and we cannot expect everyone to understand a complicated model. The goal is, thus, to make it accessible to everyone.

Therefore, the model needs simplifying. A limited amount of variables can be controlled as control inputs and the rest of them need to be fixed to semi-arbitrary values. In this case, the setting is a 2×2 treatment, with two control variables that take two different values each.

The experimental model is built over groups of 5 players. Two of them have the role of *Public Official* and the other three are *Regular Citizens*. We change the name from *Politician* to *Public Official*, because it could lead to strong presumptions. Nevertheless, in Barr and Serra [2009], it is proved that framing[4] does not have an effect in bribery experiments. The decision of using three *Regular Citizens* instead of a single one has two reasons. The first one is to introduce noise in the voting process, by dealing with three independent[5] votes instead of just one. The second one is that the distribution of payoffs with the desired properties is more simple when three voters instead of one are taken into account. Finally, the decision of using an odd number of voters is obviously made so that there are no ties in the results.

This simplified model maintains the two-period structure of the original one. These periods will be broken into parts in the Subsections 2.1 and 2.2, and in Figure 1.

In Section 2.4, the 2×2 treatment setup will be addressed. Finally, Section 2.5 contains an analysis of the players' incentives and a prediction of their behavior.

2.1 First Period

There are two main stages in the first period: the bribery stage and the voting stage.

At the beginning of the period (and the round), we consider that the two *Public*

[4] *framing*: a particular way of presenting information in behavioral economics setups that could lead to inducing certain behavior in the participants. Source: `https://www.behavioraleconomics.com/resources/mini-encyclopedia-of-be/framing-effect/`.

[5] As will be seen in Section 3, the players are not allowed to communicate with each other, so that their decisions are made independently.

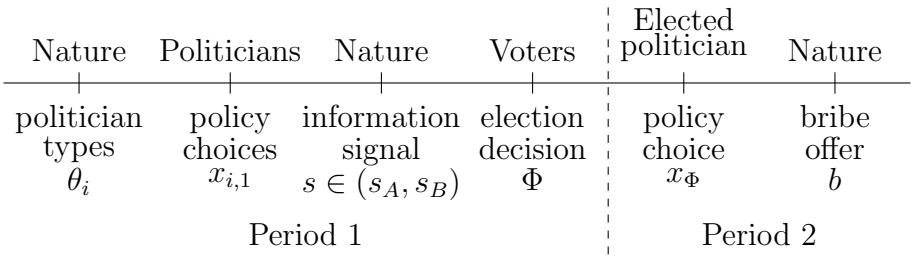


Figure 1: Timeline for the experimental model.

Officials already have a relevant role in the society, just as in the theoretical model. Thus, them taking a bribe results in a negative impact on the *Regular Citizens*' lives. These effects are reflected on the players' payoffs (see Section 2.1.1).

Just like in the theoretical model, each *Public Official* is offered a bribe by a third party. This bribe increases their personal payoff but reduces the *Regular Citizens*'. The more *Public Officials* take the bribe, the more the *Regular Citizens*' payoff is reduced.

Once the *Public Officials* have made their decision about whether to take the bribe that has been offered to them or not, there is an information signal about it. With probability p_{info}, all the members of the group are informed about whether each *Public Official* was corrupt or not. The information is either displayed truthfully for all the members of the group or is not presented to anyone. Also, the signal contains the information about both of the *Public Officials* or neither of them.

After receiving the signal, the *Regular Citizens* are required (and obliged) to vote for one *Public Official* to be elected for the next period. This works like in the theoretical model, except that the voters are not considered as one single player. The *Public Official* that obtains the most votes is the one elected for the second period.

At the end of the period, all the players are perfectly informed about which *Public Official* will be in office in the next period.

2.1.1 First Period Payoff

The basic payoff if both of the *Public Officials* are not corrupt is $25pts$. Thus, under this circumstances, everyone in the group receives the same payoff: $25pts$.

The *Public Officials* are offered to raise their payoff from $25pts$ to $45pts$ if they take the bribe. If only one of them accepts it, each of the three *Regular Citizens*' payoff is reduced from $25pts$ to $15pts$, and the other *Public Official* remains with the basic $25pts$. When both of the *Public Officials* are corrupt, each of them receives $45pts$ and the *Regular Citizens* obtain $5pts$ each (see Section 2.5 for more).

First period payoff (in *pts*)	$x_A = 0$ $x_B = 0$	$x_A = 1$ $x_B = 0$	$x_A = 0$ $x_B = 1$	$x_A = 1$ $x_B = 1$
Public Official A	25	45	25	45
Public Official B	25	25	45	45
Regular Citizen	25	15	15	5

Table 1: Payoff for the first period in points. B_i represents whether Public Official i took the bribe or not, for $i \in \{A, B\}$.

Table 1 summarizes the payoff system for the first period. Each column represents one of the four possible scenarios for the first period and each row is the respective payoff of a player with the indicated role in the corresponding scenario.

It is necessary to mention that the information about the first period is not displayed until after the *Regular Citizens* have submitted their vote. That way, there is no leak of the information of the bribery activity that affects the decision in case it is not made public.

2.1.2 First Period Utility Functions

The utility functions for the first period for both the *Public Officials* and *Regular Citizens* can be expressed as a function of the *Public Officials*' decisions. The suggested equations are the following:

$$u_{1,PO}(\theta, x) = r + x \cdot (B - \theta) \tag{8}$$

$$u_{1,RC}(x_A, x_B) = r - (x_A + x_B) \cdot p \tag{9}$$

where:

- $u_{1,PO}$: utility function of a *Public Official* for the first period of the game.
- $u_{1,RC}$: utility function of a *Regular Citizen* for the first period of the game.
- x: takes value 1 if the *Public Official* whose utility function is being computed takes the bribe and 0 otherwise.
- x_i: takes value 1 if *Public Official* i takes the bribe and 0 otherwise, for $i \in \{A, B\}$. These variables are used exclusively for the utility functions of the *Regular Citizens*.
- r: basic payoff (reward) for the period.
- B: monetary raise of the *Public Official*'s payoff when they take a bribe.

- θ: personal regret of the *Public Official* for taking a bribe (and its consequences).
- p: quantity by which the *Regular Citizens*' payoff is reduced for every *Public Official* that takes a bribe, in absolute value.

In our implementation of the model, r, B, and p take the following values:

$$r = 25pts \qquad b = 20pts \qquad p = 10pts$$

When it comes to θ, the value of the personal regret for taking a bribe, we cannot assume it to have any particular value. If the game only consisted of one period, we could deduce that $\theta > B$ for those *Public Officials* that do not take the bribe ($x = 0$). Analogously, we would say that $\theta < B$ for those who take it ($x = 1$). However, as we will see in Section 2.5, this is not the utility function that the players aim to maximize, so further information needs to be considered before making the last assumption. In other words, taking or not taking the bribe will no longer be an indicator of whether $\theta > B$ or $\theta < B$. We can ignore the case where $\theta = B$, and assume that it is randomized to 50% $x = 1$ and 50% $x = 0$.

Conversely, x, x_A, and x_B are observed variables. Just as explained in the variables' definition, x is either x_A or x_B, depending on which *Public Official*'s utility function is being computed, so it is simply an abuse of notation. We will use the value of these variables to study the behavior of the subjects. They take values of 1 or 0, being the first the action of accepting the bribe and the latter, rejecting it.

The utility of the *Public Officials* (eq. 8) consists of two components. The first one, r, represents the fix reward and is a constant. It is what the player would get if she decides to reject the bribe. The variable reward, $x \cdot (B - \theta)$, consists of the *Public Official*'s strategy (x) and the cost of making the decision to take the bribe ($B - \theta$). It is the amount by which she could increase her fix reward if she chose to take the bribe ($x = 1$). It is important to take into account not only the monetary payoff (B), but also the personal regret of the *Public Official* (θ) in the utility function.

The *Regular Citizens*' utility (eq. 9) depends only in the actions of the *Public Officials*. Again, the first term, r, is the fixed reward that each of the *Regular Citizens* would get if their *Public Officials* were honest. On the contrary, the more corruption the *Public Officials* are in, the more the variable term, $(x_A + x_B) \cdot p$, increases and the more the *Regular Citizens*' utility decreases.

2.2 Second Period

The second period of the game starts once everyone has been informed about which *Public Official* has been elected. Henceforth, this player will be addressed as *Elected Public Official* and the remaining *Public Official* as *Non-elected Public Official*.

The *Non-elected Public Official* receives a small bonus (see Section 2.2.1 for payoff details) and is alien to the game until the next round. Thus, we can consider that the round is over for this player.

Just like in the theoretical model, the *Elected Public Official* is offered a bribe again. The difference is that, in this experimental model, the bribe is offered with probability p_{bribe}. Before knowing if the bribe is being offered or not, the *Elected Public Official* makes a decision about whether she is willing to take it or not, in case it is offered.

In this period, the effect of the bribery activity on the players' payoffs is only seen when the *Elected Public Official* is willing to engage in corruption endeavors and is actually offered a bribe. In this case, the payoff of the *Elected Public Official* is given a raise and the payoff of the *Regular Citizens* is reduced. Under any other circumstances, the payoff associated to the second period does not undergo any changes.

2.2.1 Second Period Payoff

The basic payoff for this period is again $25pts$. Nevertheless, the *Non-elected Public Official* receives $5pts$, so that there is an intrinsic bonus of $20pts$ for the *Public Official* that gets elected. Thus, when the *Elected Public Official* chooses not to engage in corrupt activities, all three *Regular Citizens* and the *Elected Public Official* receive a payoff of $25pts$.

On the other hand, when the *Elected Public Official* is willing to take the bribe, there is a p_{bribe} probability that her payoff raises from $25pts$ to $45pts$. If this happens, analogously to the first period, the *Regular Citizens*' payoff is reduced from $25pts$ to $15pts$. In the remaining of cases, i.e. when the *Elected Public Official* is willing to take part in corruption but is not offered a bribe (which happens with probability $1 - p_{bribe}$), the payoff of the *Elected Public Official* and *Regular Citizens* stays at $25pts$.

Table 2 summarizes the payoff system for the second period. Each column represents one of the two possible scenarios for the second period and each row is the respective payoff of a player with the indicated role in the corresponding scenario.

Second period payoff	$x_\phi = 0$ $b = 0$	$x_\phi = 0$ $b = 1$	$x_\phi = 1$ $b = 0$	$x_\phi = 1$ $b = 1$
Elected Public Official	$25pts$	$25pts$	$25pts$	$45pts$
Non-elected Public Official	$5pts$	$5pts$	$5pts$	$5pts$
Regular Citizen	$25pts$	$25pts$	$25pts$	$15pts$

Table 2: Payoff for the second period in points. x_ϕ represents the willingness of the Elected Public Official to take the bribe. b indicates whether the bribe was offered or not.

2.2.2 Second Period Utility Functions

Before studying the utility functions, it is necessary to address the elections that took place at the end of the first period. Let $v_1, v_2, v_3 \in \{A, B\}$ the votes of the three *Regular Citizens* that take values in $\{A, B\}$ depending on which *Public Official* they vote for. Let $\phi : \{A, B\}^3 \longrightarrow \{A, B\}$ a function that determines the winner of the election. Thus, henceforth, the sub-index ϕ indicates the *Elected Public Official*.

When it comes to the utility functions for the second period, they depend on both the results of the election and the bribery activity of the *Elected Public Official*. Indeed, they could be expressed as:

$$u_{2,PO_i}(\theta, \phi(v_1, v_2, v_3), x_\phi) = r - (1 - \delta_{i\phi}) \cdot \tilde{p} + \delta_{i\phi} \cdot x_\phi \cdot b \cdot (B - \theta) \tag{10}$$

$$u_{2,RC}(x_\phi) = r - x_\phi \cdot b \cdot p \tag{11}$$

where:

- u_{2,PO_i}: utility function of *Public Official* i for the second period of the game. $i \in \{A, B\}$
- $u_{2,RC}$: utility function of a *Regular Citizen* for the second period of the game.
- $\phi(v_1, v_2, v_3)$: function determining the winner of the election. Takes values in $\{A, B\}$. v_1, v_2, and v_3 also take values in $\{A, B\}$ and represent the votes of each of the *Regular Citizens*.
- x_ϕ: takes value 1 if the *Elected Public Official* chooses to take the bribe and 0 otherwise.
- r: basic payoff (reward) for the period.
- δ_{ij}: Kronecker delta function. $\delta_{ij} = 1 \Leftrightarrow i = j$ and $\delta_{ij} = 0$ otherwise.

- $\tilde{p}$: absolute value of the difference between the basic reward for the second period and the reward that the *Non-elected Public Official* receives.
- $b \sim Bern(p_{bribe})$: random variable following a Bernoulli distribution of probability p_{bribe}, indicating whether the *Elected Public Official* was offered the bribe or not.
- B: monetary raise of the *Elected Public Official*'s payoff when the bribe is offered and she has chosen to take it.
- θ: personal regret of the *Elected Public Official* for taking a bribe.
- p: quantity by which the *Regular Citizens*' payoff is reduced if the *Elected Public Official* takes the bribe, in absolute value.

In our implementation of the model, r, $\tilde{p}$, B, and p take the following values:

$$r = 25pts \qquad \tilde{p} = 20pts \qquad b = 20pts \qquad p = 10pts$$

In this period, the input variables for the utility functions that are determined by the subjects are v_1, v_2, v_3, and x_ϕ. Indeed, they make decisions about who they want to vote for or whether they want to take a bribe when they are elected. Notice that v_1, v_2 and v_3 are decisions made after receiving information with probability p_{info} about the results of the first period bribery (details about p_{info} in Section 2.4).

b is also an input variable, but it is determined randomly, following the distribution of a Bernoulli of probability p_{bribe} (see Section 2.4 for further comments on p_{bribe}). Finally, the analysis of θ can be realized analogously to the one in Section 2.1.2.

The utility of the *Public Officials* (eq. 10) has three terms in this period. The first one, r, is the fix reward that the *Public Official* would obtain if she was elected and decided to not engage in corrupt actions. The second term, $(1-\delta_{i\phi})\cdot\tilde{p}$, represents the penalization of not being elected. It could be split in two factors: $(1 - \delta_{i\phi}$ takes the value 1 when the *Public Official* is not elected and 0 when it is, and $\tilde{p}$ is the monetary penalization for not being elected and not participating in the round anymore. In our model, we do not consider a factor of shame or regret for not being elected. The last term, $\delta_{i\phi} \cdot x_\phi \cdot b \cdot (B-\theta)$, has four different factors. $\delta_{i\phi}$ takes a value of 1 if the *Public Official* was elected and 0 otherwise, x_ϕ represents the strategy of the *Elected Public Official* regarding taking the bribe (1) or not (0), and b indicates whether the bribe was offered (1) or not (0). Thus, the last factor, $(b - \theta)$, will not be accounted for the *Public Official*'s utility unless the three other factors are equal

and take a value of 1. This last factor represents the monetary benefit of taking the bribe (B) minus the personal regret of engaging in corrupt activities (θ).

Finally, the *Regular Citizens*' utilities (eq. 11) is more simple than the *Public Officials*'. It consists of two terms. The first term, r, is again the winnings of the *Public Officials* when the *Elected Public Official* rules non-corruptly or is not offered a bribe. The other one, $x_\phi \cdot b \cdot p$, is the penalty imposed over the *Regular Citizens*' payoff when the *Elected Public Official* accepts the bribe ($x_\phi = 1$) and is offered one ($b = 1$).

2.3 Game Tree

See next page.

2.4 2×2 Treatments

This experiment has a 2×2 design. Indeed, the variables p_{info} (probability that the information about the *Public Officials*' engagement in corruption in the first period is made public) and p_{bribe} (probability that a bribe is offered to the *Elected Public Official* in the second period) both take values in $\{0.5, 1\}$.

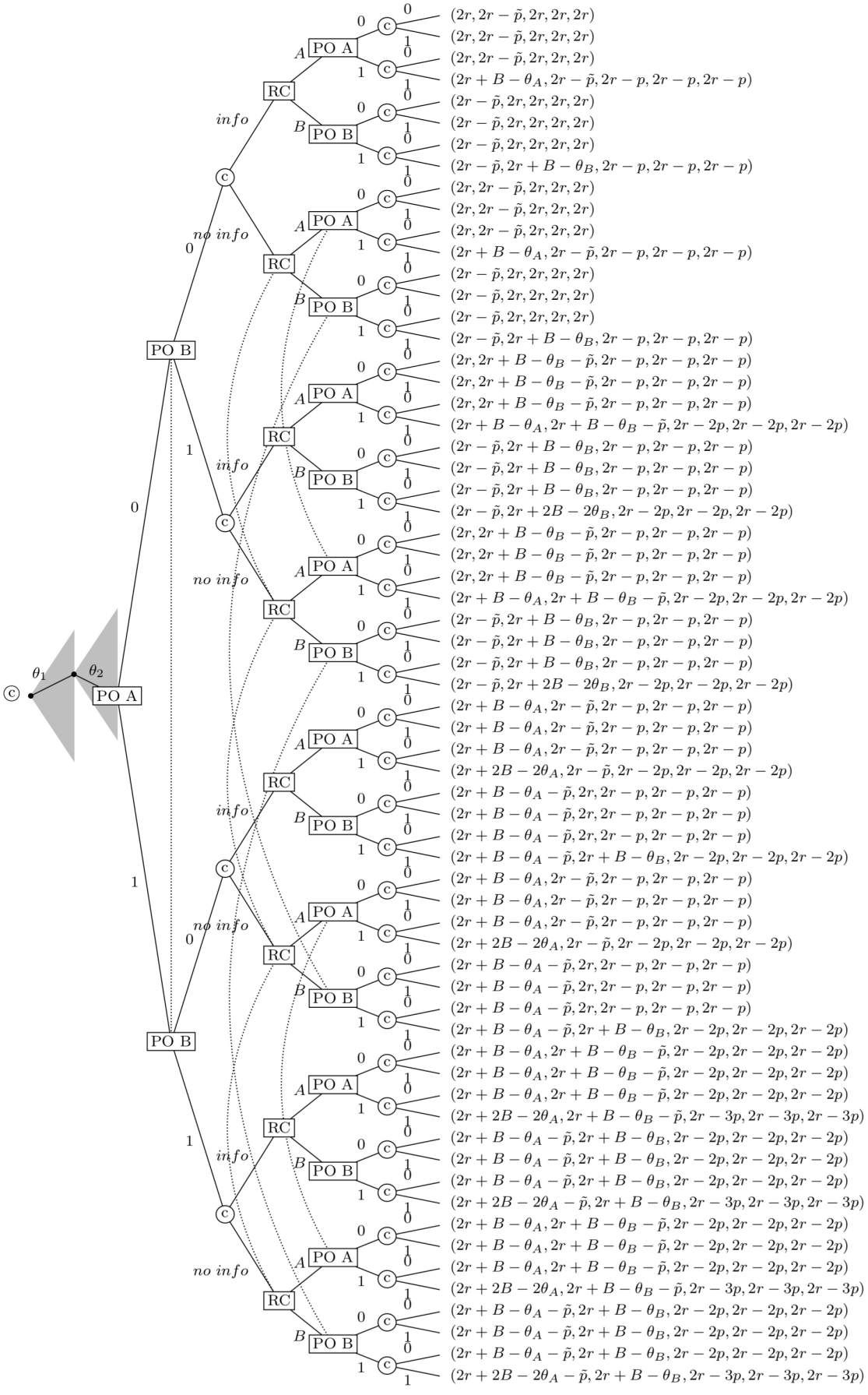
$$(p_{info}, p_{bribe}) \in \{0.5, 1\}^2$$

One of the main goals of this experiment is to study whether these different treatments induce a change in the subjects' behavior.

The names we give to the treatments and their abbreviations are summarized in Table 3.

Treatments	$p_{info} = 1$	$p_{info} = 0.5$
$p_{bribe} = 1$	Perfect Information 100% Bribe (PI100)	Imperfect Information 100% Bribe (II100)
$p_{bribe} = 05$	Perfect Information 50% Bribe (PI50)	Imperfect Information 50% Bribe (II50)

Table 3: Names (and abbreviations) of the 2×2 treatments of the game.


c
θ1
θ2
PO A
0
1
PO B
0
1
PO B
c
no info
info
RC
A
B
PO A
PO B
c
0
1
(2r, 2r − p̃, 2r, 2r, 2r)
(2r + B − θA, 2r − p̃, 2r − p, 2r − p, 2r − p)
(2r − p̃, 2r + B − θB, 2r − p, 2r − p, 2r − p)
(2r + B − θA, 2r + B − θB − p̃, 2r − 2p, 2r − 2p, 2r − 2p)
(2r − p̃, 2r + 2B − 2θB, 2r − 2p, 2r − 2p, 2r − 2p)
(2r + 2B − 2θA, 2r − p̃, 2r − 2p, 2r − 2p, 2r − 2p)
(2r + B − θA − p̃, 2r, 2r − p, 2r − p, 2r − p)
(2r + B − θA − p̃, 2r + B − θB, 2r − 2p, 2r − 2p, 2r − 2p)
(2r + 2B − 2θA, 2r + B − θB − p̃, 2r − 3p, 2r − 3p, 2r − 3p)
(2r + B − θA − p̃, 2r + B − θB, 2r − 2p, 2r − 2p, 2r − 2p)
(2r + 2B − 2θA − p̃, 2r + B − θB, 2r − 3p, 2r − 3p, 2r − 3p)

2.5 Analysis & Predictions

As explained in Section 1.2, the players of the game are expected to maximize their utility functions and base their strategy on that. For each of the roles, the total utility functions takes the following form:

$$u_{PO_i}(x, \theta, \phi, x_\phi, b) = r + x \cdot (B - \theta) + \delta[r - (1 - \delta_{i\phi}) \cdot \tilde{p} + \delta_{i\phi} \cdot x_\phi \cdot b \cdot (B - \theta)] \quad (12)$$

$$u_{RC}(x_A, x_B, x_\phi, b) = r - (x_A + x_B) \cdot p + \delta[r + x_\phi \cdot b \cdot p] \quad (13)$$

where $B, x_\phi, b, x_A, x_B \in \{0, 1\}$, $\Phi \in \{A, B\}$, and δ is the temporal discount. The description of the variables and the interpretation of the different terms can be found in Sections 2.1.2 and 2.2.2.

This is a *signaling game* where the information set for the *Public Official* i, $i \in \{A, B\}$, in the first period is:

$$\mathcal{I}_{1,PO_i} = \theta_i \in \mathbf{R}^+ \cup \{0\} \quad (14)$$

i.e. their type. Once the *Public Officials* decide the values of x_i, $i \in \{A, B\}$, the information set for the *Regular Citizens* is:

$$\mathcal{I}_{RC} = (s_A, s_B) \in \{0, 1\} \times \{0, 1\} \bigcup \emptyset$$

where $\mathcal{I}_{RC}$ is in $\{0, 1\} \times \{0, 1\}$ with probability p_{info} and in $\emptyset$ with probability $1 - p_{info}$. When $p_{info} < 1$ we will say that we are in an *Imperfect Information* treatment, contrary to the case $p_{info} = 1$ that has *Perfect Information*. Finally, the set of information for *Public Official* i, $i \in \{A, B\}$, in the second period includes their type, their choice of policy for the first period, the information (or absence of it) about the choice of information of the other *Public Official* in the first period, and the result of the election:

$$\mathcal{I}_{2,PO_i} = (\theta_i, x_i, s_{j \neq i}, \delta_{i,\Phi}) \in \left\{\mathbf{R}^+ \cup \{0\}\right\} \times \{0, 1\} \times \{\emptyset, 0, 1\} \times \{0, 1\}.$$

In this extensive form game, a mixed strategy for *Public Official* i is a pair of mappings $\sigma_i = (\sigma_{1,i}, \sigma_{2,i})$ such that:

$$\begin{array}{rccl} \sigma_{1,i}: & \mathcal{I}_{1,PO_i} & \longrightarrow & [0,1] \\ & \theta_i & \mapsto & x_i \equiv \sigma_{1,i}(\theta_i) \end{array} \qquad \begin{array}{rccl} \sigma_{2,i}: & \mathcal{I}_{2,PO_i} & \longrightarrow & [0,1] \bigcup \emptyset \\ & I_{2,i} & \mapsto & x_\Phi \equiv \sigma_{2,i}(I_{2,i}) \end{array} \quad (15)$$

where $I_{2,i} = (\theta_i, x_i, s_{j \neq i}, \delta_{i,\Phi})$. Of course, if $\delta_{i,\Phi} = 0$, $\sigma_2 \in \emptyset$, as the *Non-elected Public Official* has no saying in whether she wants to take the second bribe or not, as

she is never given the opportunity to do so. The strategies map to an interval $[0,1]$, as they represent the probability of choosing the pure strategy $x_i = 1$ (analogously $x_\Phi = 1$) over $x_i = 0$ (analogously $x_\Phi = 0$).

When it comes to the *Regular Citizens*, their pure strategies are which *Public Official* they are going to vote for to be elected. Thus, we define a *Regular Citizen*'s mixed strategy as a mapping from their information set to the probability of her voting for *Public Official A* (p_A) and assume that the probability of her voting for *Public Official B* is $p_B = 1 - p_A$. The mapping looks like:

$$\begin{array}{rccl} \nu: & \mathcal{I}_{RC} & \longrightarrow & [0,1]^2 \\ & (s_A, s_B) & \mapsto & (p_A, p_B) = (p_A, 1-p_A) \equiv \nu(s_A, s_B) \end{array} . \tag{16}$$

The *Public Officials* have beliefs about the state of the world that might condition their behavior. In our case, the state of the world is whether the information is going to be displayed or not and whether the bribe in the second period is going to be offered or not. Moreover, the *Regular Citizens* have beliefs about the *Public Officials*' types (θ_i, $i \in \{A, B\}$).

The belief system of the *Public Officials* is defined as follows:

$$\begin{array}{rccl} \beta_{PO_i,SW}: & \mathcal{I}_{1,PO_i} & \longrightarrow & \mathcal{F} \times \mathcal{F} \\ & \theta_i & \mapsto & f(\theta_i)(\cdot,\cdot) \end{array} \tag{17}$$

were $\beta_{PO_i,SW}$ is the belief of *Public Official* i about the state of the world, and $\mathcal{F} := \{f : [\mathbf{0},\mathbf{1}] \longrightarrow \mathbf{R}^+ \cup \{0\} \mid f \text{ is Lebesgue-measurable and } \int_{[0,1]} f = 1\}$ is the family of probabilistic density functions in $[0,1]$. Notice that we can restrict $f(\theta_i)(x_1, \cdot)$ after the *Public Officials* discover whether the information has been displayed. The two dimensions of f are probabilistically independent if the subject is rational.

When it comes to *Regular Citizens*, it needs to be mentioned that they do not need to build a belief system for what they expect the other *Regular Citizens* to vote. The reason is that there are only two candidates for the election, so it voting for a *Public Official* that is not their own preference does not improve the outcome (given their beliefs). Similarly, they do not need a belief system for the state of the world, because the display of information takes place before they perform any action and the possibility that the bribe is offered or not does not affect their vote. The reason of this last statement is that we assume them to maximize their utility function, so they want to vote for the *Public Official* that they believe will not choose to take the bribe, no matter if it is offered or not.

The belief system that we define for the *Regular Citizens* is the following:

$$\begin{array}{rccc} \beta_{RC,PO_i}: & \mathcal{I}_{RC} & \longrightarrow & \mathcal{G} \\ & (s_A, s_B) & \mapsto & (\hat{\theta}_i(s_A, s_B))(\cdot) \end{array} \tag{18}$$

were β_{RC,PO_i} is the belief of a *Regular Citizen* about the type of *Public Official* i, and $\mathcal{G} := \{g : \mathbf{R} \longrightarrow \mathbf{R}^+ \cup \{0\} \mid g \text{ is Lebesgue-measurable and } \int_{\mathbf{R}} g = 1\}$ is the family of probabilistic density functions in the real numbers. Thus, depending on the information that they have about each *Public Official*, they will form some beliefs about their type, i.e. their personal regret for taking the bribe.

2.5.1 Equilibria

For this problem, we analyse two types of equilibria: Bayesian equilibrium and Quantal Response equilibrium. The first theory is a classic theory introduced by John C. Harsany in the late 1960s and the second one was developed by Richard D. McKelvey and Thomas R. Palfrey in the 1990s. As explained in Section 1.2, the second notion of equilibrium is based on the first one, but includes a factor we call "*attention*" in the picture, that makes them choose an action that does not correspond to their best strategy with a certain probability.

In this section we will calculate both of the equilibria for our game. By doing so, we will try to predict what behavior the real players will have in the laboratory and see which theory explains the results more accurately.

The technique we use to calculate the equilibria is the establishment of a cutoff strategy type. We can do this thanks to the fact that we assume a continuous distribution of types in $\mathbf{R}$ and the fact that all the actions that the players can make are discrete. Thus, we predict that the *Public Officials* will take an action or another depending on whether their type is higher or lower than the cutoff. We find that we do not need to assume any particular distribution of the types. Instead, the cutoff will only depend on the probability that the other *Public Official* takes a bribe in the first period.

Bayesian Equilibrium

Given a Bayesian game, we define a *Perfect Bayesian Equilibrium* as a profile of strategies $(\sigma_A, \sigma_B, \nu_1, \nu_2, \nu_3)$, and a belief system $(\beta_{PO_A}, \beta_{PO_B}, \beta_{RC_1}, \beta_{RC_2}, \beta_{RC_3})$ for *Public Officials* and *Regular Citizens* such that:

a) The *Regular Citizens*' strategy ν is optimal given their beliefs β.

b) The *Regular Citizens*' beliefs are consistent with the *Regular Citizens*' strategies. The belief about each *Public Official* is derived from the prior beliefs about her and her strategy, using Bayes' rule to update the belief after observing signals of her actions.

c) The *Public Officials*' strategies σ_i, $i \in \{A, B\}$ are optimal given the other *Public Official*'s strategy and the *Regular Citizens*' strategies.

As mentioned at the beginning of this section, we will define the best strategies using cutoff values for the *Public Officials*. For the second period, the *Elected Public Official* only worries about maximizing the payoff function in Equation 10. Thus, the *cutoff* value for this action, $\sigma_2^*(\theta)$ is the value of the bribe, B.

$$\sigma_2^*(\theta) = \begin{cases} 1 & \text{if } \theta < B \\ 0 & \text{if } \theta \geq B \end{cases} . \tag{19}$$

Continuing by backward induction, we determine the best strategy for the *Regular Citizens* when they vote. Given that there is only one type of voters and that they all receive the same information about the game, we can consider that all the *Regular Citizens* have the same best strategy $\nu^*(s_A, s_B)$ for voting. Indeed, they seek to elect a *Public Official* that will not take the bribe in the second period. Thus, they intend to vote for a *Public Official* with a type higher than B ($\theta_\Phi > B$). As we will see in the following paragraph, the cutoff type for the first period is lower than B, so knowing that a *Public Official* has not taken the bribe in the first period $s_i = 0$ does not ensure that $\theta_i > B$. However, if θ_i is not higher than the cutoff for the first period, it will definitely not be higher than B. Thus, the best strategy for the *Regular Citizens* is to vote for the *Public Official* with $s_i = 1$, if there is a *Public Official* with $s_{-i} = 0$. In the cases where $s_A = s_B$, as the *Regular Citizens* do not possess any further information about the *Public Officials*' types, the best strategy is to vote for each of them with probability 50%.

$$\nu^*(s_A, s_B) = \begin{cases} (0.5, 0.5) & \text{if } s = \emptyset \text{ or } s_A = s_B \\ (1, 0) & \text{if } s_A < s_B \\ (0, 1) & \text{if } s_A > s_B \end{cases} . \tag{20}$$

Analogously to the second period case, there also exists a cutoff value $\bar{\theta}$ for the bribe-taking action in period 1. In this case, the *Public Officials* value tricking the *Regular Citizens* into thinking that they are honest politicians more than the value of the bribe. Therefore, the cutoff value correspond to that player that is indifferent between taking the first bribe and sending a bad signal, and waiting until the second

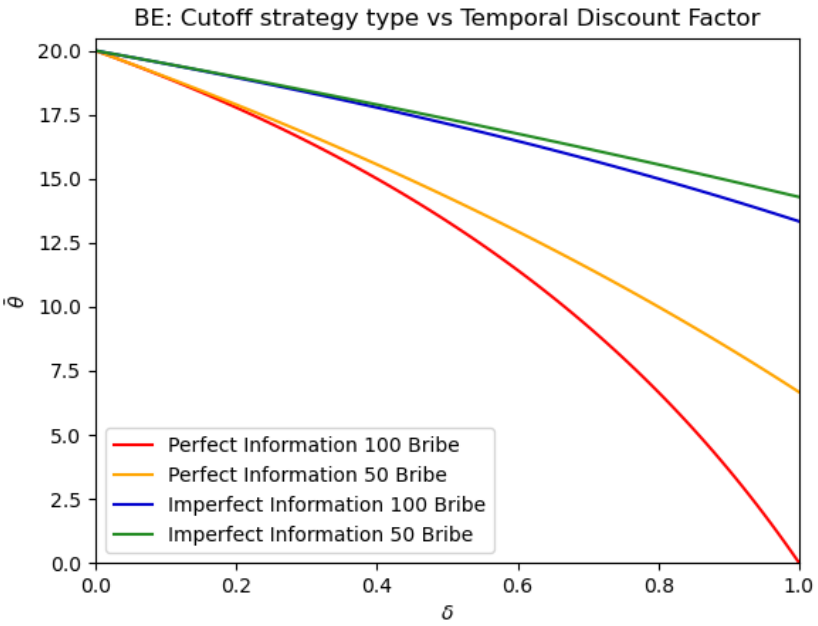


Figure 2: Bayesian Equilibrium, first period cutoff strategy type: dependence with the temporal discount factor.

period to take the bribe, increasing her possibility of election. Notice that $\bar{\theta} < B$, because all of them pretend to take the second bribe.

We compute $\bar{\theta}$ by equating $\mathbf{E}[u_{PO}(x = 1, \theta = \bar{\theta})] = \mathbf{E}[u_{PO}(x = 0, \theta = \bar{\theta})]$ (see Equation 12). The value of the cutoff type is

$$\bar{\theta} = B - \frac{\delta p_{info}}{2 - \delta p_{info} p_{bribe}} \tilde{p} \tag{21}$$

. It depends on the value of the bribe B, the punishment term for not being elected $\tilde{p}$, the probability that the information about the bribery is made public p_{info}, the probability that a bribe is offered in the second period p_{bribe}, and the temporal discount factor δ (see Figure 2).

Using this value, we can define the optimal action for the best strategy in the first period for the *Public Officials* as

$$\sigma_1^*(\theta) = \begin{cases} 1 & \text{if } \theta < \bar{\theta} \\ 0 & \text{if } \theta \geq \bar{\theta} \end{cases} . \tag{22}$$

Notice that the limit case where $\theta = B$ or $\theta = \bar{\theta}$ is included in the action that

does not take the bribe. In real life, players of such types should opt for each of the actions with a 50% probability, unless they have other factors into account. However, we will only be able to determine if $\theta > B$ or $\theta < B$ (analogously $\theta > \bar{\theta}$ or $\theta < \bar{\theta}$) from the observable measures, so we do not need to care about these cases.

One of the *other factors* that the *Public Officials* with a type $\theta = \bar{\theta}$ or $\theta = B$ could take into account is efficiency. Indeed, to mirror real life, taking a bribe is an inefficient action in terms of total payoff. This means that there are leakages of budget:

- In the first period, if the *Public Officials* do not take any bribe, the monetary reward distributed is $25 + 25 + 25 + 25 + 25 = 125pts$. However, for each *Public Official* that takes a bribe, there is a loss of $10pts$ in the budget. The worst-case scenario is when both of the *Public Officials* take the bribe, where the group reward is reduced to $45 + 45 + 5 + 5 + 5 = 105pts$.

- In the second period, the situation is analogous, except that there is only one *Public Official* that can cause this inefficiency. The decrease in the group payoff would then go from $25+5+25+25+25 = 105pts$ to $45+5+15+15+15 = 95pts$.

Agent Quantal Response Equilibrium

When the players of the game have imperfect perceptions of what is best for them, we can use AQRE. This equilibrium theory allows us to model a factor of *attention* to the game that indicates how properly players make decisions when choosing their actions.

Following the definitions in Section 1.2, we assume the error vector to have a distribution f_λ such that the probability of improvement of a strategy with that would be considered optimal in the Bayesian equilibrium is $\lambda \in [0, 1]$. As every action in the game has only two choices, the probability of improvement of a strategy with the action that would not belong to the best strategy in a Bayesian equilibrium is $1 - \lambda$. Thus, we call this factor λ the **attention factor** of the players.

Notice that this time no pure strategies can be the best strategy for the game. Indeed, unless $\lambda = 1$ or $\lambda = 0$, all we will get are mixed strategies. Notice that $\lambda = 1$ would correspond to the Bayesian equilibrium (perfect attention). Another significant case is $\lambda = 0.5$, where the game would be completely randomized and the types or preferences over the different outcomes of the game would not have any effect on the choice of actions of the players. Finally, a situation where $\lambda < 0.5$ would not be convenient for the players, since they would be choosing more often the *wrong* action instead of the *right* one.

The distribution function of the disturbance (ε) f_λ generates independent and identically distributed random variables for each player at each situation where they have to make a decision. Therefore, by the way it is defined, the distribution function is *admissible*, which means that an AQRE exists for the game[6].

Just like in any other extensive game, the same way we did in the Bayesian equilibrium section, we will break the game into subgames to solve the equilibrium, starting from the end. Therefore, we will analyze the second period bribe-taking action first, then the election, and finally the first period policy choice. This way, we will build the AQRE best strategy.

Keeping the cutoff technique, we can define two strategies over the bribe-taking action in the second period. Following the definition of f_λ, *Public Officials* with $\theta < B$ should want to take the bribe with probability λ and not take with probability $1-\lambda$. On the contrary, the *Public Officials* with $\theta \geq B$ should not be willing to take the second bribe with probability λ and only agree to take it in case it is offered with probability $1-\lambda$. The reasoning about the equality case $\theta = B$ can be done analogously to the Bayesian equilibrium case. Check the previous section for clarity on this matter.

$$\sigma_2^*(\theta, \lambda) = \begin{cases} \begin{bmatrix} 1 & \text{with prob} = \lambda \\ 0 & \text{with prob} = 1-\lambda \end{bmatrix} & \text{if } \theta < B \\ \\ \begin{bmatrix} 1 & \text{with prob} = 1-\lambda \\ 0 & \text{with prob} = \lambda \end{bmatrix} & \text{if } \theta \geq B \end{cases} \tag{23}$$

Moving back in the game timeline and knowing how the *Public Officials* are predicted to behave in the second period, the *Regular Citizens* need to vote to elect one of the politicians to the higher rank office. Again, they are assumed to have an attention factor λ that affects their decision-making skills. Thus, basing our reasoning in the Bayesian equilibrium best strategies for the *Regular Citizens*, if there is a signal from the first period bribe-taking activities, the *Regular Citizens* should expect them to be less likely to take a bribe in the second period when their signal indicates that they did not take a bribe in the first period (although they might have done so by mistake).

Thus, when they receive two signals that are different, the probability that $s_i = 0$ means $\theta_i \geq \bar{\theta}$ is λ, and the probability that $s_j = 1$ means $\theta_j \geq \bar{\theta}$ is $1-\lambda$. Therefore, in this case the strategy of the *Regular Citizens* ν^* is to vote for the *Public Official* with

[6]This is Theorem 1 in Section 3 of McKelvey and Palfrey (1998). Check the original paper about quantal response equilibrium in extensive form games McKelvey and Palfrey [1998] for the proof of this theorem.

a signal that indicates she did not take a bribe in the first period with probability λ and for the one that has a bad signal with probability $1-\lambda$.

In the other cases, i.e. when there is no information about the bribery or the signals are equal, the best strategy of the *Regular Citizens* should not differ from the one in the Bayesian equilibrium analysis. Being their goal to maximize the probability not electing a *Public Official* that will take a bribe in the second period (because her doing so would lead to a reduction in the voters' payoff), when they do not have anything to tell the *Public Officials* apart, the *Regular Citizens* cannot do better than to elect either of the politicians with 50% probability.

$$\nu^*(s_A, s_B) = \begin{cases} (0.5, 0.5) & \text{if } s = \emptyset \text{ or } s_A = s_B \\ (\lambda, 1-\lambda) & \text{if } s_A < s_B \\ (1-\lambda, \lambda) & \text{if } s_A > s_B \end{cases} . \tag{24}$$

Finally, assuming all this, the *Public Officials* can build the part of their best strategy corresponding to the first-period bribe-taking action. Calculating the cutoff type $\bar{\theta}$ is more complicated than the Bayesian equilibrium, but the procedure is the same. The *Public Officials* are divided in two categories. Those with a type $\theta \geq \bar{\theta}$ will not take the bribe with probability λ but will do with probability $1-\lambda$. On the other hand, the *Public Officials* whose type is $\theta < \bar{\theta}$ will choose to take the bribe with probability λ and not to take it with probability $1-\lambda$.

$$\sigma_1^*(\theta, \lambda) = \begin{cases} \begin{bmatrix} 1 & \text{with prob} = \lambda \\ 0 & \text{with prob} = 1-\lambda \end{bmatrix} & \text{if } \theta < \bar{\theta} \\ \\ \begin{bmatrix} 1 & \text{with prob} = 1-\lambda \\ 0 & \text{with prob} = \lambda \end{bmatrix} & \text{if } \theta \geq \bar{\theta} \end{cases} . \tag{25}$$

Computing again the value of θ for which the expected value of the disturbed payoff, $\mathbf{E}[\hat{u}]$, is equal no matter whether they take the bribe with probability λ and reject it with probability $1-\lambda$ or vice versa. This time, the cutoff value depends not only on the time discount factor δ, the probability that the offered p_{bribe}, but also on the attention factor λ (see Figures 3 and 4).

$$\bar{\theta}(\lambda) = B - \frac{\delta p_{info}}{2 - \delta p_{info} \lambda p_{bribe}} \tilde{p}. \tag{26}$$

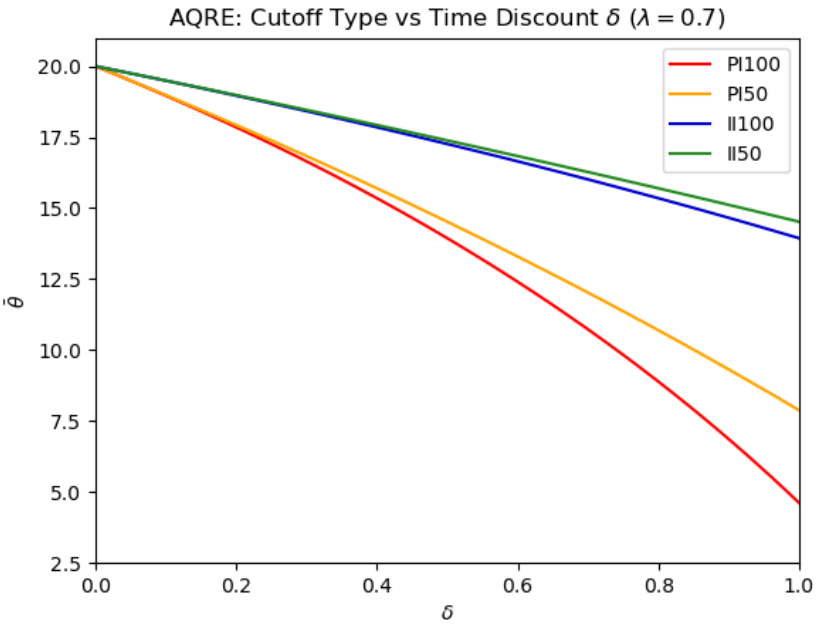


Figure 3: Agent Quantal Response Equilibrium, first period cutoff strategy type: dependence with the temporal discount factor.

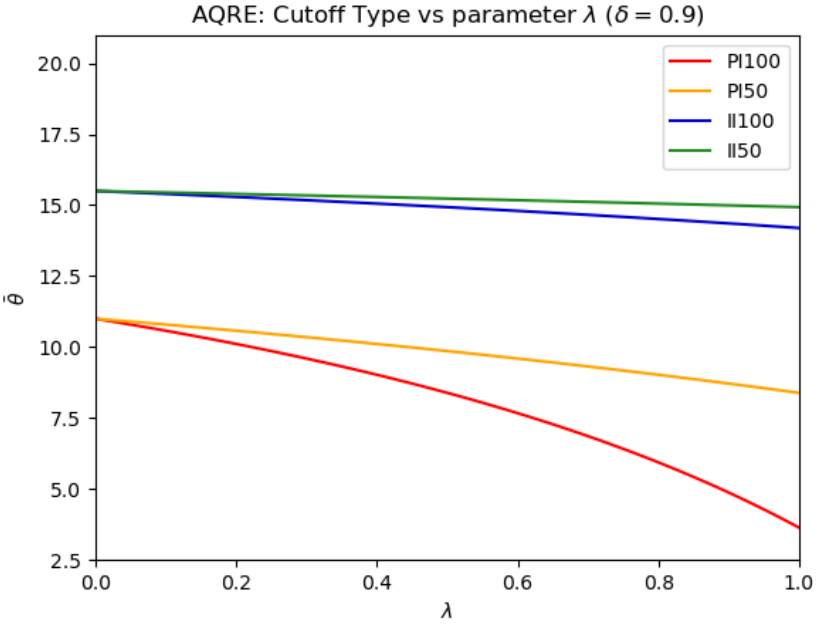


Figure 4: Agent Quantal Response Equilibrium, first period cutoff strategy type: dependence with the attention factor.

3 Experimental Setup

This section contains the details about how the experiment took place in the laboratory. Details about the software tools and the laboratory sessions are also included in Sections 3.1 and 3.2.

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 participants were members of the Mason community. The funds needed to carry out these experiments were also provided by ICES and GMU. The procedures to carry out this experiment follow the indications in Smith [1982].

Every session lasts under 60 minutes and is composed by a number of participants that is a multiple of 5, and varies from 10 to 20. These subjects are recruited using the recruiting system by Sona Systems from GMU.

If the subjects come to the laboratory on time, they receive a show-up fee of $10. Their additional earnings depend on their decisions during the game, and can vary from $5 to $30. These additional earnings are introduced in *pts* (points) currency throughout the game and are converted to US dollars at the end of the session with a conversion rate of

$$5pts = \$1.$$

At the beginning of the session, the subjects are given an anonymous label that we will use to record their behavior. Once they have their card, they are assigned a seat and are 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 to them how to play the game and how their rewards will be calculated. The software used for the making of this video was `Google Slides` from `Google, LLC`, `SimpleScreenRecorder` from `ubuntu-focal-universe`, and `Shotcut` from `Meltytech, LLC`.

There are 4 versions of the instructions (and the rest of the aspects of the session) that correspond to the 4 treatments of the experiment. The participants are not told about the existence of other versions of the game.

After watching the instructions, all participants are required to take a short quiz about them. The sole purpose of this quiz is to make sure that they understood properly what they are required to do.

Once all the quizzes are checked and they have resolved all their questions about the procedure of the experiment, they are taken to an `otree` app, see Chen et al. [2016]. This app needs to be deployed to a server in order to be accessible from computers other than the experimenter's. The one we use is `Heroku`, a `Salesforce`

platform. Further details of our app can be found in Section 3.1.

Then, the subjects play 10 rounds of the game as explained in Section 2 using the recently mentioned app. We make sure that there is no sort of communication between the different participants during the session to obtain independent observations.

At the end of the game, there is a last bonus question that consists of a list of random lottery pairs, as described in Harrison and Rutström [2008], to assess the risk aversion of the participants, and a short survey. We find that the outcome of the risk question is incoherent and we discard this information. On the contrary, the survey questions are quite useful for some analysis.

Once the session is over, the participants receive their payoff in US dollar bills when they give the card with their label back. The total mean payment per subject for this one-hour session was $19.45. Finally, `otree` integrates a system of conversion of the data into *.cvs* files that can be directly downloaded and used for analysis of the participants' interaction.

3.1 Game App

In order to conduct the experiment, it is necessary to develop an app so that the participants can play the game in it. Thus, the main objective of this app is to create the situation that we want to put the subjects in and to allow them to interact with one another in a controlled way.

The only requirement for the program to work is that the number of players is a multiple of five. The reason for this is, as explained before in Section 2, that the groups of players are made of two politicians and three regular citizens.

To access the app, the subjects are required to type in their participant label. Then, they are taken to a welcome screen that summarizes the instructions they saw in the video. There are two sections in the session: the first one is the game, and the second one is the risk assessment question together with the short survey.

In the first section they play 10 rounds of the game. At the beginning of each round, the participants are assigned a new role. Over the 10 rounds, each participant plays as a public official four times and six times as a regular citizen. Although their role assignment follows a certain pattern, the group assignment is totally random. This way, the players do not get used to the behavior of their group and the measures can be considered independent.

Then the *Public Officials* are taken to a screen where they are informed that they have been offered a bribe and are asked to select what they wish to do: take it or not (Figure 5). Meanwhile, the rest of the players see a *wait page*. These waiting screens are used every time the participants need to wait for others to take action.

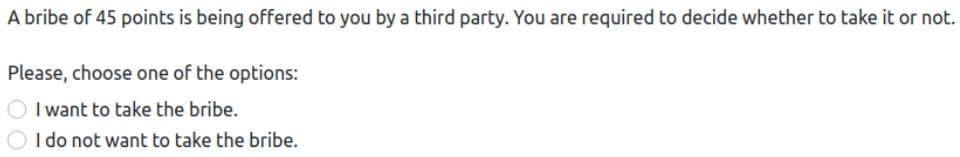


Figure 5: Screenshot of the first period bribe-taking decision.

The next screen is shown to everybody in the group, and might contain a information about the bribery or not (Figure 6). With probability $1 - p_{bribe}$, the players receive a message stating that there is no information to be displayed about the bribery. Also included in the screen is a line informing the players of their reward (in *pts*) for the first period (because the rest of the actions of the first period do not affect their first period payoff). As seen in Table 1, the *Regular Citizens* can get the value of $x_A + x_B$ just from seeing the period (where x_i takes value of 1 if *Public Official* i took the bribe in the first period and 0 otherwise). Despite this fact, in a case where they did not get explicit information about who engaged in corrupt activities and who did not, they cannot tell apart the actions of the different *Public Officials*. Therefore, seeing the reward on the screen does not bias their voting decision.

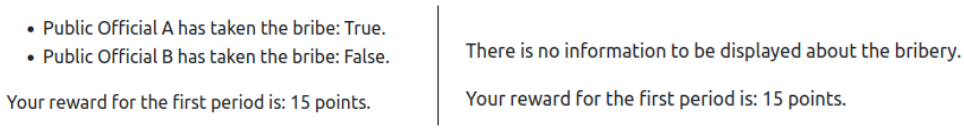


Figure 6: Left: Screenshot of the first-period bribery results when the information is displayed. Right: Screenshot of the first-period bribery results when the information is not displayed.

Without the option to go backwards, the next button takes the *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 (Figure 7). The options are *Public Official A* and *Public Official B* but we don't see any tendency to vote more for any of them when there is a situation where we expect the voters to be indifferent. These situations would be when there is no information displayed or when $s_A = s_B$. In these 302 cases, *Public Official A* is voted for 51.66% of the times and *Public Official B* 48.34%. The p-value of the statistical t-test is 0.3192, so we do not have enough evidence to reject the hypothesis that there are as many votes for A as for B when the *Regular Citizens* are indifferent about who to vote for. Despite this, the app was developed before knowing about these results. Thus, to be fair with the participants, it was prepared so that they all play *Public Official A* twice and *Public Official B* twice.

After the voting, everyone gets a message with the results. The *Public Officials*

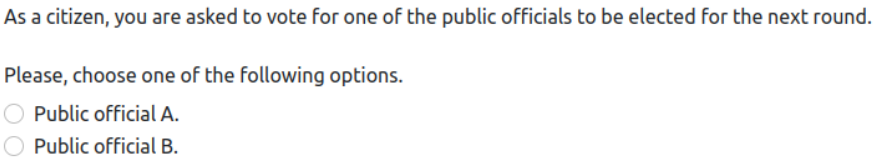


Figure 7: Screenshot of the voters' election screen.

are told whether they were elected or not and the *Regular Citizens* are informed about the tag of whoever won the election.

In the second period, the *Elected Public Official* is taken to a screen that has a question stating "If you are offered a bribe, will you want to take it?" for the treatments with $p_{bribe} = 0.5$ and a question saying "You have been offered a bribe, do you want to take a bribe?" when $p_{bribe} = 1$ (Figure 8). The *Public Official* is required to choose before moving on, while the rest of the participants wait.

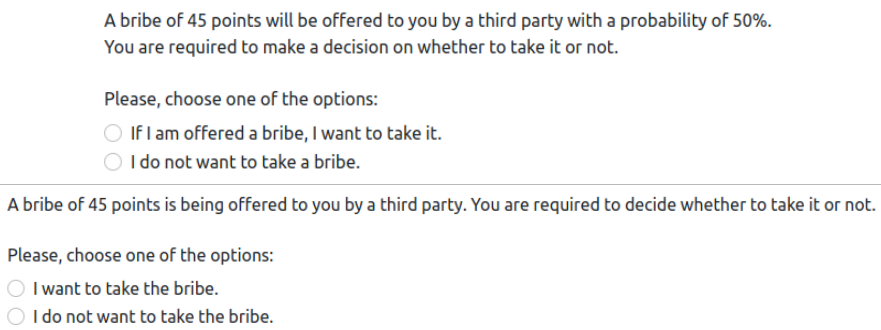


Figure 8: Screenshot of the second period bribe-taking decision. Top: $p_{bribe} = 50\%$. Bottom: $p_{bribe} = 100\%$.

Once the *Elected Public Official* has made a decision, the bribe is offered with probability p_{bribe}. The last screen of the period includes a message informing about whether the *Elected Public Official* has received (accepted and been offered) the bribe or not (Figure 9). Similarly to corresponding screen in the first period, there is also a line with the reward for the second period.

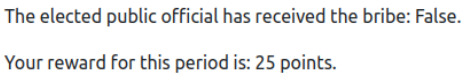


Figure 9: Screenshot of the second-period bribery results.

Finally, everyone gets a message with their total reward for this round. If they click on the `Next` button, they are taken to the role page again to begin a new round.

After the 10 rounds, one of them is randomly selected to determine the payoff of the participants. All the rounds have the same probability of being selected. At this point, the players are taken to a screen that informs them about which round was selected and what is their reward converted to US dollars.

The app also contains the risk lotteries question and a short survey at the end. Thus, after the result of the game, the participants are taken to a screen with the instructions of this last part. The risk question has a monetary motivation, so the last page of the section is the results and additional reward for this part. The reward of the game is chosen randomly.

The risk assessment screen consists of 9 cases where they need to choose in which of the two lotteries that they are presented they want to take part. The survey asks what their gender is, their level of studies, and their major.

At the end, they get a screen with instructions to wait in their seats until they are taken to the payment room.

3.2 Observations

As presented in Table 4 we got more than 50 observations for each treatment, although it was not possible to obtain an even number across treatments, due to the irregular attendance rate[7]. Subjects sign up for the sessions online and are reminded that they have a session, but some of them are late, forget to come, or even decide not to come.

	Perfect Information 100% 2^{nd} Bribe	Perfect Information 50% 2^{nd} Bribe	Imperfect Information 100% 2^{nd} Bribe	Imperfect Information 50% 2^{nd} Bribe	**Total**
Women	26	28	30	22	106
Men	31	26	18	31	106
Non binary	1	0	0	2	3
Undergrad	36	33	32	37	138
Graduate	23	22	16	16	77
Total	60	55	50	55	**220**

Table 4: Distribution of observations.

[7]Regarding Table 4, there were 5 people that chose not to disclose their gender and 5 people that did not reveal their *Level of Studies*.

4 Results

This section aims at giving insights about the experimental results that we got in the laboratory of economics. Thus, we will see if the equilibrium predictions from Section 2.5 are satisfied by the human subjects' behavior.

In order to study the behavior of the subjects, we look at their actions round per round. We do not include a mechanism to study learning across rounds, so, for general analysis, we take each round as a separate game. That means that, unless we are looking at round-per-round behavior, we will have repeated observations of the same subjects. In other words, as we use percentages and not absolute values for our analysis, what we are actually doing is averaging the subjects' behavior across rounds.

First, we will look at the behavior of the *Public Officials* in terms of the bribe-taking activity in Section 4.1. Then, in Section 4.2, we will study the *Regular Citizens*' behavior when it comes to voting. We will analyze the entanglements of these strategies, see that they could be explained with our model, and decide whether the equilibrium theories fit our data.

4.1 Behavior of Public Officials

In this section we study the behavior of *Public Officials* in our game experiment. Their more or less corrupt attitudes are reflected in their decisions regarding the bribe-taking actions.

As mentioned in Section 2.1.2, the type of the *Public Officials* θ is a private value to them. With our analysis of their actions, we will have some insight about this parameter's distribution, although its exact density function will not be possible to induce.

4.1.1 Selection of Data

The first issue that we encounter when studying the behavior of *Public Officials* is that only half of them get elected, so we only have complete information about their behavior in half of the cases. One could think that Figure 10 is be an accurate representation of reality. However, as we will see in Section 4.2, the tendency to elect *Public Officials* that took the first bribe is lower than to elect those who did not take it. Thus, Figure 10 is missing on all those *Public Officials* that did not get elected (but also represent human behavior).

In consonance with this, we decide to use the information that we have. For the first-period bribe-taking action x_1 we have complete information from all the

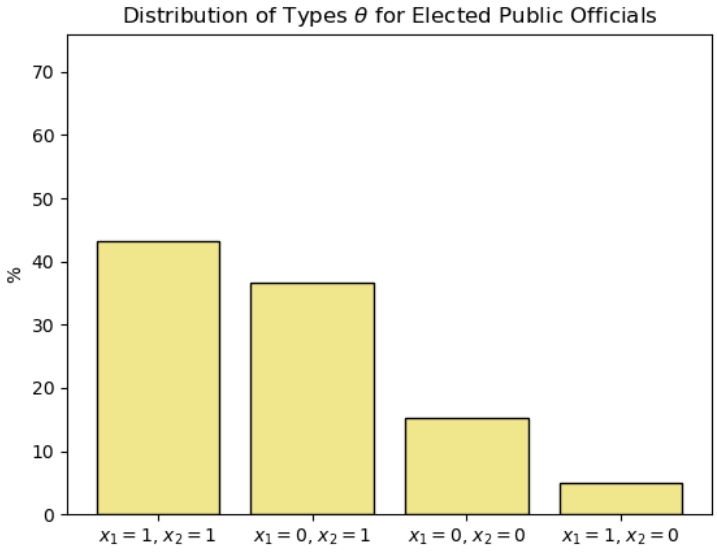


Figure 10: Distribution of the behavior of the Elected Public Officials.

subjects that are *Public Officials*. For the second-period bribery activity x_2 we only have observations from the *Elected Public Officials*. It must be said that this x_2 represents their willingness to take the bribe, independently of whether they finally take it or not. Therefore, using the information of the second-period bribery combined with the first-period actions of the same *Public Officials*, we induce the most likely behavior of the *Non-elected Public Officials* in the second period. We use the same ratio of *Non-elected Public Officials* that want to take or not the second bribe according to what they did for the first bribe as the ratio of *Elected Public Officials* under the same circumstances. The final distribution of behavior is represented in Figure 11.

It is clear that when we take into account the *Non-elected Public Officials*, the distribution of strategies changes noticeably. Obviously, not in the ratio between $(x_1 = \cdot, x_2 = 0)$ and $(x_1 = \cdot, x_2 = 1)$, but in the ratio between first-bribe takers and first-bribe non-takers.

4.1.2 General Distribution of Types

It has already been mentioned that the exact distribution of θ cannot be induced from observing the *Public Officials*' actions. However, there is much information that can be extracted from them that gives us leads about this distribution.

According to the Bayesian equilibrium theory, the results in Figure 11 can be

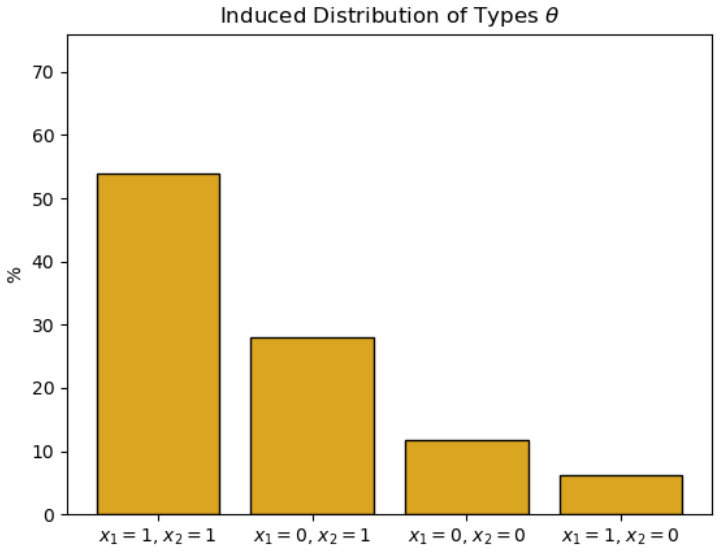


Figure 11: Distribution of behavior of the Public Officials after inducing the second-period behavior of the Elected Public Officials on the Non-elected Public Officials.

$\theta < \bar{\theta}$	$\bar{\theta} \leq \theta < B$	$B \leq \theta$
475	247	103

Table 5: Amount of subjects observed for each region of the domain of θ if the results are interpreted according to a Bayesian equilibrium.

explained as follows. First of all, the 6.25% of subjects that did not manifest to want the second bribe after taking the first one cannot be explained with this theory. The reason for this is that the cutoff strategy for the first period $\bar{\theta}$ is smaller than the one for the second period B, i.e. $\theta_{BE} = B - \frac{\delta p_{info}}{2-\delta p_{info} p_{bribe}} \tilde{p} < B$. Thus, everyone that took the first bribe is also expected to want the second one. Using this theory, we need to consider this 6.25% as noise that pollutes our data.

On the other hand, one of the advantages of the Bayesian equilibrium theory is that, if we ignore this noise, we can extract information about the types of the *Public Officials*. The results can be read in Table 5.

Statistically speaking, we can consider the probability that a *Public Official* takes the first bribe as $\int_0^{\bar{\theta}} f(\theta) d\theta$ where $f(\theta)$ is the probability density function of the *Public Officials*' types. Analogously, we could calculate the probability that a *Public Official* is willing to take the second bribe as $\int_0^{B} f(\theta) d\theta$ with the same $f(\theta)$. Using our results, we can estimate this probabilities and get confidence intervals

($\alpha = 0.05$) as follows:

$$\begin{array}{c} \int_0^{\bar{\theta}} f(\theta)d\theta = 57.576\% \pm 3.373\% \\ \int_0^{B} f(\theta)d\theta = 87.515\% \pm 2.256\% \end{array} . \tag{27}$$

On the contrary, we could consider that the attention factor is not perfect $\lambda < 1$ and use the Agent Quantal Response Equilibrium theory. From the data in Figure 11 we can extract the following information:

$$C_1 := \lambda \cdot \int_0^{\bar{\theta}} f(\theta)d\theta + (1-\lambda) \cdot \int_{\bar{\theta}}^{\infty} f(\theta)d\theta = 60.227\% \tag{28}$$

where 60.23% is the sum of the percentages of *Public Officials* who took the first bribe. Equivalently, we could have written:

$$H_1 := \lambda \cdot \int_{\bar{\theta}}^{\infty} f(\theta)d\theta + (1-\lambda) \cdot \int_0^{\bar{\theta}} f(\theta)d\theta = 39.773\% \tag{29}$$

but we would not get any additional information, as everything adds up to 1. The meaning of this Equation 28 is that those *Public Officials* that have an incentive to take the first bribe ($\int_0^{\bar{\theta}} f(\theta)d\theta$) and make the right choice (λ) and those who do not have an incentive to take the bribe ($\int_{\bar{\theta}}^{\infty} f(\theta)d\theta$) but make a mistake ($1-\lambda$) are the ones that adopted a corrupt behavior (C_1). In the same line, Equation 29 could be read as the percentage of *Public Officials* that chose the *honest* option (not taking the bribe, H_1) in the first period are those who did not have an incentive to take it ($\int_{\bar{\theta}}^{\infty} f(\theta)d\theta$) and made the right choice (λ) and those that actually had an incentive to take it ($\int_0^{\bar{\theta}} f(\theta)d\theta$) but ended up not doing so by mistake ($1-\lambda$).

Analogously, we can write:

$$C_2 := \lambda \cdot \int_0^{B} f(\theta)d\theta + (1-\lambda) \cdot \int_{B}^{\infty} f(\theta)d\theta = 82.062\% \tag{30}$$

and

$$H_2 := \lambda \cdot \int_{B}^{\infty} f(\theta)d\theta + (1-\lambda) \cdot \int_0^{B} f(\theta)d\theta = 17.938\% \tag{31}$$

for the second-period bribe-accepting actions.

Although $\int_0^{\bar{\theta}} f(\theta)d\theta + \int_{\bar{\theta}}^{\infty} f(\theta)d\theta = 100\%$ (analogously $\int_0^{B} f(\theta)d\theta + \int_{B}^{\infty} f(\theta)d\theta = 100\%$), these are consistent independent systems. Thus, although we know that for any value of $\int_0^{\bar{\theta}} f(\theta)d\theta$ there exists a value of the attention factor λ that would explain our data, we need more information to solve the system. In this case, we

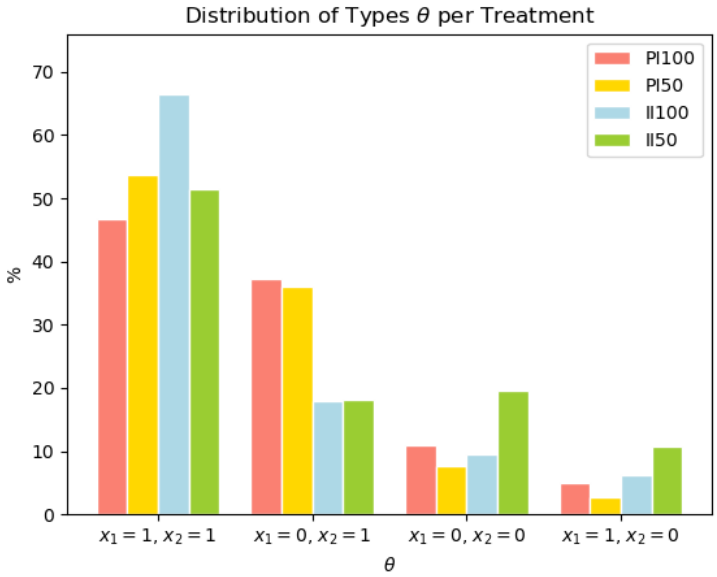


Figure 12: Distribution of behavior of the Public Officials after inducing the second-period behavior of the Elected Public Officials on the Non-elected Public Officials separated per treatment.

will use the results obtained in Equation 27 to calculate the most likely value of λ.

Using the values from the Bayesian equilibrium and a propagation of error does not work, since the fact that the percentage of *Public Officials* that take the first bribe is higher in the AQRE case than in the BE case leads to a conclusion where $\lambda > 1$.

This led us to acknowledge that there are limitations (or restrictions) to our reasoning. We need to impose that both the integrals and λ are ≤ 1. In other words, for the system to be consistent we need that $\int_0^{\bar{\theta}} f(\theta)d\theta \leq C$.

4.1.3 Distribution of Types per Treatment

In this section we aim at refining the analysis from Section 4.1.2 by bringing it to the treatments level.

The distribution of actions of the *Public Officials* separated by treatment can be found in Figure 12[8].

In order to understand these results, we shall compare them with the predictions from the Bayesian equilibrium theory. First of all, Figure 2 indicates that we should find a higher percentage of first-period bribery cases in the cases with imperfect

[8]See Table 3 for the names of the treatments.

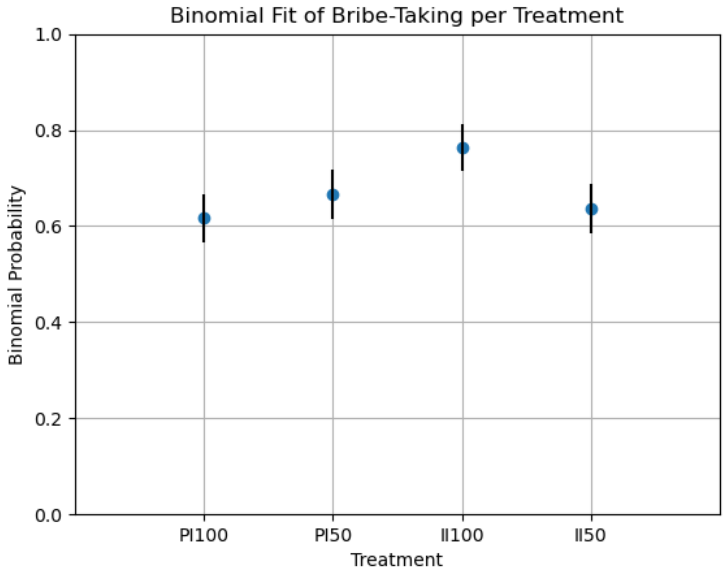


Figure 13: Binomial fit of the bribe-taking action in different treatments with a confidence interval of $\alpha = 0.05$.

information than in those with perfect information. We see actually see this the $p_{bribe} = 100\%$ cases but it is not clear for those where $p_{bribe} = 50\%$.

We can see this fact in Figure 13, where there is a plot of the binomial fit of the bribery. It represents what the probability of taking (or wanting if it is the second period) a bribe is for the different treatments. We find a significant effect of the information (p_{info}, perfect information vs. imperfect information) when the second bribe is given in all cases ($p_{bribe} = 100\%$), as the 95% confidence intervals are disjoint. On the contrary, there is not enough evidence to affirm any other relation between the individual treatments regarding bribery.

However, if we combine the treatments, we can study the effects of p_{info} and p_{bribe} separately, as in Figure 14. In Figures 2, 3, and 4 we predict that with imperfect information more first-period bribes will be taken, and with a higher probability of being offered a bribe in the second period less bribes will be expected to be taken. When we looking at the effect of the probability of the second bribe, we do not find a significant effect in the bribe-taking activities of the first period. Indeed, the $z-$value of the comparison is 0.87, which lays under 1.96, that is the threshold for an $\alpha = 0.05$ significance. What is more, there appears to be more bribes taken in the 100% second-period bribe treatments than in the 50% ones, so we can say that our model does not explain this fact.

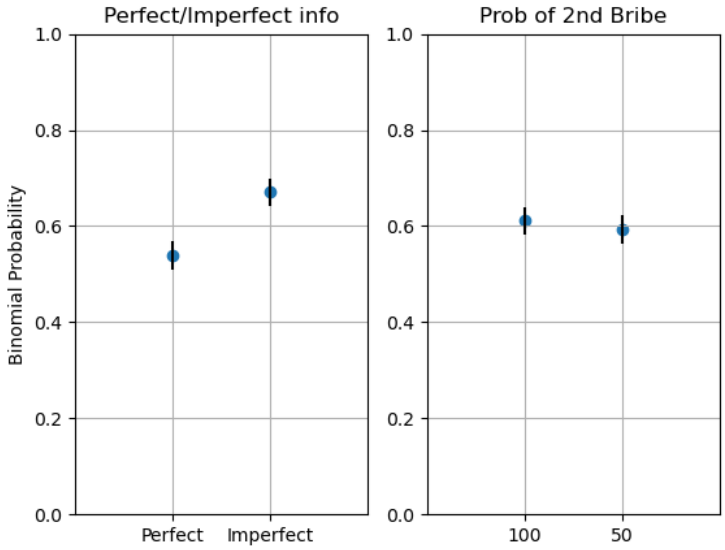


Figure 14: Binomial fit of the bribe-taking action in different combinations of treatments with a confidence interval of $\alpha = 0.05$.

On the contrary, if we look at the effects of the probability that everyone is informed about first-period bribery results, we do find an increase in the amount of first-period bribes taken when the probability of informing is lower. Indeed, we get a $z-$value of 6.33, which is greater than 1.96, so we can affirm that the probabilities are different with statistical significance.

Taking a closer look at the behavior of the *Public Officials* per treatments in the different rounds of the game, we see that the distribution of their actions is not uniform. Figure 15 shows these differences. The three thin bars that can be observed over each one of the big average bars show the actions in round 1 (left), rounds 2 to 5 (middle), and rounds 6 to 10 (right). Generally, the amount of bribes taken in the first round is smaller than in the rest of the rounds. This happens in the treatment Perfect Information 50%, and both of the Imperfect Information treatments.

However, if we dig into the statistical fit of this matter, we do not find such clear evidence. In Figure 16, it is possible to see how these differences are not statistically significant. The reason of this is that the confidence intervals of the binomial fit for the bribery in each round overlap. In order to do this, we look at each round separately so that our analysis is more delicate and subtle.

Another subject to analyze is the amount of *Public Officials* that would be

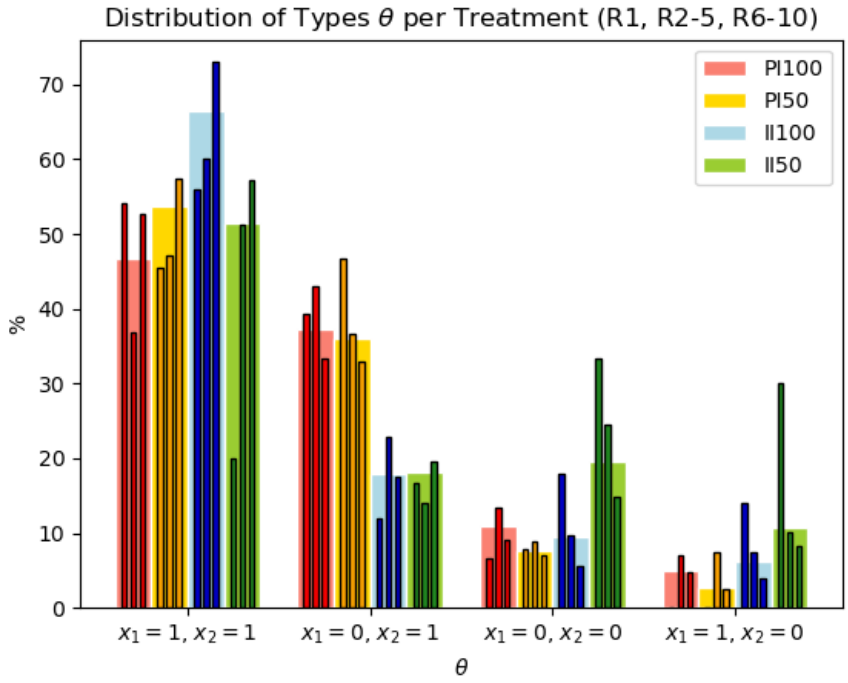
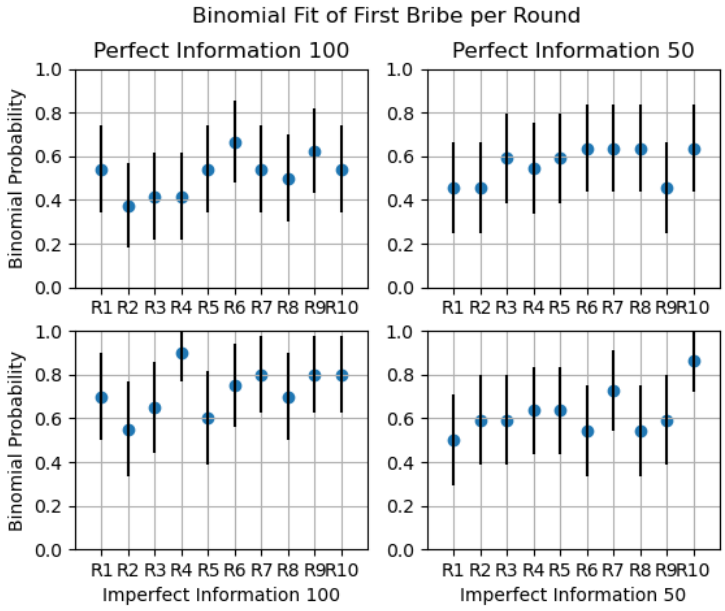


Figure 15: Distribution of behavior of the Public Officials after inducing the second-period behavior of the Elected Public Officials on the Non-elected Public Officials separated per treatment and per rounds. Each triple of thin bars represent the average for round 1, rounds 2-5, and rounds 6-10.

considered as *noise* according to the Bayesian equilibrium theory in Figure 15. In other words, the players that do not manifest to want the second bribe after having taken the first one.

We find a peak of this event in the Imperfect Information 50% bribe treatment ($(p_{info}, p_{bribe}) = (0.5, 0.5)$), which can give us a hint about the reason why this happens. Indeed, using AQRE, we can attribute this behavior to a lower attention factor $\lambda < 1$. The Imperfect Information 50% bribe treatment is the treatment with the most complicated instructions, so a reduction in the attention factor can be naturally attributed to this fact. On the other edge, the Perfect Information treatments are the ones with the most simple instructions. Thus, the focus in the first round can be set in making the *right* choices (there are no cases of taking the first bribe but rejecting the second one for these treatments in the first round).

As the other rounds of the game take place, there is a learning effect that we cannot measure but could explain the reduction in the number of subjects that adopt this particular strategy. Nevertheless, this effect could be counteracted by a more relaxed attitude of the players caused by the repetition of the same actions across



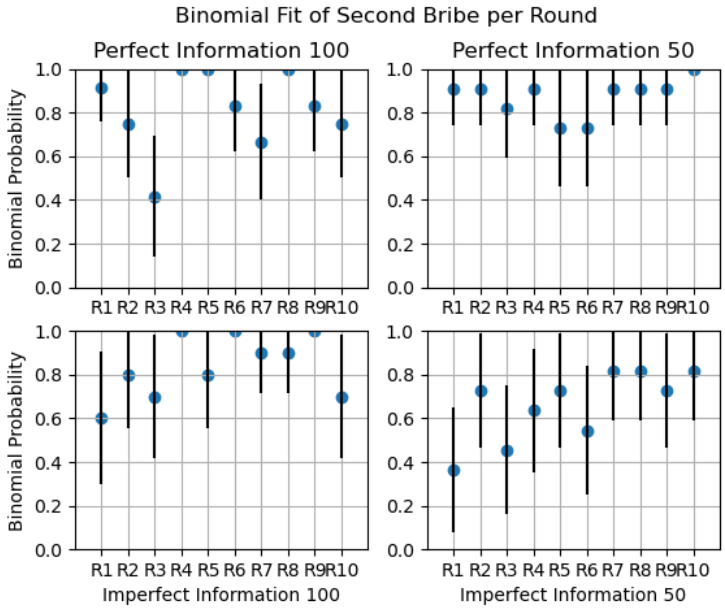


Figure 16: Binomial fit of the bribe-taking (or bribe-wanting) actions in the first (top) and second (bottom) periods. Confidence intervals with $\alpha = 0.05$.

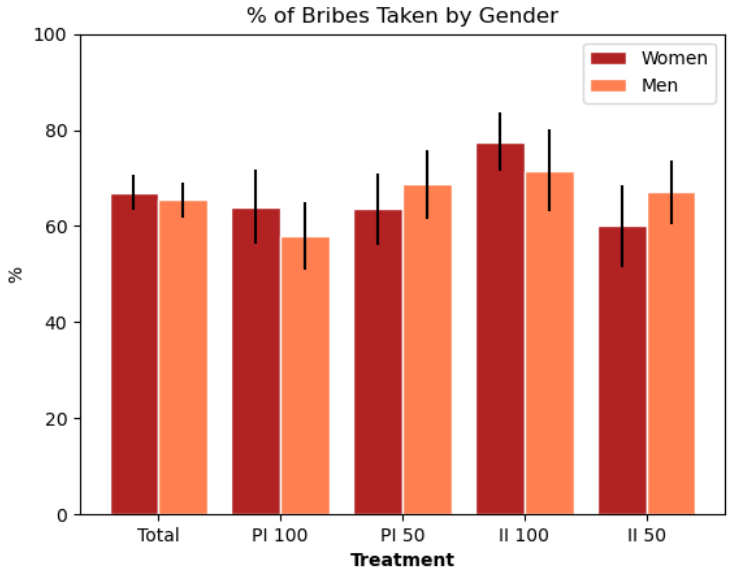


Figure 17: Percentage of bribes accepted out of the offered ones, by gender and treatment.

the rounds. This could lead to a reduction in the attention factor, whilst the learning would increase their capacity of increasing the value of λ (a better understanding of the game allows them to make better decisions but repeating the same game over and over leads to boredom, disinterest, and a reduction in the attention they pay to their actions).

One of the ways we try to mitigate the disinterest in repetition issue is by using a random round to represent their real final payoff. The subjects do not know which round will be the one they will be payed for, so it is in their best interest to treat each one of the rounds as if it was the one they will be payed for. At the end of the session, a random variable generates a number between 1 and 10 for the selection.

4.1.4 Gender Bias

Serra and Wantchekon [2012] realize a laboratory experiment with college students to study bribery tendencies and finds evidence that women take equally or less bribes than men. Our results on the matter can be observed in Figure 17.

It is clear that our evidence indicates that, a priori, gender and bribe-taking are two independent variables. The reason of this is that we get opposite results depending on the treatment that we look at.

It could be thought, however, that there is some kind of interaction between the variables of the treatment and the gender of the subjects. In particular, between the

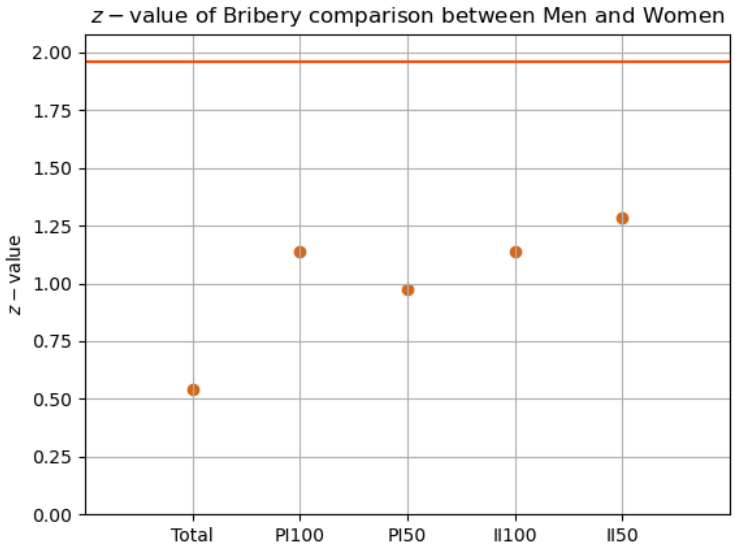


Figure 18: $z-$value of the hypothesis test that tests whether the probabilities of the binomial fits of the bribe-taking activities of men are women are considered statistically equal (H_0) or not. The results are separated by treatment.

certainty that a second bribe is going to be offered and the gender. When women know that the second bribe is going to be offered with certainty, they tend to take more bribes, and for men act like so when the probability of being offered a second bribe is reduced.

However, if we look at Figure 18 we do not find any statistically significant evidence of a difference in the probability of taking a bribe between men and women. The line at $z = 1.96$ represents the threshold $z-$value when the significance coefficient $\alpha = 0.05$. As none of the points appear higher than this line, we cannot discard our hypothesis about the two probabilities being equal.

4.2 Behavior of Regular Citizens

Someone could think that the *Regular Citizens* do not play a big role in this game because their actions do not have a direct impact on their payoff. Nothing further from reality: their strategies offer numerous insights about human behavior regarding trust, accountability, and beliefs.

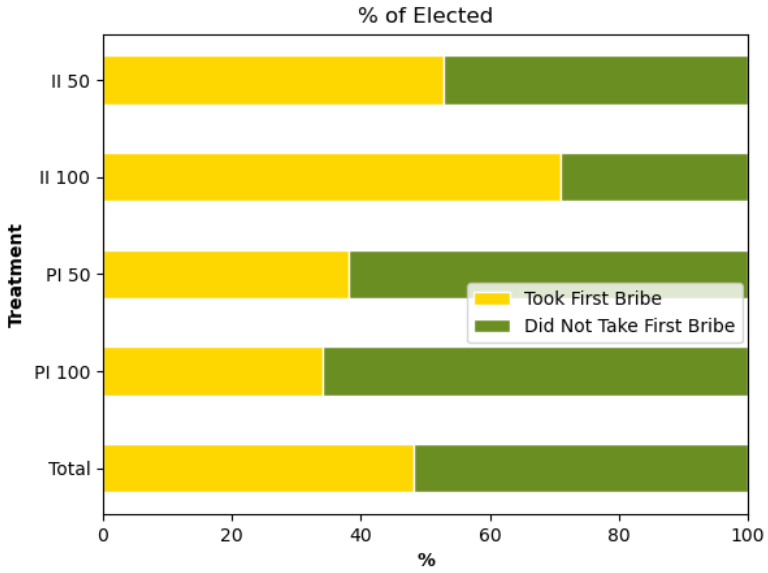


Figure 19: Percentage of Elected Public Officials that had taken the bribe in the first period or not, per treatment.

4.2.1 Election Results

The way to understand the *Regular Citizens* behavior is by looking at how they vote. Information about who they trust and how much attention they give to voting can be found there.

According to the Bayesian equilibrium theory, *Regular Citizens* are expected to vote for the *Public Officials* that do not have taken a bribe in the first period. Indeed, these *Public Officials* have a type $\theta \geq \bar{\theta}$, so the probability that $\theta \geq B$ is non-negative. On the contrary, as $B \geq \bar{\theta}$, if a *Regular Citizen* takes the first bribe, then her type satisfies $\theta < \bar{\theta} \leq B$, so it is not convenient to vote for her, since the goal of the *Regular Citizens* is expected to be maximizing their payoff (which happens when the *Public Official* do not take bribes).

Figure 19 shows the percentage of *Elected Public Officials* that had taken the first-period bribe. This means that the *Regular Citizens* voted for them knowing that it was very likely that they would take the second-period bribe too. For the average case, we observe that a 48.2% of the *Elected Public Officials* took the first bribe. For the treatments, the results are a 34.2% for the Perfect Information 100% Bribe treatment, 38.2% for Perfect Information 50% Bribe, 71.0% for Imperfect Information 100% Bribe, and a 52.7% for the Imperfect Information 50% Bribe case. All these results, especially the last two, do not coincide with what our model

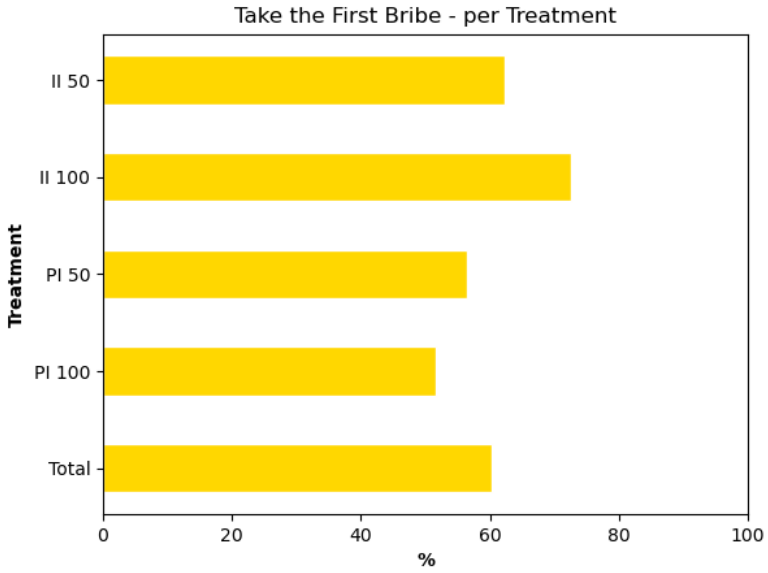


Figure 20: Percentage of the Public Officials that took the first bribe per treatment.

predicts, which is why we need to dig deeper into the matter.

First of all, we need to look at the amount of first-period bribes taken in each treatment, which is represented in Figure 19. These data suggest that, in many cases, the *Regular Citizens* did not have the option to choose a candidate that had not taken the first bribe, i.e. both candidates had $\theta < \bar{\theta} \leq B$.

When the voters have to choose between two candidates that both rejected or both accepted the first-period bribe, the outcome of the election does not say much about the behavior of the *Regular Citizens*. Therefore, to properly understand the behavior of the voters, we need to look at those elections where they can choose between a *Public Official* that took the first bribe and one that did not take it.

That is what we see in Figure 21. The rate of electing a corrupt *Public Official* when she was running against someone that did not take the first bribe is 11.1% in the Perfect Information 100% Bribe treatment, and 11.5% in the Perfect Information 50% Bribe treatment, but 46.5% and 28.6% in the Imperfect information 100% Bribe and 50% Bribe respectively. This gives an average of 48.2% of the cases, which is neither what the Bayesian equilibrium theory predicted. Although the perfect information treatments appear to be reasonable, the noise is still very significant in the rest of them. This is why we need to have a look at those election where the voters have been informed about the actions of their *Public Officials* in the first period of the game.

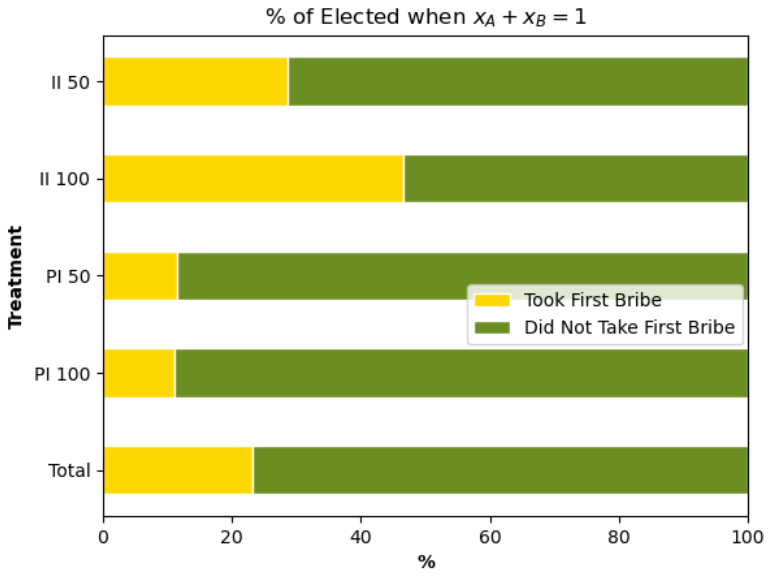


Figure 21: Percentage of the Public Officials that, having taken the first-period bribe, won the election against a Public Official that had rejected it.

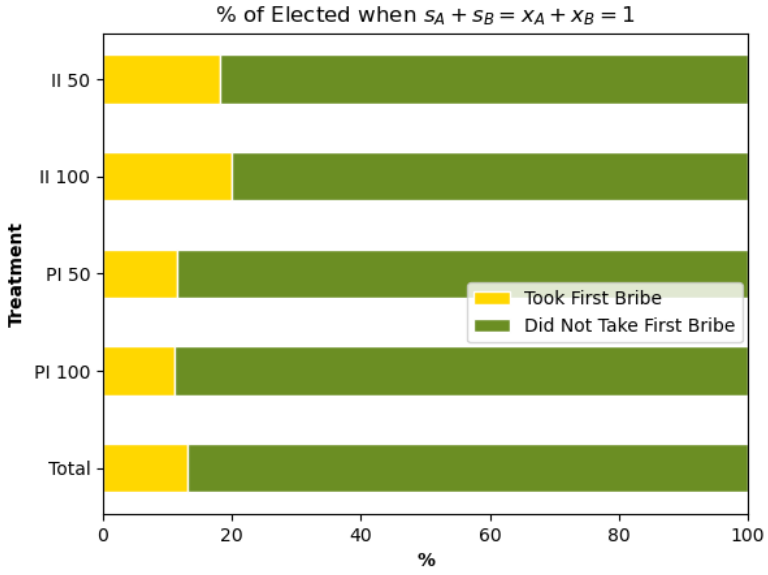


Figure 22: Percentage of the Public Officials that, having taken the first-period bribe, won the election against a Public Official that had rejected it in elections, after a public display of the information about the first-period bribery.

In Figure 22 we find the results regarding the percentage of *Elected Public Officials* that had taken a bribe in the first period in scenarios where the *Regular Citizens* were informed (before voting) about the bribery actions in the first period. In average, 13% of the elected candidates had taken the first bribe. For the different treatments, we find a value around 11% for the perfect information ones, and of 20.0% and 18.2% for the Imperfect Information 100% Bribe and 50% Bribe treatments respectively.

These values could be considered as noise in a Bayesian equilibrium, but would be better as lack of attention in an Agent Quantal Response Equilibrium theory model. Contrary to the case of the *Public Officials*, here we can estimate the value of the attention factor λ. Since the set of instructions for each treatment is different, the attention factor shall be a function of the treatment.

We should not extract λ from Figure 22, as it shows the results of the elections and λ is defined as the probability of the *Regular Citizens* **voting** for the *Public Official* that took the bribe having the option to vote for one that did not under the condition that they were informed about the bribery. Thus, Figure 23 shows the statistics of the votes for each treatment. From these results, we can estimate the attention factor as it is read in Table 6. The confidence intervals are given doing a binomial fit of the measure.

Treatment	**Attention Factor**	**Confidence Interval**
Perfect Info 100% Bribe	$\lambda = 0.79$	$(0.73, 0.85)$
Perfect Info 50% Bribe	$\lambda = 0.74$	$(0.68, 0.81)$
Imperfect Info 100% Bribe	$\lambda = 0.77$	$(0.62, 0.92)$
Imperfect Info 50% Bribe	$\lambda = 0.73$	$(0.62, 0.83)$

Table 6: Estimate of the attention factor λ for the Regular Citizens' voting strategy using a binomial fit.

It is clear that the attention the voters pay to who they vote for decreases both with p_{info} (the probability of receiving information about the first-period bribery) and p_{bribe} (the probability that the second bribe is offered). Moreover, the confidence intervals are noticeably wider for the imperfect information treatments. The reason for this is the amount of observed cases that we have, being 162 for the Perfect Information 100% Bribe treatment, 162 for the Perfect Information 50%, and 30 and 66 for the Imperfect Information 100% Bribe and 50% Bribe respectively. We get this numbers because of two reasons. The first one is that we only consider those cases where the information about the bribery is displayed, and that only happens with a 50% probability in the imperfect information treatments, so we automatically discard half of the observations. The second reason can be found in Figure 20, and is

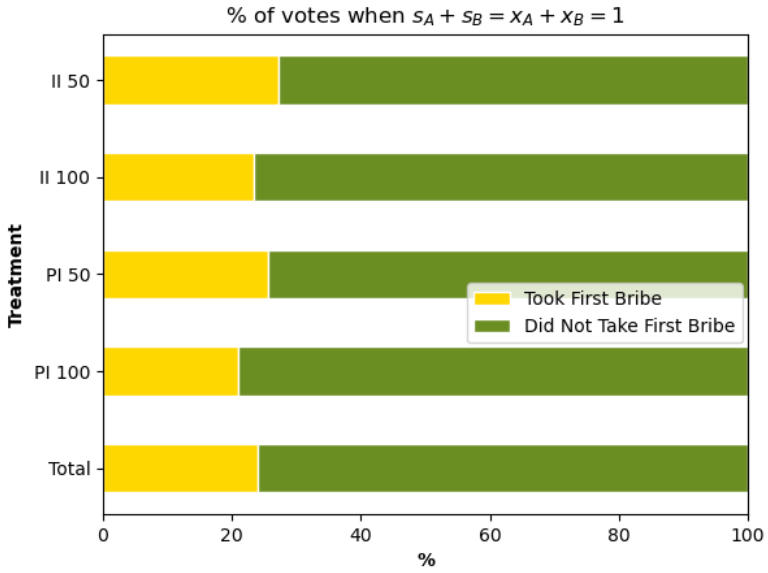


Figure 23: Percentage of the votes that go Public Officials that took the first bribe or did not take it, when the Regular Citizens were informed about the bribery decisions and there is a honest politician running against a corrupt one.

that we people take statistically more bribe in the imperfect information treatments than in the perfect information ones.

Finally, we seek to understand how the behavior of the *Regular Citizens* evolves throughout the rounds of the game. In order to do this, we analyze the results in Figure 24. We find that, in general, the *Regular Citizens* tend to elect the *Public Official* that took the first bribe more often in the first rounds of the game than in the following ones.

There is one significant exception in the Perfect Information 50% Bribe treatment. There, the rate diminishes between the first round and rounds 2 to 5, but increases again in the last 5 rounds of the game (although it is a smaller rate than in the first round).

In the rest of the cases, we need to analyze the size of our sample. Actually, all 3 cases where we get a 100% rate of accuracy in the election of *Public Officials* that rejected the bribe in the first period we have very few observations: 3 for the first round of the Perfect Information 100% Bribe treatment, 2 for the first round of the Imperfect Information 100% Bribe treatment, and 1 for rounds 2 to 5 of the latter treatment.

In conclusion, we can think that the attention factor of the *Regular Citizens* evolves throughout the rounds of the game. We do not have enough evidence to

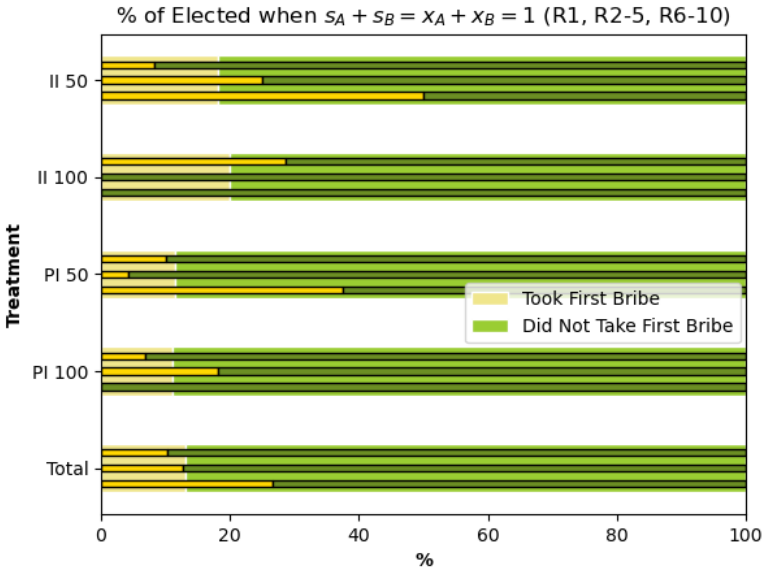


Figure 24: Percentage of the Public Officials that, having taken the first-period bribe, won the election against a Public Official that had rejected it in elections, after a public display of the information about the first-period bribery. Specifications per rounds: round 1 (lower bar), rounds 2 to 5 (middle bar), and rounds 6 to 10 (top bar). Average per treatment (light background bar).

affirm it, because of a lack of observations, but the results of the elections indicate so, because winning an election means having at least 2 of the 3 *Regular Citizens* vote for you.

5 Conclusion

We have adapted the model presented in Martinelli [2022] to make it suitable for a lab experiment. After doing a pilot to test if the values of our constants were adequate, we ran sessions of the experiment at the laboratory of experimental economics of the Interdisciplinary Center for Economic Science at George Mason University.

The model uses game theory to study accountability in election systems, career incentives in politics, and corruption. It represents a scenario of promotion by election where two *Public Officials* compete for a position in a higher-rank office and three *Regular Citizens* vote for one of them to get the job. There are two periods in the model timeline: in the first one, there is an opportunity of engaging in corruption for the *Public Officials*, and then the election takes place; and, in the second one, the *Elected Public Official* usually has another opportunity of taking a bribe.

We used two equilibrium theories to study this model. The first one, Bayesian equilibrium, predicted that the *Regular Citizens* would vote for the *Public Officials* that seemed less prone to take bribes, and that the *Public Officials* would be in 3 categories: those that take the both bribes, those who wait until the second period to take the bribe to trick the voters into electing them, and those that do not have an interest in engaging in corruption and do not take any bribes. The results that we get show some noise in relation to these predictions. On the one hand, there are *Regular Citizens* that vote for corrupt *Public Officials* while having the option to choose another that did not took the first bribe. On the other hand, we find that there are *Public Officials* that choose to reject the second bribe after taking the first one.

The second theory is Agent Quantal Response equilibrium. Using it, we predict that the players will make mistakes and deviate from the best strategies defined by the Bayesian equilibrium with a certain probability. We represent this probability with an *attention factor*. We find that this theory has potential to explain better the results that we observe, because sometimes the players do not act as expected or make mistakes. On the cons side, adding a new parameter complicates things. We actually found that for certain cases it was complicated to estimate this attention factor.

Moreover, we establish 4 treatments of the game. They are distinguished depending on the probability that the information about the first period is displayed before the election, which can be either 0.5 or 1, and the probability that a bribe is offered to the *Elected Public Official* in the second period, which can also be either 0.5 or 1.

Our predictions predict that there will be more first-period bribery not only in the treatments with imperfect information than in those with perfect information, but also more when there is more uncertainty that the second bribe will be offered. Our results prove the first case with statistical significance, but would need more evidence to be able to justify the second.

All in all, this first experimental model about this matter has still some tuning that could improve it, but is already able to prove that Martinelli's theoretical model is likely to represent a good explanation of human behavior.

6 Extensions

This project was programmed to have a fixed duration in time, so it was inevitable to think of ways of extending or improving it that could not be put into practice. This section will contain some of these ideas.

First of all, as seen in Section 4.2.1, we need more observations of a certain kind in order to study the voting behavior of *Regular Citizens* across rounds. This kind of observations are those where the voters have been informed about the first period bribery and have the option to choose between a *Public Official* that took the bribe in the first period and one that did not. Moreover, we could also use more observations to refine the statistical significance of our results in Section 4.1.3 regarding differences between treatments.

Some aspects that could enrich our understanding of the picture are studying the value of the time discount and introducing a method to study learning between rounds. Studying the time discount δ helps us determine $\bar{\theta}$ and, thus, the distribution of the types of the *Public Officials* θ. Introducing a method to study learning would enlighten the results that we found that varied across the rounds of the game, and see how much of that learning process affects the players' decisions.

Another extension could be considering a variable attention factor λ. It could depend on the role of the players, since *Regular Citizens* appear to be more distracted than *Public Officials*. Not only that, but it could also be a characteristic of each particular subject. Thirdly, the attention factor could also change between the first and the second periods, as the *responsibility* that the *Elected Public Official* has is higher than the one the *Public Officials* have in the first period.

The type of the *Public Officials* is something that we were not able to make them reveal explicitly with our design of this experiment. Obviously, achieving to do so would be a great improvement to the model.

Finally, two last things that could be added to expand the model. The first one is introducing a probability of punishment when a *Public Official* takes a bribe, to simulate how real justice works. Of course, this would have to be adjusted to different scenarios, but could help us understand different behaviors of politicians across countries where justice has more or less public authority. The second one is to increase the competition and add more candidates to the elections to see if corruption increases or decreases.

References

César Martinelli. Accountability and Grand Corruption. 2022.

Claudio Ferraz and Frederico Finan. Exposing corrupt politicians: the effects of Brazil's publicly released audits on electoral outcomes. *The 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.

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.

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

César Martinelli and Thomas R Palfrey. Communication and information in games of collective decision: A survey of experimental results. In *Handbook of Experimental Game Theory*. Edward Elgar Publishing, 2020.

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

Danila Serra. Combining top-down and bottom-up accountability: evidence from a bribery experiment. *The journal of law, economics, & organization*, 28(3): 569–587, 2012.

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.

Danila Serra and Leonard Wantchekon. Experimental research on corruption: Introduction and overview. In *New advances in experimental research on corruption*. Emerald Group Publishing Limited, 2012.

Martin J Osborne and Ariel Rubinstein. *A course in game theory*. MIT press, 1994.

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

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

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.

Glenn W Harrison and E Elisabet Rutström. Risk aversion in the laboratory. In *Risk aversion in experiments*. Emerald Group Publishing Limited, 2008.