1-13-2008

Essays on Corruption and Preferences

Angelino Casio Viceisza

PERMISSION TO BORROW

NOTICE TO BORROWERS

All dissertations deposited in the Georgia State University Library must be used only in accordance with the stipulations prescribed by the author in the preceding statement.

The author of this dissertation is:
Angelino Casio Giovanni Viceisza
215 Piedmont Avenue NE, Apt. 302
Atlanta, GA 30308

The director of the dissertation is:
Dr. James C. Cox
Andrew Young School of Policy Studies
Georgia State University
P. O. Box 3992
Atlanta, GA 30302-3992

Users of this dissertation not regularly enrolled as students at Georgia State University are required to attest acceptance of the preceding stipulations by signing below. Libraries borrowing this dissertation for the use of their patrons are required to see that each user records here the information requested.

Name of User	Address	Date	Type of use (Examination only or copying)

ESSAYS ON CORRUPTION AND PREFERENCES
BY
ANGELINO CASIO GIOVANNI VICEISZA

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree
of
Doctor of Philosophy
in the
Andrew Young School of Policy Studies
of
Georgia State University

GEORGIA STATE UNIVERSITY
2007

KORUPSHON

BY

ANGELINO CASIO GIOVANNI VICEISZA

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree
of
Doctor of Philosophy
in the
Andrew Young School of Policy Studies
of
Georgia State University

GEORGIA STATE UNIVERSITY
2007

ACCEPTANCE

This dissertation was prepared under the direction of the candidate's Dissertation Committee. It has been approved and accepted by all members of that committee, and it has been accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics in the Andrew Young School of Policy Studies of Georgia State University.

Dissertation Chair:	Dr. James C. Cox
Committee:	Dr. Susan K. Laury
	Dr. Jorge L. Martinez-Vazquez
	Dr. Charles N. Noussair
	Dr. Ragan A. Petrie
	Dr. Mary Beth Walker
	Dr. Yongsheng Xu

Electronic Version Approved:
James R. Alm, Dean
Andrew Young School of Policy Studies
Georgia State University
December 2007

ACKNOWLEDGMENTS

In any work of this magnitude, there is an endless list of individuals and entities that one could thank. Unfortunately, this is not possible due to limited space. So, I would like to start by thanking those people that have supported me over the years, whom I am unable to thank by name. You all know who you are.

Having said that, there is a selected group of people that has been instrumental in you being able to read the current text. I would like to start with my dissertation committee. First of all, a special thanks to the chair of my committee–Dr. James C. Cox–who supported me in numerous ways. Jim: You have been great to work with both academically and personally. And, my only regret is that we did not get to work with each other for a longer period! However, as my father likes to say: *"The night is still young and so are we!"* To Dr. Ragan A. Petrie, you have inspired me more than you will ever know! You have invested an enormous amount of time in my overall development as well as the development of this text. Suffice it to say that words are only the beginning of a proper gratitude! To Dr. Susan K. Laury, your precision and detail when providing comments has been uncanny and I am much obliged. I also thank you for four great years as my GRA advisor. To Dr. Jorge L. Martinez-Vazquez, your expertise in numerous areas of the economics literature (not just public economics and finance) has been and continues to be admirable. I thank you for having lended your expertise to me both prior to as well as during my dissertation days. To Dr. Mary Beth Walker, I thank you for having kept me econometrically and generally sound. I especially thank you for our chats both on and off econometrics-related topics. To Dr. Yongsheng Xu, your knowledge is impeccable. You are the only professor from whom I took four courses while at Georgia State. I have learned so much from you over the years and I look forward to continue learning more! Finally, to Dr. Charles N. Noussair, you may have been physically far; however, your comments and knowledge were not. I thank you for having given me the opportunity to learn from you. We shall meet again.

In addition to my committee, there are some additional academics and staff that have had a direct influence on the current state of this dissertation. Dr. Robin Boadway, Dr. Marco Castillo, Dr. Andreas Ortmann, Dr. Vjollca Sadiraj, Dr. Máximo Torero, Dr. Lise Vesterlund, Dr. Sally Wallace and Dr. Frans van Winden; I thank you all for your well-received comments. I also thank seminar participants at the Federal Reserve Bank of Atlanta's Finance Team, the Georgia State University Economics Department, the International Food Policy Research Institute (IFPRI), the 2007 joint LACEA-LAMES meetings and the First World Meeting of the Public Choice Societies for helpful comments. I am also indebted to Krawee Ackaramongkolrotn, Dr. Artidiantun Adji, Magdalena Benza, Roberta Calvet, Cesar Ciudad (and GRADE, Perú), Dr. Urs Fischbacher, Dr. Benjamin Greiner, Daniel Hall, William Holmes, Sarah Jacobson, Dr. Christine Moloi, Dr. Carmit Segal, Amanda

Wilsker and Ming-Hung Yao for their assistance with realizing the experiments. To George Pearson and John MacKendrick at MacKichan, thanks! My LaTeX-skills are elementary and, the two of you always made time to answer my questions.

To those who made this financially possible. While at Georgia State, I would explicitly like to thank: Dr. James C. Cox and the Experimental Economics Center for overall support; Dr. Ragan A. Petrie and Dr. Máximo Torero (at IFPRI) for support in realizing the mail experiment; Dr. Sally Wallace and the Fiscal Research Center for helping me "survive" during the second year of the doctoral program; the Georgia State Economics Department for continued support over the years and the Georgia State dissertation grant program for special dissertation support. Furthermore, I would like to thank the Dean and Faculty of Social and Economic Studies at the University of the Netherlands Antilles (in particular, drs. Renny Maduro) for their support during the summer of 2005. Before my time at Georgia State, a special thanks goes to: Mr. Gregory Elias and the Gregory Elias Foundation, and Mr. Tyrone Alberto and United Telecommunication Services. Both these individuals and their institutions made my days at Boston University possible and without those days there could not have been days at Georgia State!

On a final note, I would like to thank family and friends who have provided broad support over the years. Renny Maduro, you are thanked for convincing me that a doctorate in economics was indeed the right place for me. I am done; now it is your turn! Wilfred "Freddy" Curiel and Joke "Mrs. Joke" Curiel; you and your whole family have always been supportive. You are great! Dr. Michael G. Papaioannou (at the International Monetary Fund) and Dr. Daniel C. Weiner (at Boston University); you are both recognized for your undoubted support and advice. Both of you trusted in me when I did not necessarily trust in myself. We all need this type of support at some point in life. To families Viceisza, Virginia, Romondt, Richards, Villanueva, Griffith, Magdalena, Calmes, Chirino and Gonet; may God bless you richly! You may not have been here in person at all times; however, your positive energy was always heart-felt. To Deka and Nahim, thanks for having supported my father when I was unable to. You will be rewarded! To Glenn, Nicholaus, Ashley and Adan, thanks for the happy times and thanks for bringing joy to my sister's life. Before you know it, we will be sending the kids off to college! Johnny and Mary: You have not been forgotten! Thanks for your support over the years. The movies, dinners, laughs and tears all made Atlanta feel like home!

Finally, I would to thank four individuals that were there from the very start. First, to my grandmother, Aura Richards-Villanueva–the only "square" woman in the world! You have always wanted to write a book. We may not have been able to write your book yet, but I have done my best at writing my own. After all, any of my accomplishments are as much yours as they are mine! To my sister, Fiorella Viceisza Griffith. What can I say! You have always been great and you will always be great! Thanks for the laughs, the advice, the support, the attention, the talks, you name it!...We shall talk on the phone in about five minutes. To my father, Gio-

vanni Arturo Viceisza, Ven: As you always correctly point out, you have been there through it all. This is undoubtedly your dissertation as much as it is mine. You deserve great things and I know that one day soon you will receive them! You are my father and my friend and you will forever remain both. Finally, I would like to dedicate this dissertation to my mother, the late Gladijs Felomena Viceisza-Virginia. You may no longer be with us in person, but you will *forever* be with us in spirit.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

ESSAYS ON CORRUPTION AND PREFERENCES

BY

ANGELINO CASIO GIOVANNI VICEISZA

December 2007

Committee Chair: Dr. James C. Cox

Major Department: Economics

This dissertation comprises three essays. The theme that unifies them is "experiments on corruption and preferences." The first essay (chapter 2) reports theory-testing experiments on the effect of yardstick competition (a form of government competition) on corruption. The second essay (chapter 3) reports theory-testing experiments on the effect of efficiency and transparency on corruption. Furthermore, this essay revisits the yardstick competition question by implementing an alternative experimental design and protocol. Finally, the third essay (chapter 4) reports a theory-testing randomized field experiment that identifies the causes and consequences of corruption.

The first essay finds the following. Theoretically, the paper derives a main proposition which suggests that institutions with more noise give rise to an increase in corrupt behavior and a decrease in voter welfare. Empirically, the paper finds a few key results. First, there are an initial nontrivial proportion of good incumbents in the population. This proportion goes down as the experiment session progresses. Secondly, a large proportion of bad incumbents make theoretically inconsistent choices given the assumptions of the model. Third, overall evidence of yardstick competition is mild. Yardstick competition has little effect as a corruption-taming mechanism when the proportion of good incumbents is low. Namely, an institution that is characterized by a small number of good incumbents has little room for

yardstick competition, since bad incumbents are likely to be replaced by equally bad incumbents. Thus, incumbents have less of an incentive to build a reputation. This is also the case in which (1) yardstick competition leads to non-increasing voter welfare and (2) voters are more likely to re-elect bad domestic incumbents. Finally, a partitioning of the data by gender suggests that males and females exhibit different degrees of learning depending on the payoffs they face. Furthermore, male voter behavior exhibits mild evidence of yardstick competition when voters face the pooling equilibrium payoff.

The second essay finds the following. First, efficiency is an important determinant of corruption. A decrease in efficiency makes it more costly for incumbents to "do the right thing." This drives them to divert maximum rents. While voters retaliate slightly, voters tend to be worse off. Secondly, increased lack of a particular form of transparency (as defined in terms of an increase in risk in the distribution of the unit cost) leaves corrupt incumbent behavior unchanged. In particular, if the draw of the unit cost is unfavorable, incumbents tend to be less corrupt. Third, there is strong evidence of yardstick competition. On the incumbent's side, yardstick competition acts as a corruption-taming mechanism if the incumbent is female. On the voter's side, voters are less likely to re-elect the incumbent in the presence of yardstick competition. Specifically, voters pay attention to the difference between the tax signal in their own jurisdiction and that in another. As this difference increases, voters re-elect less. This gives true meaning to the concept of "benchmarking." Finally, the analysis sheds light on the role of history and beliefs on behavior. Beliefs are an important determinant of incumbents' choices. If an incumbent perceives a tax signal to be associated with a higher likelihood of re-election, he is more likely to choose it. On the voter's side, history tends to be important. In particular, voters are more likely to vote out incumbents as time progresses. This suggests that incumbents care about tax signals because they provide

access to re-elections while voters use the history of taxes and re-elections in addition to current taxes to formulate their re-election decisions.

Finally, the third essay finds the following. First, 19.08% of mail is lost. Secondly, money mail is more likely to be lost at a rate of 20.90% and this finding is significant at the 10% level. This finding suggests that loss of mail is systematic (non-random), which implies that this type of corruption is due to strategic behavior as opposed to plain shirking on the part of mail handlers. Third, we find that loss of mail is non-random across other observables. In particular, middle-income neighborhoods are more likely to experience lost (money) mail. Also, female heads of household in low-income neighborhoods are more likely to experience lost mail while female heads of household in high-income neighborhoods are much less likely to experience lost (money) mail. Finally, this form of corruption is costly to different stakeholders. The sender of mail bears a direct and an indirect cost. The direct cost is the value of the mail. The indirect cost is the cost of having to switch carriers once mail has been lost. Corruption is also costly to the intended mail recipient as discussed above. Finally, corruption is costly the mail company (SERPOST) in terms of lost revenue and to society in terms of loss of trust. Overall, the findings suggest that public-private partnerships need not increase efficiency by reducing corruption; particularly, when the institution remains a monopoly. Increased efficiency in mail delivery is likely to require (1) privatization and (2) competition; otherwise, the monopolist has no incentive to provide better service and loss of mail is likely to persist.

INTRODUCTION

This dissertation comprises three essays. The theme that unifies them is "corruption and preferences." In recent decades, corruption has taken a primary role on the agenda of policymakers, international non-governmental organizations and academia. Economic analyses of corruption have warranted a broad-based approach. From theories of corruption (Aidt 2003) to empirical models of corruption (Martinez-Vazquez, Arze and Boex 2007 and Tanzi 1998), the literature has grown in important and significant ways.

The first essay in this dissertation studies the extent to which yardstick competition acts as a corruption-taming mechanism in an experimental environment. It contributes to the literature in a few significant ways. First, it provides experimental data on the effect of yardstick competition on incumbent and voter behavior and voter welfare. This is done in a controlled laboratory setting such that econometric identification is relatively simple. In particular, the results can be compared with those of Besley and Case, Dincer, Ellis and Waddell and Johnson in order to say something about internal versus external validity of the theory.

Secondly, the study suggests ways in which to fortify the theory in order to use it as a more meaningful tool for prediction. By conducting a strict theory-testing experiment, I am able to identify some key areas in which the theory can be extended to better capture agent behavior. Specifically, the results suggest that theoretical assumptions on preferences and types (in games) are crucial in deriving meaningful predictions.

Third, the study contributes to the literature on experimental signaling (possibly in games), which includes but is not limited to studies on the plausibility of sequential equilibria in signaling games (see for example Brandts and Holt 2005), studies exploring the lemons phenomenon in markets (see for example Miller and

Plott 1985) and other studies on signaling in different contexts such as voluntary contributions (see for example Potters, Sefton and Vesterlund forthcoming), reputation building (see for example Grosskopf and Sarin 2006) and board composition and behavior (see for example Gillette, Noe and Rebello 2003).

Finally, the study contributes to our understanding of yardstick competition as a mechanism for reducing corruption. This is relevant, since it informs policy design and analysis not only in public economics contexts but also in other contexts. Namely, yardstick competition is a mechanism–sometimes also referred to as "benchmarking"–that can be applied in many other contexts (see for example Holmstrom 1982 and Schleifer 1985).

The main findings are the following. First, there is an initial nontrivial proportion of good incumbents in the population. This proportion goes down as the experiment session progresses. Secondly, a large proportion of bad incumbents make theoretically inconsistent choices given the assumptions of the model and the manner in which preferences are theoretically specified. Finally, there is mild evidence of yardstick competition. In particular, an institution with low proportion of good incumbents has little room for yardstick competition, since bad incumbents are likely to be replaced by equally bad incumbents. This is also the case in which (1) yardstick competition leads to non-increasing voter welfare (as is observed in these experiments) and (2) voters are more likely to re-elect bad domestic incumbents due to the presence of equally bad foreign incumbents.

The second essay in this dissertation uses laboratory experiments to address three main questions related to the determinants and control of corruption. First, the study asks whether an increase in the marginal cost of public funds gives rise to more or less corruption. Secondly, the study asks whether a decrease in transparency (as modeled by a mean-preserving spread of the distribution of the unit cost of public good provision) affects corrupt behavior. Finally, the study returns to

the main question addressed in Viceisza (2007a), which is whether an experimental environment with yardstick competition gives rise to more or less corruption.

To understand the contributions of this study, it is useful to set the context by focusing on two of Aidt's (2003) main claims. Aidt distinguishes between four types of theoretical studies on corruption: (1) those that model efficient corruption, (2) those that model corruption in the presence of a benevolent principal (e.g., the voter) that delegates decision-making power to a non-benevolent agent (e.g., the incumbent), (3) those that model corruption with a non-benevolent principal and (4) those that model self-reinforcing corruption in that history may play a role. Also, Aidt posits that two important considerations unite these four categories of models of corruption: (1) the degree of benevolence of the government official and (2) the role of institutions versus history as a determinant of corruption.

These arguments are cited here because they provide the main rationale for conducting the experiments reported in this study. Namely the theoretical model that guides the experimental design is a prime example of studies pertaining to Aidt's second category. In particular, Besley and Smart's (2007) theoretical political agency model is one in which a principal (the voter) delegates decision-making authority to a potentially non-benevolent agent (the incumbent or the challenger) by means of an election. So, these theory-testing experiments provide laboratory experimental data that is relevant to a large class of models as categorized by Aidt.

Furthermore, by experimentally operationalizing Besley and Smart's model, this study explores the validity of the two considerations that unite Aidt's four categories. With regard to the first consideration, the experimental treatments explore the extent to which an agent's benevolence (i.e., the incumbent politician's corruptibility) is affected by two key changes: (1) an increase in the marginal cost of public funds, which represents an increase in tax inefficiency and (2) an increase in risk (noise) in the distribution of the unit cost of the public good, which represents a

form of lack of transparency on the incumbent's side. With regard to the second consideration, the experimental treatments explore the rivalry between a potential corruption-taming institution (i.e., yardstick competition) and "history."

These experimental treatments are also of further independent interest since the data resulting from them have policy implications for developing countries. Since the marginal cost of public funds–which is a measure of tax efficiency–is closely related to the choice of tax instruments within a country (Dahlby 1998), the experimental marginal cost treatment is interesting since it sheds light on the extent to which an external increase in the marginal cost of public funds–possibly caused by a change in tax mix–affects corruptibility. This has policy implications for developing countries since these are known to have relatively different tax mixes compared to developed countries and under certain circumstances can be hypothesized to have higher marginal costs of public funds (Warlters and Auriol 2005.)

Furthermore, if the distribution of the unit cost of public good provision is taken to represent transparency in the political system, analyzing the effects of a less transparent institution on corrupt behavior is clearly of interest to developing economies. So, the experimental treatments that are characterized by a mean preserving spread of the distribution of the unit cost of the public good (i.e., a noisier distribution and therefore, a less transparent institution) address the question how a certain type of transparency affects corruption.

Finally, we explore the interaction between yardstick competition and "history," which is relevant to all economies–developing and developed alike. While Viceisza (2007a) has reported some experiments addressing this question, the experimental treatments conducted in this study are based on a different design and protocol.

In any type of theory-testing experiment, there are numerous ways to operationalize the assumptions of the theory and very few are the theories that provide sufficient (institutional) detail such that all ambiguities with regard to their imple-

mentation is ruled out. Viceisza (2007a) found that using a within-subjects design to test the yardstick hypothesis does not lead to significant treatment effects. As discussed in the conclusion to that study, part of this may be due to the within-subjects nature of the experimental design.

So, this new set of experiments asks whether a between-subjects design gives rise to any significant treatment effects. Furthermore, the experimental design and protocol are amended to reflect some crucial changes. First, these experiments operationalize the distribution of the unit cost of the public good using a physical device. Secondly, the experiments reduce the information set arising from yardstick competition to one random choice made by another politician as opposed to a distribution of choices made by all politicians. Finally, the experiments implement a trial, quiz and summary of the task to enhance subject understanding in addition to a post-questionnaire that can be used to control for it.

The study finds the following main results. First, tax inefficiency is an important determinant of corrupt incumbent behavior. In particular, an increase in the marginal cost of public funds makes it more costly for incumbents to equalize first-period payoffs. This drives them to separate and divert maximum rents in the first period. While voters retaliate slightly by ousting incumbents, they are worse off.

Secondly, we find that increased lack of a particular form of transparency (as defined in terms of an increase in risk in the distribution of the unit cost) leaves corrupt incumbent behavior unchanged. If the draw of the unit cost is unfavorable, incumbents tend to be less corrupt. So, the results suggest that lack of transparency (as defined in Viceisza 2007a) need not always make voters worse off.

While this finding may seem counterintuitive, it is not given the parameterization of the experiments. First, we must not ignore the importance of assumptions on types and preferences as discussed in Viceisza (2007a). Contrary to Besley and Smart's (2007) assumption that incumbents are either good or bad, incumbents are

known to behave strategically. Since the mean-preserving spread in the distribution of the unit cost reduces the cost of equalizing first-period payoffs in the favorable state, it makes sense why incumbents behave less corruptly if the unit cost is low. Furthermore, Viceisza (2007a) models lack of transparency on the incumbent side. This can be contrasted with lack of transparency on the voter's side as discussed in Besley and Smart (2007).

Third, the experiments find strong evidence of yardstick competition. On the incumbent's side, yardstick competition acts as a corruption-taming mechanism if the incumbent is female. On the voter's side, voters are less likely to re-elect the incumbent in the presence of yardstick competition. Specifically, voters pay attention to the difference between the tax signal in their own jurisdiction and the signal in another jurisdiction. As this difference increases, voters re-elect less.

Fourth, history is an important determinant of corruption and of re-election decisions. Incumbents are likely to make choices as they did in previous repetitions and voters are likely to vote out increasingly as the repetitions go by. In other words, they distrust the political system more significantly as time goes by and yardstick competition does not affect that. Finally, we find that individual-specific factors such as gender and beliefs play a significant role in incumbent behavior. In particular, female incumbents are more likely to divert rent when the unit cost is unfavorable. Furthermore, incumbents' beliefs are more important than voters' beliefs in decision-making. Voters focus mainly on taxes charged (i.e., payoffs).

Finally, the third essay in this dissertation reports a theory-testing randomized field experiment that makes use of an existing public institution–the postal system in Lima, Perú–to identify and measure corruption, its causes and its consequences. Furthermore, the findings suggest policy implications that could be effective in controlling this type of corruption. Thus, our study contributes to the literature by providing contextual answers to four of the above five questions.

The study is in the same area of the literature as Olken (2007) and Bertrand et al. (forthcoming); namely, randomized field experiments that aim at identifying and measuring the causes and consequences of corruption. Thus, it contributes to a fast-emerging literature on field experiments that are of interest to policymakers in general with a particular focus on development (Duflo 2006) and experimental economists (Harrison and List 2004.)

Specifically, we implement a (2x2)-design in which each household within a stratified sample of households across Lima is sent four envelopes: (1) an envelope with money with same sender and recipient last name, (2) an envelope without money with same sender and recipient last name, (3) an envelope with money with distinct sender and recipient last name and (4) an envelope without money with distinct sender and recipient last name. The envelopes are sent in four batches (on average, one month apart) and for each household the four envelopes are randomly assigned to batch. The envelopes are also randomly assigned other systematic characteristics, which are elaborated upon below.

We compare receipt of money and non-money envelopes to learn whether loss of mail is nonrandom. Furthermore, we compare receipt across same and different sender/recipient last names to learn whether or not family mail is more likely to be lost because its content is perceived to be of higher value. Also, we control for household-specific characteristics (such as neighborhood welfare and head of household's gender) to learn whether loss of mail affects social groups differently. Finally, we quantify the costs of corruption and loss of efficiency based on the value of lost mail and the costs of an alternatively popular carrier–Western Union.

We find the following. First, 19.08% of mail is lost. Secondly, money mail is more likely to be lost at a rate of 20.90% compared to no-money mail which disappears at a rate of 14.37% (this difference is statistically significant at the 10% level). This finding suggests that loss of mail is systematic (non-random), which im-

plies that this type of corruption is due to strategic behavior as opposed to plain shirking on the part of mail handlers.

Third, we find that loss of mail is non-random across other observables. Corruption is more costly to certain societal groups. Middle-income neighborhoods are more likely to experience lost (money) mail (these differences are statistically significant at the 10% level). These findings are informed by the predictions of the model, which suggest that mail handlers may perceive the benefits (costs) of lost mail to be higher (lower) in middle-income neighborhoods compared to low- and high-income neighborhoods.

We also find that female heads of household in low-income neighborhoods are more likely to experience lost mail. This is not necessarily a surprising result, since this group may be perceived to be vulnerable (this difference is statistically significant at the 10% level). Furthermore, female heads of household in high-income neighborhoods are much less likely to experience lost (money) mail (these differences are statistically significant at the 5% level).

Finally, this form of corruption is costly to different stakeholders. The sender of mail bears a direct and an indirect cost. Corruption is also costly to the intended mail recipient. Middle-income neighborhoods and female heads of household in low-income neighborhoods are more likely to suffer. Finally, corruption is costly the mail company (SERPOST) in terms of lost revenue and to society in terms of loss of trust as a form of social capital (Fukuyama 1995). Overall, the findings suggest that public-private partnerships need not increase efficiency by reducing corruption; particularly, when the institution remains a monopoly. Increased efficiency in mail delivery is likely to require (1) privatization and (2) competition.

The remainder of this dissertation is organized as follows. Chapters 2, 3 and 4 represent the first, second and third essays respectively. The back matter of the dissertation comprises the conclusion, appendices, references and vita.

AN EXPERIMENTAL INQUIRY INTO THE EFFECT OF YARD-STICK COMPETITION ON CORRUPTION

Introduction

According to the Organization for Economic Co-operation and Development, the Nigerian President recently estimated the total cost of corruption at a quarter of Africa's total income.[1] Nigeria is not alone. As suggested by Transparency International's indices, many other countries (particularly, developing countries) all over the world are coping with this problematic phenomenon.[2]

The nature and causes of corruption have been studied in many contexts. Some general survey pieces include Martinez-Vazquez, Arze and Boex (2004), Aidt (2003), Tanzi (1998) and Bardhan (1997). Much like any other form of (economic) behavior, corrupt behavior can be seen as the result of an interaction between an agent's environment and the institution in which such agent makes decisions.[3] In particular, it is interesting to know how (changes in) institutions affect corrupt behavior in key areas of society such as politics and the economy.

Recently, economists have become interested in almost self-correcting mechanisms for (political) corruption that fall under the category of fiscal decentralization (see for example Bardhan and Mookherjee 2005). The main question addressed in this study is related to one such mechanism. We ask whether an experimental insti-

[1] The World Bank (1997) defines corruption as: "The abuse of public office for private gain."

[2] Corruption is termed problematic because there is generally widespread consensus among (social) scientists across many disciplines that corruption is costly to society, since it hinders economic growth and thus promotes poverty and income inequality. See for example Mauro (1995).

[3] There are many ways to define the terms environment and institution. This paper adopts Smith's (1989) definitions. The environment consists of the collection of all agents' characteristics; that is, tastes and technology, which in traditional economics are represented by utility or preference functions, resource endowments and production or cost functions. The institution defines the language (messages or actions) of communication. The institution also specifies the order in which economics agents move, or that there is no form (moves are free form), and the rules under which messages become contracts and thus allocations. Culture can also be seen as part of an agent's environment. Even though culture may have an effect on corrupt behavior, this is not a central question in this paper.

tution with yardstick competition (a form of government competition) has an effect on corruption and if so, in what direction.[4] In particular, we adopt the Besley and Smart (2007) political agency model to conduct a theory-testing experiment on this main question.

The main hypothesis is that since yardstick competition gives rise to an information externality that amplifies the voter's information set, the voter is generally better able to judge incumbent behavior in the presence of yardstick competition. The bad incumbent being aware of this will restrain corrupt behavior in an attempt to gain re-election. Thus, it can be argued under reasonable assumptions that yardstick competition reduces corruption and in most cases, increases voter welfare.

Theoretically, the link between yardstick competition and corruption has been addressed in several studies including but not limited to Besley and Smart (2007), Belleflamme and Hindriks (2005), Bodenstein and Ursprung (2004) and Bordignon et al. (2004). The results do not always agree. For example, Besley and Smart and Belleflamme and Hindriks both study the main question using a political agency model of elections with the possibility for moral hazard and adverse selection. Yet, they find opposite effects.

Since their models are setup to study the combined effect of adverse selection and moral hazard, the role for an election is twofold: First, to restrain rent diversion by bad incumbents as in the pure moral hazard problem (the discipline effect) and secondly, to separate bad incumbents from good incumbents as in the pure ad-

[4]Yardstick competition is a form of government competition as reviewed in Viceisza (2004). The government competition literature in public economics includes models of tax, tax base, fiscal and yardstick competition. For a published review on models of tax competition, see for example Wilson (1999). Besley and Case (1995a) were one of the first to define yardstick competition in public economics. *Yardstick competition between incumbent politicians arises when voters in one jurisdiction use the performance of incumbents in other jurisdictions as a benchmark.* In this study, we are neither concerned with the accuracy of the information transmission mechanism nor with the mechanism itself. In particular, we are not concerned with how information is distributed to voters or whether or not such information is correct. Since these separate questions lead into the role of media objectivity and media transparency which are beyond the scope of this study, we control these aspects (flow and accuracy of information) as part of the experimental design.

verse selection case (the selection effect). In the Besley and Smart model, these can be two competing effects. While "mimicking" reduces rent-seeking in the short run (the discipline effect), it also reduces selection between good and bad politicians in the long run (the selection effect). In the Belleflamme and Hindriks model, however, the electoral incentive and the desire for re-election on the part of the bad incumbent politician are themselves the source of inefficiency. As a consequence, the two effects are reinforcing and yardstick competition always improves voter welfare by taming corrupt behavior.

Empirically, the link between yardstick competition and corruption has been addressed using field data; however, it has not been addressed experimentally. Besley and Case (1995a), Dincer, Ellis and Waddell (2006) and Johnson (2006) have all studied yardstick competition using U.S. field data. While the first two studies found evidence of yardstick competition, the third study finds only very limited evidence of yardstick competition. So, the empirical results are mixed.

Furthermore, since field data are–by their very nature–noisy and based on an unknown data-generating process, it is not straightforward to econometrically identify a causal effect of yardstick competition. This poses a significant difficulty when pinpointing as subtle a phenomenon as yardstick competition, which can relatively easily be confounded with other forms of government competition.

Experimentally, as reviewed in Abbink (2005b) and Dusek, Ortmann and Lizal (2005), no studies have addressed the main question, which is the effect of yardstick competition on corruption. While Potters et al. (2004) address the question of collusion under yardstick competition, they do not address the main question as such.

This study thus contributes to the literature in a few significant ways. First, it provides experimental data on the effect of yardstick competition on incumbent and voter behavior and voter welfare. This is done in a controlled laboratory setting such that econometric identification is relatively simple. In particular, the re-

sults can be compared with those of Besley and Case, Dincer, Ellis and Waddell and Johnson in order to address internal versus external validity of the theory.

Secondly, the study suggests ways in which to fortify the theory in order to use it as a more meaningful tool for prediction. By conducting a strict theory-testing experiment, we are able to identify some key areas in which the theory can be extended to better capture agent behavior. Specifically, the results suggest that theoretical assumptions on preferences and types (in games) are crucial in deriving meaningful predictions.

The general notion of operationally meaningful theorems is of course not new and dates back to Samuelson (1947). However, recently it has been addressed by Weibull (2004) and Cox (2004) in the context of testing game theory when using experimental methods. Both authors allude to–among other issues–the importance of assumptions on preferences when interpreting experimental data and results. This is particularly relevant in other-regarding-preferences-like experiments.[5]

Third, the study contributes to the literature on experimental signaling (possibly in games), which includes but is not limited to studies on the plausibility of sequential equilibria in signaling games (see for example Brandts and Holt 2005), studies exploring the lemons phenomenon in markets (see for example Miller and Plott 1985) and other studies on signaling in different contexts such as voluntary contributions (see for example Potters, Sefton and Vesterlund forthcoming), reputation building (see for example Grosskopf and Sarin 2006) and board composition and behavior (see for example Gillette, Noe and Rebello 2003).

Finally, the study contributes to our understanding of yardstick competition as a mechanism for reducing corruption. This is relevant, since it informs policy design and analysis not only in public economics contexts but also in other con-

[5] Cox (2004) defines the term other-regarding preferences as opposed to self-regarding preferences (i.e., traditional economic man preferences). Basically, experiments on other-regarding preferences include but are not limited to experiments on fairness, trust and reciprocity; the so-called social preferences literature.

texts. Namely, yardstick competition is a mechanism–sometimes also referred to as "benchmarking"–that can be applied in many other contexts (see for example Holmstrom 1982 and Schleifer 1985).

Despite these main contributions, the study has some limitations–some of which are addressed in Viceisza (2007c). First, the experimental design is based on a within-subject design.[6] Secondly, we have not conducted any sensitivity-type treatments.[7] Finally, the experimental design does not induce types. In other words, experimental subjects are allowed to enter the experiments with their own homegrown types.[8] This may impact the effect of yardstick competition and will be elaborated upon later in the paper.

The main findings are the following. First, there is an initial nontrivial proportion of good incumbents in the population. This proportion goes down as the experiment session progresses. Secondly, a large proportion of bad incumbents make theoretically inconsistent choices given the assumptions of the model and the manner in which preferences are theoretically specified. Finally, there is mild evidence of yardstick competition. In particular, an institution with low proportion of good incumbents has little room for yardstick competition, since bad incumbents are likely to be replaced by equally bad incumbents. This is also the case in which (1) yardstick competition leads to non-increasing voter welfare (as is observed in these experiments) and (2) voters are more likely to re-elect bad domestic incumbents due to the presence of equally bad foreign incumbents.

The remainder of the paper is organized as follows: The next section discusses the theory of yardstick competition proposed by Besley and Smart and a main proposition on transparency. Then, we discuss the experimental design and how it relates to the theory. Thereafter, we summarize the findings and conclude.

[6] This is opposed to a between- or across-subject design (see Viceisza 2007c).
[7] The data from these treatments are reported by Viceisza (2007c).
[8] An induced-type treatment may be part of an upcoming paper.

Theoretical Model

Game Description: Players, Information, Actions and Payoffs

Players and Periods. The theory is based on a game theoretic model of elections that is cast in a principal-agent framework. There are two "active" players–a principal (the voter) and a first agent (the first-period incumbent), and one "passive" player–a second agent (the challenger). An agent's type (i) can be good (g) or bad (b). Each agent knows his or her own type; however, the principal and the other agent do not. The model comprises two periods. At the end of the first period, there is an election. The second agent is "passive," since this agent only plays a role during elections and during the second period if elected.[9]

The above describes a dynamic game of incomplete information. It is dynamic because it encompasses two periods and decisions made in the first period affect outcomes in the second period. There is incomplete information, since at least one player is uncertain about another player's payoffs. In particular, the voter is uncertain about the politicians' payoffs, since such payoffs are type-dependent.

Information Revelation and Actions. Following Harsanyi (1967, 1968), the game can be transformed into a dynamic game of imperfect information by letting the timing and information revelation proceed as follows.

First Period

1. Nature draws the types of the incumbent and the challenger from i.i.d. distributions such that $\Pr(i = g) = \pi$. This information is common knowledge.

2. Nature reveals the politician's type only to the respective politician.

3. Nature draws the unit cost of public good provision θ from i.i.d. distributions

[9] If elected, the challenger will behave in one of the following ways: If good, the challenger will maximize voter welfare; if bad, the challenger's behavior will be dictated by the lame-duck effect.

such that $\Pr(\theta = H) = q$, $\Pr(\theta = L) = 1 - q$, $H > L$. This information is common knowledge.

4. Nature reveals the level of θ to the incumbent politician but not to any other player.

5. The "active" agent's action set. After observing θ, the incumbent politician chooses the level of the public good (G) and the amount of rent diversion (r), which together determine the total tax collection, $t = \theta G + r$. Under the assumption that there is a maximum feasible level of government tax collections (T) in any period, the incumbent politician's action set A_I is $A_I \equiv \{(G, r) \,|\theta G + r = t \in [0, T], G \geq 0, r \geq 0\}$. This sets the maximum level of rent diversion at $\overline{r} = T$.

6. The voter observes the level of the public good (G) and the total tax collection (t) and pays t. The voter does *not* observe the rent diverted.

7. The principal's action set. At the end of the first period, after the voter has observed (G, t), there is an election. The voter chooses to re-elect or not by comparing the signals received from the "active" agent (first-period incumbent) with the prior belief that the "passive" agent (challenger) is good or bad.[10] So, the voter's action set $A_V \equiv \{\text{re-elect, not re-elect}|\Psi\}$, where Ψ is some information set of the form $\{G, t, \pi, q\}$.

8. If the voter re-elects, the incumbent politician is in office again. If the voter does not re-elect, the challenger is elected to office.

[10] The model assumes that *one* voter determines whether or not the incumbent is re-elected to office. While this may seem contrary to standard majority-voting models, we can interpret this model as a median voter or representative voter type model in which the median or representative voter is decisive.

Second Period

1. The type of the second-period incumbent politician is determined.
2. Nature draws the unit cost of public good provision θ as described previously.
3. Nature reveals the level of θ only to the second-period incumbent.
4. The second-period incumbent chooses (G, r).
5. The voter observes (G, t), pays t and does not get to vote.
6. The game comes to an end.

Players' Payoffs. The players' payoff functions are as follows:

1. Voter welfare is given by $W(G,t) = G - \mu C(t)$, where G, t are defined as before, μ is an exogenous parameter indexing the marginal cost of public funds and C is a strictly convex and increasing function with $C(0) = 0$. [11,12]
2. Politician welfare is type dependent:[13]
   (a) A good politician only cares about voter welfare and thus she has the same welfare function as the voter, $W(G,t) = G - \mu C(t)$.[14]

[11] Browning (1976) defines the marginal cost of public funds as the direct tax burden plus the marginal welfare cost produced in acquiring the tax revenue.

[12] Intuitively, the assumption that the tax cost function is increasing makes sense. The higher the tax rate, the higher the cost that is imposed on the voter. However, why must the tax cost function also be strictly convex? This is related to the marginal cost of taxation. The higher the tax rate, the higher the marginal (added) cost of each unit taxed. Mathematically, this gives rise to well behaved (strictly convex) preferences for the voter and the good type.

[13] A politician's complete type is determined by the pair (i, θ).

[14] Lockwood (2005) partly discusses what might happen if the good type also behaves strategically. In particular, he finds that the hybrid equilibrium need not be stable in the Cho and Kreps (1987) sense.

(b) A bad incumbent is not concerned with the provision of the public good and always diverts rent for private purposes.[15]

First-period bad incumbent. Given a discount factor of $0 < \beta < 1$ and a probability of re-election equal to ψ, the incumbent's payoff function is $R = r_1 + \beta\psi r_2$, where R is the discounted expected total rent diversion by the incumbent if re-elected and r_j denotes the rent diverted in period j for $j = 1, 2$.[16]

Second-period bad incumbent. The incumbent's payoff function is equal to the rent diverted in period 2, r_2.

Solution to the Game: Equilibrium Strategies

The game is most easily solved by some type of backward induction argument. This leads to perfect Bayesian equilibria (PBE) of the kind discussed by Fudenberg and Tirole (1991).

Good Incumbent. A good second-period incumbent chooses rent diversion equal to zero and solves:[17]

$$\max_G W\left(G, t\right) \text{ s.t. } t = \theta G \implies \max_G \left[G - \mu C\left(\theta G\right)\right]. \tag{1}$$

[15] This can be motivated as follows: A priori and regardless of the level of θ, the bad incumbent faces an alternative between setting $r_1 = 0$ or $r_1 = \bar{r}$. If the bad incumbent sets $r_1 = 0$, the total two-period payoff is $\beta\bar{r}$ if re-elected and zero otherwise. If the bad incumbent sets $r_1 = \bar{r}$, the total two-period payoff is $(1 + \beta)\,\bar{r}$ if re-elected and $\bar{r}$ otherwise, which is greater than $\beta\bar{r}$ since β is assumed to be strictly less than one. Hence, it is always in the bad type's best interest to choose $r_1 > 0$.

[16] ψ represents the probability of re-election when the voter observes high tax spending, t_H.

[17] Note that under this interpretation an incumbent (good or bad) is not concerned with the choice of t. The level of tax collection is indirectly determined–via budget balancing–by the choice of (G, r), rendering the latter two variables the choice variables.

This maximization problem has the following first-order and second-order conditions:

$$1 - \mu\theta \frac{\partial C}{\partial t}(\theta G) = 0 \quad (2)$$

$$-\mu\theta^2 \frac{\partial^2 C}{\partial t^2}(\theta G) < 0. \quad (3)$$

Statements 2 and 3 imply:[18]

$$G(\theta, \mu) \equiv G_\theta(\mu) = \arg\max_G \left[G - \mu C(\theta G)\right]. \quad (4)$$

By the implicit function theorem:

$$\frac{dG_\theta}{d\theta} = -\frac{-\mu \frac{\partial C}{\partial t}(\theta G) - \mu\theta G \frac{\partial^2 C}{\partial t^2}(\theta G)}{-\mu\theta^2 \frac{\partial^2 C}{\partial t^2}(\theta G)} < 0 \text{ and} \quad (5)$$

$$\frac{dG_\theta}{d\mu} = -\frac{-\theta \frac{\partial C}{\partial t}(\theta G)}{-\mu\theta^2 \frac{\partial^2 C}{\partial t^2}(\theta G)} < 0. \quad (6)$$

So, a good second-period incumbent chooses as follows in the second period and maximizes voter welfare:

$$\{G_L, r_2 = 0 | \theta_2 = L\} \implies t_L = LG_L \quad (7)$$

$$\{G_H, r_2 = 0 | \theta_2 = H\} \implies t_H = HG_H. \quad (8)$$

The associated equilibrium welfare levels are:

$$W_L \equiv W(G_L, t_L) = G_L - \mu C(LG_L) \quad (9)$$

$$W_H \equiv W(G_H, t_H) = G_H - \mu C(HG_H). \quad (10)$$

[18] As a matter of notation, G_θ is associated with $r_j = 0$ and thus a good type. Any other G (such as $G_{-\theta}, G_T$) is associated with $r_j > 0$ and thus a bad type. Any tax level associated with a good type is seen as the true low or true high levels (t_L, t_H respectively) except for when type (b, L) mimicks type (g, H) and spends t_H. In any other circumstance, tax levels associated with bad types are either t_{-H} or T.

A good first-period incumbent behaves identically to a good second-period incumbent. The voter and good-type equilibrium levels can be analyzed graphically as shown in figure 1. Fix $W_\theta = \bar{W}_\theta$, and let $\bar{W}_\theta = G_\theta - \mu C(\theta G_\theta)$ such that $G_\theta = \bar{W}_\theta + \mu C(\theta G_\theta)$. Recalling that $\theta G_\theta = t_\theta$, we can plot such indifference curve in (G, t)-space.

A few important properties can be derived for such indifference map:

1. The indifference curve in (G, t)-space can be represented as an increasing and strictly convex function of t since $C(.)$ is increasing and strictly convex.

2. $W_H \lesseqgtr W_L$ depending on the trade-off between the level of θ and G_θ. These represent two competing forces.[19]

3. Given preferences of the good and the bad type, signaling is more costly for (g, H) than for (g, L). This results in a single-crossing property like that of Spence (1973).[20]

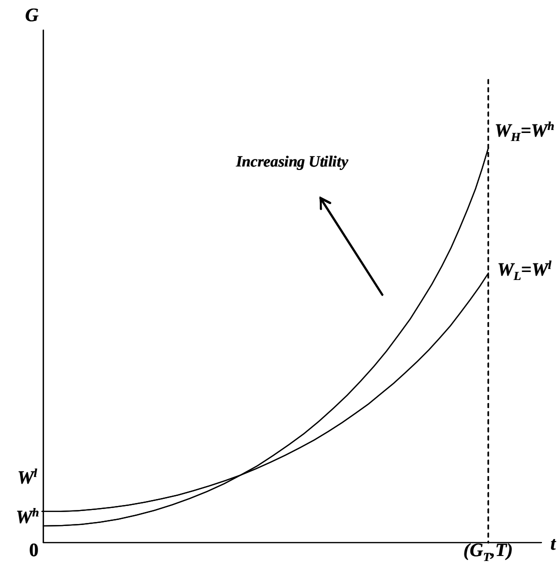
Bad Incumbent and Voter Beliefs. The behavior of a bad second-period incumbent is dictated by the lame-duck effect. The incumbent diverts maximum rents by setting $G = G_T = 0$ and $\bar{r}_2 = t = T$.[21] The associated equilibrium welfare

[19] Recall that $W_H = G_H - \mu C(t_H)$ and $W_L = G_L - \mu C(t_L)$ such that $W_H - W_L = G_H - \mu C(t_H) - (G_L - \mu C(t_L))$, which equals $W_H - W_L = G_H - G_L - (\mu (C(t_H) - C(t_L))) \lesseqgtr 0$ as $t_H \gtreqless t_L$, which in turn depends on the levels of H versus L and G_H versus G_L.

[20] This single-crossing property comes from the fact that the bad type always diverts some rent. So, the economically rational voter knows that whenever she observes (G_L, t_L)–which leads to an equilibrium welfare level of W_L–the incumbent must be of type (g, L). So, it is implicitly more costly for type (g, H) to send a signal than it is for type (g, L). So, the indifference curves associated with equilibrium level W_H are steeper than those associated with W_L.

[21] There is an implicit assumption that a bad incumbent's welfare is unaffected by the choice of G even though he or she lives in the society and will benefit from the public good. This should be clear from the specification of preferences discussed previously. This assumption allows us to set the level of the public good when type b diverts maximal rents (G_T) equal to zero as a result of equilibrium; i.e., since the bad type's welfare is unaffected by the public good, the incumbent loses nothing by providing zero of the public good.

Figure 1: Indifference Map for Voter and Good-type Welfare



level for the voter is:

$$W_T \equiv W\left(G_T = 0, T\right) = -\mu C\left(T\right). \tag{11}$$

This point is represented in figure 1 by the point with coordinates $(G_T = 0, t = T)$. At this point, the voter has minimum utility as calculated above. Note that the voter has no retaliation measure (strategic move) in the second period, since the game comes to an end. The voter pays the associated maximum tax level and endures the associated "disutility."

The behavior of a bad first-period incumbent is cost-dependent. When the cost is high (i.e., $\theta = H$), type (b, H) will charge higher taxes than necessary for provision of the associated public good. While the politician chooses a public good provision level of G_H, a tax higher than t_H must be charged in order for the incumbent to divert rent. The voter observes a tax level that is "too high" and concludes with certainty that a bad type is in office. Thus, in equilibrium the voter does not reelect the incumbent since $t > t_H$ while $G = G_H$. Type (b, H) realizes this and

consequently diverts maximal rents (T) in the first period in anticipation of being voted out of office. The corresponding public good provision level is thus $G_T = 0$.

On the other hand, when the unit cost is low (i.e., $\theta = L$), the bad incumbent will never engage in low spending of t_L since it is always in his or her best interest to divert some rent. Consequently, the voter will re-elect whenever (G_L, t_L) is observed. In other words, $\Pr(i = g|t_L) = 1$ since low tax collections (t_L) must be the result of the good type being in office. The more complicated question is how much rent type (b, L) will divert. On the one hand, type (b, L) may want to restrain first-period rent diversion (corruption) with some probability λ in an attempt to fool the voter by pooling with type (g, H). If so, the incumbent will choose a high public good provision G_H, charge high taxes t_H and divert first-period rents equal to $r_1 = t_H - LG_H$. On the other hand, type (b, L) may decide to divert maximum rents in the first period if the incumbent believes chances of re-election to be small or if the incumbent is impatient.

These alternative scenarios are part of the PBE strategy profile discussed below. Whether or not type (b, L) exercises restraint ($\lambda > 0$) and thus whether or not the equilibrium will be strictly pooling depends on the first-period incumbent's payoff function. In particular, type (b, L) restrains in the first period only if

$$r_1 + \beta\psi r_2 \geq \bar{r} \iff r_1 + \beta\psi\bar{r} \geq \bar{r} \iff r_1 \geq (1 - \beta\psi)\,\bar{r}. \tag{12}$$

When inequality 12 is strict, type (b, L) exercises restraint with certainty ($\lambda = 1$). Notice that the expression in 12 depends on ψ and β.

The probability of re-election ψ depends on the voter's posterior belief that such spending was generated by type (g, H). The posterior probability that spending t_H

was generated by such type is calculated by the voter using Bayes' rule as follows:

$$\begin{aligned} p &= \Pr(g|t_H) = \frac{\Pr(t_H|g)}{\Pr(t_H|g) + \Pr(t_H|b)} \\ &= \frac{\pi q}{\pi q + (1-\pi)(1-q)\lambda}. \end{aligned} \tag{13}$$

The voter re-elects with positive probability ($\psi > 0$) only if

$$\begin{aligned} p &= \Pr(g|t_H) \geq \pi \iff \frac{\pi q}{\pi q + (1-\pi)(1-q)\lambda} \geq \pi \\ &\iff \lambda \leq \frac{q}{1-q}. \end{aligned} \tag{14}$$

When this inequality is strict, the voter re-elects with certainty. Notice that this expression depends on the incumbent's restraint probability λ.

Thus, inequalities 12 and 14 jointly characterize the possible equilibrium configurations when a bad type faces low costs. Inequality 12 determines the bad politician's rule for exercising restraint and is a function of the voter's re-election probability ψ. In turn, inequality 14 determines the voter's re-election rule upon observing t_H and is a function of the incumbent's restraint probability λ. Besley and Smart characterize the possible equilibrium configurations in the following lemma.

Lemma 1 *An equilibrium exists for all values of parameters and is generally unique.*

1. *A pooling equilibrium, with $\lambda = \psi = 1$, exists iff*

$$q \geq \frac{1}{2} \;\&\; r_1(\mu) \geq (1-\beta)\bar{r} \tag{15}$$

2. *A separating equilibrium, with $\lambda = 0$ and $\psi = 1$, exists iff*

$$r_1(\mu) \leq (1-\beta)\bar{r} \tag{16}$$

3. A hybrid equilibrium, with $\lambda = \frac{q}{1-q}$ *and* $\psi = \frac{(\bar{r}-r_1)}{\beta\bar{r}}$*, exists iff*

$$q < \frac{1}{2} \;\&\; r_1(\mu) \geq (1-\beta)\,\bar{r} \tag{17}$$

Proof. See Besley and Smart. ∎

Intuitively, these equilibrium configurations say the following:

1. The pooling equilibrium corresponds to the case in which it is optimal for type (b, L) to exercise restraint with certainty $(\lambda = 1)$ since r_1 is high enough or β is high enough (i.e., the incumbent is sufficiently patient). Both types send the same message and the voter observes high taxes t_H regardless of the type of politician that is in office. The voter re-elects with certainty when observing t_H such that $\psi = 1$.

2. The separating equilibrium corresponds to the case in which r_1, β or ψ is low enough so that it is optimal for type (b, L) to divert maximal rents $(\lambda = 0)$ in the first period. The voter observes high taxes t_H if type (g, H) is in office and maximal taxes T if type (b, L) is in office. The bad incumbent is detected with certainty ex post. If the voter observes t_H, the voter re-elects with certainty, since such spending will only be chosen by type (g, H).

3. The hybrid equilibrium corresponds to a mix. In the hybrid equilibrium, type (b, L) adopts a strictly mixed strategy between restraint and maximal rent diversion. The voter will observe (t_H, t_H) some of the time and (t_H, T) some of the time.[22]

[22] If the incumbent has a very low discount factor β, the pooling and hybrid equilibria are less likely to occur and the separating equilibrium is more likely to occur. So, implicitly, it is assumed that β is high enough such that all three possible equilibria configurations can occur.

Propositions

Four main propositions are discussed below. The first two are Besley and Smart's main propositions on the effect of yardstick competition on incumbent behavior and on voter welfare. The third and the fourth are related to two key parameters of interest: the marginal cost of public funds (μ) and the unit cost of the public good (θ).

Before stating the yardstick propositions, it is useful to discuss what it means to model yardstick competition. In order to consider yardstick comparisons, one must introduce another jurisdiction–the foreign jurisdiction.

In particular, Besley and Smart focus on symmetric equilibria among incumbents and voters in the two jurisdictions by assuming that the joint probability mass function of cost shocks $\Pr(\theta, \theta')$ is symmetric. Furthermore, they assume that the cost shocks in the two jurisdictions are positively correlated by letting $\Pr(H, H) = \Pr(L, L) = \frac{\rho}{2}$ and $\Pr(H, L) = \Pr(L, H) = \frac{1-\rho}{2}$ for $\rho > \frac{1}{2}$.

Under such conditions, the voter's strategy involves yardstick competition when re-election occurs with positive probability if spending is high in both jurisdictions, but the probability of re-election is zero if domestic spending is high and foreign spending is low; i.e., $\psi(t_H, t_H) = \psi$ for some $\psi > 0$ and $\psi(t_H, t_L) = 0$.

Proposition 2 *Suppose that $r_1(\mu) > (1-\beta)T$. Then, voters use yardstick competition in equilibrium. A pooling equilibrium exists if and only if $\pi \geq \frac{1}{2}$, and a hybrid equilibrium exists if and only if $\pi < \frac{1}{2}$.*

Proof. See Besley and Smart. ■

Intuitively, this proposition tells us that the case with yardstick competition differs from the one jurisdiction case in three essential ways:

1. A bad incumbent may not be re-elected when she chooses t_H, if the foreign incumbent is type (g, L).

2. A good domestic incumbent is retained in office when costs are high, and the foreign politician chooses maximal rents.

3. Pooling may no longer be optimal for incumbents when the foreign incumbent has a poor initial reputation.

Proposition 3 *There exist parameters* $0 < \tilde{\pi}_a < \tilde{\pi}_b < 1/2$ *such that voter welfare is lower when yardstick comparisons are available than when they are not if* $\pi < \tilde{\pi}_a$*, and the converse is true if* $\pi > \tilde{\pi}_b$*.*

Proof. See Besley and Smart. ■

This result shows that voters who are better informed about the fiscal environment may be worse off in equilibrium, since bad incumbents will make less of an effort to build a reputation when they first take office. Namely, in some cases–in particular, when π is low–voters would be better off if they could commit to ignoring the fiscal performance in the other jurisdiction in the course of a domestic election. Yardstick competition is welfare decreasing when politicians' reputations are poor since rents are increased with little advantage from the improved information generated as most politicians who are voted out of office are replaced by an incumbent of the same type.

Proposition 4 *Increased inefficiency in the tax institution (as represented by an increase in* μ*) that leaves equilibrium strategies* (λ, ψ) *unchanged reduces voter welfare, even if it reduces rent diversion by bad politicians.*

Proof. See Besley and Smart. ■

The key assumption underlying the above result is that the equilibrium strategies remain unchanged. Yet, this need not be the case. The type of equilibrium may induce a change in strategy, which in turn can give rise to competing selection and discipline effects. In particular, we may find that an increase in μ (depending on the value of μ) induces a shift from pooling or hybrid to separating.

Proposition 5 *A less transparent public goods provision institution (as represented by an increase in risk of θ in the second-order stochastic dominance sense) is associated with an increase in expected equilibrium rent diversion for the whole game. Furthermore, a less transparent institution (as defined above) is associated with a decrease in equilibrium voter welfare in all possible equilibrium configurations–if they exist–as long as the marginal social cost of taxation is less than the inverse of the product of the marginal cost of public funds and the unit cost of the public good.*

Proof. See appendix A. ■

Intuitively, this proposition says that in a less transparent institution–i.e., an institution in which θ has a noisier distribution–equilibrium rent diversion is expected to increase and equilibrium voter welfare is expected to decrease as long as the marginal cost of taxation is low enough.

Experimental Design

From Theory to Experiments

The experimental treatments discussed in this paper are theory-testing in the sense that their designs are dictated by the theory of corruption and yardstick competition presented in the previous section. Furthermore, the questions they are designed to answer follow directly from the propositions (and lemma) discussed previously. This section explains how we make the transition from theory to experiments and thus how the experimental design relates to or departs from the aforementioned theory. In the process, we shall indicate some limitations of the experimental design. Some of those will be addressed in upcoming experiments.

We construct a dynamic experimental signaling game to capture the agency model of elections discussed previously.[23] Brandts and Holt (1992), Banks, Camerer

[23] Unlike in the main text, we maintain neutral terminology throughout the subject instructions. Incumbents are termed "Player X" and voters are termed "Player Z."

and Porter (1994) and more recently, Brandts and Holt (2005) discuss the possible complications associated with testing equilibrium behavior in experimental signaling games. In particular, the latter reference finds that there is mixed evidence about the adequacy of equilibrium refinements to predict what players will do in signaling games. They find that the deviation of subjects' decisions from equilibrium predictions is explained by a simple Bayesian learning process. This may be relevant when implementing an experimental test of the model discussed here, since the model assumes that players Bayesian update. Grether (1980) also discussed issues relating to a lack of Bayesian updating on the part of subjects.

We are slightly less concerned about such issues in these experiments due to two main reasons. First, since subjects in the experiments engage in several rounds of decision-making, we expect them to learn over the course of the experiment. Consequently, averaged decisions (as considered here) may be less noisy than repitition-based decisions. Secondly, since we are ultimately interested in the differential effect between the baseline treatment (Treatment B) and the yardstick treatment (Treatment Y), any lack of Bayesian updating (if applicable) is likely to be present in both treatments and affect both treatments at similar rates. So, we implement an experimental signaling game as discussed below.

An incumbent's type is assumed to be determined by a pair (i, θ). We start by discussing the first component. The first component (i) is either "good (g)" or "bad (b)" and is allowed to be homegrown. In other words, the experimental environment does not induce upon subjects that they be good or bad. We choose not to induce this component upon subjects for two main reasons. First, while the primary aim of the experiment is to test a model of corruption and yardstick competition that rests on the assumption that there is some positive proportion of good types in the population, a secondary aim is to test whether this assumption is empirically valid to begin with. In fact, in accordance with parallelism to the naturally

occurring world, it is sensible to ask whether assuming existence of a positive proportion of good types in the population is a plausible assumption. For if it is not, then the model discussed in the previous section must be called into question, since its predictions crucially rely on the proportion of good types in the population being greater than zero. So, data resulting from the current design give a verdict as to the empirical validity of this assumption.

Secondly, an induced-type environment represents a different way of testing the model. While such environment still enables us to study of the effect of yardstick competition on the voter side, it rules out the possibility of inferring incumbents' types from the population. In particular, it rules out the ability to study the effect of yardstick competition on corrupt incumbent behavior, which is obviously a question of interest in this study. The future research agenda is likely to include experiments based on an induced-type environment.

The second component of this pair θ represents the unit cost of public good provision and is controlled ("induced") as part of the experimental environment. This is done for numerous reasons. First, to be consistent with the information revelation described previously. As the game describes, θ is determined exogenously by nature. So, controlling the determination of θ is consistent with the theory. Secondly, not inducing the value of θ as part of the experimental environment would result in a relatively complicated game for subjects. Finally, by inducing (controlling) θ in this sense, we guarantee that the correlation between the levels of θ in different jurisdictions (i.e., between the different subject pairs) is as required by the theory–recall the condition that shocks be positively correlated. Specifically, subjects are informed of the following. Incumbents are choosing from two possible screens. The screen the incumbent observes is determined by a virtual flip of a coin. If the coin lands "heads" (i.e., θ is low), the incumbent chooses from one screen and if the coin lands "tails" (i.e., θ is high), the incumbent chooses from another screen.

So, the draw of θ is revealed to the incumbent before he or she makes a choice. The voter is informed of the distribution of θ and $\Pr(\theta, \theta')$, however, he or she is not informed of the exact realization of θ and θ'. While the current experimental protocol does not physically portray the determination of θ in the baseline and yardstick treatments, the discussion is implicit. This was done to maintain the subject instructions and experimental protocol relatively short and simple. While this is one advantage of this protocol, a possible disadvantage is that the current revelation of cost shocks is not sufficiently transparent in order for voters to meaningfully understand the information that is conveyed to them in the yardstick competition treatments (Treatment Y). It remains an empirical question whether a more explicit discussion of the the cost determination, $\Pr(\theta)$ and $\Pr(\theta, \theta')$, will enable voters to make a more informed decision. Viceisza (2007c) takes on this question.

Given the above design, the incumbent knows the complete characterization of his or her type–i.e., (i, θ)–prior to making a decision. However, since part of the incumbent's type is homegrown, the complete characterization of the subject's type is a priori unknown to the experimenter. This implies that the incumbent's preferences are also unknown a priori. So, the experiment is designed such that the experimenter can learn incumbents' types from observing subjects' choices. This is important since–contrary to an induced-type environment–the present environment enables us to test a compound hypothesis that relates to (1) the proportion of types in the incumbent population and (2) the effect of yardstick competition on incumbent and voter behavior. Furthermore, a homegrown-type environment implies that voter subjects are a priori uninformed about the proportions of good and bad types in the incumbent population. They learn this as the experiment progresses.

Specifically, the payoff pairs presented to the experimental subjects are based on the following parameterizations. The bad-type incumbent's per-period payoff function is characterized by $r = t - \theta G_\theta$, where all notation is as defined previously.

So, if we observe such a choice, the subject is inferred to behave "as if he is bad." The good-type incumbent's per-period payoff function is the same as the voter's, which is an explicit form of the voter's theoretical welfare function. Contrary to the assumption that the tax-cost function be strictly convex, we choose a linear tax-cost parameterization and let $C(t) = t$. This parameterization is chosen because it provides sufficient behavioral separation and gives rise to tractable payoffs from a practical standpoint. In particular, this parameterization in combination with the condition that $G_T > 0$ (see below) avoid having to deal with negative payoffs based on the choice of signals. Thus, the voter's payoff function is characterized by $W(G, t) = G - \mu t$, where all notation is as defined previously.[24] So, if we observe such choice, the subject is inferred to behave "as if he is good."

As discussed below, the treatments of interest (B and Y) consist of five repetitions of the one-shot play of the theoretical game. One such signaling game consists of two periods.[25] In the first period, the incumbent chooses a signal payoff pair (p_x, p_z), which is based on a signal pair of the form (t, G_θ). All payoff pairs are common information. Once the incumbent makes his or her choice, the voter observes his or her payoff (p_z) and the possible payoffs for the incumbent. The voter then responds by either accepting (re-electing) or rejecting (not re-electing) this payoff. Regardless of whether the voter accepts or rejects, first-period payoffs are determined by the incumbent's choice. However, the voter's decision to accept or reject affects payoffs in the second period of the repetition. If the voter accepts, the

[24] A linear tax-cost function affects the good-type's maximization problem in expression 1. In particular, we get $\max_G [G - \mu\theta G]$, which reduces to $\max_G [G(1 - \mu\theta)]$. If the parameter $(1 - \mu\theta) \geq 0$, then $G^* > 0$. If not, then $G^* = 0$. So, in order to have an interior solution, we need μ and θ to satisfy the following relation: $\theta \leq \frac{1}{\mu}$. This condition is satisfied in all parameterizations. A linear tax-cost function also leaves conditions 5 and 6 unaffected. Namely, suppose the condition $\theta \leq \frac{1}{\mu}$ holds. Now, consider a slight increase in either θ or μ such that θ becomes strictly greater than $\frac{1}{\mu}$. Then, G^* goes from strictly positive to zero in a discrete manner. The relationships captured by expressions 5 and 6 are maintained.

[25] Note that the first (second) *part* in the experimental game corresponds to the first (second) *period* in the theoretical game. One repetition of such game is called a period in the experimental game. This is due to z-tree's terminology.

incumbent chooses between another set of alternative payoff pairs. That choice determines the payoffs in the second period and the voter does not get to accept or reject this second-period choice. If the voter rejects, the incumbent does not get to play in the second period and he or she gets zero payoff. In turn, the "challenger's type" is determined at random and modeled as a move of nature. In other words, the challenger is not an actual person in the experimental game. This is also why the challenger was termed "passive" in the theoretical game. If the challenger's type is (g, L)–i.e., if the two coin tosses land $(heads, heads)$–the voter gets a high payoff with twenty five percent chance. If the challenger's type is (g, H)–i.e., if the two coin tosses land $(heads, tails)$–the voter gets a medium payoff with twenty five percent chance. If the challenger's type is bad–i.e., if the coin tosses land either $(tails, heads)$ or $(tails, tails)$–the voter gets a low payoff with fifty percent chance.

The following additional parameterizations and theoretical guidelines are relevant for the treatments discussed below.

1. The marginal cost of public funds is strictly positive; $\mu = \frac{1}{4} > 0$.

2. The unit cost of the public good (θ) equals one or two with equal probability.

3. There is a maximal level of tax collection, T. In particular, $T = 4.00$.

4. There is a negative relationship between the unit cost of public good provision and the level of public good provision; i.e., $\frac{\Delta G_\theta}{\Delta \theta} < 0$ within the same treatment.

5. The payoff pairs presented in the experimental game are on the equilibrium path according to the specified preferences. In other words, the incumbent and the voter are not faced with alternatives that are off the equilibrium path.

Finally, one additional theoretical implication guides the experimental design. It departs slightly from the theory and thus is described separately. In the theoreti-

cal game, G_T is zero. This follows from the assumption that the bad type does not care about public good provision. In particular, when the bad type diverts maximal rents of T, is it optimal and economically rational to set public good provision equal to zero (i.e., $G_T = 0$). From a practical standpoint, this is problematic since $G_T = 0$ automatically implies negative payoffs for the voter whenever the incumbent makes a choice associated with such signal. Since negative experimental payoffs tend to introduce complications for subjects, we would rather avoid such parameterization. Consequently, in the experimental game, the minimum level of public good is set equal to one, i.e., $G_T = 1$.

What does this imply for the theoretical model? Well, $G_T = 0$ is the result of an equilibrium condition. In other words, the incumbent that "separates" will never set $G_T > 0$ without an additional assumption in the model. So, we introduce the additional theoretical assumption that the bad type must provide at least some level of the public good when diverting maximum rents. This can be seen as a complementary assumption to that of maximum rents equal to T. So, let $G_T = \delta$, for some $\delta > 0$. What restrictions–if any–should be placed on δ? Well, theoretically δ must be small enough such that all equilibrium configurations still exist. In other words, δ cannot be so large that it rules out the possibility of a separating equilibrium occurring a priori.[26] Experimentally, δ must be large enough such that voter welfare is nonnegative in Treatment B and Treatment Y. G_T equal to one satisfies this condition given the above parameterizations.

[26] To derive bounds for δ, consider the restraint rule in expression 12. Replace $\bar{r} = T$ by $\bar{r} = T - \delta\theta$ for $\theta \in \{H, L\}$, where $H > L$. The pooling equilibrium exists as long as $\delta \geq \frac{(1-\beta)T - r_1}{H(1-\beta)}$ while the separating equilibrium exists as long as $\delta \leq \frac{(1-\beta)T - r_1}{L(1-\beta)}$. Thus, $\frac{(1-\beta)T - r_1}{H(1-\beta)} \leq \delta \leq \frac{(1-\beta)T - r_1}{L(1-\beta)}$.

Treatments

The experimental design is based on two treatments:[27]

1. A Baseline treatment (Treatment B).[28] This treatment tests the basic predictions of the theory as captured by the discussion pertaining to lemma 1.

2. A Yardstick competition treatment (Treatment Y). This treatment tests the predictions arising from the model as captured by propositions 2 and 3.

Treatment B. During the first period of the repetition, the incumbent is presented with a choice between three or two alternatives. Whether the incumbent chooses between three or two alternatives depends on the level of θ. If $\theta = 1$, the incumbent chooses between three alternatives. If $\theta = 2$, the incumbent chooses between two alternatives. The rationale comes from the fact that we only consider choices that are on the equilibrium path according to the specified preferences.

If $\theta = 1$, the incumbent can behave according to one of the following types: (1) the incumbent can behave as type (g, L), set rents equal to zero and thus equalize payoffs or (2) the incumbent can behave as type (b, L) and play according to the pooling or separating equilibrium depending on the incumbent's discount factor and expected probability of acceptance. If $\theta = 2$, the incumbent can behave according to one of the following types: (1) the incumbent can behave as type (g, H) by setting rents equal to zero and thus equalizing payoffs or (2) the incumbent can behave as type (b, H) and divert maximum rents in equilibrium under the expectation of being rejected.

[27] Viceisza (2007c) reports two additional treatments: (1) a Marginal cost of public funds treatment (Task M) and a stochastic (unit) cost of public goods or Risk treatment (Task R). Treatment M is the same as Treatment B with the only exception that the marginal cost of public funds (μ) has increased. The treatment tests the predictions arising from the model as captured by proposition 4. Treatment R is the same as Treatment B with the only exception that the unit cost of the public good (θ) has undergone an increase in risk in the second-order stochastic dominance sense. This treatment tests the predictions arising from the model as captured by proposition 5.

[28] In the subject instructions (appendix B), the term "Treatment" is replaced by "Task."

Each alternative that the incumbent faces consists of a payoff pair of the form (p_x, p_z). The payoffs for a good incumbent are dictated by preferences as specified, $p_x^g = p_z = G - \frac{1}{4}t$ under the assumption that rent is equal to zero. The payoffs for a bad incumbent are dictated by preferences as specified, $p_x^b = t - \theta * G$.

Table 1 shows the first-period payoff pairs faced by type (i, θ) if this type sends a signal of the form (t, G). Subjects face these payoffs in the experiment sessions (see subject instructions).[29] From the table it is clear that the incumbent can equalize payoffs if good or behave according to the pooling or separating equilibrium if bad depending on (1) the draw of θ, (2) the subject's (subjective) discount factor (β) and/or (3) the subject's expected probability of acceptance (re-election).

It should be noted that even though it seems costless for type (b, L) to choose Alternative 1 rather than Alternative 2 when $\theta = 1$, these payoff pairs are based on completely different signal pairs according to the specified preferences. In particular, preferences as specified by the model dictate that a bad incumbent never equalizes payoffs, since payoffs can only be equal when $r = 0$. So, a choice that gives rise to equal payoffs for incumbents and voters is inherently the result of a characteristic of the good type by assumption of the model. It is important to keep this in mind when the results are discussed.

Furthermore, it should be clear that the pooling equilibrium is only an option when $\theta = 1$. Again, this is the result of the experimental design which only takes into account actions that are on the equilibrium path. In particular, the pooling equilibrium corresponds to the case in which the incumbent of type (b, L) chooses A_2 in the first period by "pooling" with type (g, H), "fools" the voter, gets accepted (re-elected) and "slams" the voter in the second period by choosing A_5 (behaves like a lame duck).

[29] All payoffs shown here–as well as those shown in the instructions–are in experimental dollars. Each experimental dollar is a quarter of a real U.S. dollar. This was done to maintain the payoffs predicted by the signaling game–given the current parameterization–while still enabling affordable subject payments.

On the other hand, the separating equilibrium corresponds to the case in which the incumbent of type (b, L) or type (b, H) diverts maximum rents in the first period of the repetition with the expectation of being rejected. If accepted, both types will "slam" the voter in the second period of the repetition. In table 1 the separating equilibrium is represented by the incumbent choosing A_3 or A_2 depending on the draw of θ with the expectation of being rejected but the intent to "slam" the voter by choosing A_5 or A_4 if accepted.

Once the incumbent makes a choice, the voter observes his or her payoff. The voter is informed of the range of possible payoffs for the incumbent. The voter then accepts or rejects the incumbent's choice. Either way, the incumbent and the voter get the first-period payoffs chosen by the incumbent. However, the voter's acceptance or rejection decision affects payoffs in the second period of the repetition. If the voter accepts the incumbent's choice, the incumbent gets to choose another alternative which will determine second-period payoffs. These alternatives are shown in table 2.[30]

If the voter rejects the incumbent's choice, the incumbent gets zero payoff for the second period of the repetition and the voter's payoffs are randomly determined as follows: (1) with twenty five percent chance, the voter's payoff is 1.50 experimental dollars, (2) with twenty five percent chance, the voter's payoff is 0.75 experimental dollars or (3) with fifty percent chance, the voter's payoff is 0.16 experimental dollars.

[30]The alternatives in table 2 are termed A_i or A_j, for $j = i + 1$ since the subscript depends on the draw of θ in the first period of the repetition. For example, if θ was 1 in the first period, then the incumbent had a choice between three alternatives $\{A_1, A_2, A_3\}$ in the first period. Consequently, if accepted, the incumbent would face a choice between Alternative 4 (A_4) or Alternative 5 (A_5) in the second period of the repetition. On the other hand, if θ was 2 in the first period, then the incumbent would face a choice between Alternative 3 (A_3) or Alternative 4 (A_4) in the second period.

Table 1: First Period Payoffs (Treatments B and Y)

	$\theta = 1$		$\theta = 2$
Alternative	(g, L) sends $(2, 2)$	Alternative	(g, H) sends $(3, 1.50)$
A_1	$(1.50, 1.50)$	A_1	$(0.75, 0.75)$
	(b, L) sends $(3, 1.50)$		
A_2	$(1.50, 0.75)$		not applicable*
	(b, L) sends $(3.35, 1)$		(b, H) sends $(3.35, 1)$
A_3	$(2.35, 0.16)$	A_2	$(1.35, 0.16)$
* No comparable alternative due to "on the equilibrium path" assumption.			

Table 2: Second Period Payoffs (Treatments B and Y)

	$\theta_{2,L}= 1$		$\theta_{2,H}= 2$
Alternative	(g, L) sends $(2, 2)$	Alternative	(g, H) sends $(3, 1.50)$
A_3 or A_4	$(1.50, 1.50)$	A_3 or A_4	$(0.75, 0.75)$
	(b, L) sends $(3.35, 1)$		(b, H) sends $(3.35, 1)$
A_4 or A_5	$(2.35, 0.16)$	A_4 or A_5	$(1.35, 0.16)$

The first and second periods of the experimental game together constitute a repetition. Treatment B consists of five such repetitions that comprise two periods. Each period θ is randomly drawn and revealed to the incumbent accordingly. The voter knows the distribution of θ each and every period. Note that repetition is introduced to avoid complications as discussed in Brandts and Holt (2005). As alluded to previously, they find that subjects follow a Bayesian learning process when playing signaling games. One way to partly exploit this finding is to introduce repetition. The choice of five periods is based on a trade-off between "optimality of subject understanding and learning" and length of the experiment sessions. Since the theory is not one of repeated play, subjects are paired with different partners each and every period. They are informed accordingly before the experiment begins (see subject instructions).

Treatment Y. This treatment represents the yardstick competition treatment. As suggested by the model, we enable voters to use yardstick comparisons by

amplifying their information set. In particular, the voter gets to see the distribution of first-period choices made by all incumbents in the previous treatment (Treatment B). This additional information can be seen as the amplification of the voter's information set that arises due to yardstick competition. The subject instructions indicate exactly how this information is presented to subjects.

This treatment thus differs from the previous only for the voter. It can be questioned whether the information conveyed in Treatment Y is inconsistent with the theory of yardstick competition. Namely, the theory of yardstick competition discussed previously assumes existence of one other ("foreign") jurisdiction from which information is observed. Nonetheless, the model is not explicit whether such "foreign" jurisdiction must be one other jurisdiction or an aggregate (representative-type) jurisdiction. we chose to convey aggregate offers/choices as opposed to *one* other incumbent's offer/choice, since the latter signal seems relatively uninformative and inconsistent with what a voter would observe in the naturally occurring environment. Regardless, to perform a stricter test of the theory, we plan to conduct some sessions in which voters are shown the offer/choice made by just *one* other randomly chosen incumbent.

Finally, it is important to note that due to the nature of the main question, it is necessary to have a distribution of the incumbents' decisions in order to run Treatment Y. In other words, if the voter's information set is to be amplified, the question "where does this information come from" must be addressed. From a practical standpoint, there are a couple of possibilities: (1) The voter can observe a distribution of first-period aggregated choices made by some external group of incumbents that participated in Treatment B or (2) the voter can observe the distribution of first-period aggregated choices made by incumbents in a preceding Treatment B in which he or she actually participated as a voter.

The first possibility entails a design in which voters in Treatment Y observe

a distribution of incumbent first-period choices that comes from a session other than their own. Even though this possibility gives rise to the ability to implement a standard BY-YB design for decomposing learning effects from treatment effects, it is more intricate to motivate to voter subjects why they will be observing decision-making from an experiment session other than their own. Furthermore, the nature of the question and how it relates to the setup of the theoretical model really requires a subject to first participate in Treatment B and then take part in Treatment Y. So, we chose for a within-subjects design.

While this facilitates motivation of Treatment Y to subjects, it also rules out the possibility of adopting the above mentioned BY-YB design. So, in order to tease apart learning/sequencing effects from treatment effects, an alternative design is needed. The experimental design in this paper is based on a within-subject design in which subjects participate in four treatments, two Treatments B and two Treatments Y. In particular, subjects participate in the four treatments in the following order, $B_1Y_1B_2Y_2$. So, in total, subjects participate in twenty repetitions, each consisting of two periods.

By comparing behavior in Treatment B_1 (B_2) with behavior in Treatment Y_1 (Y_2) and separately comparing behavior in Treatment B_1 (Y_1) with behavior in Treatment B_2 (Y_2), treatment effects can be partly separated from learning effects. As part of the experimental protocol, subjects are informed that they will participate in four treatments; however, they are not informed as to the nature of all four treatments before beginning the experiment (see subject instructions). The nature of each treatment (task) is only revealed to subjects as they are about to enter the respective treatment.

While the above within-subjects design has its advantages, it also has its disadvantages. Namely, one might expect the effect of yardstick competition to be mitigated by this design. Specifically, by the time voters enter Treatments Y_1 and

Y_2, they may have sufficient experience facing different incumbents–due to random re-pairing–that they no longer respond to additional information arising from yardstick competition.

To see whether this aspect of the experimental design is indeed clouding the treatment effect, we plan to run across-subject experiments. While such design may give rise to a treatment effect, it also has its relative disadvantages. The main one is that if subjects in Treatment Y did not participate in Treatment B, they are less likely to believe the yardstick information conveyed to them.

Brief Recapitulation

Before proceeding to the experimental findings, it is necessary to stress two issues. First, it should be noted that the unit cost of public good provision θ plays an important role in the general game as well as in the interpretation of the experimental results. So, the reader should familiarize him or herself with this random variable in order to properly consume the experimental findings.

Secondly, it is important for the reader to keep in mind the steps involved in one repetition of the game. Thus, we briefly recapitulate the steps below. Essentially, Treatments B and Y differ only by the information given to voters (see subject instructions). Both treatments consist of five repetitions each consisting of two periods. By random re-pairing, we aim to reduce any repeated-play-of-the-game effects. One repetition involves the following steps:

1. Subjects are randomly assigned to be an incumbent or a voter. Once assigned a role, the subject maintains this role throughout the experiment.

2. Subjects are allowed to behave according to their homegrown types i, which can be "good (g)" or "bad (b)."

3. The computer virtually tosses a coin and the unit cost of public good provi-

sion θ is revealed to incumbents by means of the screen that they observe.

4. The voters are implicitly informed of the distribution of θ and the correlation between the shocks.

5. The incumbent chooses a payoff pair that represents a tax-public good signal.

6. The voter observes his or her part of the payoff and the range of possible payoffs for the incumbent.

7. The voter chooses to accept or reject such payoff.

8. If the voter rejects such payoff, the incumbent is not re-elected to office and the challenger's type is determined by means of a lottery.

9. If the voter accepts such payoff, the incumbent moves to the second period.

10. The second-period incumbent's alternatives are determined by a draw of θ.

11. The voter's second-period payoffs are determined by the incumbent's choice. The voter has no acceptance/rejection decision.

Implementation

The experiments took place in the experimental laboratory at the Experimental Economics Center (ExCEN) at Georgia State University. Subjects were recruited using ExCEN's online recruiter system, which contains names of students taking courses in many different areas including but not limited to accounting, economics, finance, geology, nursing and political science. The experiments were programmed and conducted with the software z-Tree (Fischbacher 2007).

Four experiment sessions were run. Two sessions consisted of twenty four subjects, one session consisted of twenty two subjects and one session consisted of eighteen subjects. In each session, half of the subjects were incumbents and half of the

subjects were voters. Subjects were randomly assigned to seats as they entered the experimental laboratory. This implied random assignments of roles as well. Once assigned to a role, subjects maintained that role throughout the whole experiment. Subjects were informed accordingly (see subject instructions).

An experiment session lasted one hour and a half and paid on average of $19.09 (standard deviation $2.28). Of these payoffs, $10 was certain and determined by a $5 show-up fee, a $2 payment for participating in a trial and a $3 payment for completing a post-questionnaire (see elaboration below). The remaining payoffs were determined by participating in the sequence of four tasks described previously.

Recognizing that the nature of the experimental game is relatively complicated, two additional steps were undertaken to promote and test for subject understanding. First, subjects participated in a trial. This was a five minute opportunity for subjects to familiarize themselves with the software (i.e., the game). Subjects "played" with themselves and got to see the screens that both the incumbent and the voter would see once the treatments started.

After subjects completed the trial they were asked if there were any questions. If so, these were addressed person by person before Task B_1 started. Subjects were informed before initiating the trial that the trial paid one fixed fee. In particular, they were informed that payment was not dependent on how fast they went through the trial and that they should pay close attention to the screens in the trial (see subject instructions).

Secondly, to further test and control for subject understanding, subjects participated in a post-questionnaire. The post-questionnaire prompted the subjects for understanding using specific questions. It indicated that a relatively small proportion of subjects were confused ex-post. Excluding these subjects' responses from the analysis did not affect any of the main results discussed below.

Experimental Findings

This section is dedicated to the discussion of the main experimental findings. First, it discusses issues in aggregating the data. Secondly, it presents a graphical overview of subjects' choices. Third, it considers a one-shot analysis of the game. Fourth, it addresses the main experimental findings arising from Treatments B and Y. Finally, it discusses some gender differences that are distinguishable in the data.[31]

The following demographic characteristics describe the average subject profile. Almost 55% of subjects are female. The average subject age is 20.02 years with a standard deviation of 1.62 years. The maximum annual income range reported is $10,001 to $30,000. Finally, 7.95% of the subjects are economics majors.

Data Aggregation

As mentioned previously, four experiment sessions were run in order to gather sufficient data for statistical significance. Before aggregating the data to do within-subject comparisons, we tested whether the variables of interest were comparable across sessions. The main variables of interest are Choice, Accept and Choice2. "Choice" represents the first-period choices made by the incumbent, "Accept" represents the acceptances by the voter and finally, "Choice2" represents the second-period choices made by the incumbent if the voter accepted the first-period choice.

An Epps-Singleton and a Kolmogorov-Smirnov test on these three variables (at the session level) suggested no statistically significant differences across sessions one, three and four. However, for session two, we rejected the null hypothesis that the "Choice" observations were from the same distribution as the observations from

[31] Other demographic characteristics such as income (wealth) and major of study have also been shown to have an effect on corrupt behavior. See for example Frank and Schulze (2000), Abbink and Ellman (2005) and Chakrabarti and Subramanian (2006). We do not consider such issues here due to a lack of variation in the sample.

the other sessions. Consequently, we addressed each of the questions below using both the complete aggregate data set that includes session two and the aggregate data set that excludes session two. It turns out that the main results are the same irrespective of which data set is used. So, we use the complete aggregate data set that includes session two to support the conclusions.

Graphical Overview of Subjects' Choices

Before proceeding to the main analyses, it is useful to look at an overview of subjects' choices. The figures below summarize the progression of three types of behavior over time. First, figure 2 summarizes incumbents' averaged first-period choices conditional on θ being low. Secondly, figure 3 summarizes incumbents' averaged first-period choices conditional on θ being high. Finally, figure 4 summarizes voters' averaged acceptance rates conditional on the incumbents' offers.

In interpreting the graphs, it is important to keep the following in mind. Repetitions 1 through 5 represent Treatment B_1, repetitions 6 through 10 represent Treatment Y_1 and so on. The treatments are separated by vertical lines. Furthermore, in any given period the draw of θ was either high or low. So, there are periods in which–for example–the 1.50 payoff could not have been offered/accepted, since it was not available (recall the payoff pairs discussed in table 1).

Similarly, the 0.75/0.16 payoffs could have arisen due to the fact that θ was low or due to the fact that θ was high. This is particularly relevant when considering the voters' decision-making. In order to compare the relevant points on the graph, figure 4 conditions the relevant periods on the draw of θ by assigning them the same color.

The following patterns seem to be distinguishable. Figures 2 and 3 both show strong convergence between offering the high and low payoffs conditional on the draw of θ. In other words, figures 2 and 3 suggest that as the experiment session

progresses, incumbents are becoming more corrupt by offering lower payoffs. As for voter behavior, figure 4 suggests that voters are accepting the 1.50 payoff at lower rates. On the other hand, conditional on the draw of θ acceptance rates seem to be relatively stable for the 0.75 and 0.16 payoffs, with a mild downward trend.

Figure 2: Incumbents' First-Period Choices When Unit Cost is Low

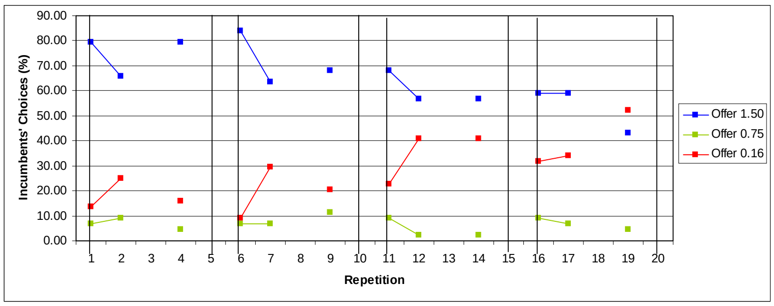
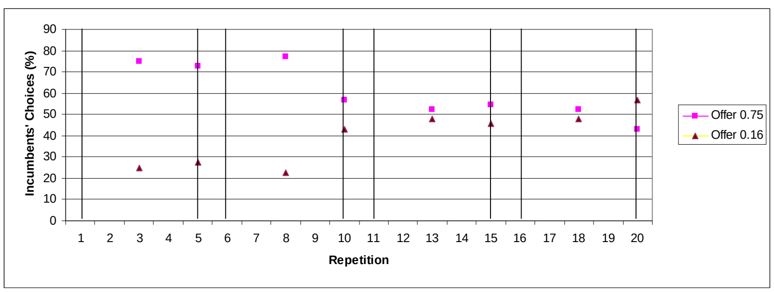


Figure 3: Incumbents' First-Period Choices When Unit Cost is High

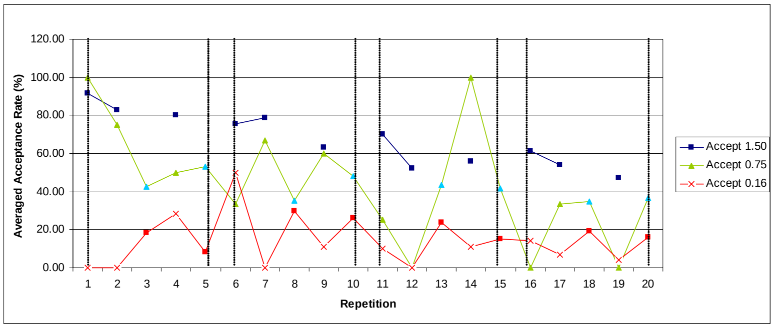


One-Shot Analysis of The Game

Essentially, a strict test of the theoretical model entails conducting a one-shot play of the game. In other words, a game in which subjects would only participate in Treatments B_1 and Y_1 both consisting of just one period. The future research agenda may comprise a "one-shot-play" design of this type, which would serve as a control treatment.

However, as explained previously, the current design involves repetition to mitigate subject confusion. As a result, two important disclaimers are necessary. First,

Figure 4: Voters' Acceptance Rates Conditional on Offer and Unit Cost



in the sections below several results will be interpreted empirically since the theoretical model cannot explain behavior beyond the first repetition. Secondly, to reduce the noise of possible subject confusion and/or dependence on past experiences, we consider subjects' averaged decisions when analyzing the data and when teasing apart learning and treatment effects. Regardless of repetition, it is important to note that random re-matching of subjects does reduce the concern about repeated play. Nonetheless, the subjects are learning something about the population of types as the experiment session progresses.

In the current design, subject behavior that is most consistent with the theoretical game seems to be first-period behavior in Treatments B_1 and Y_1. These repetitions are the closest approximations to a one-shot play of the game with and without yardstick competition. A Mann-Whitney test comparing the first periods of Treatments B_1 and Y_1 rejects the null that the voters' acceptances in the two periods are the same when voters face the 1.50 and 0.16 payoffs (p-values are 0.0751 and 0.0662 respectively).

However, this result is neither replicated for the 0.75 payoff nor for first-period decision-making in Treatments B_2 and Y_2. The latter is supported by the Mann-Whitney tests conducted in the next section on subjects' averaged choices across all repetitions.

To further tease apart a treatment effect, we compare behavior in the fourth

repetition of Treatment B_1 with behavior in the first repetition of Treatment Y_1. We choose to study fourth-repetition behavior in Treatment B_1, since later repetitions are expected to be less affected by learning. However, since an end-of-treatment effect is expected to affect behavior in the last repetition (repetition five), we consider fourth- as opposed to fifth-repetition behavior. It is reasonable to expect that differences between behavior in the fourth repetition of B_1 and the first repetition of Y_1 are due to yardstick competition.

A Mann-Whitney test on choices made during the aforementioned two repetitions suggests that there is no evidence of yardstick competition when voters make decisions with respect to any of the proposed payoffs (p-values are all greater than 0.49). Furthermore, a Mann-Whitney test conducted on the first and fourth repetitions of B_1 suggests that behavior in the two repetitions is not the same. These two results suggest that learning rather than yardstick competition is driving the differences in first-repetition behavior across Treatments B_1 and Y_1. Finally, it should be noted that the incumbent's first-repetition behavior does not seem to differ from Treatments B_1 (B_2) and Y_1 (Y_2).

Treatment B

To further test the basic theoretical model, we analyze the results arising from Treatment B. In doing so, the following main questions are addressed:

1. Are there good incumbents in the subject population and if so, what is the proportion of good types π?

2. Are there bad incumbents in the population and if so, do they play according to the equilibria in lemma 1?

3. Do voters play according to the equilibria in lemma 1?

The incumbent: Good-Type Behavior. To obtain an estimate of the proportion of good types in the population, it is necessary to be clear what is meant by a good type. In the theoretical game, the good type is assumed to set rents equal to zero and therefore, he or she equalizes payoffs in both periods conditional on the level of θ. In particular, when θ is low, the preferences of the good type dictate that he or she charges low taxes of t_L. As a consequence, the voter infers that observing a low tax (t_L) must be the result of the good type being in office, i.e., $P(i = g|t_L) = 1$.

In the empirical game, it turns out that this is no longer the case. Given the assumption on preferences, the above reasoning would predict that when θ is low a bad incumbent would "pool" or "separate" and never equalize payoffs. However, this prediction fails. It turns out that the incumbent goes to Alternative 1 in the first period of the repetition and acts as if he or she is good, gets accepted most of the time and then "slams" the voter in the second period.

While this strategy will be elaborated upon at a later point, for now it is relevant since this empirical strategy forces us to re-think the concept of a good versus a bad type in the empirical game. In particular, the incumbent is now of the good type if and only if he or she equalizes payoffs in both periods of the repetition. This is in contrast to the theoretical game in which the incumbent's type can be inferred from observing only the first period choice.

Only given this understanding is it sensible to ask whether there is a positive proportion of good types in the population, since the good type is now characterized by an incumbent that equalizes payoffs in both periods. Since θ is induced, we can infer the incumbent's homegrown type $(i = \{g, b\})$ from the subject's choices.

A final noteworthy comment is that since the game on which the theory is based is a two-period game, the theory has no prediction as to how subjects (in particular, incumbents) will behave beyond the first repetition. In other words, the theory

has no prediction as to how subjects will behave in repetitions two through twenty. So, to remain in line with testing the theory, we proceed in a couple of ways. First, to get an estimate of the proportion of types in the population without confounding learning or treatment effects, we use the data from the first two periods (i.e., the first repetition) of all sessions to infer incumbents' types. Secondly, we ask the stronger empirical question to what extent there is a positive proportion of incumbents that remain good throughout the whole experiment; rather than just one given repetition.

As for the first question, we find that 25% of incumbents chose equal payoffs during both periods of the first repetition in treatment B_1; i.e., $\hat{\pi} = \frac{1}{4}$, which suggests that 25% of incumbents are good in the sense that they maximize societal welfare in an equitable manner when re-elected. For purposes of comparison, table 3 contains proportions of good types for the nth repetition, where n represents the first and last repetition in treatments B_1, B_2, Y_1 and Y_2 respectively. These proportions answer the question what happens to $\hat{\pi}$ as the experiment session progresses. Note that this is an empirical question, since the theory has no prediction in this respect.

The reasoning for calculating these proportions is discussed next. First, $\hat{\pi}$ is calculated for the beginning of the treatment, since it is expected that there is some type of "restart" effect in the first repetition of a new treatment. Furthermore, $\hat{\pi}$ is calculated for the end of the treatment, since subjects know this is the last repetition and are less likely to "play fictitiously." Table 3 confirms these expectations. The level of $\hat{\pi}$ goes down as a treatment progresses and the level of $\hat{\pi}$ jumps upward when a new treatment begins. More specifically, $\hat{\pi}$ goes down to 2.27% in the final repetition.

It is worth mentioning that there seems to be a "structural break" in the pro-

portion of good types when Treatment B_2 is initiated.[32] In fact, this result is recognizable in many of the tables that are upcoming. We discuss possible causes in the Treatment Y section below. Finally, it should be noted that the theory rules out the possibility that incumbents "turn bad" (and vice versa) by assuming that players have a "fixed" type. These results suggest otherwise; in particular, subjects "change types" (i.e., alter their behavior) as they advance through the experiment.

While this finding can be disclaimed by the fact that the theory is not one of multiple repetitions, it should be noted that this behavior is nonetheless exhibited in the first repetition already. One way to control for this is to have subjects play a one-shot version of this game. If incumbents are found to behave like the good type in the first period and "slam" the voter in the second period in a one-shot version of this game as well, such finding would optimally support the conclusion that subjects are "changing types."[33]

Table 3: Proportion of Good Types within a Repetition

$N = 44$								
Repetition	1	5	6	10	11	15	16	20
$\hat{\pi}$	25.00%	6.82%	22.73%	4.55%	6.82%	6.82%	4.55%	2.27%

We now turn to the second question posed above, which asks to what extent there is a positive proportion of incumbents that remain good throughout the experiment session (i.e., Treatments B_1, Y_1, B_2 and Y_2). Table 4 answers this question.

In this table, $\hat{\pi}'_K$ represents the proportion of incumbents that equalized payoffs during both periods of a repetition more than $K\%$ of the time given the opportunity to do so. For example, $\hat{\pi}'_{75}$ is the percentage of incumbents that equalized

[32] The term "structural break" has a technical definition in the econometrics literature; in particular, the time series literature. In this paper, the term is used loosely to indicate a situation in which subject behavior undergoes a significant "jump"–i.e., change.

[33] This same disclaimer applies for upcoming sections where we relate the findings of the twenty-repetition game with the predictions from the model.

payoffs in both periods of a repetition more than 75% of the time given the opportunity to do so. So, the proportion $\hat{\pi}'_{75}$ includes the proportion of incumbents who equalized payoffs 100% of the time as well. The same holds for the other proportions.

The table shows that 4.55% of incumbents equalized payoffs during the whole experiment whenever they had the opportunity to do so. Empirically, these are the truly "good" incumbents, since they equalize payoffs during the whole experiment session. The table also reports the complement proportion to $\hat{\pi}'_0$. This is the proportion of incumbents that never equalized payoffs in both periods even though they had the opportunity to do so. These can be considered the "truly bad types," since–at no point during the experiment session–did they behave like a good type. This proportion is equal to 50% of the subjects; i.e., 50% of subjects never chose to equalize payoffs in both periods of a repetition.

Table 4: Proportion of Good Types within a Session

$N = 44$						
Proportion	$\hat{\pi}'_{100}$	$\hat{\pi}'_{75}$	$\hat{\pi}'_{50}$	$\hat{\pi}'_{25}$	$\hat{\pi}'_0$	$1 - \hat{\pi}'_0$
%	4.55%	11.36%	13.64%	36.36%	50%	50%

The incumbent: Bad-Type Behavior. Thus far, we have mainly focused on the proportion of good types in the subject population. Even though this enables inference on the proportion of bad types, it is necessary to look at the following issues: (1) the progression of bad types and (2) their behavior. In doing so, it is important to address both first- and second-period behavior.

First-period behavior can indicate whether an incumbent is a bad type and if so, whether he or she plays according to the theoretical equilibria–i.e., whether he or she "pools" or "separates." Second-period behavior conditional on first-period behavior indicates whether an incumbent is a good type or a bad type that "pools", "separates" or makes theoretically inconsistent choices.

Table 5 presents first-period behavior for the incumbent. In particular, the table conveys the percentage of times that incumbents made (un)equal choices during the first period of all repetitions conditional on the draw of θ. The table reflects a few key aspects. First, many incumbents choose to equalize payoffs in the first period. On average, incumbents choose to equalize payoffs 75% of the time in Treatment B_1 when $\theta = 1$. This proportion goes down to 53.79% in Treatment Y_2. It seems that incumbents are moving from equalizing to "separating" as the experiment session progresses. This can be confirmed by observing the progression of $P(A_3|\theta = 1)$. This finding is robust to the draw of θ as illustrated by the table. Furthermore, on average incumbents are "pooling" 3.32% of the time and "separating" 33.76% of the time. These values are not shown in the table.

Finally, it can also be inferred from table 5 that the majority of incumbents (on average, 62.92%) chose equal payoffs during the first period of the repetition. As discussed previously, the theory–at least for the first period–would conclude that those subjects were good. However, table 6 leads to a different conclusion.

Table 5: Distribution of Incumbents' First-Period Choices

$N = 44$	Treatment (Task)			
	B_1	Y_1	B_2	Y_2
Probability	[1-5]*	[6-10]	[11-15]	[16-20]
$P(A_1\|\theta = 1)$	75.00**	71.97	60.61	53.79
$P(A_2\|\theta = 1)$	6.82	8.33	4.55	6.82
$P(A_3\|\theta = 1)$	18.18	19.70	34.84	39.39
$P(A_1\|\theta = 2)$	73.86	67.05	53.41	47.73
$P(A_2\|\theta = 2)$	26.14	32.95	46.59	52.27
* The values in brackets represent repetitions.				
** These values are in percentages.				

Table 6 presents the percentage of times that incumbents made equal or unequal choices during the second period of the repetition conditional on having been accepted in the first period and conditional on the draw of θ in the first period. The

table suggests many actions that are off the equilibrium path.

First, a main result is that many incumbents choose to equalize payoffs in the first period of the repetition and then "slam" voters in the second period of the repetition. These are termed theoretically inconsistent choices. They are inconsistent with the theory, since the theory requires that a bad type divert rent in both periods of the repetition as opposed to just one period of the repetition. In particular, the theory predicts that the conditional probabilities $P(A_5|A_1, \theta_{1,L})$ and $P(A_4|A_1, \theta_{1,H})$ will be equal to zero.[34] Clearly, this is not the case.

The main explanation seems to be that the assumption that the incumbent is either good or bad and that the incumbent maintains such type within the two-period game fails. This further implies that a key aspect of the theoretical and experimental environment–namely, that the incumbent's preference structure is specified by either p_x^g or p_x^b (defined previously)–is called into question. Note that this should be disclaimed by the fact that this finding may not be robust to alternative parameterizations of the experiment. For example, when $\theta_1 = 1$, if Alternative 2 (A_2) in table 1 were "costly" to the incumbent in the sense that the payoff pair were $(2.00, 1.00)$ as opposed to $(1.50, 0.75)$ and Alternative 1 conditional on $\theta_1 = 2$ were $(1.00, 1.00)$, it would be an empirical question whether this finding would still arise.[35]

Nonetheless, it should be recognized that the specification of preferences in the theoretical model may need to be re-thought. Since theoretically payoff equalization is a characteristic of the good type, the sole fact that we observe bad incumbents choosing to equalize payoffs, is an indication that the incumbent's preferences cannot just be characterized by either a good- or a bad-type's preferences. Instead, a "correct" characterization is most likely a mixture of these two types of preferences.

[34] Recall that $\theta_{1,L}$ represents the first-period level of θ which is low–i.e., $L = 1$. Similarly, $\theta_{1,H}$ indicates that θ is high ($H = 2$) in the first period.

[35] This question will be addressed in a future set of experiments.

This issue is partly addressed by Lockwood (2005).

Finally, the table suggests that some good incumbents are "turning bad." This can be concluded by observing the decrease in $P(A_4|A_1, \theta_{1,L})$ and $P(A_3|A_1, \theta_{1,H})$ as the experiment session progresses. Partially, this finding also goes against the theory, since the theory assumes that–at least in the first repetition–incumbents would have a fixed type. The theory is further violated in that subjects "switch" types in other directions as well. Namely, the theory predicts that the conditional probabilities $P(A_4|A_2, \theta_{1,L})$, $P(A_4|A_3, \theta_{1,L})$ and $P(A_3|A_2, \theta_{1,H})$ will be zero. However, this is not the case. In other words, some incumbents "pool" or "separate" in the first period of the repetition and equalize thereafter.

Table 6: Second-Period Conditional on First-Period and on Theta

$N = 44$	Treatment (Task)			
	B_1	Y_1	B_2	Y_2
Probability	[1-5]	[6-10]	[11-15]	[16-20]
$P(A_4\|A_1, \theta_{1,L})$	33.33	28.99	20.83	17.95
$P(A_5\|A_1, \theta_{1,L})$	66.67	71.01	79.17	82.05
$P(A_4\|A_2, \theta_{1,L})$	14.29	0.00	50.00	100.00
$P(A_5\|A_2, \theta_{1,L})$	85.71	100.00	50.00	0.00
$P(A_4\|A_3, \theta_{1,L})$	0.00	0.00	33.33	0.00
$P(A_5\|A_3, \theta_{1,L})$	100.00	100.00	66.67	100.00
$P(A_3\|A_1, \theta_{1,H})$	32.26	25.00	40.00	20.00
$P(A_4\|A_1, \theta_{1,H})$	67.74	75.00	60.00	80.00
$P(A_3\|A_2, \theta_{1,H})$	0.00	0.00	12.50	12.50
$P(A_4\|A_2, \theta_{1,H})$	100.00	100.00	87.50	87.50

Voter Behavior. Thus far, the voter's behavior has not been explicitly addressed. While the voter's behavior can be inferred from the discussion so far, this section is explicitly dedicated to the voters' acceptance (A) and rejection decisions. A couple tables guide the discussion.

Table 7 shows the percentage acceptances conditional on the first-period choices that were available to the incumbent based on the draw of θ. The probabilities in

the table should be interpreted as follows: $P(A, 1.50|\theta_{1,L})$ is the probability of accepting a payoff of 1.50 when 1.50 was available since the draw of θ was low in the first period. Similarly, $P(A, 0.75|\theta_{1,L})$ is the probability of accepting a payoff of 0.75 when 1.50 was available since θ was low. Similar interpretations can be associated with the remaining probabilities.

The table suggests a few key results: First, the majority of voters initially accept the 1.50 payoff, as the theory would predict. However, as the experiment session progresses, voters accept this payoff at lower rates. In this sense, voters most definitely play according to the theoretical equilibrium in the first period.

Secondly, voters fail to play according to the equilibrium prediction when the incumbent "separates." Equilibrium behavior would predict that $P(A, 0.16|\theta_{1,L})$ and $P(A, 0.16|\theta_{1,H})$ be equal to zero. Yet, they are not.[36]

Finally, voters seem to be better at detecting the bad incumbent when he or she "pools" according to the theoretical equilibrium only as the experiment session progresses. This effect can be seen by observing $P(A, 0.75|\theta_{1,L})$, which decreases to 11.11% by the end of the experiment session. This is consistent with findings in the literature. On the other hand, $P(A, 0.75|\theta_{1,H})$ does not change much through the course of the experiment session.

It is also worth noting that contrary to the theoretical game in which the level of π is common knowledge so that voters can update according to Bayes' rule, the experimental game is silent in this respect. This is due to the fact that the experimental design does not induce good or bad behavior upon the subjects. In particular, voters need to "somehow" create a belief with respect to the incumbent's type since the incumbent is allowed to have his or her homegrown type. From table 7 it is clear that voters initially perceive π as being higher than later in the experiment session.

[36] The questionnaire indicates that these decisions are partly due to a conscious choice by some voters to accept all first-period choices.

Table 7: Distribution of Voters' Acceptances Conditional on theta

$N = 44$	Treatment (Task)			
	B_1	Y_1	B_2	Y_2
Probability	[1-5]	[6-10]	[11-15]	[16-20]
$P(A, 1.50\|\theta_{1,L})$	84.85	72.63	60.00	54.93
$P(A, 0.75\|\theta_{1,L})$	77.78	54.55	33.33	11.11
$P(A, 0.16\|\theta_{1,L})$	8.33	11.54	6.52	7.69
$P(A, 0.75\|\theta_{1,H})$	47.69	40.68	42.55	35.71
$P(A, 0.16\|\theta_{1,H})$	13.04	27.59	19.51	17.39

The set of probabilities discussed in the previous table give rise to the possibility to gauge whether the voter is able to detect bad incumbents and thus whether the voter behaves according to the "pooling" equilibrium. A more intuitive way of thinking about the voters' decisions is to consider the percentage of acceptances that are conditional on the voter's payoff rather than the draw of θ. After all, such probabilities are more consistent with the voter's information set in the sense that–when accepting or rejecting a first-period choice–the voter does not know the draw of θ. Table 8 displays this alternative set of probabilities. The patterns remain unaltered. Acceptance rates continue to go down for payoffs of 1.50 and 0.75 while they tend mildly upwardly for the payoff of 0.16.

Table 8: Distribution of Voters' Acceptances Unconditional on theta

$N = 44$	Treatment (Task)			
	B_1	Y_1	B_2	Y_2
Probability	[1-5]	[6-10]	[11-15]	[16-20]
$P(A, 1.50)$	84.47	75.38	59.52	52.08
$P(A, 0.75)$	52.03	42.06	35.19	31.82
$P(A, 0.16)$	9.60	23.04	12.29	9.72

Treatment Y

This section is dedicated to the main question, which is to what extent yardstick competition has an effect on corruption. The first part discusses some stylized

facts and trends that are distinguishable from the data tables discussed thus far. The second part revisits Besley and Smart's two main propositions on the effect of yardstick competition on voter and incumbent behavior as well as voter welfare. Finally, the third part attempts to tease apart learning from treatment effects by exploiting the B_1Y_1-B_2Y_2 experimental design.

Stylized Facts Across Treatments. Thus far, the paper has not explicitly paid much attention to the time trends arising in the data. This section addresses that by summarizing a set of stylized facts. These are useful when thinking about the effect of yardstick competition on corruption. In doing so, it is important to keep in mind that the effect of yardstick competition should reflect itself in the incumbent's and the voter's decision-making. The following stylized facts can be distinguished.

1. The incumbent–Good Type. The proportion of good types in the subject population ($\hat{\pi}$) is decreasing as the experiment session progresses (see table 3). In particular, only 4.55% of subjects are "truly good" from an empirical standpoint in the sense that they equalize payoffs throughout the whole experiment session when given the opportunity to do so (see table 4). Furthermore, there seems to be a significant drop in the percentage of first-period good types when making the transition from Treatment B_2 to Treatment Y_1. This was previously termed a "structural break" in behavior. While the "restart effect" is noted when transitioning from B_1 to Y_1, it is not noted when transitioning from Y_1 to B_2.

2. The incumbent–Bad Type in First Period. It is very clear that as the experiment session progresses, incumbents are starting to show their true colors. In expectation of being rejected more often, incumbents start to "separate" more often. This is shown in table 4 by observing the progression of $P\left(A_3|\theta=1\right)$

and $P(A_2|\theta = 2)$. Again, the "structural break" in behavior is exhibited when transitioning from Y_1 to B_2.

3. The incumbent–Bad Type in Second Period. There seems to be some evidence of "slamming" the voter more frequently as the experiment session progresses. An interesting pattern arising from table 6 is that bad incumbents seem to "slam" the voter more frequently in the yardstick competition treatments than in the baseline treatments. This is generally true for both draws of θ and may have to do with the fact that incumbents retaliate for being rejected more often in the Y treatments. Furthermore, the "structural break" remains present when making the transition from Y_1 to B_2.

4. The voter. As voters "learn" and observe other incumbent's choices arising from yardstick competition, they start to reject more frequently (see table 7). An interesting pattern is that as the session progresses, more voters tend to accept the low payoff of 0.16. This is very much a violation of the theory on a repetition by repetition basis. Further disaggregated analysis and responses in the post questionnaire suggest that this result is not due to subject confusion.

5. "Structural Break in Behavior." It is worth asking what the cause of this so-called "structural break" in behavior may be. Recalling that this occurs when transitioning from Y_1 to B_2, we take this as mild evidence of yardstick competition. How so? Well, once the third treatment is announced as another Treatment B, incumbents and voters may realize that the fourth treatment will most likely be another Treatment Y. Therefore, decisions made in Treatment B_2 will very much have an effect on the voters' information sets in Treatment Y_2. Consequently, incumbents may alter their behavior. This finding should be disclaimed by the fact that subjects may be getting tired at this point in the experiment session. So, part of the "structural break" may also

> be caused by their desire to end the game. If so, incumbents are expected to "separate" more often and voters are expected to reject more often. Regardless, responses in the post questionnaire regarding the length of the experiment suggest that this "structural break" in behavior cannot be explained by subject "fatigue" alone.

Propositions 2 and 3 Revisited. The theoretical model discussed previously has some clear predictions with respect to the effect of yardstick competition. These were established in the form of propositions 2 and 3.

Proposition 2 has several theoretical predictions for the case with yardstick competition. First, the proposition claims that in the presence of yardstick competition we observe a pooling equilibrium if $\hat{\pi} > \frac{1}{2}$ and a hybrid equilibrium otherwise. From tables 3 and 4, it is safe to infer that $\hat{\pi} < \frac{1}{2}$. So, if the theory were correct, incumbents should be adopting a mixed strategy over pooling and separating in the presence of yardstick competition.

From a relative standpoint, this seems to be the case. On average incumbents play according to the pooling and separating equilibria more often in Y_1, Y_2 in comparison to B_1, B_2. In particular, in Y_1, incumbents "pool" 8.33% of the time and "separate" 19.70% of the time as opposed to 6.82% and 18.18% respectively in B_1. Similarly, they "pool" 6.82% of the time and "separate" 39.39% of the time in Y_2 in comparison to 4.55% and 34.84% in B_2. In turn, voters are able to react accordingly by accepting less often as illustrated by table 7.

Furthermore, proposition 2 also has the implication that the theoretical pooling equilibrium will occur less often when the foreign incumbent has a poor initial reputation. This aspect is more difficult to infer from the data, since incumbents seem to violate this type of equilibrium to begin with. Yet, there is some evidence to suggest that–as the experiment session progresses, which also implies that $\hat{\pi}$ is going

down (i.e., the foreign incumbent has a worse reputation)–the theoretical pooling equilibrium is less likely to occur. At the same time, support for proposition 2 is found by considering the empirical "analogue" of the theoretical pooling equilibrium. Namely, as the (foreign) incumbents' reputations worsen, incumbents are less likely to make the theoretically inconsistent choices discussed previously and "separate" by diverting maximum rents in the first period of the repetition. Voters are also more likely to reject such theoretically inconsistent choices as compared to before; i.e., they reject the 1.50 payoff more often as shown by table 7.

Proposition 3 basically predicts that if $\hat{\pi}$ is low, voter welfare will be lower when yardstick comparisons are available than when they are not. The rationale is that bad incumbents will do little to build a reputation. They will be voted out with little gain for the voter, since they are replaced by politicians of the same type. This proposition also seems to be supported in the data. As shown in table 9, average voter welfare is decreasing in the presence of yardstick competition. However, standard Mann-Whitney tests performed on average incumbent payoffs between treatments find that the differences between the average payoffs are not statistically significant (p-values are 1.000). In other words, voter welfare is statistically unchanged in the presence of yardstick competition.

This result is in line with two main findings. First, recall that $\hat{\pi}$ is relatively low. Even at the initial stage of the game, $\hat{\pi}$ was estimated to be 25%. So, proposition 3 would suggest that voter welfare is non-increasing with yardstick competition. Secondly, as is discussed further below, the learning effect dominates the treatment (i.e., yardstick competition) effect in the experiment sessions. This is most likely due to (1) the small proportion of good types in the population and (2) the learning on the part of voters that this is the case.

Table 9: Voters' Total Average Welfare

$N = 44$	
Treatment	Average Payoffs
B_1	7.56*
Y_1	7.35
B_2	6.87
Y_2	6.69
* Payoffs are in experimental dollars.	

The above discussion suggests that even though subjects do not play according to the theoretical equilibria in Treatment B, minor modifications of the theoretical hypotheses enable interpretations of the case with yardstick comparisons. In particular, the experimental yardstick games are characterized by a key environment parameter: a level of $\hat{\pi}$ that is below 50% and low enough that yardstick comparisons leave welfare levels unchanged. This is also the case in which yardstick comparisons have little effect on corrupt incumbent behavior.

Finally, it is worth noting that these results suggest some evidence against the Belleflamme and Hindriks (2005) model. While their model is not an explicit object of study in the paper, the fact that welfare seems to be non-increasing in the presence of yardstick competition provides some evidence against their model.

Learning or Treatment Effects: Averaged Analysis. Previously, we considered treatment and learning effects for first-period behavior in isolation. In this section, we consider averaged choices over all repetitions. Two tables support the main conclusion that yardstick competition has little effect on incumbent and voter behavior. All tables report Mann-Whitney p-values for the two variables of interest–in this case "Choice" and "Accept"–in order to distinguish any treatment effects.

Both tables suggest that the main effects observed in the data are due to learning and not due to yardstick competition. Table 10 reports Mann-Whitney tests for

the incumbent's averaged choices conditional on the draw of θ.[37] The table suggests uniformly that there are no statistically significant differences between Treatments B_1 (B_2) and Y_1 (Y_2), but that there *are* statistically significant differences between B_1 (Y_1) and B_2 (Y_2). This implies that the incumbent's behavior is unaffected by yardstick competition, but strongly affected by learning.

Table 11 reports the same test for the voter's averaged choices conditional on the voter's payoff resulting from the incumbent's first-period choice. This table is slightly less conclusive. For the payoff of 1.50, the results align with those above. In particular, there is a strong learning effect that mitigates the treatment effect. However, when the payoff is 0.75, there is only marginal evidence of learning. The treatment effect remains non-existent. Finally, when the payoff is 0.16 there is neither evidence of learning nor of yardstick competition.

These results should be interpreted carefully; especially in light of the previous discussion on the one-shot analysis of the game. As suggested, there seems to be an evidence of yardstick competition, which is cluttered by a learning effect. This gives rise to the immediate question: If we observe a partial treatment effect when comparing first-period behavior in Treatments B_1 and Y_1, why do we not observe such effect when comparing averaged choices across all repetitions in the respective treatments?

There are a couple of possibilities. First, when comparing first-period behavior, there seems to be a learning effect in addition to this treatment effect. It is possible that when decisions are averaged across repetitions, the learning effect dominates the treatment effect. Secondly, as the game progresses, it is possible that subjects gather sufficient information so that they are less likely to use the information arising from yardstick competition. In other words, voters are much more likely to reject the incumbents' offers by the time they enter Treatments B_2 and Y_2 without

[37] These results are supported by t-tests.

paying attention to the yardstick information.

So, the above does not imply that yardstick competition has no effect in general, but it does imply that the current theoretical model may be an intricate manner of representing a model of corruption and yardstick competition in a laboratory environment. Furthermore, it is possible that the evidence of yardstick competition is mitigated by too much repetition (five repetitions) and a "break-down" of the game. In other words, by playing the game repeatedly, voters learn that there is a sufficiently large proportion of bad types among the incumbent population and thus they ignore the additional information provided in Treatment Y_2. Besley and Smart's proposition on the effect of yardstick comparisons on voter welfare (proposition 3) seems to be partly informative in this respect given the current level of $\hat{\pi}$ that is to be inferred from the incumbent population (recall tables 3, 4 and 9).

This suggests a couple of areas for future research: First, a design in which subjects participate in the trial (possibly, for a longer period) and then proceed to playing the treatments for a total of for example eight repetitions; i.e., let each treatment consist of only two repetitions as opposed to five. Secondly, a design in which subjects participate in the trial and then play a one-shot version of the game as dictated by the theoretical model. Such treatments should then shed some light on whether the effects observed in the current data are due to the treatment, the sequencing (learning) or the repetitive nature of the experimental design.

Table 10: Incumbent's First-Period Choice Conditional on theta

$N = 44$	
Comparison	Mann-Whitney Test (p-values)
$B_1Y_1\|\theta_{1,L}$	0.5826
$B_1B_2\|\theta_{1,L}$	0.0335
$B_2Y_2\|\theta_{1,L}$	0.4872
$Y_1Y_2\|\theta_{1,L}$	0.0183
$B_2Y_1\|\theta_{1,L}$	0.0857
$B_1Y_1\|\theta_{1,H}$	0.2472
$B_1B_2\|\theta_{1,H}$	0.0116
$B_2Y_2\|\theta_{1,H}$	0.5211
$Y_1Y_2\|\theta_{1,H}$	0.0315
$B_2Y_1\|\theta_{1,H}$	0.1027

Table 11: Voter Acceptance Conditional on Proposed First Period Payoff

$N = 44$	
Comparison	Mann-Whitney Test (p-values)
$B_1Y_1\|1.50$	0.4580
$B_1B_2\|1.50$	0.0132
$B_2Y_2\|1.50$	0.4992
$Y_1Y_2\|1.50$	0.0214
$B_2Y_1\|1.50$	0.0929
$B_1Y_1\|0.75$	0.2657
$B_1B_2\|0.75$	0.1049
$B_2Y_2\|0.75$	0.4620
$Y_1Y_2\|0.75$	0.1485
$B_2Y_1\|0.75$	0.5267
$B_1Y_1\|0.16$	0.2195
$B_1B_2\|0.16$	0.7796
$B_2Y_2\|0.16$	0.7384
$Y_1Y_2\|0.16$	0.1366
$B_2Y_1\|0.16$	0.2312

Learning or Treatment Effects: Disaggregated Analysis. As a final check for yardstick competition, we return to disaggregated data from the four experiment sessions. While several of the discussions that follow are based on statistically insignificant results, they are still interesting. Table 13 recapitulates by session (i.e., 1–4) and treatment (i.e., B, Y) the average first-period decisions made by incum-

bents (i.e., Choice) and voters (Accept). These distributions are not conditioned on the draw of the unit cost θ, since it is useful to look at these results from the voter's perspective. Namely, the "Choice" distributions in Treatments B_1 and B_2 represent the actual information conveyed to voters in Treatments Y_1 and Y_2.

First, note that the proportion of incumbents that equalize first-period payoffs with certainty is always less than 50%. Even though voters do not know this information a priori, they observe it once they are in the Y-treatments. Thus, according to proposition 3 we would expect voters to be weakly worse off in the yardstick sessions than in the baseline sessions. This finding was established previously.

Secondly, while voter welfare is lower in the presence of yardstick competition due to the proportion of types in the incumbent population, this finding is also driven by voters' decision-making. In particular, table 13 indicates that except for one treatment (viz. session 4, Treatment Y_2), voters tend to re-elect more frequently in the presence of yardstick comparisons upon observing the lowest payoff of 0.16. While these differences are not statistically significant (lowest p-value is 0.12), this is a striking result. Why would voters re-elect bad incumbents at any higher rates in the Y-treatments unless they are confused?

To rule out the latter as a possibility, we perform Epps-Singleton tests comparing the distribution of voters' actual re-election decisions in the Y-treatments with random distributions of re-election decisions. Generally, these tests reject the nullhypothesis that voters' re-election decisions are distributed randomly (p-values range from 0.00 to 0.007). In other words, voters' re-election decisions are not random. So, the above finding is taken as evidence of yardstick competition. In particular, this finding supports the claim that bad domestic incumbents are more likely to be retained in office when voters observe equally bad foreign incumbents. Finally, it should be noted that with the exception of Treatment Y_1 in session 3, yardstick competition never seems to push incumbents towards better behavior.

Table 12: Distribution of Choices and Re-elections by Session and Treatment

$N = 24$	Session 1, B_1		Session 1, Y_1		Session 1, B_2		Session 1, Y_2	
Payoff	Choice	Accept	Choice	Accept	Choice	Accept	Choice	Accept
1.50	50*	86.67	45	70.37	35	66.67	33.33	8
0.75	38.33	34.78	30	27.77	20	25	25	20
0.16	11.67	0	25	13.33	45	7.41	41.67	8
$N = 18$	Session 2, B_1		Session 2, Y_1		Session 2, B_2		Session 2, Y_2	
1.50	42.22	78.95	33.33	80	22.22	80	26.67	41.67
0.75	26.67	50	26.67	58.33	26.67	50	13.33	16.67
0.16	31.11	21.43	40	22.22	51.11	13.04	60	14.81
$N = 22$	Session 3, B_1		Session 3, Y_1		Session 3, B_2		Session 3, Y_2	
1.50	45.45	84	49.09	74.07	45.45	56	38.18	38.10
0.75	23.64	76.92	30.91	58.82	21.82	50	20	58.33
0.16	30.91	5.88	20.00	27.27	37.73	16.67	41.82	21.73
$N = 24$	Session 4, B_1		Session 4, Y_1		Session 4, B_2		Session 4, Y_2	
1.50	41.67	88	43.33	69.23	40	50	30	55.55
0.75	43.33	53.85	38.33	34.78	28.33	41.18	31.67	42.11
0.16	15	11.11	18.33	22.22	31.67	15.79	38.33	4.35
* All values are in percentages. $T = 20$ in all cases.								

Gender Differences

It is well-known that gender (or sex) tends to affect subjects' choices. This finding has been discussed in corruption studies (see for example Frank and Schulze 2000) and in other types of studies (see for example Cox and Deck 2006).

The tables below suggest the following main results. For both incumbents and voters yardstick competition does not seem to have a significant effect on subjects' choices. However, in terms of learning, the results are disjoint. While male incumbents exhibit learning in Treatments Y conditional on $\theta = 1$, female incumbents exhibit learning in Treatments B and Y conditional on $\theta = 2$. There is also a case in which neither male nor female incumbents exhibit learning (see $B_1 B_2 | \theta_{1,L}$).

As for the voter, the effects of learning seem to be mixed as well. Female voters only learn in Treatment B when conditioned on 1.50. However, male voters exhibit learning in Treatment Y and B conditional on respectively 1.50 and 0.75. When the voter's payoff is 0.16, neither male nor female voters exhibit learning.

Finally, table 17 suggests some mild evidence of yardstick competition for male voters when transitioning from Treatment B_1 to Treatment Y_1 when the payoff is 0.75. This is also captured by table 15. However, table 18 shows that the effect of yardstick competition here is not economically significant. This is reflected by the mild change in averaged payoffs from 1.08 to 0.99. It should be noted that one of Besley and Smart's yardstick propositions (proposition 2) informs the statistically significant finding. In particular, the proposition claims that voters will use yardstick competition in equilibrium when "pooling" as long as $\pi \geq \frac{1}{2}$. Thus, the fact that male voters do not "pool" as frequently (i.e., reject more frequently) in Treatment Y_1 as compared to Treatment B_1 may suggest that voters are using yardstick comparisons to update their beliefs regarding the level of π. Consequently, they reject the 0.75 payoff more frequently in the presence of yardstick competition.

Table 13: Distribution of Incumbents' First-Period Choices

$N = 44$	Treatment (Task)							
	Male				Female			
	B_1	Y_1	B_2	Y_2	B_1	Y_1	B_2	Y_2
Probability	[1-5]*	[6-10]	[11-15]	[16-20]	[1-5]	[6-10]	[11-15]	[16-20]
$P(A_1 \vert \theta = 1)$	82.35**	74.51	64.71	54.90	70.37	70.37	58.02	53.09
$P(A_2 \vert \theta = 1)$	1.96	9.80	3.92	7.84	9.88	7.41	4.94	6.17
$P(A_3 \vert \theta = 1)$	15.69	15.69	31.37	37.26	19.75	22.22	37.04	40.74
$P(A_1 \vert \theta = 2)$	82.35	73.53	67.65	58.82	68.52	62.96	44.44	40.74
$P(A_2 \vert \theta = 2)$	17.65	26.47	32.35	41.18	31.48	37.04	55.56	59.26
* The values in brackets represent repetitions.								
** These values are in percentages.								

Table 14: Second-Period Choices Conditional on First-Period Choices and theta

$N=44$	Treatment (Task)							
	Male				Female			
	B_1	Y_1	B_2	Y_2	B_1	Y_1	B_2	Y_2
Probability	[1-5]	[6-10]	[11-15]	[16-20]	[1-5]	[6-10]	[11-15]	[16-20]
$P(A_4 \vert A_1, \theta_L)$	31	43	33	31	35	18	11	9
$P(A_5 \vert A_1, \theta_L)$	69	57	67	69	65	82	89	91
$P(A_4 \vert A_2, \theta_L)$	0	0	100	100	17	0	0	n/a
$P(A_5 \vert A_2, \theta_L)$	100	100	0	0	83	100	100	n/a
$P(A_4 \vert A_3, \theta_L)$	n/a	0	50	0	0	0	0	0
$P(A_5 \vert A_3, \theta_L)$	n/a	100	50	100	100	100	100	100
$P(A_3 \vert A_1, \theta_H)$	46	30	447	14	22	21	36	25
$P(A_4 \vert A_1, \theta_H)$	54	70	56	86	78	79	64	75
$P(A_3 \vert A_2, \theta_H)$	n/a	0	20	0	0	0	0	14
$P(A_4 \vert A_2, \theta_H)$	n/a	100	80	100	100	100	100	86
n/a means this "option" was never chosen.								

Table 15: Distribution of Voter Acceptances Unconditional on theta

$N=44$	Treatment (Task)							
	Male				Female			
	B_1	Y_1	B_2	Y_2	B_1	Y_1	B_2	Y_2
Probability	[1-5]	[6-10]	[11-15]	[16-20]	[1-5]	[6-10]	[11-15]	[16-20]
$P(A, 1.50)$	77.54	81.88	54.55	50.00	92.06	68.25	65.00	54.63
$P(A, 0.75)$	60.00	38.46	35.71	23.68	45.83	47.92	39.47	35.71
$P(A, 0.16)$	17.39	27.08	9.38	13.96	6.25	15.91	14.52	7.78

Table 16: Incumbents' First-Period Choices

$N=44$	Mann-Whitney Test (p-values)	
Comparison	Male	Female
$B_1Y_1 \vert \theta_{1,L}$	0.6828	0.7207
$B_1B_2 \vert \theta_{1,L}$	0.1530	0.1138
$B_2Y_2 \vert \theta_{1,L}$	0.5103	0.7096
$Y_1Y_2 \vert \theta_{1,L}$	0.0777	0.1037
$B_2Y_1 \vert \theta_{1,L}$	0.2593	0.1903
$B_1Y_1 \vert \theta_{1,H}$	0.4497	0.4167
$B_1B_2 \vert \theta_{1,H}$	0.2471	0.0233
$B_2Y_2 \vert \theta_{1,H}$	0.5739	0.6776
$Y_1Y_2 \vert \theta_{1,H}$	0.3336	0.0380
$B_2Y_1 \vert \theta_{1,H}$	0.6714	0.0635

Table 17: Voter Acceptance Conditional on Proposed First Period Payoff

$N = 44$	Mann-Whitney Test (p-values)	
Comparison	Male	Female
$B_1Y_1\|1.50$	0.5858	0.1379
$B_1B_2\|1.50$	0.1192	0.0326
$B_2Y_2\|1.50$	0.7545	0.5296
$Y_1Y_2\|1.50$	0.0231	0.3565
$B_2Y_1\|1.50$	0.0527	0.6851
$B_1Y_1\|0.75$	0.1060	0.8865
$B_1B_2\|0.75$	0.0863	0.5609
$B_2Y_2\|0.75$	0.3346	0.8536
$Y_1Y_2\|0.75$	0.2290	0.3522
$B_2Y_1\|0.75$	0.8297	0.5193
$B_1Y_1\|0.16$	0.3776	0.3015
$B_1B_2\|0.16$	0.7006	0.3520
$B_2Y_2\|0.16$	0.8133	0.4601
$Y_1Y_2\|0.16$	0.2066	0.3815
$B_2Y_1\|0.16$	0.1419	0.8573

Table 18: Voter Average Payoffs Conditional on First Period Payoff

$N = 44$					
Condition	Average Payoffs*		Condition	Average Payoffs*	
Accept (A)	Male	Female	Reject (R)	Male	Female
$A, 1.50\|B_1$	2.06	1.96	$R, 1.50\|B_1$	2.27	2.33
$A, 1.50\|Y_1$	1.96	1.91	$R, 1.50\|Y_1$	2.26	2.21
$A, 1.50\|B_2$	1.92	1.87	$R, 1.50\|B_2$	2.06	2.22
$A, 1.50\|Y_2$	1.97	1.76	$R, 1.50\|Y_2$	2.03	2.14
$A, 0.75\|B_1$	1.08	1.08	$R, 0.75\|B_1$	1.54	1.48
$A, 0.75\|Y_1$	0.99	1.06	$R, 0.75\|Y_1$	1.49	1.30
$A, 0.75\|B_2$	1.12	1.18	$R, 0.75\|B_2$	1.65	1.32
$A, 0.75\|Y_2$	1.03	1.14	$R, 0.75\|Y_2$	1.40	1.50
$A, 0.16\|B_1$	0.32	0.32	$R, 0.16\|B_1$	0.79	0.88
$A, 0.16\|Y_1$	0.32	0.32	$R, 0.16\|Y_1$	0.99	0.99
$A, 0.16\|B_2$	0.59	0.42	$R, 0.16\|B_2$	0.88	0.93
$A, 0.16\|Y_2$	0.32	0.47	$R, 0.16\|Y_2$	0.99	0.99
* Payoffs are in experimental dollars					

Conclusion

This study is a contribution to the literature both theoretically and empirically. Theoretically, the study contributes a main proposition on the effect of a mean-preserving spread in the distribution of the unit cost of the public good (θ) on incumbent behavior and voter welfare. In particular, the study finds that a public-goods institution with more noise gives rise to more corruption and lower voter welfare.

Empirically, the study is a contribution in the following ways: (1) it provides experimental data on the effect of yardstick competition on incumbent and voter behavior and voter welfare; (2) it suggests ways in which to fortify the theory in order to use it as a more meaningful tool for prediction; (3) it contributes to the literature on experimental signaling (in games) and (4) it contributes to our understanding of yardstick competition as a potential corruption-taming mechanism.

Specifically, the study finds the following. First, there is an initial nontrivial (positive) proportion of good incumbents in the population. This proportion goes down as the experiment sessions progress. Secondly, a large proportion of bad incumbents make theoretically inconsistent choices given the assumptions of the model and the manner in which preferences are theoretically specified. In particular, such choices are inconsistent with the theory, since they are the result of a subject that "chooses to be good" in the first period of the repetition and then "chooses to be bad" in the second period of the repetition.

This has a few main implications: (1) the definition of a good and bad type must be re-thought, (2) the theoretical model is unable to handle this empirical phenomenon and must therefore be reformulated and (3) since the model is unable to account for these theoretically inconsistent outcomes, it can be called into question as a completely predictive tool for subject behavior in the baseline and yardstick games.

Third, evidence of yardstick competition is mild. Averaged choices across all repetitions suggest that yardstick comparisons have little effect on incumbent and voter behavior. In particular, the results indicate that these experiments are characterized by an institution with low proportion of good incumbents, which leaves little room for yardstick competition. This finding is informed by one of Besley and Smart's main propositions, which is recalled in this paper as proposition 3.

In this type of an institution with low π, there is little opportunity for yardstick comparisons since bad incumbents are likely to be replaced by equally bad incumbents. So, the voter would be better off if he or she could ignore the information arising from yardstick competition. This also turns out to be the case in which yardstick competition leads to unchanged or decreasing voter welfare as observed in the experiments.

Further evidence of yardstick competition is indicated by the fact that subjects are more likely to re-elect a truly bad incumbent in the yardstick sessions as opposed to the baseline sessions. In other words, if voters observe a sufficiently large proportion of foreign incumbents separating (i.e., offering a first-period payoff of 0.16), they are more likely to re-elect domestic incumbents that separate. They do so at their own cost.

Finally, a partitioning of the data by gender suggests that male and female subjects learn in different settings. Furthermore, there seems to be mild evidence of yardstick competition for male voters when they face a payoff of 0.75. This gives some support to Besley and Smart's other main yardstick proposition–proposition 2. In particular, there seems to be some evidence that male voters use yardstick comparisons to update their beliefs when facing the pooling equilibrium payoff of 0.75. They tend to reject it more frequently in the presence of yardstick comparisons.

The results of the experiments should be disclaimed by the limitations of the study, which are intimately related to some of the extensions considered in Viceisza (2007c). First, the yardstick competition effect may be mitigated by the within-subjects design. This can be "remedied" by running a between-subjects design. Secondly, the theoretically inconsistent choices may be overwhelming because the current payoff structure makes it non-costly for type (b, L) to equalize payoffs. This suggests an alternatively parameterized treatment.

Third, voters may currently be unclear how the unit cost of public good provision is determined in the sense that the process lacks transparency. This suggests a treatment in which the distribution of the unit cost is operationalized by means of physical devices. Furthermore, voters may be unclear how to interpret a distribution of first-period offers as opposed to just one such offer. This suggests conducting a treatment in which a voter observes just *one* other incumbent's offer chosen at random. Finally, these findings may not be robust to alternative parameterizations of the experiment. This suggests conducting Treatments M and R (see Viceisza 2007c).

EFFICIENCY, TRANSPARENCY AND YARDSTICK COMPETITION

Introduction

In recent decades, corruption has been placed highly on the economic policy agenda. Specifically, international organizations such as the International Bank for Reconstruction and Development, the International Monetary Fund and the Organization for Economic Co-operation and Development all have supported and continue to support numerous anti-corruption initiatives across the globe.

While at some point there was debate in the economics literature whether or not corruption actually enhances economic efficiency by "greasing the wheels of society," recent literature seems to agree on two main arguments. First, weak institutions are at the heart of corruption be it "good" or "bad" (Kaufmann, Kraay and Mastruzzi forthcoming). Secondly, there are cases in which corruption can be argued to be "bad" in the sense that it reduces growth (Mauro 1995) and increases poverty and inequality (see Gupta, Davoodi and Alonso-Teme 1998) and in such cases, anti-corruption mechanisms are necessary.

With a phenomenon that is so common (Transparency International's 2007 Corruption Perceptions Index) and that has been argued to be efficiency-reducing both theoretically (Rose-Ackerman 1978, Klitgaard 1991, Shleifer and Vishny 1993, Djankov et al. 2002) and empirically (previous references, Olken 2007, Betrand et al. forthcoming and Viceisza 2007b), it is important that academics, policymakers and the general public have a thorough understanding of its determinants, consequences and possible mechanisms for taming it.

Thus, a scientifically comprehensive approach to analyze corruption is warranted. In other words, it is useful to study corruption theoretically, empirically

using field data and empirically using experimental data.[38] While the majority of inquiries into corruption have been theoretical or empirical using field data, one of the principal roles for laboratory experiments on corruption is to test theories of corruption more strictly than is otherwise possible by either using field experiments or standard applied econometric techniques. In turn, such laboratory experimental data can suggest ways in which theories of corruption can be improved for doing more meaningful policy development and analysis.

This study uses laboratory experiments to address three main questions related to the determinants and control of corruption. First, the study asks whether an increase in the marginal cost of public funds gives rise to more or less corruption.[39] Secondly, the study asks whether a decrease in transparency (as modeled by a mean-preserving spread of the distribution of the unit cost of public good provision) affects corrupt behavior. Finally, the study returns to the main question addressed in Viceisza (2007a), which is whether an experimental environment with yardstick competition gives rise to more or less corruption.

To understand the contributions of this study, it is useful to set the context by focusing on two of Aidt's (2003) main claims. Aidt distinguishes between four types of theoretical studies on corruption: (1) those that model efficient corruption, (2) those that model corruption in the presence of a benevolent principal (e.g., the voter) that delegates decision-making power to a non-benevolent agent (e.g., the incumbent), (3) those that model corruption with a non-benevolent principal and (4) those that model self-reinforcing corruption in that history may play a role. Also, Aidt posits that two important considerations unite these four categories of models of corruption: (1) the degree of benevolence of the government official and (2) the

[38] Some general survey pieces are: Aidt (2003) for theories of corruption, Martinez-Vazquez, Arze and Boex (2007) and Tanzi (1998) for field empirics on corruption and Abbink (2005b) for experiments on corruption.

[39] Browning (1976) defines the marginal cost of public funds as the direct tax burden plus the marginal welfare cost produced in acquiring the tax revenue.

role of institutions versus history as a determinant of corruption.

These arguments are cited here because they provide the main rationale for conducting the experiments reported in this study. Namely the theoretical model that guides the experimental design is a prime example of studies pertaining to Aidt's second category. In particular, Besley and Smart's (2007) theoretical political agency model is one in which a principal (the voter) delegates decision-making authority to a potentially non-benevolent agent (the incumbent politician or the challenger) by means of an election. So, these theory-testing experiments provide laboratory experimental data that is relevant to a large class of models as categorized by Aidt.

Furthermore, by experimentally operationalizing Besley and Smart's model, this study explores the validity of the two considerations that unite Aidt's four categories. With regard to the first consideration, the experimental treatments explore the extent to which an agent's benevolence (i.e., the incumbent politician's corruptibility) is affected by two key changes: (1) an increase in the marginal cost of public funds, which represents an increase in tax inefficiency and (2) an increase in risk (noise) in the distribution of the unit cost of the public good, which represents a form of lack of transparency on the incumbent's side. With regard to the second consideration, the experimental treatments explore the rivalry between a potential corruption-taming institution (i.e., yardstick competition) and "history."

These experimental treatments are also of further independent interest since the data resulting from them have policy implications for developing countries. Since the marginal cost of public funds–which is a measure of tax efficiency–is closely related to the choice of tax instruments within a country (Dahlby 1998), the experimental marginal cost treatment is interesting since it sheds light on the extent to which an external increase in the marginal cost of public funds–possibly caused by a change in tax mix–affects corruptibility. This has policy implications for devel-

oping countries since these are known to have relatively different tax mixes compared to developed countries and under certain circumstances can be hypothesized to have higher marginal costs of public funds (Warlters and Auriol 2005.)

Furthermore, if the distribution of the unit cost of public good provision is taken to represent transparency in the political system, analyzing the effects of a less transparent institution on corrupt behavior is clearly of interest to developing economies. So, the experimental treatments that are characterized by a mean preserving spread of the distribution of the unit cost of the public good (i.e., a noisier distribution and therefore, a less transparent institution) address the question how a certain type of transparency affects corruption.

Finally, we explore the interaction between yardstick competition and "history of play," which is relevant to all economies–developing and developed alike. While Viceisza (2007a) has reported some experiments addressing this question, the experimental treatments conducted in this study are based on a different design and protocol. This is motivated below.

In any type of theory-testing experiment, there are numerous ways to operationalize the assumptions of the theory and very few are the theories that provide sufficient (institutional) detail such that all ambiguities with regard to their implementation is ruled out. Viceisza (2007a) found that using a within-subjects design to test the yardstick hypothesis does not lead to significant treatment effects. As discussed in the conclusion to that study, part of this may be due to the within-subjects nature of the experimental design.

So, this new set of experiments asks whether a between-subjects design gives rise to any significant treatment effects. Furthermore, the experimental design and protocol are amended to reflect some crucial changes. First, these experiments operationalize the distribution of the unit cost of the public good using a physical device. Secondly, the experiments reduce the information set arising from yardstick

competition to one random choice made by another politician as opposed to a distribution of choices made by all politicians. Finally, the experiments implement a trial, quiz and summary of the task to enhance subject understanding in addition to a post-questionnaire that can be used to control for it.

The study finds the following main results. First, tax inefficiency is an important determinant of corrupt incumbent behavior. In particular, an increase in the marginal cost of public funds makes it more costly for incumbents to equalize first-period payoffs. This drives them to separate and divert maximum rents in the first period. While voters retaliate slightly by voting incumbents out of office, they are worse off.

Secondly, we find that increased lack of a particular form of transparency (as defined in terms of an increase in risk in the distribution of the unit cost) leaves corrupt incumbent behavior unchanged. If the draw of the unit cost is unfavorable, incumbents tend to be less corrupt. So, the results suggest that lack of transparency (as defined in Viceisza 2007a) need not always make voters worse off.

While this finding may seem counterintuitive, it is not given the parameterization of the experiments. First, we must not ignore the importance of assumptions on types and preferences as discussed in Viceisza (2007a). Contrary to Besley and Smart's (2007) assumption that incumbents are either good or bad, incumbents are known to behave strategically. Since the mean-preserving spread in the distribution of the unit cost reduces the cost of equalizing first-period payoffs in the favorable state, it makes sense why incumbents behave less corruptly if the unit cost is low. Furthermore, Viceisza (2007a) models lack of transparency on the incumbent side. This can be contrasted with lack of transparency on the voter's side as discussed in Besley and Smart (2007).

Third, the experiments find strong evidence of yardstick competition. On the incumbent's side, yardstick competition acts as a corruption-taming mechanism if

the incumbent is female. On the voter's side, voters are less likely to re-elect the incumbent in the presence of yardstick competition. Specifically, voters pay attention to the difference between the tax signal in their own jurisdiction and the signal in another jurisdiction. As this difference increases, voters re-elect less.

Fourth, history is an important determinant of corruption and of re-election decisions. Incumbents are likely to make choices as they did in previous repetitions and voters are likely to vote out increasingly as the repetitions go by. In other words, they distrust the political system more significantly as time goes by and yardstick competition does not affect that. Finally, we find that individual-specific factors such as gender and beliefs play a significant role in incumbent behavior. In particular, female incumbents are more likely to divert rent when the unit cost is unfavorable. Furthermore, incumbents' beliefs are more important than voters' beliefs in decision-making. Voters focus mainly on taxes charged (payoffs) and the history of those taxes.

The remainder of the paper is organized as follows. The following section discusses the design, protocol and implementation. Then, we report the main findings. The final section concludes.

Experimental Design

Experimental Game: Basics, Players, Information, Actions and Payoffs

The theoretical model that underlies the experimental design follows Besley and Smart (2007). It is a game theoretic model of elections that is cast in a principal-agent framework. There are two "active" players–a principal (the voter) and a first agent (the first-period incumbent), and one "passive" player–a second agent (the challenger). An agent's type (i) can be good (g) or bad (b). Each agent knows his or her own type; however, the principal and the other agent do not. The model is comprised of two periods. At the end of the first period, there is an election. The

second agent is "passive," in the sense that this agent only plays a role during elections and during the second period if elected.

This describes a dynamic game of incomplete information. It is dynamic because it encompasses two periods and decisions made in the first period affect outcomes in the second period. There is incomplete information, since at least one player is uncertain about another player's payoff. In particular, the voter is uncertain about the politicians' payoffs, since such payoffs are type-dependent. Following Harsanyi (1967, 1968), the game is transformed into a dynamic game of imperfect information by introducing moves of nature that determine types. For a detailed explanation of the theoretical game, see Viceisza (2007a). The experimental game described below is a strict adaptation of this theoretical game.

The experimental game is a ten-time repeated one-shot dynamic signaling game in which incumbents and voters face alternatives that are "on the equilibrium path." In order to minimize repeated-game effects, subjects are told and guaranteed that they will not interact with the same player for more than one repetition (see implementation and protocol section as well as subject instructions.) Subjects are randomly assigned to be an incumbent or a voter. Once assigned such role, a subject maintains the same role throughout the entire experiment.

The following steps describe one full repetition:

1. Nature moves and the incumbent's type (i) is determined to be good (g) or bad (b).

2. Nature moves again and the unit cost of public good provision (θ) is determined to be high or low and revealed to the incumbent.

3. The incumbent's type–which is private information–is determined by two components (i, θ), i.e., whether he or she is by nature good or bad and whether the unit cost of public good provision is by nature high or low.

4. The voter knows the distribution of good and bad types and the distribution of the unit cost of the public good. He or she does not know the actual draw of the unit cost.

5. Having observed his or her type and preferences, the incumbent chooses a payoff pair from a set of alternatives. Each alternative represents a tax-public good (t, G) signal.

6. The voter observes his or her payoff and the range of possible payoffs for the incumbent.

7. The voter re-elects (accepts) or votes out (rejects) the incumbent.

8. If the incumbent is voted out, he or she gets a second-period payoff of zero and the challenger's type is determined by a lottery.[40] The result of the lottery determines the voter's second-period payoff.

9. If the incumbent is re-elected, the incumbent is in office for a second period.

10. The second-period alternatives that he or she chooses from are affected by another draw of the unit cost of the public good θ as determined by nature. Second-period payoffs for both players are determined by the incumbent's second-period choice.

11. There is no election at the end of the second period.

12. This entire process is repeated ten times. In each repetition, the incumbent faces a different voter and vice versa.

In operationalizing this game, some main issues in experimental design need to be addressed. First, we must answer the question whether or not types and preferences are to be induced. On the voter side, this is less of an issue. While voters are

[40] Since the challenger is a move of nature and therefore a "passive" player in the theoretical game, the challenger is introduced experimentally as a lottery.

theoretically assumed to have the same preferences as the good incumbent, experimentally voters are allowed to have their own homegrown preferences. This is part of what we want to learn from conducting the experiments.

On the incumbent side, the question is slightly more involved. As explained above, the incumbent's type is determined by two components–whether he or she is good or bad and whether the unit cost is high or low. The latter is induced by physically tossing a coin and inputting the result into the computer. The former however is not induced. The rationale for not inducing good or bad behavior is explained in detail in Viceisza (2007a) and hinges on what we hope to learn from conducting the experiments. In particular, by not inducing the incumbent to be good or bad, we are able to learn the proportion of good and bad types in the population and thus we are able to test one of the most basic assumptions of the theory (viz. that the proportion of good types is strictly greater than zero.)

Allowing incumbents to behave according to their homegrown types has another important implication: Contrary to the assumption that theoretical voters know the exact distribution of types within the population, the voters in the experiment do not. In particular, voter subjects must create their own beliefs with regard to the proportion of good versus bad types in the population. Aggregate re-election (acceptance) rates enable us to infer such beliefs from the data. Furthermore, incumbents' actual choices enable us to infer the true proportion of good and bad types within the subject population. This can in turn be compared to voters' beliefs.

Secondly, we must decide how to specify payoffs and parameterize the experiments. While we summarize these parameterizations here, we refer the reader to Viceisza (2007a) for further motivation. Table 19 provides a summary of the parameterizations for the baseline game. All treatments reported in this experiment rely on these parameterizations with possibly some minor modifications as explained in the next section.

Table 19: Baseline Parameterizations

Basic Definitions: t = tax, G_t = public good provision associated with tax (t)	
Good-Type/Voter Payoff	$W(G,t) = G - \mu t^{1.001}$
Bad-Type Payoff	$R = t - \theta * G_t$
Unit Cost (θ)	With equal probability, $\theta = 1$ (low) or $\theta = 2$ (high)
Marginal Cost (μ)	$\mu = \frac{1}{4}$
Max. Tax Collection (T)	$T = 4$
Min. Public Good (G_T)	$G_T = 1$

Treatments

To address the three main questions posited in the introduction, we must first conduct a baseline treatment from which we can infer deviations. This leads to a total of four treatments: (1) the Baseline treatment (Treatment B), (2) the Marginal cost treatment (Treatment M), (3) the unit cost Risk treatment (Treatment R) and (4) the Yardstick treatment (Treatment Y.) As mentioned previously, all four treatments are based on the above game and parameterization with possibly some minor modifications.

This section serves two purposes: First, to describe the theoretical propositions that inform each of these treatments. Secondly, to describe how each of these treatments differ from each other in terms of parameterizations and payoff alternatives faced by the subjects. As a final comment, it should be noted that all experimental treatments were conducted using a between-subjects design. In other words, any given pair of subjects (politician-voter) participated in one and only one of these treatments. This is a primary difference between the current design and that implemented in Viceisza (2007a).

The Baseline Treatment (Treatment B). Besley and Smart (2007) derive one main proposition that can be used to make inferences from the data resulting from Treatment B. Given the specified game form and payoffs, they derive three equilibrium configurations–a pooling, separating and hybrid equilibrium.

Intuitively, these say the following:

1. The pooling equilibrium corresponds to the case in which it is optimal for the bad incumbent that faces low unit cost to exercise restraint in the first period. In this case, this type sends the same message as the good type that faces high cost and the voter observes high taxes regardless of the type of politician that is in office. The voter re-elects with certainty upon observing this level of taxes.

2. The separating equilibrium corresponds to the case in which it is optimal for the bad type that faces low unit cost to divert maximal rents in the first period. The voter observes high taxes if the good type that faces high cost is in office and maximal taxes if the bad type is in office. The bad incumbent is detected with certainty ex post. Thus, if the voter observes high taxes, the voter re-elects with certainty; however, upon observing maximum taxes, the voter does not re-elect the incumbent to office.

3. The hybrid equilibrium corresponds to a mix between these two equilibria. In the hybrid equilibrium, the bad type that faces low cost adopts a strictly mixed strategy between restraint and maximal rent diversion. The voter observes high taxes some of the time and maximum taxes some of the time.

The data resulting from Treatment B will be tested against these predictions. Treatment B is based on the parameterizations in table 19. Given these parameterizations, incumbents and voters face the following alternatives during the experimental treatments (see table 20). These alternatives apply to the first period of each repetition and represent the incumbent's and voter's payoffs respectively. For second-period alternatives, see subject instructions.

Table 20: First-Period Alternatives faced by Incumbent in Treatment B

Alternative	If Unit Cost is Low ($\theta = 1$)	Alternative	If Unit Cost is High ($\theta = 2$)
A_1	$(1.50, 1.50)$	A_1	$(0.75, 0.75)$
A_2	$(1.50, 0.75)$		not applicable*
A_3	$(2.35, 0.16)$	A_2	$(1.35, 0.16)$
* There is no comparable alternative due to the "on the equilibrium path" assumption.			

The Marginal Cost Treatment (Treatment M). Besley and Smart (2007) derive one main proposition with regard to the marginal cost of public funds. In particular, they show that an increase in the marginal cost of public funds–which represents increased inefficiency in the tax institution–reduces voter welfare even if it reduces corrupt incumbent behavior. This main result rests on the assumption that the incumbent and voter equilibrium strategies remain unchanged, which is not always the case.

The main difference between Treatment M and Treatment B is that the former is based on a higher marginal cost of public funds than the latter. Specifically, as stated in table 19 the marginal cost of public funds for Treatment B is $\mu = \frac{1}{4}$. In Treatment M this parameter is increased to one half and this has a direct effect on the alternatives faced by the incumbent. Table 21 summarizes these alternatives accordingly.

Table 21: First-Period Alternatives faced by Incumbent in Treatment M

Alternative	If Unit Cost is Low ($\theta = 1$)	Alternative	If Unit Cost is High ($\theta = 2$)
A_1	$(0.60, 0.60)$	A_1	$(0.40, 0.40)$
A_2	$(0.60, 0.40)$		not applicable*
A_3	$(2.35, 0.00)$	A_2	$(1.35, 0.00)$
* There is no comparable alternative due to the "on the equilibrium path" assumption.			

The Unit Cost Risk Treatment (Treatment R). Viceisza (2007a) derives a main proposition with regard to the unit cost of the public good. In particular, it is shown that a less transparent public goods provision institution (as represented

by an increase in risk of θ in the second-order stochastic dominance sense) is associated with an increase in expected equilibrium rent diversion. Furthermore, a less transparent institution is associated with a decrease in equilibrium voter welfare as long as the marginal social cost of taxation is less than the inverse of the product of the marginal cost of public funds and the unit cost of the public good.

Intuitively, this proposition says that in a less transparent institution–i.e., an institution in which the unit cost of the public good (θ) has a noisier distribution–equilibrium corruption is expected to increase and equilibrium voter welfare is expected to decrease as long as the marginal cost of taxation is low enough.

The main difference between these two treatments is that Treatment R is based on a unit cost of the public good that has a riskier distribution than the unit cost in Treatment B. In particular, the high level of the unit cost is now two and one half and the low level of the unit cost is one half. So, while the mean of the distribution of the unit cost has remained the same, the variance is higher. As anticipated, this has a direct effect on the alternatives faced by the incumbent. The specific alternatives are discussed in table 22.

Table 22: First-Period Alternatives faced by Incumbent in Treatment R

Alternative	If Unit Cost is Low ($\theta = \frac{1}{2}$)	Alternative	If Unit Cost is High ($\theta = 2\frac{1}{2}$)
A_1	$(2.63, 2.63)$	A_1	$(0.49, 0.49)$
A_2	$(2.63, 0.49)$		not applicable*
A_3	$(2.85, 0.16)$	A_2	$(0.85, 0.16)$
* There is no comparable alternative due to the "on the equilibrium path" assumption.			

The Yardstick Treatment (Treatment Y). Besley and Smart focus on symmetric equilibria among incumbents and voters in the two jurisdictions by assuming that the joint probability mass function of unit cost shocks $\Pr(\theta, \theta')$ is symmetric. Furthermore, they assume that the cost shocks in the two jurisdictions are positively correlated. Under such conditions, the voter's strategy involves yardstick

competition when re-election occurs with positive probability if spending is high in both jurisdictions, but the probability of re-election is zero if domestic spending is high and foreign spending is low.

Their first yardstick proposition intuitively tells us that the case with yardstick competition differs from the one jurisdiction case in three essential ways:

1. A bad incumbent may not be re-elected when he or she charges high taxes, if the foreign incumbent is good and faces a low unit cost.

2. A good domestic incumbent is retained in office when costs are high, and the foreign politician charges maximum taxes.

3. Pooling may no longer be optimal for the domestic incumbent when the foreign incumbent has a poor initial reputation.

Their second yardstick proposition shows that voters who are better informed about the fiscal environment may be worse off in equilibrium, since bad incumbents will make less of an effort to build a reputation when they first take office. Namely, in some cases–in particular, when there are relatively few good incumbents–voters would be better off if they could commit to ignoring the fiscal performance in the other jurisdiction in the course of a domestic election.

Yardstick competition is welfare decreasing when politicians' reputations are poor, since rents are increased with little advantage from the improved information generated as most politicians who are voted out of office are replaced by an incumbent of the same type.[41] These yardstick results guide the data analysis when comparing Treatment Y to Treatment B.

A final comment is necessary: This treatment is based on the same parameterization as Treatment B. The difference between this treatment and Treatment B is the information observed by the voters–as explained in the subject instructions. In

[41] Some evidence of this was observed in the experiments reported by Viceisza (2007a).

particular, voters in Treatment Y observe additional information over and above what the voters in Treatment B observe. In each repetition, Y-voters observe a random first-period choice made by some incumbent in a B-session. This is the information arising from yardstick competition and it is randomly chosen, since the concept of a neighboring jurisdiction is not necessarily well-defined in this context.

Implementation and Protocol

The experiments took place in the experimental laboratory at the Experimental Economics Center (ExCEN) at Georgia State University. Subjects were recruited using ExCEN's online recruiter system, which contains names of students taking courses in many different areas including but not limited to accounting, actuarial science, biology, business administration, chemistry, economics, finance, geology, geography, mathematics, nursing, political science and sociology. The experiments were programmed and conducted with the software z-Tree (Fischbacher 2007).

In each experiment session, half of the subjects were incumbents and half of the subjects were voters. Each experiment session had at least twenty subjects in order to guarantee that a given politician and voter were never paired for more than one repetition. In other words, a given politician did not interact with a given voter for more than one repetition and vice versa. So, re-matching during the experiment was pre-determined according to the "two-ships-passing-in-the-night" design. Subjects were informed accordingly.

Each experiment session consisted of instructions, a five-minute trial, a quiz, a summary of the treatment, the treatment and a post-questionnaire.[42] The experiments lasted an average of ninety minutes and average payoffs were $14.13,

[42] Note that the appendices comprise subject instructions only for Treatments B and Y, since the instructions for Treatments M and R are minor variations of B in terms of payoffs. The subject instructions consist of a description of the task (treatment), a description of the trial and quiz and a summary of the task (treatment). Note also that the appendices do not contain the actual quizzes. The interested reader should contact the author.

with the highest-paying session having an average payoff of \$15.95 and the lowest-paying session having an average payoff of \$12.70. Subjects were paid \$5.00 for showing up, \$3.00 for completing the trial and quiz and \$2.00 for completing the post-questionnaire.

The procedures during the experiment were as follows. Subjects were assigned a number at sign-in. These numbers were used to randomly enter subjects into the experimental laboratory. Random entry also determined random assignment to a fixed role and different pairs during the main treatment.

After entering the experimental laboratory, subjects were handed paper-based instructions, which were read out loud by the experimenter. They were then put through the five-minute trial. This was an opportunity for subjects to interact with the software and practice making decisions that did not affect their payoffs. Thereafter, subjects were put through the quiz. Contrary to the trial, the quiz was not timed. Subjects were informed that they would make three dollars for completing the trial and quiz regardless of how many questions they answered correctly. They were asked to pay attention to the screens observed in the trial and to the questions asked in the quiz.

During the quiz, the software informed subjects whether or not they answered a particular question correctly. In both cases, the quiz gave an overview of the correct answer and referred subjects to the instructions. After the quiz was completed, subjects were asked whether they had any questions that they wanted clarified in private. If so, those were clarified. The experimenter then summarized the treatment. In doing so, the experimenter elaborated on the main issues and any particular issues that seemed to be problematic based on the responses in the quiz. The problematic issues were uniform across sessions.

Once this process was completed, subjects were ready to start the main treatment. At this point, subjects were informed whether they would be Player X (a

politician) or Player Z (a voter). Subjects were told to note that the room was divided into two "sections" by means of a blank column of computer stations running from front to back. This blank column divided the room into incumbents and voters. Everyone on one side of the room was randomly assigned to be an incumbent and everyone on the other side was randomly assigned to be a voter. Subjects were reminded that they would keep the same role throughout the entire experiment.

In addition, within each section, subjects were separated from each other by means of dividers. This guaranteed that subjects could make their decisions in private. Furthermore, the voter side had higher dividers such that it would be impossible for a particular voter to observe the result of the coin toss. This also reinforced the nature of the information asymmetry that is crucial to the game. Subjects were told not to communicate with each other during the experiment.

The treatment procedures were as follows. At the beginning of each repetition and period within a repetition, the experimenter flipped a coin. This represented the move by nature that determined the unit cost of the public good (θ), which was induced. The coin toss took place in front of the first two politicians. They observed and verified the coin toss and its result. Upon verification, the experimenter input the result into the computer and the respective period was conducted. Upon conclusion of the two periods (i.e., a given repetition), the process was repeated until all ten repetitions were concluded. After ten repetitions, subjects completed the post-questionnaire.

Experimental Findings

Aggregation and Demographics

A total of eight experiment sessions were conducted: two B-sessions, two M-sessions, two R-sessions and two Y-sessions. The main choice (endogenous) variables of interest are first-period choices by incumbents (this variable is named Choice)

and re-election (acceptance) decisions by voters (this variable is named Accept).[43] Full distribution tests (Kolmogorov-Smirnov, KS and Epps-Singleton, ES) suggest that there are no statistically significant differences between the sessions for these main variables of interest.[44]

Aggregation across experiment sessions leads to the following number of subjects for the respective treatments: (1) Treatment B has a total of 30 politicians and 30 voters, (2) Treatment M has a total of 22 politicians and 22 voters, (3) Treatment R has a total of 22 politicians and 22 voters and (4) Treatment Y has a total of 28 politicians and 28 voters. These numbers represent average number of observations, since not all sessions within the treatments were of the same size. In particular, Treatment M and Treatment R both had a session of minimum size–i.e., twenty subjects. KS tests and Mann-Whitney tests suggest that behavior in these sessions is not statistically significantly different from behavior in the larger sessions (p-values are greater than 0.26).

The following descriptive statistics describe the average subject profile. Female subjects constitute 56.03% of the sample. The average age is 22.11 years with a standard deviation of 4.04 years. About 13% of subjects are economics majors, 12% are biology majors, 8% are accounting majors, 4% are political science majors. The remaining subjects comprise miscellaneous majors including but not limited to film and video, sociology, gerontology, chemistry and english. Finally, the maximum self-

[43] Incumbents' second-period choices if re-elected are also of potential interest. This variable is named Choice2 and it is studied descriptively in the summary tables.

[44] All KS p-values are greater than 0.358. All ES p-values are greater than 0.13. The only exception is for the variable Choice when comparing the R-sessions. However, after "unconditioning" first-period choices on the draw of the unit cost (θ)–i.e., re-coding decisions to be either "equalize" or "not equalize"–we cannot reject the null that incumbents' first-period decisions in the two sessions are the same (p-value is 0.196.) Furthermore, if we separate first-period decisions in the R-sessions by the level of the unit cost, we are once again unable to reject the null that first-period choices in the two R-sessions are the same (p-value is 0.270.) This implies that one of the following techniques should capture the discrepancy between the R-sessions when running regressions: (1) partitioning the data into "equalize/not equalize", (2) controlling for the draw of the unit cost in the regressions or (3) partitioning the data by the level of the unit cost. As a final control measure, we also include session-level dummy variables for the respective R-sessions (R_1, R_2) in the regressions. The findings are robust to all these separate controls.

reported annual income range is \$30,001 to \$60,000 with a mode annual income range of \$0 to \$1,000.

Average Behavior: Tabular Overview

In this subsection, we look at average incumbent and voter decisions for the respective treatments. In particular, we compare average rates across Treatments B, M, R and Y. We start by discussing the progression of incumbents' types during the treatment. We then focus on three main variables: Choice (incumbents' first-period choices), Accept (voters' re-election decisions) and Choice2 (incumbents' second-period choices conditional on having been re-elected (accepted) for a second term).

Incumbents' Types: Good versus Bad. Besley and Smart's (2007) model defines a good incumbent as one that equalizes payoffs in both periods. It is interesting to study what percentage of incumbents actually behave in this manner on a repetition basis. Table 23 answers this question. In particular, the table compares the progression of good types (in percentages) across treatments conditional on the unit cost of the public good (θ) by repetition.

The following patterns are distinguishable. First, with some fluctuation the percentage of good types tends to be higher the first half of the treatment. This finding is relatively robust across all treatments and draws of the unit cost. Secondly, the proportion of good types is overwhelmingly lower in Treatment M compared to any other treatment. This proportion is relatively higher in Treatment R. Finally, the proportion of good types is fairly persistent in Treatments B and Y–particularly when the unit cost is high.

Table 23: Proportion of Good Incumbents Across Repetitions and Treatments

	Repetition									
	1	2	3	4	5	6	7	8	9	10
Treatment	Unit Cost Low									
B	55.56	40.00	40.00	*	25.00		41.67		33.33	28.57
M	25.00	0.00	0.00	28.57	14.29	25.00	14.29	0.00	0.00	
R	71.43	75.00	62.50		16.67		33.33	50.00	50.00	
Y			30.77	37.50	16.67	16.67	14.29	10.53	33.33	
	Unit Cost High									
B	33.33	20.00		21.43	33.33	33.33		27.27	25.00	
M	75.00	0.00	20.00	0.00	0.00			0.00		16.67
R	50.00	28.57	62.50	33.33	75.00	87.50	0.00		20.00	0.00
Y	33.33	28.57		33.33	0.00	16.67	33.33		16.67	22.22
All numbers are in percentages. $22 \leq N \leq 30$, $T = 10$										
* Blank cells imply that the draw of the unit cost was not realized in a repetition.										

Incumbents' First- and Second-Period Choices. Another interesting statistic is the average percentage of incumbents that equalize first-period payoffs within a treatment conditional on the draw of the unit cost of the public good. Table 24 looks at this. The following patterns that reinforce previous results seen in table 23 are distinguishable. First, incumbents are relatively less likely to equalize first-period payoffs in Treatment M. Furthermore, incumbents in Treatment Y equalize first-period payoffs at similar rates regardless of the level of the unit cost. These rates are lower than those in Treatment B. Finally, while incumbents in Treatment R are more likely to equalize first-period payoffs when the unit cost is low, they are also less likely to do so when the unit cost is high. This is in comparison to Treatment B.

Table 24: Incumbents' Average First-Period Choices Across Treatments

$22 \leq N \leq 30$, $T = 10$	Equalize	Not Equalize	Equalize	Not Equalize
Treatment	Unit Cost Low		Unit Cost High	
B	64.00	36.00	66.67	33.33
M	33.58	66.42	40.70	59.30
R	82.14	17.86	64.71	35.29
Y	58.27	41.73	57.52	42.48
All numbers are in percentages.				

As discussed previously, Besley and Smart's (2007) model defines a good type as one that always equalizes payoffs and a bad type as one that always diverts rent– i.e., behaves corruptly. Furthermore, the model assumes that agents are either good or bad. So, it is interesting to ask whether there are incumbents who behave as if they are good in the first period and as if they are bad in the second period. Not only does this mitigate the plausibility of the aforementioned assumption; however, it shows clear strategic behavior on the part of the incumbent. Viceisza (2007a) termed choices that follow this pattern theoretically inconsistent according to Besley and Smart's definition of types.

Table 25 shows this descriptive across the treatments. In particular, the table shows the percentage of incumbents that equalize payoffs in the first period and divert maximum rents in the second period. The following patterns are noteworthy. First, incumbents in Treatment M are much more likely to divert maximum rents in the second period after having equalized in the first period. This holds regardless of the level of the unit cost. Secondly, a similar pattern is noticeable for Treatment Y. Third, incumbents in Treatment R are less likely to slam the voter in the second period after having equalized in the first. Finally, with the exception of Treatment B incumbents make these theoretically inconsistent choices at similar rates regardless of the draw of the unit cost.

Table 25: Bad in Second Period Given Good in First Period

$22 \leq N \leq 30$, $T = 10$		
Treatment	Unit Cost Low	Unit Cost High
B	56.94	68.75
M	77.42	78.26
R	47.27	48.78
Y	74.58	71.67
All numbers are in percentages.		

Voters' Re-election Decisions. Thus far, we have mainly studied incumbent behavior. The question remains to what extent voters behave differently across the different treatments. Table 26 draws this comparison, where equalize, pool and separate are defined according to altenatives 1, 2 and 3 in tables 20, 21 and 22. In particular, recall that alternative 1 represents good-type behavior in the sense that it leads to equal first-period payoffs for both players. Alternatives 2 and 3 represent bad-type behavior in the form of pooling or separating respectively.

The following trends are noticeable. First, voters tend to re-elect at higher rates in Treatment Y. Secondly, voters are more likely to re-elect in the pooling case in Treatment M. Finally, voters in Treatment R are least likely to re-elect in the pooling/separating case.

Table 26: Voters' Re-election Decisions

$22 \leq N \leq 30$, $T = 10$			
Treatment	Equalize	Pool	Separate
B	61.89	19.52	11.80
M	67.54	29.40	11.74
R	66.40	9.42	10.00
Y	76.99	21.25	15.48
All numbers are in percentages.			

As a final exercise to conclude this section, we take a look at what we term the "fooling factor." The fooling factor compares voters' actual "beliefs" about incumbents' natures with incumbents' actual natures. Suppose we want to obtain an idea

of voters' beliefs with regard to incumbents' natures; i.e., what percentage of incumbents are perceived to be good or bad by voters. Then, a reasonable measure to capture this is the average re-election rate.

We can in turn compare these rates to incumbents' actual choices. In particular, we can compare voters' re-election decisions to incumbents' second-period decisions conditional on their first-period choice. We can thus come up with a fooling factor, which is a measure that captures the extent to which incumbents are able to fool voters and get re-elected when they are not supposed to.

A comparison of the averages in table 25 and the complement probabilities of those in table 26 (i.e., non-re-election rates) indicates that incumbents are able to fool voters in all treatments. In other words, voters re-elect more frequently than they are supposed to on average. Treatment B has a fooling factor of 18.83 percentage points.[45] Treatment M has a fooling factor of 44.96 percentage points. Treatment R has a fooling factor of 11.67 percentage points. Finally, Treatment Y tops the list with a fooling factor of 51.57 percentage points.

Treatment Effects

This section focuses on the main treatment effects. We discuss the findings for each treatment (i.e., Treatment M, Treatment R and Treatment Y) separately. In each case, we discuss incumbents' and voters' regressions. We specify the model under consideration by (1) expressing the estimation equation and (2) defining the variables of interest. Then, we perform estimation and inference.

Four main comments apply to the analyses below. First, throughout the discussion we focus on random-effects (RE) estimators. While it is understood that such

[45] These numbers are calculated as follows. Consider Treatment B. Conditional on equalization, voters re-elect (accept) at a rate of 61.89%, which in turn implies that they vote out (reject) at a rate of 38.11%. Given the behavior in table 25, however, they should vote out at a rate of 56.94%. Thus, voters are fooled at a rate of 18.83 percentage points in this treatment. The other factors are calculated similarly.

estimators have more stringent assumptions associated with them than fixed-effects estimators, the choice for RE estimation is dictated by the fact that essentially all covariates of interest are individually invariant. Thus, if estimation were to occur via FE, it would wipe out the main partial effects of interest (Wooldrige 2002).

Secondly, a potentially important variable "history" is not completely controlled for in the regressions below. History of play should not be a major explanatory variable in our regressions, since the experiments were carefully designed to avoid repeated-play effects. In particular, subjects were informed that they would never interact with the same person for more than one repetition.

Nonetheless, we might expect subjects to play fictitiously (Brown 1951, Robinson 1951 and others).[46] If this is the case, we should control for history accordingly. This would require instrumental-variable (IV) estimation, since the history of incumbents' and voters' choices are endogenous. Since there are no proper instruments, we partially control for history by including (1) a time trend and (2) the history of the unit cost (θ_{t-1}) as explanatory variables in the regressions. The time trend is intended to capture any upward or downward trends that may be present in the data. The history of the unit cost acts as an imperfect proxy for history of choices in both the incumbent and voter regressions.

The historical value of the unit cost is exogenous, since it is randomly determined by the experimenter by means of a coin flip. Furthermore, it directly impacts incumbents' past first-period choices and thus it indirectly affects the set of signals that were observable by the voter in the previous period. So, in this sense, the history of the unit cost satisfies both the conditions for an IV or proxy variable (Wooldridge 2002).

However, contrary to the history of incumbents' and voters' choices, the his-

[46]The modern game-theoretic usage of the term "fictitious play" can be different from the definitions first discussed by Brown (1951) and Robinson (1951). Here we use the term to indicate a situation in which subjects play as if they are paired with the same player every repetition.

tory of the unit cost is not individually variant. In other words, the history of the unit cost is imperfect either as a proxy variable or as an IV. So, the question becomes whether to use this variable as a weak instrument or as an imperfect proxy. We choose the latter approach, since we expect past choices to have relatively little effect on incumbent and voter behavior due to the re-pairing design discussed previously.

Third, as discussed in detail in Viceisza (2007a) and as summarized in the present work, there are four propositions that inform the experimental designs in this paper. Thus, it is logical to call upon those propositions when formulating hypotheses for the purpose of statistical inference. Since we have designed the experiments according to a specific structure (i.e., a logical game form), we know to a great extent the data generating processes.

In particular, since the data are generated according to a *sequential* game of incomplete information by construction (i.e., by design of the experiments), we can rule out with certainty the possibility that decisions are being made simultaneously by the incumbent and the voter. Thus, we estimate separate equations of interest for each treatment–one for the first mover (the incumbent) and one for the second mover (the voter).

Finally, it should be noted that there are some further limitations to the analyses below. First, we can model learning in more formal ways. For example, we can follow models such as those developed in Roth and Erev (1995) or Camerer and Ho (1999). Furthermore, we can apply estimation techniques such as those discussed in Aguirregabiria and Mira (2007) or Arellano and Bond (1991). However, as discussed in the original papers, these models have their own set of complications including feasibility of estimation and sample size requirements. So, we limit our analysis to the models below.

Treatment M. An individual incumbent's (i) main choice variable in any given repetition (t) is his first-period choice, $Choice_{it}$. There are two possible ways to define this variable. The first is to let $Choice_{it}$ be a dummy variable that is equal to one if the incumbent chooses to equalize first-period payoffs (i.e., behave as if good). The second alternative is to let $Choice_{it}$ have a range of three possible values according to separation, pooling or equalization. We report findings based on the first formulation, since it lends itself to easier interpretations of the coefficients; although the main treatment effects are robust to the alternative formulation.

We estimate the following equation:

$$Choice_{it} = \beta_0 + \beta_1 gender_i + \beta_2 \theta_t + \beta_3 d_M + \beta_4 beliefs_{it} + \beta_5 \theta_{t-1} + \beta_6 t + \beta_7 I + \varepsilon_{it}, \quad (18)$$

where $i = 1, ..., 52$, $t = 1, ..., 10$, β_0 represents a constant term, $gender_i$ represents a dummy variable that is equal to one if the individual is female, θ_t represents the current draw of the unit cost of the public good, d_M represents a dummy variable that is equal to one if the individual is part of the treatment group, θ_{t-1} is the past lag of the unit cost as a proxy for history, t is a time trend, I is a set of interaction terms consisting of $(gender_i * \theta_t)$, $(gender_i * d_M)$, $(beliefs_{it} * d_M)$, $(\theta_{t-1} * d_M)$, $(gender_i * \theta_t * d_M)$ and $(\theta_t * \theta_{t-1})$ and ε_{it} has a one-way error component structure of the form $\varepsilon_{it} = \alpha_i + \nu_{it}$, where α_i represents a vector of unobserved individual heterogeneities and ν_{it} satisfies the following strict exogeneity assumption, $E(\nu_{it}|X_{it}, \alpha_i) = 0$, where X_{it} is the set of explanatory variables in expression 18.

We also include a proxy variable for beliefs in the right-hand side of equation 18. It is expected that when an incumbent makes a first-period choice, he anticipates a particular re-election decision to follow such choice. In other words, an incumbent has some belief towards re-election associated with taxes charged and this belief is likely to influence his first-period choice. So, beliefs can be a statistically

significant explanatory variable in the incumbent regressions.

Since beliefs are an inherent characteristic of the incumbent, they are unobserved. Furthermore, there is no reason to expect beliefs to be correlated with any of the other explanatory variables in equation 18. So, the correct approach in controlling for beliefs is to use a proxy variable that satisfies the following conditions: (1) irrelevance for explaining the dependent variable and (2) correlation with beliefs (Wooldridge 2002).

We claim that $Accept_{it}$ constitutes a valid proxy for beliefs. First, this variable satisfies requirement (1) since it is chosen by the voter and thus, it should not be relevant in explaining the incumbent's choice. Secondly, the variable satisfies requirement (2) since it is highly correlated with beliefs. This is testable. Under the assumptions that the regressors in equation 18 are exogenous and capture all other unexplained variation in $Choice_{it}$, the least-squares residuals from the estimation of equation 18 represent consistent estimates of the error term and thus of beliefs. A regression of these estimates on $Accept_{it}$ finds that this variable is strongly significant (p-value 0.000).

Table 27 reports the estimates of a linear probability model (LPM).[47] The unit cost of the public good affects incumbent behavior; in particular, if the unit cost is high, incumbents are more likely to equalize payoffs in the baseline and marginal cost treatments. Overall, incumbents in Treatment M are less likely to equalize first-period payoffs in comparison to Treatment B. This is not surprising, since the marginal cost treatment makes it more costly to equalize first-period payoffs.

Beliefs are also important. Specifically, if an incumbent associates a higher probability of re-election with a particular first-period payoff, he is more likely to choose it. In other words, incumbents choose taxes with an eye on re-election. Finally, note that female incumbents facing a high cost are more likely to behave corruptly.

[47] The results are robust to alternative estimation methods (i.e., logit and probit).

Table 27: Incumbent Regression (Treatment M)

Dependent Variable: $Choice_{it}$ (1=equalize, 0=not equalize)	
$N = 52$, $T = 10$	
$constant$	0.544 (0.000)*
$gender_i$	-0.031 (0.667)
θ_t	0.170 (0.030)
d_M	-0.382 (0.000)
$beliefs_{it}$	0.387 (0.000)
$history\ (\theta_{t-1})$	-0.052 (0.472)
t	-0.007 (0.353)
$gender_i{*}\theta_t$	-0.151 (0.101)
$gender_i{*}d_M$	0.158 (0.122)
$beliefs_{it}{*}d_M$	0.102 (0.209)
$\theta_{t-1}{*}d_M$	-0.038 (0.670)
$gender_i{*}\theta_t{*}d_M$	-0.047 (0.656)
$\theta_t{*}\theta_{t-1}$	0.027 (0.764)
overall R^2	0.3193
* p-values in parentheses.	

In other words, when it is costly to equalize, female incumbents are more likely to pool or separate and divert rents. This finding is robust across Treatments B and M.

An individual voter's (i) main choice variable in any given repetition (t) is his re-election decision, $Accept_{it}$, which is a dummy variable that is equal to one if the voter re-elects (accepts) the incumbent. We estimate the following equation:

$$Accept_{it} = \beta_0 + \beta_1 gender_i + \beta_2 equal_{it} + \beta_3 pool_{it} + \beta_4 d_M + \beta_5 \theta_{t-1} + \beta_6 t + \beta_7 I + \varepsilon_{it}, \quad (19)$$

where $i = 1, ..., 52$, $t = 1, ..., 10$, β_0 represents a constant term, $gender_i$ represents a dummy variable that is equal to one if the individual is female, $equal_{it}$ represents a dummy variable that is equal to one if the voter observes a payoff that signals equal first-period payoffs, $pool_{it}$ represents a dummy variable that is equal to one if the voter observes a pooling first-period payoff, d_M represents a dummy variable that is

equal to one if the individual is part of the treatment group, θ_{t-1} is the past lag of the unit cost as a proxy for history, t is a time trend, I is a set of interaction terms consisting of $(gender_i * d_M)$, $(gender_i * equal_{it})$, $(gender_i * pool_{it})$, $(\theta_{t-1} * d_M)$ and $(t * d_M)$ and ε_{it} has a one-way error component structure of the form $\varepsilon_{it} = \alpha_i + \nu_{it}$, where α_i represents a vector of unobserved individual heterogeneities and ν_{it} satisfies the following strict exogeneity assumption, $E(\nu_{it}|X_{it}, \alpha_i) = 0$, where X_{it} is the set of explanatory variables in expression 19. Note that in the case of the voter, we are less concerned about controlling for beliefs, since the signal that is observed by the incumbent (i.e., the first-period payoff) is in itself a proxy for beliefs. After all, the incumbent and the voter are playing a signaling game by construction.

Table 28 reports the estimates of a linear probability model (LPM).[48] As expected, the higher the first-period payoff, the more likely the incumbent's probability of re-election. Furthermore, note that voters are more likely to vote out the incumbent as time goes by. Also, while gender is irrelevant in the baseline treatment, it is not in the marginal cost treatment. In particular, female voters are less likely to re-elect the incumbent in the marginal cost treatments.

Also, female incumbents are more likely to re-elect when having observed the pooling equilibrium payoff. Finally, note that while history is irrelevant in the baseline treatment, it seems to matter slightly in the marginal cost treatment. In particular, voters are more likely to re-elect in Treatment M if they re-elected in the previous repetition.

---

[48] These findings are robust to alternative estimation methods (i.e., logit and probit).

Table 28: Voter Regression (Treatment M)

Dependent Variable: $Accept_{it}$ (1=re-elect, 0=vote out)	
$N = 52$, $T = 10$	
constant	0.171 (0.102)*
$gender_i$	0.155 (0.186)
$equal_{it}$	0.615 (0.000)
$pool_{it}$	0.234 (0.006)
d_M	0.228 (0.328)
history (θ_{t-1})	-0.059 (0.245)
t	-0.017 (0.064)
$gender_i * d_M$	-0.522 (0.018)
$gender_i * equal_{it}$	0.037 (0.764)
$gender_i * pool_{it}$	0.233 (0.046)
$\theta_{t-1} * d_M$	0.311 (0.101)
$t * d_M$	0.032 (0.174)
overall R^2	0.3012
* p-values in parentheses.	

Treatment R. We follow a similar estimation procedure as the one discussed for the incumbent in Treatment M. We estimate the following equation:

$$Choice_{it} = \beta_0 + \beta_1 gender_i + \beta_2 \theta_t + \beta_3 d_{R_1} + \beta_4 d_{R_2} + \beta_5 beliefs_{it} + \beta_6 \theta_{t-1} + \beta_7 t + \beta_8 I + \varepsilon_{it}, \tag{20}$$

where $i = 1, ..., 52$, $t = 1, ..., 10$, β_0 represents a constant term, $gender_i$ represents a dummy variable that is equal to one if the individual is female, θ_t represents the current draw of the unit cost of the public good, d_{R_1} represents a dummy variable that is equal to one if the individual is part of the first risk session (R_1), d_{R_2} represents a dummy variable that is equal to one if the individual is part of the second risk session (R_2), $beliefs_{it}$ are captured by the proxy variable $Accept_{it}$, θ_{t-1} is the past lag of the unit cost as a proxy for history, t is a time trend, I is a set of interaction terms consisting of $(gender_i * \theta_t)$, $(gender_i * d_{R_1})$, $(gender_i * d_{R_2})$, $(beliefs_{it} * d_{R_1})$, $(beliefs_{it} * d_{R_2})$, $(\theta_{t-1} * d_{R_1})$, $(\theta_{t-1} * d_{R_2})$, $(gender_i * \theta_t * d_{R_1})$, $(gender_i * \theta_t * d_{R_2})$, $(\theta_t * d_{R_1})$, $(\theta_t * d_{R_2})$ and $(\theta_t * \theta_{t-1})$ and ε_{it} has a one-way error component structure of the form $\varepsilon_{it} = \alpha_i + \nu_{it}$, where α_i represents a vector of

unobserved individual heterogeneities and ν_{it} satisfies the following strict exogeneity assumption, $E(\nu_{it}|X_{it}, \alpha_i) = 0$, where X_{it} is the set of explanatory variables in expression 20.

Table 29 reports the estimates of a linear probability model (LPM).[49] Notice that the unit cost of the public good seems to play a significant role in the unit cost risk treatments. This is not surprising, since these treatments test the effect of a mean-preserving spread on incumbent and voter behavior. In particular, the result suggests that the higher the unit cost, the more likely the incumbent is to equalize first-period payoffs. This confirms the result seen previously.

Overall, the increase in risk seems to leave corrupt behavior unchanged when comparing the risk and baseline sessions. While the signs of the risk dummies point in opposite directions for the two sessions, both coefficients are statistically insignificant. In particular, incumbents are neither more nor less likely to behave in a corrupt manner. Finally, beliefs remain a significant determinant of incumbent behavior in both the baseline and the unit cost risk treatments.

[49] The results are robust to alternative estimation methods (i.e., logit and probit).

Table 29: Incumbent Regression (Treatment R)

Dependent Variable: $Choice_{it}$ (1=equalize, 0=not equalize)	
$N = 52$, $T = 10$	
$constant$	0.460 (0.000)*
$gender_i$	-0.031 (0.682)
θ_t	0.223 (0.030)
d_{R_1}	0.218 (0.146)
d_{R_2}	-0.142 (0.461)
$beliefs_{it}$	0.400 (0.000)
$history\ (\theta_{t-1})$	-0.009 (0.911)
t	-0.000 (0.980)
$gender_i{*}\theta_t$	-0.151 (0.143)
$gender_i{*}d_{R_1}$	-0.163 (0.256)
$gender_i{*}d_{R_2}$	0.208 (0.261)
$beliefs_{it}{*}d_{R_1}$	0.063 (0.549)
$beliefs_{it}{*}d_{R_2}$	0.037 (0.713)
$\theta_{t-1}{*}d_{R_1}$	-0.037 (0.735)
$\theta_{t-1}{*}d_{R_2}$	0.028 (0.868)
$gender_i{*}\theta_t{*}d_{R_1}$	0.056 (0.785)
$gender_i{*}\theta_t{*}d_{R_2}$	-0.156 (0.468)
$\theta_t{*}d_{R_1}$	-0.240 (0.159)
$\theta_t{*}d_{R_2}$	0.028 (0.868)
$\theta_t{*}\theta_{t-1}$	-0.057 (0.554)
overall R^2	0.2508
* p-values in parentheses.	

We follow a similar estimation procedure as the one discussed for the voter in Treatment M. We estimate the following equation:

$$Accept_{it} = \beta_0+\beta_1 gender_i+\beta_2 equal_{it}+\beta_3 pool_{it}+\beta_4 d_{R_1}+\beta_5 d_{R_2}+\beta_5\theta_{t-1}+\beta_6 t+\beta_7 I+\varepsilon_{it}, \tag{21}$$

where $i = 1, ..., 52$, $t = 1, ..., 10$, β_0 represents a constant term, $gender_i$ represents a dummy variable that is equal to one if the individual is female, $equal_{it}$ represents a dummy variable that is equal to one if the voter observes a signal that indicates equal first-period payoffs, $pool_{it}$ represents a dummy variable that is equal to one if the voter observes a pooling first-period payoff, d_{R_1} represents a dummy variable

that is equal to one if the individual is part of the first treatment group (R_1), d_{R_2} represents a dummy variable that is equal to one if the individual is part of the second treatment group (R_2), θ_{t-1} is the past lag of the unit cost as a proxy for history, t is a time trend, I is a set of interaction terms consisting of $(gender_i * d_{R_1})$, $(gender_i * d_{R_2})$, $(gender_i * equal_{it})$, $(gender_i * pool_{it})$, $(\theta_{t-1} * d_{R_1})$, $(\theta_{t-1} * d_{R_2})$, $(t * d_{R_1})$ and $(t * d_{R_2})$ and ε_{it} has a one-way error component structure of the form $\varepsilon_{it} = \alpha_i + \nu_{it}$, where α_i represents a vector of unobserved individual heterogeneities and ν_{it} satisfies the following strict exogeneity assumption, $E(\nu_{it}|X_{it}, \alpha_i) = 0$, where X_{it} is the set of explanatory variables in expression 21.

Table 30 reports the estimates of a linear probability model (LPM).[50] As expected, voters are more likely to re-elect as first-period payoffs increase. Also, seem less likely to re-elect as time goes by in the baseline treatment and in the first risk session. Finally, while gender seems to be insignificant overall, it plays a role in two specific contexts: (1) female voters are less likely to re-elect in the second risk session and (2) female voters are more likely to re-elect upon observing the pooling equilibrium payoff.

Treatment Y. We follow a similar estimation procedure as discussed previously. We estimate the following equation:

$$Choice_{it} = \beta_0 + \beta_1 gender_i + \beta_2 \theta_t + \beta_3 d_Y + \beta_4 beliefs_{it} + \beta_5 \theta_{t-1} + \beta_6 t + \beta_7 I + \varepsilon_{it}, \quad (22)$$

where $i = 1, ..., 58$, $t = 1, ..., 10$, β_0 represents a constant term, $gender_i$ represents a dummy variable that is equal to one if the individual is female, θ_t represents the current draw of the unit cost of the public good, d_Y represents a dummy variable that is equal to one if the individual is part of the treatment group, $beliefs_{it}$

[50] These findings are robust to alternative estimation methods (i.e., logit and probit).

Table 30: Voter Regression (Treatment R)

Dependent Variable: $Accept_{it}$ (1=re-elect, 0=vote out)	
$N = 52$, $T = 10$	
$constant$	0.166 (0.130)*
$gender_i$	0.156 (0.223)
$equal_{it}$	0.615 (0.000)
$pool_{it}$	0.241 (0.004)
d_{R_1}	0.248 (0.451)
d_{R_2}	0.351 (0.220)
$history\ (\theta_{t-1})$	-0.057 (0.247)
t	-0.017 (0.058)
$gender_i{*}d_{R_1}$	0.054 (0.838)
$gender_i{*}d_{R_2}$	-0.486 (0.072)
$gender_i{*}equal_{it}$	0.040 (0.735)
$gender_i{*}pool_{it}$	0.228 (0.045)
$\theta_{t-1}{*}d_{R_1}$	-0.087 (0.683)
$\theta_{t-1}{*}d_{R_2}$	-0.156 (0.205)
$t * d_{R_1}$	-0.069 (0.127)
$t * d_{R_2}$	0.007 (0.727)
overall R^2	0.2684
* p-values in parentheses.	

are captured by the proxy variable $Accept_{it}$, θ_{t-1} is the past lag of the unit cost as a proxy for history, t is a time trend, I is a set of interaction terms consisting of $(gender_i * \theta_t)$, $(gender_i * d_Y)$, $(beliefs_{it} * d_Y)$, $(\theta_{t-1} * d_Y)$, $(gender_i * \theta_t * d_Y)$ and $(\theta_t * \theta_{t-1})$ and ε_{it} has a one-way error component structure of the form $\varepsilon_{it} = \alpha_i + \nu_{it}$, where α_i represents a vector of unobserved individual heterogeneities and ν_{it} satisfies the following strict exogeneity assumption, $E\left(\nu_{it} | X_{it}, \alpha_i\right) = 0$, where X_{it} is the set of explanatory variables in expression 22.

Table 31 reports the estimates of a linear probability model (LPM).[51] First, note that in general incumbents behave in a more corrupt manner in the presence of yardstick competition. Secondly, beliefs are an important determinant of incumbents' choices both in the baseline and in the yardstick treatments. Third, while gender is insignificant overall, it matters in specific contexts. In particular, female

[51] The results are robust to alternative estimation methods (i.e., logit and probit).

incumbents are less likely to equalize when the unit cost is high. Finally, female incumbents are more likely to equalize first-period payoffs in the yardstick treatments. So, while the average incumbent increases corrupt behavior in the presence of yardstick competition, the average female incumbent reduces corrupt behavior.

Table 31: Incumbent Regression (Treatment Y)

Dependent Variable: $Choice_{it}$ (1=equalize, 0=not equalize)	
$N = 58$, $T = 10$	
$constant$	0.485 (0.039)*
$gender_i$	0.112 (0.421)
θ_t	0.119 (0.343)
d_Y	-0.313 (0.023)
$beliefs_{it}$	0.372 (0.000)
$history$ (θ_{t-1})	-0.097 (0.441)
t	-0.004 (0.523)
$gender_i$*θ_t	-0.146 (0.077)
$gender_i$*d_Y	0.330 (0.052)
$beliefs_{it}$*d_Y	0.145 (0.049)
θ_{t-1}*d_Y	-0.008 (0.915)
$gender_i$*θ_t*d_Y	-0.056 (0.561)
θ_t*θ_{t-1}	0.037 (0.618)
overall R^2	0.2883
* p-values in parentheses.	

We follow a similar estimation procedure as discussed previously with some minor modifications. In particular, we estimate the following equation:

$$Accept_{it} = \beta_0 + \beta_1 gender_i + \beta_2 pool_{it} + \beta_3 yardstick_{it} + \beta_4 \theta_{t-1} + \beta_5 I + \varepsilon_{it}, \quad (23)$$

where $i = 1, ..., 28$, $t = 1, ..., 10$, β_0 represents a constant term, $gender_i$ represents a dummy variable that is equal to one if the individual is female, $pool_{it}$ represents a dummy variable that is equal to one if the voter observes a pooling first-period payoff, θ_{t-1} is the past lag of the unit cost as a proxy for history, I is a set of interaction terms consisting of $(gender_i * yardstick_{it})$ and $(gender_i * pool_{it})$ and ε_{it} has

a one-way error component structure of the form $\varepsilon_{it} = \alpha_i + \nu_{it}$, where α_i represents a vector of unobserved individual heterogeneities and ν_{it} satisfies the following strict exogeneity assumption, $E(\nu_{it}|X_{it}, \alpha_i) = 0$, where X_{it} is the set of explanatory variables in expression 23.

There are some subtle differences between the voter equations discussed previously and equation 23. All of these can be traced back to the yardstick nature of the Y-sessions. First, the yardstick sessions are identical to the baseline sessions except for one minor modification: In addition to their first-period payoffs, voters see one randomly drawn first-period payoff from another B-session. So, the immediate question arises how to model this subtle change if we want to compare the baseline and yardstick data.

One alternative is to include a standard dummy variable that is equal to one if the subject is part of the treatment group and zero otherwise. It turns out that if we use this formulation, we are led to the conclusion that there is no evidence of yardstick competition. Another alternative is to define a yardstick variable that takes into account the nature of the yardstick information. However, under those circumstances we are unable to collapse the data from the baseline and yardstick treatments, since any yardstick variable that incorporates the explicit nature of the yardstick signal is undefined in the baseline treatment.

In other words, if we want to appeal to the yardstick nature of the Y-sessions by defining a yardstick variable, the above regression will be based only on data from the Y-sessions (i.e., $N = 28$). So, the question we ask is whether the yardstick variable captures any treatment effects given the assumption that Treatments B and Y are otherwise identical. This is of course not an unreasonable assumption, since the baseline and yardstick treatments were designed to be identical except for the yardstick component. So, if we control for a similar set of covariates in equation 23 as we would if we were estimating a baseline regression, we would expect the

yardstick variable to capture any relevant treatment effects.

The second question is how to define the yardstick variable. We tried several alternative definitions and it turns out that the findings are not robust to all definitions. In particular, treatment effects arise if we define the yardstick variable as deviations from the yardstick signal. In other words, $yardstick_{it}$ in equation 23 is defined as the difference between the randomly drawn signal (i.e., 1.50, 0.75 or 0.16) and the actual first-period payoff faced by the voter.

We thus estimate equation 23 for three separate cases. First, we estimate the equation for the case when the signal is equal to 1.50. Then, we estimate the equation for the case when the signal is 0.75. Finally, we estimate the equation when the signal is 0.16. Each such estimation is based on a subset of repetitions, since in any given repetition the yardstick signal can only take on one particular value.

Table 32 reports the estimates of a linear probability model (LPM) for the three cases.[52] The main result arising from this table is that there is strong evidence of yardstick competition. In particular, voters pay attention to the difference between the tax rate in another jurisdiction and their own. Furthermore, any deviation from the signal that indicates that they are worse off in comparison to the other jurisdiction is punished by voting out (rejecting) the incumbent.

To interpret the yardstick competition coefficient, recall that the yardstick variable is defined as the difference between the yardstick signal and the voter's first-period payoff. So, suppose the signal equals 1.50. Then, the yardstick variable has a nonnegative range. In particular, the variable can take on the values 0.00 (1.50 minus 1.50), 0.75 (1.50 minus 0.75) or 1.34 (1.50 minus 0.16). So, an increase in the yardstick variable from zero to positive automatically tells us that the first-period payoff is below the yardstick signal. Thus a negative sign on the $yardstick_{it}$ coefficient tells us that the voter is more likely to vote out the incumbent when his first-

[52] These findings are robust to alternative estimation methods (i.e., logit and probit).

period welfare is below that of another jurisdiction. A similar reasoning holds when the signals are 0.75 and 0.16 respectively.

Finally, it should be noted that while gender is overall insignificant, female voters remain sensitive to treatment effects. In particular, while female incumbents are more likely to re-elect the pooling equilibrium payoff, they are less likely to re-elect in the presence of yardstick competition when the signal is 0.75.

Table 32: Voter Regression (Treatment Y)

Dependent Variable: $Accept_{it}$ (1=re-elect, 0=vote out)			
	1.50	0.75	0.16
$N = 28$	$T = 2$	$T = 4$	$T = 4$
$constant$	0.870 (0.001)*	0.069 (0.650)	-0.118 (0.478)
$gender_i$	-0.048 (0.769)	0.021 (0.859)	-0.064 (0.659)
$pool_{it}$	0.084 (0.699)	0.143 (0.332)	0.128 (0.234)
$yardstick_{it}$	-0.522 (0.000)	-0.235 (0.034)	-0.476 (0.000)
$history\ (\theta_{t-1})$	-0.033 (0.801)	0.283 (0.001)	0.180 (0.019)
$gender_i * yardstick_{it}$	-0.051 (0.774)	-0.309 (0.037)	-0.076 (0.669)
$gender_i * pool_{it}$	0.165 (0.546)	0.203 (0.289)	0.057 (0.706)
overall R^2	0.4591	0.3827	0.3219
* p-values in parentheses.			

Comparison with Viceisza (2007a)

This section presents a main table (table 33) that draws comparisons between the results in this paper and those discussed in Viceisza (2007a). The table draws comparisons in two dimensions: (1) experimental implementation and (2) results. The major difference is the fact that the present experimental design gives rise to yardstick competition both on the incumbent's and the voter's side. It should be noted, however, that there are some subtle similarities such as the fact that voter welfare tends to be non-increasing with yardstick competition. This is likely due to the fact that there is a relatively low proportion of good incumbents in the population as discussed by Besley and Smart (2007).

Table 33: Comparison of Current Findings with Viceisza (2007a)

	Current	Viceisza (2007a)
Implementation		
Within	–	Yes
Between/Across	Yes	–
Trial	Yes	Yes
Post-Trial Quiz	Yes	–
Task Summary	Yes	–
Post-Questionnaire	Yes	Yes
Unit Cost Operationalization	Physical Devices	Virtual
Results		
Low proportion of good incumbents	Yes	Yes
Incumbents theoretically inconsistent	Yes	Yes
Yardstick: Incumbent	Yes	No
Yardstick: Voter	Yes	Limited
Yardstick leaves welfare unaffected	Yes	Yes
Voters less likely to re-elect as $t \uparrow$	Yes	Yes
Gender: Females sensitive to treatment	Yes	No

Economic Significance

Thus far, the discussion has mainly focused on statistical significance. A separate question is whether any statistically (in)significant effects are economically significant. In fact, it turns out that all treatments are economically significant with the exception of Treatment Y. Namely, both in Treatment B and Y voters are worse off in comparison to incumbents. However, voters are no worse off in Treatment Y than they are in Treatment B. In both cases, voters suffer on average \$2.00 compared to the case in which payoffs were equalized throughout the whole experiment.

For the other treatments, it is a different story. In Treatment M, voters are economically worse off than they are in Treatment B. They are worse off by an average of \$2.00. In absolute terms, incumbents are worse off as well by an average of \$0.25; however, in relative terms they are better off. These results support the results found in previous sections.

Finally, in absolute terms both incumbents and voters are better off in Treatment R; however, voters are better off than incumbents in relative terms. Namely, voters are better off by an average of \$1.50 and incumbents are better off by an average of \$0.50 in comparison to Treatment B. This is the case even though we did not find any clear-cut statistically significant treatment effects.

Conclusion

This paper reports theory-testing experiments that answer three main questions. First, the extent to which tax inefficiency–as represented by an increase in the marginal cost of public funds–affects corrupt behavior. Secondly, the effect of lack of a particular form of transparency (as modeled by a mean-preserving spread in the distribution of the unit cost of public goods) on corrupt behavior. Finally, whether yardstick competition acts as a corruption-taming mechanism; particularly, in the presence of historical forces.

We find the following results. First, tax inefficiency is an important determinant of corrupt incumbent behavior. In particular, an increase in the marginal cost of public funds makes it more costly for incumbents to equalize first-period payoffs. This drives them to separate and divert maximum rents in the first period. While voters retaliate slightly by voting incumbents out of office, voters tend to be worse off.

Secondly, we find that increased lack of a particular form of transparency (as defined in terms of an increase in risk in the distribution of the unit cost) leaves corrupt incumbent behavior unchanged. If the draw of the unit cost is unfavorable, incumbents tend to be less corrupt. So, the results suggest that lack of transparency (as defined in Viceisza 2007a) need not always make voters worse off.

Third, the experiments find strong evidence of yardstick competition. On the incumbent's side, yardstick competition acts as a corruption-taming mechanism if the

incumbent is female. On the voter's side, voters are less likely to re-elect the incumbent in the presence of yardstick competition. Specifically, voters pay attention to the difference between the tax signal in their own jurisdiction and the signal in another jurisdiction. As this difference increases, voters re-elect less. Despite these changes in behavior, voter welfare tends to be non-increasing in the presence of yardstick competition. This is likely due to the low proportion of good incumbent present in the population. This is further discussed by Besley and Smart (2007).

Fourth, history can be an important determinant of corruption and of re-election decisions. Incumbents are likely to make choices as they did in previous repetitions and voters are likely to vote out increasingly as the repetitions go by. In other words, voters tend to distrust the political system more as time goes by depending on the institution.

Fifth, we find that gender plays a significant role in incumbent and voter behavior. In particular, female incumbents tend to be more sensitive to treatment effects both on the incumbent and the voter side. Finally, beliefs tend to be an important factor in decision-making. Incumbents are likely to choose first-period payoffs with which they associate a higher probability of re-election. Voters tend to re-elect the higher the first-period payoff, which is a signal of the incumbent's nature.

CORRUPTION IN THE MAIL: EVIDENCE FROM A RANDOMIZED FIELD EXPERIMENT IN LIMA, PERU

Introduction

Corruption is an age-old question (see discussions by Aidt 2003 and Bardhan 1997.) While folk wisdom has long recognized its existence, some related questions have proven difficult to answer scientifically. First, there are the questions of definition and existence. Namely, what actually constitutes corruption? Secondly, what are its causes? Third, how do we measure it? Fourth, does corruption have any (negative) consequences and if so, what are they? Finally, if corruption is costly to society, what can we do to control it?

The first question was addressed long ago by for example Kautiliya (see Bardhan's 1997 account of Kautiliya's writings.) However, recently The World Bank (1997) formally answered this question by defining corruption as the abuse of public office for private gain.[53] Although this is a start, the quest to identify corruption in everyday settings in order to answer the remaining questions has sparked a large literature on economic approaches to (political) corruption–even before a formal definition was available.[54]

While finding answers to the remaining questions has proven to be more involved, the literature seems to be in widespread consensus on a couple of issues.

[53] A similar definition was provided by Jain (2001).

[54] Some general pieces include Martinez-Vazquez, Arze and Boex (2007), Bardhan and Mookherjee (2005), Aidt (2003) and Bardhan (1997). Furthermore, the following authors – just to name a few – have focused on answering the five corruption-related questions. Tanzi (1998) looked at causes. Kaufmann, Kraay and Mastruzzi (2006) and Olken (2006) worked on measurement. Considerable time has been spent on consequences. The debate has focused on whether corruption is efficiency-enhancing in that it "greases the wheels of society" (Leff 1964, Huntington 1968, Lui 1985) or whether it is "costly" in that it is efficiency-reducing (Rose-Ackerman 1978, Klitgaard 1991, Shleifer and Vishny 1993, Djankov et al. 2002.) Furthermore, when it is costly, the question has been raised whether it is more costly to certain social groups (Gupta, Davoodi and Alonso-Teme 1998 and Transparency International.) Finally, Klitgaard (1988) has worked on mechanisms for controlling corruption. Also see discussions by organizations such as the IMF and the World Bank.

First, improved answers to the above questions lead to a better understanding of corruption, which in turn leads to a better understanding of how to counteract corruption when necessary. Secondly, weak (public) institutions and social history are at the heart of corruption (Viceisza 2007c) and lack of economic growth (Kaufmann, Kraay and Mastruzzi forthcoming.)

This study reports a theory-testing randomized field experiment that makes use of an existing semi-public monopolist institution–the postal system in Lima, Perú–to identify and measure corruption, its potential causes and its consequences. The study thus contributes to the literature by providing contextual answers to four of the above five questions. Furthermore, the experiment provides observations on a counterfactual that is usually not observed. In particular, it is unusual to send items of value in the mail. Therefore, day-to-day behavior dictates searching less for these items in the mail. So, any treatment effects that are identified enable us to observe behavior off the "equilibrium" path and constitute a lower bound to what might actually be observed if it were customary to send items of value by mail.

The study is in the same area of the literature as Olken (2007) and Bertrand et al. (forthcoming); namely, randomized field experiments that aim at identifying and measuring corruption, its causes and consequences. Thus, it contributes to a fast-emerging literature on field experiments that are of interest to policymakers in general with a particular focus on development (Duflo 2006) and experimental economists (Harrison and List 2004.)

Specifically, we implement a (2x2)-design in which each household within a stratified quasi-random sample of households across Lima is sent four birthday-like envelopes: (1) an envelope with money with same sender and recipient last name, (2) an envelope without money with same sender and recipient last name, (3) an envelope with money with distinct sender and recipient last name and (4) an en-

velope without money with distinct sender and recipient last name. The envelopes are sent in four batches (on average, one month apart) and for each household the four envelopes are randomly assigned to batch. The envelopes are also randomly assigned other systematic characteristics, which are elaborated upon below.

We compare receipt of money and non-money envelopes to learn whether loss of mail is nonrandom.[55] Furthermore, we compare receipt across same and different sender/recipient last names to learn whether or not family mail is more likely to be lost because its content is perceived to be of higher value. Also, we control for household-specific characteristics (such as neighborhood welfare and head of household's gender) to learn whether loss of mail affects social groups differently. Finally, we quantify the costs of corruption and loss of efficiency based on the value of lost mail and the costs of an alternative carrier–Western Union.

We find the following. First, 19.08% of mail is lost. Secondly, money mail is more likely to be lost at a rate of 20.90% compared to no-money mail which disappears at a rate of 14.37% (this difference is statistically significant at the 10% level). This finding suggests that loss of mail is systematic (non-random), which implies that this type of corruption is due to strategic behavior as opposed to plain shirking on the part of mail handlers.

Third, we find that loss of mail is non-random across other observables. Corruption is more costly to certain societal groups. Middle-income neighborhoods are more likely to experience lost (money) mail (these differences are statistically significant at the 10% level). These findings are informed by the predictions of the model, which suggest that mail handlers may perceive the benefits (costs) of lost mail to be higher (lower) in middle-income neighborhoods compared to low- and high-income neighborhoods.

We also find that female heads of household in low-income neighborhoods are

[55] Note that since the mail system in Perú is a semi-public institution, (nonrandom) loss of mail is a form of corruption according to the World Bank (1997) and Jain (2001).

more likely to experience lost mail. This is not necessarily a surprising result, since this group may be perceived to be vulnerable (this difference is statistically significant at the 10% level). Furthermore, female heads of household in high-income neighborhoods are much less likely to experience lost (money) mail (these differences are statistically significant at the 5% level).

Finally, this form of corruption is costly to different stakeholders. The sender of mail bears a direct and an indirect cost. Corruption is also costly to the intended mail recipient. Middle-income neighborhoods and female heads of household in low-income neighborhoods are more likely to suffer. Finally, corruption is costly the mail company (SERPOST) in terms of lost revenue and to society in terms of loss of trust as a form of social capital (Fukuyama 1995). Overall, the findings suggest that public-private partnerships need not increase efficiency by reducing corruption; particularly, when the institution remains a monopoly. Increased efficiency in mail delivery is likely to require (1) privatization and (2) competition.

The remainder of the paper is organized as follows. The following section presents the theoretical model. Then, we discuss the experimental design and implementation. Finally, we present the results and conclude.

Theoretical Model

There are many theoretical models of corruption as categorized by Aidt (2003.) Our model can be classified under Aidt's second category–namely, models of corruption in which a benevolent principal delegates decision-making "authority" to a non-benevolent agent. In particular, our model is based on a game played between a sender of mail (to be called "the sender") and an average handler of mail (to be called "a handler.")[56] In the next, we first describe the game. We then proceed to solving the game and discussing the hypotheses that guide our experimental design.

[56] Note that the term "sender" is used here to indicate a person sending mail. It is not used in conventional terms of a signaling game.

The Mail Game: Players, Actions and Payoffs

We model the game as a variant of the matching pennies game that has been discussed in many other contexts (Fudenberg and Tirole 1992.) In particular, we model the problems facing the sender and a handler as a simultaneous-move game with complete information, which we call the "the mail game." The strategic form is in table 34.

Table 34: Strategic-Form Representation of The Mail Game

		Sender	
		No Money	Money
Handler	Deliver	$(0, 0)$	$(0, B(x))$
	Open	$(-t(x), 0)$	$(R - t(x), -C(x))$

Consider the problems faced by each player. On the one hand, the sender sends mail and faces the question whether or not to include items of value in the mail. Without loss of generality, let the value item be "money" and let the non-value item be "no money." Then, the actions available to the sender are "money" and "no money." On the other hand, a handler faces the question whether or not to open the mail. Thus, the actions available to a handler are "open" and "deliver."

With each such action there is an associated payoff conditional on what the other player does. Consider the sender. The benefit of successful delivery of money mail is $B(x) > 0$. The cost of lost money mail is $C(x) > 0$. If it is no-money mail, the benefit/cost to the sender is zero. Next, consider a handler. A handler faces a cost of being caught when mail is opened, $t(x) > 0$. Furthermore, a handler faces a potential reward $R > t(x)$ when money mail is opened. If the mail is delivered, the benefit/cost to a handler is zero.

The benefit and the cost to the sender (i.e., $B(x)$, $C(x)$) and the cost to a handler (i.e., $t(x)$) are assumed to be functions of x. The reward R is not. In the theo-

retical game, x strictly represents a set of observable sender characteristics.[57] This formulation is interesting because it allows–among other things–for taste-based discrimination on the part of a handler. For example, money may be of different value to a particular sender. This may affect a handler's perceived value of $B(x)$ and $C(x)$ and thus the likelihood of mail being opened. Furthermore, the likelihood of getting caught may be higher depending on the observed x. Thus, x sends signals to a handler and affects behavior via the perceived values of $B(x)$, $C(x)$ and $t(x)$.

Solution

We seek to characterize pure and/or mixed strategy Nash equilibria for the mail game. Since the game is a variant of the matching pennies game, it should come as no surprise that there are no pure-strategy Nash equilibria associated with this game. Namely, suppose the sender chooses "no money." Then, a handler will choose "deliver." However, knowing that a handler chooses "deliver" induces the sender to choose "money." This in turn will drive a handler to "open." So, given our nontrivial parameterization, there are no stable pure-strategy Nash equilibria associated with this game and both players randomize in equilibrium.

In order to find the mixed-strategy Nash equilibrium associated with this game, suppose we assume risk neutrality. Then, each player solves an expected payoff maximization problem. In particular, the sender chooses "money" if the expected payoff from sending money is greater than the expected payoff from not sending money. Let $(1-p)$ be the likelihood that the mail is delivered. Then, the sender is indifferent between "money" and "no money" if expression 24 is satisfied:

$$(1-p) * B(x) + p * -C(x) = p * 0 + (1-p) * 0. \tag{24}$$

[57] Empirically, x may constitute a larger set of observables such as recipient characteristics, mail characteristics and a handler's characteristics. This is further discussed below. In the context of the theoretical game, observables other than those pertaining to the sender are undefined.

Noting that p is the likelihood that the mail is opened, i.e., $\Pr(open)$, this leads to a handler's equilibrium strategy in expression 25:

$$\Pr(open) = \frac{B(x)}{B(x) + C(x)}. \tag{25}$$

Similarly, a handler chooses "open" if the expected payoff from opening the mail is greater than or equal to the expected payoff of delivering the mail. Let $(1 - q)$ be the likelihood that the mail contains money. Then, a handler is indifferent between "open" and "deliver" if expression 26 is satisfied:

$$q * -t(x) + (1 - q) * [R - t(x)] = q * 0 + (1 - q) * 0. \tag{26}$$

Noting that $(1 - q)$ is the likelihood that the mail contains money, i.e., $\Pr(money)$, this leads to the sender's equilibrium strategy in expression 27:

$$\Pr(money) = \frac{t(x)}{R}. \tag{27}$$

The above shows that both players adopt mixed strategies and thus, randomize in equilibrium. It is useful to interpret these mixed strategies. After all, the game theory literature has debated the interpretation and plausibility of mixed-strategy equilibria for a while (Fudenberg and Tirole 1992). Traditionally, two defenses of mixed strategies have been offered. Both have intuitive appeal in the context of our model. First, a mixed strategy represents different pure strategies used by a large population of players. This has intuititive appeal in the context of our model since the sender and a handler in the mail game represent a large population of senders and handlers–some of which choose "money" and some of which choose "open."

Secondly, a mixed strategy can be interpreted as the result of small, unobservable variations in players' payoffs. Formally, Harsanyi's (1973) purification theorem

shows that any mixed-strategy equilibrium can "almost always" be obtained as the limit of a pure-strategy equilibrium in a sequence of slightly perturbed games. In particular, mixed-strategy equilibria can be thought of as representing asymmetric information. Thus, the perturbed games represent Bayesian games with distributions of types in Harsanyi's formulation.

The representation of the strategic-form mail game as a Bayesian game is relevant. First, it allows for infra-marginality which is ruled out by a simple mixed-strategy equilibrium. This might be important due to the fact that best-response functions in a finite game are step functions, which implies that the infra-marginal player (if we think of the game as a "market") will have the same preferences as the marginal player. This need not be the case in applications. Furthermore, Harsanyi's formulation allows for the simple strategic-form mail game to be interpreted in a richer context–viz. Bayesian games–which also could have been used to describe the game.

Comparative Results and Hypotheses

From the sender's equilibrium strategy in expression 27, it can be noted that the higher a handler's cost of getting caught $t(x)$, the more likely the sender will send money mail. Furthermore, the higher the reward R (i.e., benefit to a handler), the less likely the sender will send money mail. Similarly, from a handler's equilibrium strategy in expression 25, we note that the higher (lower) the perceived benefit $B(x)$ (cost, $C(x)$) associated with the mail, the higher the likelihood that the mail will be opened.[58]

The following hypotheses are informed by the model and guide the experimental design in the next section.

[58] Consider Becker's (1968) model of crime and punishment as a starting point for $B(x)$ and $C(x)$. These represent stylized benefits and costs imposed on society by the commitment of a criminal act. Also, consider Becker's "supply of offenses" discussion in relation to R and $t(x)$.

- H1: Loss of mail is nonrandom and thus constitutes a corrupt activity. In other words, the likelihood of money mail being lost is higher than no-money mail. Thus, $\Pr\left(open|money\right) > \Pr\left(open|no\_money\right)$.

- H2: Certain observable characteristics serve as signals to a handler. Certain signals are associated with a higher (lower) perceived benefit (cost) of mail delivery and thus they induce a higher (lower) likelihood of certain mail being opened. Thus, $\Pr(open|x) > \Pr(open|x')$, where x and x' represent a distinct set of observable characteristics. In particular, the above implies that either $B(x)$ is larger than $B(x')$ or that $C(x)$ is smaller than $C(x')$. In other words, either a handler is more likely to associate more money with x or a handler is more likely to associate less value of money with x.

Finally, we return to the contents of x. Empirically, x can represent a vector of observables consisting of more than just sender characteristics. For example, x can represent the bait itself–i.e., whether or not the mail contains money. In this sense, H1 is a special case of H2. Furthermore, x can consist of combined sender/recipient characteristics–e.g., whether or not the sender's last name is the same as the recipient's last name.

Last, x can comprise several other characteristics such as the recipient's neighborhood (which is a representation of income or social status), the recipient's gender and possibly, a handler's own personal characteristics in relation to the aforementioned observables. For now, it is important to note that regardless of the exact information contained in x, all of these aspects may induce the relation between the above conditional probabilities to move in a particular direction. We appeal further to the implications of the model once we get to interpretation of the results.

Design and Implementation

In this section, we discuss the design and implementation. We are strongly driven by the hypotheses discussed previously in the context of the mail game (i.e., H1 and H2.) We first discuss the main control treatments that test these hypotheses. Then, we discuss some additional issues that are controlled by the experimenter. Finally, we discuss issues related to implementation of the experiments, including sample selection and data collection.

A recurring theme throughout the discussion will be randomization. We stress randomization for a couple of reasons. First, while we hypothesize that loss of mail is nonrandom, we do not know at what point in the mail handling process mail is actually "lost." In other words, mail may be lost at the central level (i.e., the main distribution center); however, it may also be lost at the district level (i.e., the district office) or at the street level (i.e., at the level of the ultimate deliverer of mail.)

So, since we do not know the identity of our treatment and control groups at any given point in time, we randomize as much as possible. This mitigates any type of sample selection bias (Banerjee et al. 2006 and Wooldridge 2002.) Secondly, randomization across observable characteristics that are controlled by the experimenter mitigates any type of bias that may arise on a handler's side. Namely, if certain types of mail are systematically likely to be characterized by certain observables, this may "tip off" a particular handler, which–in turn–may exacerbate learning and social (neighboring) effects (Duflo forthcoming.)

Design I: Main Control Treatments

To test the main hypotheses discussed in the previous section, we implement a (2x2)-design consisting of two control treatments. The first is a "bait" treatment that randomly assigns money and no money to pieces of mail. The second is a "familiar" treatment that randomly assigns same and different sender and recipient

last names to pieces of mail. We refer to the case in which the sender and the recipient have the same last name as "familiar" and the case in which they do not as "foreign."

The experimental design is within-households in the sense that each recipient household is sent four pieces of mail: (1) a familiar money envelope, (2) a familiar no-money envelope, (3) a foreign money envelope and (4) a foreign no-money envelope. This is opposed to a between-households design in which households would strictly be assigned to one of the following treatments: "money," "no money," "familiar" and "foreign."[59]

While a potential benefit of a between-households design is that it mitigates any type of order effects ("red flags"), we choose a within-households design because of the following reasons. First, a between-households design requires a substantially larger sample of recipient households; in fact, four times as much as what we currently have if households are to be randomly assigned to only one of the four treatments and if we seek balanced comparisons. Furthermore, the larger the group of recipient households, the more we need to be concerned about "tipping off" mail handlers. As discussed previously, this is a legitimate concern when mitigating bias in the data. So, we choose a within-households design.

To further mitigate any data bias resulting from order and/or subject-pool effects as well as any sample selectivity bias, we implement the following controls. First, we send four sets of envelopes to each household. Any set of envelopes to all households is sent as one batch and the four batches are sent within a range of three to five weeks apart.[60] The envelopes are identical in terms of size and appearance, and differ only by the characteristics of interest, which are (1) the treatment

[59] Note that the terms within- and between-subjects are traditionally used in the experimental literature to indicate designs in which subjects' identities are known. It should be noted that in this case, subjects' identities are completely unknown; in particular, the subjects of interest are handlers of mail whose identities are unobserved. What we observe is a reduced-form of their actions–i.e., whether or not mail is delivered.

[60] We deliberately tweak the spaces between mailings to further mitigate any type of bias.

characteristics (i.e., content and sender last name) and (2) additional observable characteristics, which include the sender's address, the color of the envelope and the sender's handwriting. We discuss the treatment characteristics below. The additional observable characteristics are discussed in the next subsection.

The main treatment characteristics are implemented as follows. First, all pieces of mail comprise a birthday-like envelope containing a card with enscription "Feliz Día" or "Happy Birthday" depending on whether the sender is familiar or foreign. Secondly, if the mail is money mail, it contains two U.S. dollar bills ($2.00) folded in half. This serves as bait. If the mail is no-money mail, it obviously does not contain money. However, it does contain a small lottery number in the lower right-hand corner. The relevance of the lottery number is explained later.

Secondly, since the mail is sent in four batches, we need four names (two familiar and two foreign) if we are to randomize (i.e., sample without replacement from) sender last names across batches. The familiar names are "L. Last Name" and "M. Last Name." The foreign names are "P. Thomson" and "J. Scott."[61] So, for example, if for a particular batch an envelope is randomly chosen to go to a household with last name Perez coming from familiar1, then the sender is L. Perez. If familiar2, then M. Perez. The enscription in the envelope is adjusted according to the sender last name.

For each household, we randomize the main treatment observables across batches by sampling without replacement and keeping in mind that each household must be sent a money and a no-money envelope from both a familiar and a foreign sender. So, ultimately, each household is sent two money and two no-money envelopes at random points in time and each envelope randomly comes from one of the four senders.

[61] The experiment was conducted in two stages–a pilot stage and a large stage. In particular, the pilot stage only comprised money mail. In the pilot stage, we did not randomize across all observables. Furthermore, the foreign name was Mike Tucker. We control for these discrepancies in the empirical analysis.

According to the theoretical model, if loss of mail is random, then there should be no difference between the probability that money is opened and the probability that no-money mail is opened; i.e., $\Pr(open|no\_money) = \Pr(open|money)$. So, by comparing the likelihood of delivery across money and no-money envelopes ceteris paribus, we are able to test H1. Furthermore, if variants of a particular observable characteristic x have no effect on the likelihood of mail delivery, then $\Pr(open|x) = \Pr(open|x')$. In particular, if x represents the sender's last name and a handler infers nothing from the fact that mail has the same sender and recipient last name (familiar) as opposed to a distinct last name (foreign), there should be no discrepancy between the above conditional probabilities controlling for other factors. So, by comparing the likelihood of delivery across familiar and foreign ceteris paribus, we are able to test a special case of H2.

Two final comments are necessary. First, note that as the experimenter we control two main aspects given what we want to learn from conducting the experiment. In particular, the theoretical model tells us that two aspects determine corruptibility–bait and an observable characteristic such as matching of sender/recipient last name. Thus, we control those and vary them systematically to tease apart any treatment effects. Under usual circumstances a sender of mail does not observe the counterfactuals, which can be a confounding effect.

Secondly, note that the theoretical model is a one-shot simultaneous-move game of complete information. So, strictly speaking, it does not allow for repeated interactions between the sender and a handler. However, with an average of one month between mail batches, we find that the model represents a reasonable approximation. Furthermore, since different senders randomly send different pieces of mail at different points in time, it is not unreasonable to expect that distinct handlers handle distinct pieces of mail. This is consistent with the mail game and with the interpretation of mixed-strategy equilibrium discussed earlier.

Design II: Other Control Issues

The theoretical model defines x as a particular observable characteristic of the sender. However, empirically, x may encompass a much larger set of observables. In particular, the following are observable to a handler over and beyond mail content and the sender's last name: (1) the sender's address, (2) the color of the envelope, (3) the sender's handwriting, (4) the recipient's neighborhood (as a sign of income and social status), (5) the recipient's gender and (6) a handler's personal characteristics.

Consider first the last observable–a handler's personal characteristics. This is less of an observable–certainly a priori. As a consequence, it is less controllable by the experimenter. In particular, we do not have a clear picture of a handler's personal characteristics, mainly because his or her identity is completely unknown. Two comments are relevant. First, once the study is completed, we can seek more consistent information on an average handler's personal characteristics by district. For now, we refrain from doing so in order not to "poison the well." Secondly, since we are ultimately interested in the differential effect between treatments, the personal characteristics of a handler are of slightly lesser importance.

As for the first five characteristics, these are a priori observable by the experimenter. Thus, they are controlled for both a priori and ex post. For now, we are concerned with a priori controls, which are instituted as follows. Given that there are four batches of envelopes, we use four variants for each observable similarly to sender name–i.e., we use four sender addresses (two in Atlanta, GA and two in Washington, D.C.), four distinct colors and four different handwritings. For each household we then randomize colors across batches by sampling without replacement. We randomize sender addresses and handwriting across batches in a similar manner. In the next subsection, we discuss how we control a priori for the remaining observables–i.e., the recipient's neighborhood and gender.

Implementation I: Sample Construction

In order to be able to implement the field experiment, we first need a representative sample of households in Lima, Perú that are willing to serve as mail recipients.[62] This sample is not the actual sample of interest; however, it represents the means to an end. In particular, this sample enables us to observe the result of mail handlers' actions. Consequently, its representativeness is of extreme importance to the unobserved sample of interest–especially, if mail is lost at the district office or street level.

Thus, controls relating to the recipient's neighborhood and gender were of primary importance when constructing the sample of recipient households. The first step to construct a random sample of households in Lima was to gather a relatively large number of potentially interested households from which we would draw a random subgroup. To do so, we hired recruiters that usually work as surveyers for Grupo de Análisis para el Desarollo (GRADE) and those recruiters were in charge of gathering potential recipient households.

The recruiters were asked to solicit willingness to participate from recipient households. In doing so, they mainly pulled from their standard survey-takers, which are representative of a large range of social strata. Both recruiters and potential recipient households were informed of the basics of the study; however, they were not informed of the exact purpose of the study. In particular, potential participants were informed that the study would entail confirming receipt or non-receipt of pieces of mail. This first step led to a total of 206 potential recipient households.

We then performed four checks for consistency and representativeness of the data. First, we used GIS to map all addresses. This had two main purposes: (1) it served as a check whether or not the addresses were correct and active and (2) it

[62] The choice of Lima, Perú was dictated by a couple of main factors. First, this is where there is a hypothesized treatment effect–based on experiences with sending mail. Secondly, this is where there is a network to realize a study of this magnitude.

gave a first look at the geographical distribution of addresses. This first test rendered fifteen (15) addresses invalid and thus brought the group of potential households to 191. These 191 households were relatively representative of Lima.

Secondly, we performed a GIS-based "radius test." Basically, we traced a one kilometer (i.e., 0.62 miles) radius around each household.[63] Then, we calculated the total number of households within each household's radius including the household itself. Each household that had five (5) or more households within its radius was dropped from the sample. This avoided that we had locational clusters within the sample of potential households and it reduced the group of potential households to 154.

Thirdly, we performed a check at the district post office level. Since each district has a fixed number of post office branches, we compared the total number of potential households within a district to the number of post offices in the respective district. In cases where the total number of post offices were small compared to the total number of households within a district, we randomly chose a subset of households to be eligible for the study.

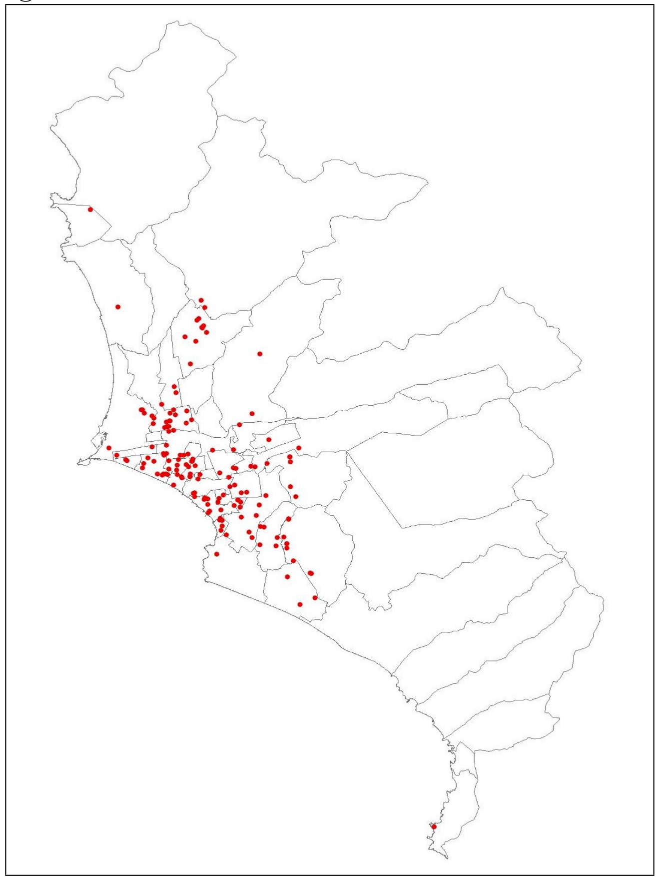
Finally, we did a manual check to see whether any neighboring households survived the above eliminations. These final two checks led to a total number of 147 eligible recipient households, all of which were willing to participate in the study.[64] So, the ultimate sample of recipient households is in some sense a stratified quasi-random sample as discussed in Banerjee et al. (2006). An approximate distribution of the households across Lima is portrayed in figure 5. Of these households, 47% of the mail recipients are male and 53% are female. These households are representative of 77.27% of Lima's postal districts.

---

[63] As a reference point, a one-kilometer radius represents a radius of approximately ten (10) blocks.

[64] The data analysis is based on $N = 138$. The discrepancy is due to some ex post attrition caused by (1) nonresponse and (2) changed/invalid addresses.

Figure 5: Distribution of Households Across Lima



Implementation II: Data Collection

All batches of mail were sent by airmail through the United States Postal Service (USPS.) After a batch was sent, each household was informed via email that they had been sent an envelope. The English translation of the email that was sent to the households is included in the appendix. In particular, they were informed of the contents of the envelope and were requested to confirm receipt or non-receipt via email within a two- to three-week period.[65] If they did not, the recruiters were requested to track down responses. Households were also asked to report the following aspects if applicable: (1) the condition and contents of the envelope, (2) the

[65]Typically, USPS and Royal Mail consider international postal mail to be lost once fifteen business days (i.e., three weeks) have past. According to the Universal Postal Union (UPU), there is no international standard for lost mail.

sender's name and address and (3) the date received.

Furthermore, they were told to keep the two dollars that were in the money mail. This was done to mitigate any incentives to lie about non-receipt. Also, households were informed that they should hold onto the lottery tickets included in no-money mail, since these lottery tickets would be used to randomly give additional rewards at the end of the study–i.e., once all four batches were sent. Households were asked to report the lottery number when confirming receipt of no-money mail and they were informed of the terms of the lottery. In particular, there would be five prizes of U.S.$ 50.00, five prizes of U.S.$ 25.00 and ten prizes of U.S.$ 10.00. Those who did not receive no-money mail were entered into a separate lottery comprising five prizes of U.S.$ 25.00.

The purpose of the lottery tickets was twofold: (1) to reward households that did not get the money-mail (since money mail is expected to be lost more frequently) and (2) to ensure that households had sufficient incentive to participate in the study from start to finish and report accordingly. Finally, households were requested not to go to the post office or to inquire about the mail with the postal carrier, since this could introduce bias in the data.

To further mitigate attrition and non-response, the recruiters followed up with their participant households. A recruiter's task was twofold: (1) to recruit potential recipient households and (2) to follow up with participating households. There was one main recruiter who was in charge of nine other recruiters. All recruiters were paid a flat fee for their services. The main recruiter was rewarded about half a month's salary. The remaining recruiters were rewarded about a third of a month's salary.

Results

The Mail in Perú

Public mail in Perú is run by Servisios Postales del Perú S.A., also known as SERPOST S.A.[66] SERPOST has a concession that obliges it to provide mail services both domestically and internationally to the Peruvian society as established in legislative decree number 685. Based on 2001 data, SERPOST is not the only mail operator in Perú; although it remains the main actor. In particular, SERPOST can be considered a semi-public monopolist.

The following descriptives (based on 2005 data) characterize SERPOST. SERPOST operates 2054 permanent post offices and has a total full-time staff of 1959 employees, of whom 41% are female. A permanent post office serves an average of 13,600 inhabitants. On a daily basis, SERPOST handles close to 32,000 letter-post items domestically and receives an average of 13,500 letter-post items from international destinations.

Treatment Effects and Interpretations

In this subsection, we identify the main treatment effects based on non-parametric means comparison tests, viz. Mann-Whitney (MW) tests. We resort to this form of analysis primarily due to nonlinearities in the data. These will become evident as we discuss the main effects below.

Identification, Measurement and Causes. We start with the question whether there is loss of mail. We find that 17.93% of mail is lost. While there are no international statistics on lost mail according to the Universal Postal Union, this percentage is substantial compared to the 0.052% reported by Royal Mail on an an-

[66] The statistics reported in this section are based on one of the following sources: (1) the Universal Postal Union (http://www.upu.int) or (2) SERPOST (http://www.serpost.com.pe.)

nual basis. Next, we ask whether loss of mail is systematic (i.e., non-random). A comparison of loss rates across money and no-money envelopes finds that 14.37% of no-money mail is lost, while 20.90% of money mail is lost. This difference is statistically significant at the 10% level (MW p-value is 0.10).

The above result indicates that loss of mail is systematic and supports H1. It indicates that there is strategic behavior taking place, since the type of mail that is sent affects the probability of loss. It should be noted that the discrepancy between loss rates allows us to distinguish this form of sophisticated or strategic behavior from the case in which all types of mail are equally likely to be lost due to for example shirking. While shirking can also be argued to be a form of corrupt behavior, it is different from the type of strategic behavior that we identify here.

Having established systematic loss of mail, we ask whether loss of mail differs across other observable characteristics. So, we now consider H2. The set of observables in x can contain many variables. We focus on the following: (1) neighborhood effects, (2) gender effects, (3) close social distance effects (i.e., whether the mail is familiar or foreign) and (4) combinations hereof. For now, we abstain from discussing other characteristics such as the color of the envelope, the sender's handwriting or the sender's address, since these are more complicated to interpret.[67]

We start with neighborhood effects. Using poverty rates across (postal) districts in Lima, we classify neighborhoods as (1) high income if they have less than ten percent poor, (2) middle income if they have between ten percent and thirty percent poor and (3) low income if they have more than thirty percent poor. We find that loss rates differ across neighborhoods. Middle-income neighborhoods are more likely to experience lost mail overall at 22.07% compared to the poor (16.95%) and the rich (13.33%). This difference is statistically significant at the 10% level (MW p-value is 0.096).

[67] Recall that these controls were mainly instituted a priori in order to avoid "red flags."

Furthermore, middle-income neighborhoods are more likely to have money mail lost. Specifically, money mail disappears at a rate of 27.27% compared to 18.18% for the poor and 15.52% for the rich. This difference is also statistically significant at the 10% level (MW p-value is 0.08). These results indicate that middle-income neigborhoods are more likely to experience lost mail, particularly if the mail contains money.

The theoretical model discussed previously lends itself to a nice interpretation of this finding. Recall that the probability of loss is driven by the benefit and the cost of opening mail as follows: $\Pr(open|x) = \frac{B(x)}{B(x)+C(x)}$. So, if handlers are more likely to open mail conditional on observing x, where x represents a middle-income neighborhood, they must perceive $B(x)$ to be higher or $C(x)$ to be lower.

The benefits of middle-income mail may be perceived to be higher than low-income mail, since middle-income mail is more likely to be valuable. Furthermore, middle-income neighborhoods may be perceived to lack proper alternatives for securely transferring money in comparison to high-income neighborhoods. In turn, the costs of opening middle-income mail may be perceived to be lower compared to high-income mail, since security is likely to play a bigger role in high-income neighborhoods. This indirectly affects the likelihood of getting caught. So, indirectly this finding tells us something about unsatisfied markets in middle-income neighborhoods.

Next, we turn to gender effects. While gender is not significant by itself, if we interact it with neighborhoods it becomes significant. In particular, low-income females are more likely to experience lost mail at a rate of 21.62% compared to low-income males (9.09%). This difference is significant at the 10% level (MW p-value is 0.08). Furthermore, high-income females are far less likely to experience mail loss at a rate of 2.63% compared to 24.66% in middle-income neighborhoods and 21.62% in high-income neighborhoods (MW p-value is 0.01). High-income females

are also more likely to receive money mail at a rate of 95.65% compared to 75.00% otherwise (MW p-value is 0.03).

Finally, we consider the close social distance aspect. We find limited evidence that familiar envelopes are more likely to be lost if they contain money at a rate of 20.65% compared to 12.36%. This finding is not significant at the 10% level (MW p-value is 0.13).

Consequences. We now consider the question why we should care about the finding that a substantial percentage of mail is lost. The fact that we find evidence of corruption is in itself interesting; however, it need not be reason for concern. After all, there are forms of corruption that tend to increase efficiency. Our main argument in this section is that this form of corruption reduces efficiency since loss of mail represents loss of information and information is important for a well-functioning market mechanism and for economic efficiency (see discussions by Stigler 1961 and Smith 1982).

Furthermore, corruption is costly to several small- and stakeholders in society. First, this form of corruption impacts the sender of mail. Suppose we consider each of the four senders of mail in our mail experiment. On average, each sender sends 68 envelopes. The cost of lost mail to each sender consists of three components: (1) the direct costs of the envelope, card and postage, (2) the direct costs of the contents (U.S.$ 2.00 for money envelopes) and (3) the costs of alternative mail services.

Components (1) and (2) are real costs and amount to U.S.$ 29.20 (23.10 plus 6.10) per sender.[68] These costs represent 19.08% of the total gross cost of the mail. Component (3) is a hypothetical cost. It is likely that a sender whose mail is lost will seek alternative modes of sending value mail. Suppose a sender compares the

[68] This amount is derived from 34 money envelopes at U.S.$ 3.25 and 34 no-money envelopes at U.S.$ 1.25 multiplied by the respective loss rates.

costs of sending an envelope containing U.S.$ 68.00 by mail or via an alternative carrier such as Western Union. The cost discrepancy is relatively substantial at U.S.$ 18.75.[69]

Secondly, this form of corruption is costly to the intended mail recipient. These costs are less problematic from an aggregate economy perspective, since any lost funds are likely to be used in other forms of economic activity. Nonetheless, they constitute a direct cost to the intended mail recipient. We saw previously that this form of corruption is more costly to female heads of low-income neighborhoods and middle-income neighborhoods. Especially, the former group can be a particularly vulnerable group to begin with.

Third, this form of corruption is costly to SERPOST. In particular, SERPOST's 2005 operating revenues were U.S.$ 4,186,764.05. If 19.08% is representative of loss of mail across Perú and this percentage resorts to alternative modes of communication, then this form of corruption costs SERPOST U.S.$ 798,834.58 of revenues on annual basis. This can also have further ramifications for the Peruvian economy if there are other competitors in the mail sector whose revenues do not necessarily benefit the Peruvian economy (e.g., Western Union, FEDEX etc.)

Finally, perhaps the greatest cost of this form of corruption is not quantifiable. It is the loss of trust that arises within and towards the Peruvian society due to this type of corruption. There have been many works on trust as a form of social capital and as a driving force of (economic) prosperity (e.g., Fukuyama 1995.) This form of corruption attacks trust within the society directly and affects not only how people perceive the mail system, but also how they perceive government, accountability, transparency and society. While these costs are not quantifiable, they may be the most significant.

[69] The costs for sending money via Western Union have been calculated using the pricing services at www.westernunion.com. The prices were calculated both for Georgia and the District of Columbia. The cost discrepancy is the same regardless of point of origin. The money transfer fee is U.S.$ 20.00 and the costs of envelope and postage are U.S.$ 1.25.

Implications and Interpretations. The findings discussed previously have further implications. In particular, they indicate that privatization may not work as an efficiency-enhancing measure. SERPOST is a public-private partnership and thus can be considered part private. It is clear that the privatized component has not led to better service; particularly for middle-income neighborhoods. This relates to arguments raised against privatization as a corruption-taming mechanism (see for example Kaufmann and Siegelbaum 1997).

At the same time, we must ask why this form of corruption persists. One possible explanation is the fact that there is little competition in the postal sector. SERPOST remains the largest actor in the mailing sector and it is essentially a monopolist by concession. Thus, poor quality of service persists, since there are no incentives to change.

Conclusion

This paper reports a theory-testing randomized field experiment that makes use of an existing semi-public monopolist institution–the postal system in Lima, Perú–to identify and measure corruption, its potential causes and its consequences. We find the following. First, 19.08% of mail is lost. Secondly, money mail is more likely to be lost at a rate of 20.90% compared to no-money mail which disappears at a rate of 14.37% (this difference is statistically significant at the 10% level). This finding suggests that loss of mail is systematic (non-random), which implies that this type of corruption is due to strategic behavior as opposed to plain shirking on the part of mail handlers.

Third, we find that loss of mail is non-random across other observables. In particular, corruption is more costly to certain societal groups. Middle-income neighborhoods are more likely to experience lost mail. They are also more likely to experience loss of money mail (both differences are statistically significant at the 10%

level). This result is informed by the predictions of the model, which suggests that mail handlers may perceive the benefits (costs) of lost mail to be higher (lower) in middle-income neighborhoods compared to low- and high-income neighborhoods.

For example, middle-income neighborhoods may be thought to be more likely to "receive" items of value in comparison to low-income neighborhoods. They may also be thought to be less likely to report lost mail compared to high-income neighborhoods. This in turn affects the likelihood of handlers getting caught. So, these two effects likely lead to (money) mail being lost at a higher rate in middle-income neighborhoods.

We also find that female heads of household in low-income neighborhoods are more likely to experience lost mail. This is not necessarily a surprising result, since this group may be perceived to be vulnerable to begin with (this difference is statistically significant at the 10% level). Furthermore, female heads of household in high-income neighborhoods are much less likely to experience lost (money) mail (these differences are statistically significant at the 5% level). This may be due to the fact that this demographic is more likely to report lost mail.

Finally, this form of corruption is costly to different stakeholders. The sender of mail bears a direct and an indirect cost. The direct cost is the value of the mail. The indirect cost is the cost of having to switch carriers once mail has been lost. Corruption is also costly to the intended mail recipient. In particular, middle-income neighborhoods and female heads of household in low-income neighborhoods are more likely to suffer. Finally, corruption is costly the mail company (SERPOST) in terms of lost revenue and costly to society in terms of loss of trust as a form of social capital (Fukuyama 1995).

Overall, the findings suggest that public-private partnerships need not increase efficiency by reducing corruption; particularly, when the institution maintains a monopoly by form of government concession. This goes back to arguments against

privatization as discussed in Kaufmann and Siegelbaum (1997). Increased efficiency in mail delivery is likely to require (1) privatization and (2) competition. Without external competition, the monopolist has no incentive to provide better service and loss of mail is likely to persist.

CONCLUSION

This dissertation consists of three essays. The first essay addresses the extent to which yardstick competition acts as a corruption-taming mechanism in an experimental environment. The study finds the following. First, there is an initial non-trivial proportion of good incumbents in the population. This proportion goes down as the experiment session progresses. Secondly, a large proportion of bad incumbents make theoretically inconsistent choices given the assumptions of the model and the manner in which preferences are theoretically specified. Finally, there is mild evidence of yardstick competition. In particular, an institution with low proportion of good incumbents has little room for yardstick competition, since bad incumbents are likely to be replaced by equally bad incumbents. This is also the case in which (1) yardstick competition leads to non-increasing voter welfare (as is observed in these experiments) and (2) voters are more likely to re-elect bad domestic incumbents due to the presence of equally bad foreign incumbents.

The second essay addresses three main questions related to the determinants and control of corruption. First, whether an increase in the marginal cost of public funds gives rise to more or less corruption. Secondly, whether a decrease in transparency (as modeled by a mean-preserving spread of the distribution of the unit cost of public good provision) affects corrupt behavior. Finally, whether an experimental environment with yardstick competition gives rise to more or less corruption. The study finds the following. First, tax inefficiency is an important determinant of corrupt incumbent behavior. In particular, an increase in the marginal cost of public funds makes it more costly for incumbents to equalize first-period payoffs. This drives them to separate and divert maximum rents in the first period. While voters retaliate slightly by voting incumbents out of office, they are worse off. Secondly, we find that increased lack of a particular form of transparency (as defined

in terms of an increase in risk in the distribution of the unit cost) leaves corrupt incumbent behavior unchanged. If the draw of the unit cost is unfavorable, incumbents tend to be less corrupt. So, the results suggest that lack of transparency (as defined in Viceisza 2007a) need not always make voters worse off. While this finding may seem counterintuitive, it is not given the parameterization of the experiments. First, we must not ignore the importance of assumptions on types and preferences as discussed in Viceisza (2007a). Contrary to Besley and Smart's (2007) assumption that incumbents are either good or bad, incumbents are known to behave strategically. Since the mean-preserving spread in the distribution of the unit cost reduces the cost of equalizing first-period payoffs in the favorable state, it makes sense why incumbents behave less corruptly if the unit cost is low. Furthermore, Viceisza (2007a) models lack of transparency on the incumbent side. This can be contrasted with lack of transparency on the voter's side as discussed in Besley and Smart (2007). Third, the experiments find strong evidence of yardstick competition. On the incumbent's side, yardstick competition acts as a corruption-taming mechanism if the incumbent is female. On the voter's side, voters are less likely to re-elect the incumbent in the presence of yardstick competition. Specifically, voters pay attention to the difference between the tax signal in their own jurisdiction and the signal in another jurisdiction. As this difference increases, voters re-elect less. Fourth, history is an important determinant of corruption and of re-election decisions. Incumbents are likely to make choices as they did in previous repetitions and voters are likely to vote out increasingly as the repetitions go by. In other words, they distrust the political system more significantly as time goes by and yardstick competition does not affect that. Finally, we find that individual-specific factors such as gender and beliefs play a significant role in incumbent behavior. In particular, female incumbents are more likely to divert rent when the unit cost is unfavorable. Furthermore, incumbents' beliefs are more important than voters' beliefs in

decision-making. Voters focus mainly on taxes charged (payoffs) and the history of those taxes.

Finally, the third essay in this dissertation reports a theory-testing randomized field experiment that makes use of an existing public institution–the postal system in Lima, Perú–to identify and measure corruption, its causes and its consequences. The study finds the following. First, 19.08% of mail is lost. Secondly, money mail is more likely to be lost at a rate of 20.90% (this is significant at the 10% level). This finding suggests that loss of mail is systematic (non-random), which implies that this type of corruption is due to strategic behavior as opposed to plain shirking on the part of mail handlers. Third, we find that loss of mail is non-random across other observables. Middle-income neighborhoods are more likely to experience lost (money) mail. Also, female heads of household in low-income neighborhoods are more likely to experience lost mail while female heads of household in high-income neighborhoods are much less likely to experience lost (money) mail. Finally, this form of corruption is costly to different stakeholders. The sender of mail bears a direct and an indirect cost. The direct cost is the value of the mail. The indirect cost is the cost of having to switch carriers once mail has been lost. Corruption is also costly to the intended mail recipient as discussed above. Finally, corruption is costly the mail company (SERPOST) in terms of lost revenue and to society in terms of loss of trust. Overall, the findings suggest that public-private partnerships need not increase efficiency by reducing corruption; particularly, when the institution remains a monopoly. Increased efficiency in mail delivery is likely to require (1) privatization and (2) competition; otherwise, the monopolist has no incentive to provide better service and loss of mail is likely to persist.

APPENDIX A: PROOF OF PROPOSITION 5

Note: This proof assumes that θ and G_θ are continuous. This is done to be consistent with expression 5. It is implied that when θ and G_θ are discrete, the appropriate definitions and notation apply.

Existence of Equilibria. The good type's equilibrium behavior is by assumption unchanged when the distribution of θ undergoes a mean preserving spread. The more complicated question is whether the bad type's–in particular, type (b, L)'s–equilibrium behavior (as in expressions 15 through 17) remains unchanged.[70]

Consider type (b, L)'s restraint formulation rule in expression 12. The magnitude of r_1 is crucial. Recall that $r_1 = t_1 - \theta_1 G_{\theta_1}$, where the subscript 1 indicates the time period. It follows that if $\theta_1 G_{\theta_1}$ is large enough, type (b, L) is left no choice but to charge maximal taxes to divert at least some rent. Furthermore, if $r_1 = 0$ a priori the game becomes trivial.

So, a necessary and sufficient condition for existence of the equilibria discussed in expressions 15 through 17 is that $r_1 > 0$, which holds if and only if $t_1 > \theta_1 G_{\theta_1}$. So, assuming a starting point for which $t_1 > \theta_1 G_{\theta_1}$, a change in θ must give rise to the relationship $\frac{dt_1}{d\theta_1} > \frac{d\left(\theta_1 G_{\theta_1}\right)}{d\theta_1}$, which is satisfied as long as the inverse relationship between θ and G_θ is sufficiently bounded. To know when this is the case, compare the interaction between a change in θ and a change in G_θ. In other words, consider the magnitude of the derivative of θG_θ. Recognizing that $\frac{d(\theta G_\theta)}{d\theta} = G_\theta + \theta \frac{dG_\theta}{d\theta}$, it is clear that the magnitude of this expression depends on the magnitude of $\frac{dG_\theta}{d\theta}$. From

[70] Note that the voter's re-election rule is affected only by λ since q is unchanged. λ–in turn–is dictated by type (b, L)'s restraint formulation.

expression 5, we have

$$\frac{dG_\theta}{d\theta} = -\frac{-\mu\frac{\partial C}{\partial t}(\theta G) - \mu\theta G\frac{\partial^2 C}{\partial t^2}(\theta G)}{-\mu\theta^2\frac{\partial^2 C}{\partial t^2}(\theta G)} = $$
$$= -\frac{1}{\theta^2}\frac{\frac{\partial C}{\partial t}(\theta G)}{\frac{\partial^2 C}{\partial t^2}(\theta G)} - \frac{1}{\theta}G =$$
$$= -\frac{1}{\theta}\left(\frac{1}{\theta}\frac{\frac{\partial C}{\partial t}(\theta G)}{\frac{\partial^2 C}{\partial t^2}(\theta G)} + G\right) < 0 \qquad (28)$$

While the sign of expression 28 is known, its magnitude is not. The magnitude will depend on the empirical parameterization of the tax cost function. However, given that $\frac{dt_1}{d\theta_1} > \frac{d\left(\theta_1 G_{\theta_1}\right)}{d\theta_1}$ must hold, some bounds for the magnitude of expression 28 can be established.

Noting that $\frac{dt_1}{d\theta_1} \gtreqless 0$ depending on the level of θ, we have the following cases:

Proof.

- Case 1: $\frac{dt_1}{d\theta_1} \leq 0$. This implies that $G_{\theta_1} + \theta_1\frac{dG_{\theta_1}}{d\theta_1} \leq 0$ such that $\left|\frac{dG_{\theta_1}}{d\theta_1}\right| \geq \frac{G_{\theta_1}}{\theta_1}$.
- Case 2: $\frac{dt_1}{d\theta_1} > 0$. This implies that $G_{\theta_1} + \theta_1\frac{dG_{\theta_1}}{d\theta_1} \gtreqless 0$ such that the above holds or $\left|\frac{dG_{\theta_1}}{d\theta_1}\right| < \frac{G_{\theta_1}}{\theta_1}$.

■

Thus, the derivations below assume that θ and G_θ are such that either case 1 or case 2 is satisfied. This will guarantee a priori existence of all three types of equilibria.

Equilibrium Rent Diversion. Expected equilibrium rent diversion in this case is given by

$$ER = (1-\pi)\,qT + (1-\pi)\,(1-q)\left[\psi\left((H-L)\,G_H + \beta T\right) + (1-\psi)\,T\right]$$
$$= (1-\pi)\left(qT + (1-q)\left[\psi\left((H-L)\,G_H + \beta T\right) + (1-\psi)\,T\right]\right) \qquad (29)$$

which increases as $(H-L)$ increases, which is the case here since $H_2 > H_1$ and

$L_2 < L_1$.

Equilibrium Voter Welfare. From Besley and Smart we know that equilibrium welfare is

Proof.

$$\begin{aligned} EW &= (1+\beta)\left[\pi q W_H + \pi(1-q)W_L + (1-\pi)W_T\right] + \\ & (1-\pi)(1-q)\lambda(W_H - W_T) + \\ & \beta\left[\pi(1-\pi)\left[q\psi + (1-q)(1-\psi\lambda)\right]\right] * \qquad (30) \\ & (qW_H + (1-q)W_L - W_T) \qquad (31) \end{aligned}$$

which–recalling that $W_\theta = G_\theta - \mu C(\theta G_\theta)$ for $\theta = \{H, L\}$–has first-order partial

$$\begin{aligned} \frac{\partial EW}{\partial \theta} &= (1+\beta)\pi q\left(\frac{\partial G_H}{\partial H} - \mu\frac{\partial C}{\partial t}(.) * \left(G_H + H\frac{\partial G_H}{\partial H}\right)\right) + \\ & (1+\beta)\pi(1-q)\left(\frac{\partial G_L}{\partial L} - \mu\frac{\partial C}{\partial t}(.)\left(G_L + L\frac{\partial G_L}{\partial L}\right)\right) + \\ & (1-\pi)(1-q)\lambda\left(\frac{\partial G_H}{\partial H} - \mu\frac{\partial C}{\partial t}(.)\left(G_H + H\frac{\partial G_H}{\partial H}\right)\right) + \\ & \beta q\left[\pi(1-\pi)\left[q\psi + (1-q)(1-\psi\lambda)\right]\right] * \\ & \left(\frac{\partial G_H}{\partial H} - \mu\frac{\partial C}{\partial t}(.) * \left(G_H + H\frac{\partial G_H}{\partial H}\right)\right) + \\ & \beta(1-q)\left[\pi(1-\pi)\left[q\psi + (1-q)(1-\psi\lambda)\right]\right] * \\ & \left(\frac{\partial G_L}{\partial L} - \mu\frac{\partial C}{\partial t}(.)\left(G_L + L\frac{\partial G_L}{\partial L}\right)\right). \qquad (32) \end{aligned}$$

The sign of expression 32 depends on the sign of any term $\left[\frac{\partial G_\theta}{\partial \theta} - \mu\frac{\partial C}{\partial t}(.) * \left(G_\theta + \theta\frac{\partial G_\theta}{\partial \theta}\right)\right]$ for $\theta = \{H, L\}$, since all coefficients are positive for the equilibrium configurations discussed. ■

Rearranging this expression, we have

$$\frac{\partial G_\theta}{\partial \theta} - \mu\frac{\partial C}{\partial t}(.) * \left(G_\theta + \theta\frac{\partial G_\theta}{\partial \theta}\right) = \frac{\partial G_\theta}{\partial \theta}\left(1 - \mu\theta\frac{\partial C}{\partial t}\right) - \mu\frac{\partial C}{\partial t}G_\theta \qquad (33)$$

which is negative as long as

$$\begin{aligned}\left(1-\mu\theta\frac{\partial C}{\partial t}\right) &> 0 \iff \\ \frac{\partial C}{\partial t} &< \frac{1}{\mu\theta}\end{aligned} \tag{34}$$

since $\frac{\partial G_\theta}{\theta} < 0$ by expression 5 and the discussion on page 18.

APPENDIX B: EXPERIMENT INSTRUCTIONS I

Note: The experiment instructions have been slightly adapted for presentation in the paper. In particular, the following changes have been made: (1) the figures have been renumbered as part of the paper and (2) the figures have been "reshaped" (i.e., made smaller).

Instructions

Preliminaries

- Hello and welcome.
- You are now taking part in an economics experiment.
- Please read these instructions carefully.
- We ask that–as of this point–you no longer talk to each other.
- Please turn off all cellular phones, two-way pagers and any other electronic devices.
- This set of instructions will explain the steps that are involved in taking part in this economics experiment. Everything you need to know to participate in this experiment is explained below.
- If you have any questions about these instructions or about any issue during the experiment, please raise your hand. We will come to you and answer your questions at your cubicle.

General

- This experiment consists of a trial, four tasks and a questionnaire.

- Each task requires interaction between two players: Player X and Player Z.
- Before you make any decisions, you will randomly be assigned a role. Once you are assigned a role, you will maintain that role throughout the whole experiment.
- So, please pay very close attention to all of the instructions.

Task B

Instructions for Player X

- There are five periods in this task. Each period has two parts.
- In this task, each period Player X will be paired with a NEW Player Z. So, Player X will NOT be paired with the same Player Z each and every period.

First Part of the Period

1. In the first part of the period, Player X makes a choice between THREE alternatives or TWO alternatives.
2. The computer will "toss a coin." If the coin lands "HEADS," Player X chooses between THREE alternatives. If the coin lands "TAILS," Player X chooses between TWO alternatives.
3. Each alternative is a payoff pair. The first number is Player X's payoff. The second number is Player Z's payoff.
4. Suppose the coin lands "HEADS." The screen Player X will see is shown in figure 6. Please take a moment to look at the layout of the screen and the payoffs. The first column indicates the respective alternative. The second and third columns indicate Player X's and Player Z's payoffs under those alternatives. For example, if Player X chooses alternative 1, Player X gets a payoff of

1.50 experimental dollars and Player Z also gets a payoff of 1.50 experimental dollars. On other hand, if Player X chooses alternative 2, Player X gets a payoff of 1.50 experimental dollars and Player Z gets a payoff of 0.75 experimental dollars. Finally, if Player X chooses alternative 3, Player X gets a payoff of 2.35 experimental dollars and Player Z gets a payoff of 0.16 experimental dollars.

Figure 6: Player X Sample Screen: First Part of Period (THREE alternatives)

5. Suppose the coin lands "TAILS." The screen with the TWO alternatives Player X chooses from is shown in figure 7. Please take a moment to look at the layout of the screen and the payoffs. The screen is similar to the previous; however, Player X now chooses between TWO rather than THREE alternatives. Also, note that the payoffs are different.

6. Once Player X decides on the preferred alternative, Player X selects the corresponding radio button at the bottom of the screen. Player X clicks OK to confirm this choice. This choice will determine both Player X's and Player Z's payoffs for the first part of the period.

Second Part of the Period

1. Player Z will get to "accept" or "reject" Player X's choice for the first part of the period. If Player Z "rejects" Player X's choice, Player X and Player

Figure 7: Player X Sample Screen: First Part of Period (TWO alternatives)

Z STILL get the payoffs that Player X chose for the first part of the period. However, Player X's payoff for the second part of the period will be ZERO and Player X will NOT get to play during the second part of the period.

2. If Player Z "accepts" Player X's choice, Player X and Player Z STILL get the payoffs that Player X chose for the first part of the period. However, IN ADDITION, Player X will get to make another choice in the second part of the period. This choice will determine Player X's and Player Z's payoffs for the second part of the period. Player Z will NOT get to "accept" or "reject" this choice at the end of the second part of the period.

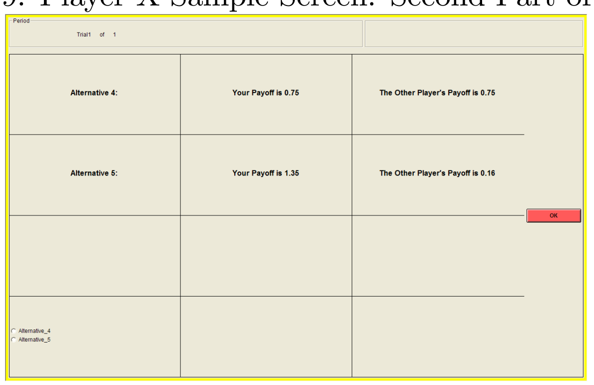
3. The alternatives that Player X chooses from in the second part of the period–IF Player Z "accepts" the first-part choice–are determined at random.

4. The computer will "toss a coin." If the coin lands "HEADS," Player X sees the screen in figure 8. Please take a moment to look at the layout of the screen and the payoffs.

5. If the coin lands "TAILS," Player X sees the screen in figure 9. Again, please take a moment to look at the layout of the screen and the payoff pairs.

Figure 8: Player X Sample Screen: Second Part of Period

Figure 9: Player X Sample Screen: Second Part of Period



Task Payoffs

Player X's payoff in each period is the sum of the payoff in the first and the second part of the period. At the end of each period, Player X will see how much he or she earned during that period and any previous periods. All payoffs shown in this task are in experimental dollars. Each experimental dollar is a quarter of a real U.S. dollar. So, for example, if Player X makes 8 experimental dollars in this task, Player X will get paid 2 real U.S. dollars for this task at the end of the experiment.

Waiting Involved...

Note that each decision round is complete when everyone participating in the experiment (i.e., all Players X and all Players Z) has submitted a decision. Only at that point will the program advance to the next decision round. This means that the delay of a single participant's decision can prevent the next round from starting. Please take time to make your decision, but please also keep the above in mind.

Questions

Are there any questions so far?

Task B

Instructions for Player Z

- There are five periods in this task. Each period has two parts.
- In this task, each period Player Z will be paired with a NEW Player X. So, Player Z will NOT be paired with the same Player X each and every period.

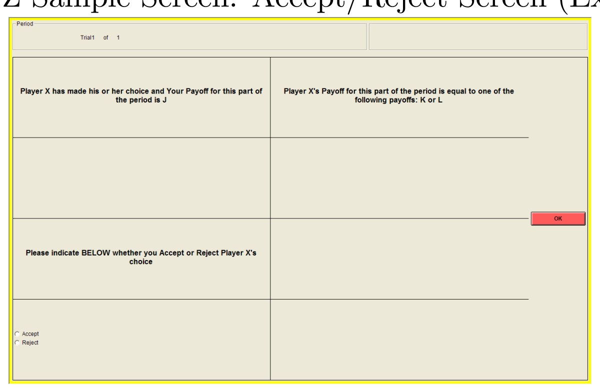
First Part of the Period

1. Player Z will NOT get to see whether Player X chose between THREE alternatives (i.e., figure 6) or TWO alternatives (i.e., figure 7).

2. Player Z sees a screen of the type shown in figure 10. Figure 10 assumes Player X chose a payoff of J experimental dollars for Player Z. In this task, J can be 1.50, 0.75 or 0.16 experimental dollars. Please take a moment to look at the layout of the screen in figure 10.

   - The first cell in the first row indicates Player Z's payoff as chosen by Player X. In this case, the payoff chosen is J.
   - The second cell in the first row indicates the possible payoffs for Player X. In this case, the possible payoffs are K and L. In the actual task, the numbers for K and L will vary depending on the alternative chosen by Player X.
   - The second row is BLANK.
   - The first cell in the third row instructs Player Z to "Please indicate your choice BELOW."
   - Finally, the first cell in the fourth row gives the option for Player Z to "accept" or "reject" Player X's choice.

Figure 10: Player Z Sample Screen: Accept/Reject Screen (Example: Payoff=J)



3. Once Player Z has selected the preferred radio button at the bottom of the screen, Player Z clicks OK to confirm this choice.

4. Remember that Player Z's payoff in the first part of the period is determined by Player X's choice REGARDLESS of whether Player Z "accepts" or "rejects" Player X's choice.

5. Player Z's decision to "accept" or "reject" Player X's choice affects payoffs in the second part of the period. This is explained further below.

Second Part of the Period

1. If Player Z "accepts" Player X's choice, Player X gets to make another choice in the second part of the period. Player Z will NOT get to "accept" or "reject" this choice in the second part.

2. If Player Z "rejects" Player X's choice, Player X does NOT get to make a choice in the second part of the period. INSTEAD, Player X gets ZERO payoff and Player Z's payoff is determined at random as follows:

    - With 25% chance, Player Z's payoff is 1.50 experimental dollars.
    - With 25% chance, Player Z's payoff is 0.75 experimental dollars.
    - With 50% chance, Player Z's payoff is 0.16 experimental dollars.

Task Payoffs

Player Z's payoffs are calculated in the same way as Player X's. Player Z's payoff in each period is the sum of the payoff in the first and the second part of the period. At the end of each period, Player Z will see how much he or she earned during that period and any previous periods. All payoffs shown in this task are in experimental dollars. Each experimental dollar is a quarter of a real U.S. dollar. So,

for example, if Player Z makes 8 experimental dollars in this task, Player Z will get paid 2 real U.S. dollars for this task at the end of the experiment.

Waiting Involved...

Remember that each decision round is complete when everyone participating in the experiment (i.e., all Players X and all Players Z) has submitted a decision. This means that the delay of a single participant's decision can prevent the next round from starting. Please take time to make your decision, but please also keep the above in mind.

Questions

Are there any questions before we discuss the Trial?

Trial

Now, we will have five minutes of what is called a trial. During this trial, you will NOT be assigned a role. This means that you will get to make choices as if you were Player X. Then, you will get to observe those choices as if you were Player Z. During this process, you will get to see the screens that both Players X and Z will get to see once Task B starts. During this trial, you will NOT be playing against any other player. So, the trial is an opportunity for you to interact and familiarize yourself with the software.

The trial will last five minutes. You will be paid 2.00 REAL U.S. dollars for this trial REGARDLESS of how many periods you complete. So, please take your time and pay attention to the screens you see in the trial.

If you have any questions during the trial period, please raise your hand. We will come to your cubicle and answer your questions.

Are there any questions before we start the Trial?

Task Y

Instructions for Player X

1. For Player X, this task is NO different than the previous task.

Instructions for Player Z

1. For Player Z, this task is SLIGHTLY different than the previous task.

2. As shown in figure 10, in this task–Task Y–the second row of the "accept/reject" screen is NO LONGER BLANK.

3. The second row now indicates two additional pieces of information:

   - First, the second row indicates the total number of first-part offers made by ALL Players X during ALL periods in the previous task. This number does NOT include second-part offers. This information is indicated in the first cell of the second row.

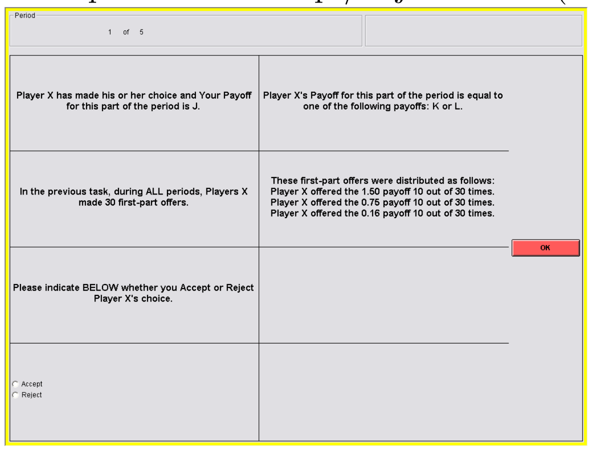
   - Secondly, the second row summarizes the distribution of these first-part offers. In other words, the second row ALSO summarizes how many times ALL Players X offered Player Z (*) a first-part payoff of 1.50, (*) a first-part payoff of 0.75 or (*) a first-part payoff of 0.16 during ALL periods in the previous task. This information is indicated in the second cell of the second row.

   - For instance, in the example below (figure 1), ALL Players X together made 30 first-part offers during ALL periods in the previous task.

   - Of these 30 first-part offers, 10 were associated with a Player Z payoff of 1.50, another 10 were associated with a Player Z payoff of 0.75 and another 10 were associated with a Player Z payoff of 0.16.

   - This is the ONLY difference between this task and the previous task!

Figure 11: Player Z Sample Screen: Accept/Reject Screen (Example, Payoff=J)



Task Payoffs

In this task, payoffs for both players are calculated in the same way as in the previous task.

Waiting Involved...

As in the previous task, RECALL that each decision round is complete when everyone participating in the experiment has submitted a decision. Only at that point will the program advance to the next decision round.

Questions

Are there any questions?

APPENDIX C: EXPERIMENT INSTRUCTIONS II

Note: The experiment instructions have been slightly adapted for presentation in the paper. In particular, the following changes have been made: (1) the figures have been renumbered as part of the paper, (2) the figures and text have been reduced in size, and (3) the preliminaries section has only been included for Task B.

Instructions

Preliminaries

- Hello and welcome.
- You are now taking part in an economics experiment.
- Please read these instructions carefully.
- We ask that–as of this point–you no longer talk to each other.
- Please turn off all cellular phones, two-way pagers and any other electronic devices.
- This set of instructions will explain the steps that are involved in taking part in this economics experiment. Everything you need to know to participate in this experiment is explained below.
- If you have any questions about these instructions or about any issue during the experiment, please raise your hand. We will come to you and answer your questions at your cubicle.

General

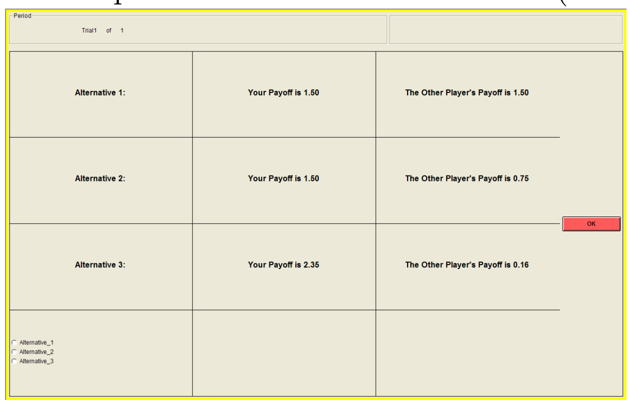
- This experiment consists of four sections: (1) a trial, (2) a quiz, (3) a task called Task B and (4) a questionnaire.
- Task B requires interaction between two players: Player X and Player Z.
- Before you make any decisions, you will be told whether you are Player X or Player Z. You will keep the same role for the entire experiment.
- Please pay very close attention to all of the instructions.

Task B

1. There are ten periods in this task. Each period has two parts, a first and a second part.
2. Each period you will be paired with a DIFFERENT person in the room. So, you will NOT interact with the same person for more than one period.
3. At the beginning of the first part of the period, a coin is flipped. Player X knows the result of the coin flip. Player Z does not. Player X should NOT reveal the result of the coin flip to Player Z. The coin flip takes place in front of the first Player X. He verifies the coin flip and I input it into the computer.
4. If the coin flip is "HEADS," Player X chooses between THREE alternatives; if it is "TAILS," TWO alternatives.
5. Each alternative is a payoff pair: The first is Player X's; the second is Player Z's.
6. Suppose the coin lands "HEADS." Player X sees the screen in figure 12. Please look at the screen and the payoffs. The first column is the respective alternative. The second and third columns are Player X's and Player Z's payoffs. If Player X chooses alternative 1, Player X and Player Z get 1.50. If Player X

chooses alternative 2, Player X gets 1.50 and Player Z gets 0.75. Finally, if Player X chooses alternative 3, Player X gets 2.35 and Player Z gets 0.16.

Figure 12: Player X Sample Screen: First Part of Period (THREE alternatives)



7. Suppose the coin lands "TAILS." Player X sees the screen in figure 13. The screen is similar to the previous; but, Player X chooses between TWO rather than THREE alternatives. Also, the payoffs are different.

Figure 13: Player X Sample Screen: First Part of Period (TWO alternatives)

8. Player X selects his choice and clicks OK to confirm. This choice determines payoffs for the first part of the period.

9. Player Z does NOT see the coin flip. So, Player Z does NOT see whether Player X chose between the THREE alternatives in figure 12 or the TWO

alternatives in figure 13.

10. Player Z sees a screen like the one in figure 14. Figure 14 is a sample screen. Please look at the screen in figure 14.

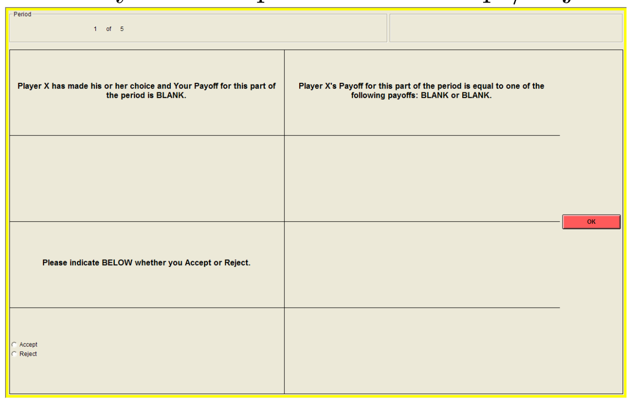
    - The first cell in the first row indicates Player Z's payoff. In the sample screen, the payoff is indicated as "BLANK." This number can be either 1.50, 0.75 or 0.16 depending on the alternative chosen by Player X.
    - The second cell in the first row indicates Player X's possible payoffs. In the sample screen, the possible payoffs are indicated as "BLANK." In the actual task, Player X's possible payoffs depend on Player Z's payoffs as shown in the following table.

	If Player Z has...	Then, Player X has...
Case a	1.50	1.50
Case b	0.75	1.50 OR 0.75
Case c	0.16	2.35 OR 1.35

    - So, if Player Z has a first-part payoff of 1.50, then Player X has a first-part payoff of 1.50 also. If Player Z has a first-part payoff of 0.75, then Player X could also have a first-part payoff of 0.75, but Player X could have a first-part payoff of 1.50 instead. This depends on the number of alternatives that Player X chose from. Finally, if Player Z has a first-part payoff of 0.16, then Player X either has a first-part payoff of 2.35 or 1.35. This depends on the number of alternatives that Player X chose from.

11. Player Z "accepts" or "rejects" at the end of the first part of the period. Player Z selects his choice at the bottom of the screen and clicks OK to confirm. The decision to "accept" or "reject" determines whether or not Player X makes another choice in the second part of the period. It does NOT determine or affect first-part payoffs for either player. First-part payoffs are whatever Player X chose them to be.

Figure 14: Player Z Sample Screen: Accept/Reject Screen



12. So, if Player Z "rejects," both players still get the payoffs chosen by Player X in the first part of the period. However, Player X does NOT make a choice in the second part of the period. Instead, Player X's payoff for the second part of the period is zero and Player Z's second-part payoff is determined randomly as indicated in the following table.

If Player Z "rejects," Player Z's Second-Part Payoff is:
1.50 with 25% chance
0.75 with 25% chance
0.16 with 50% chance

13. On the other hand, if Player Z "accepts," both players still get the payoffs chosen by Player X in the first part of the period. But, in addition, Player X makes a choice in the second part of the period and this choice determines second-part payoffs for both players. The computer will pause and I will flip the coin again. This second coin toss affects the alternatives that Player X chooses from in the second part of the period. This second coin flip will be verified in the same way as the first coin flip. Players Z will NOT see the result of this coin flip.

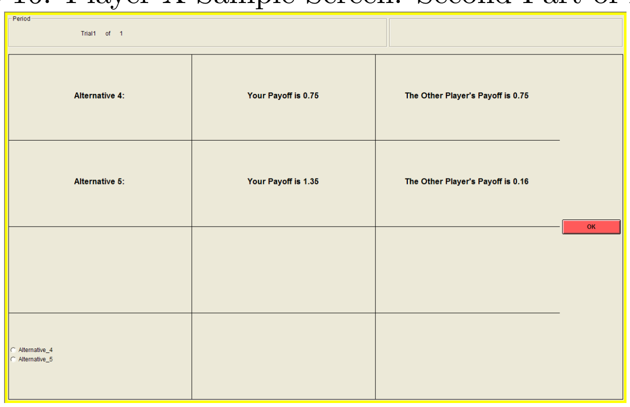
14. If the coin lands "HEADS," Player X sees the screen in figure 15. Please look at the screen and the payoffs.

15. If the coin lands "TAILS," Player X sees the screen in figure 16. Please look

Figure 15: Player X Sample Screen: Second Part of Period

at the screen and the payoffs.

Figure 16: Player X Sample Screen: Second Part of Period



16. Player X's second-part choice determines Player X's and Player Z's payoffs for the second part. Player Z will NOT get to "accept" or "reject" at the end of the second part of the period.

17. The above steps describe the two parts within a period. This task consists of ten of these periods.

Task Payoffs

Payoffs in each period are the sum of the payoff in the first and the second part of the period. At the end of each period, you will see how much you earned during

that period and any previous periods. You will be paid your total earnings for all periods. All payoffs shown are in experimental dollars. Each experimental dollar is a quarter of a real U.S. dollar. So, for example, if you make 8 experimental dollars in this task, you will get paid 2 real U.S. dollars for this task at the end of the experiment.

Waiting Involved...

Note that each decision round is complete when everyone participating in the experiment (i.e., all Players X and all Players Z) has submitted a decision. This means that the delay of a single participant's decision can prevent the next round from starting. Please take time to make your decision, but please also keep the above in mind.

Questions

Are there any questions before we discuss the Trial?

Trial & Quiz

Now, we will have five minutes of what is called a trial. During this trial, you will NOT be assigned a role. This means that you will get to make choices as if you were Player X. Then, you will get to observe those choices as if you were Player Z. During this process, you will get to see the screens that both Players X and Z will get to see once Task B starts. During this trial, you will NOT be interacting with any other player. So, the trial is a practice opportunity for you to familiarize yourself with the software and the instructions.

The trial will last five minutes. If you have any questions during the trial period, please raise your hand. We will come to your cubicle and answer your questions.

After the trial is completed, you will be given a quiz on the computer to make sure you understand the procedures and how your earnings are determined. The questions in the quiz are based on the trial and the instructions.

You will be paid 3.00 REAL U.S. dollars for participating in this trial and completing the quiz REGARDLESS of how many trial periods you complete or how many quiz questions you answer correctly. So, please take your time and pay attention to the screens you see in the trial and the questions you are asked in the quiz.

Are there any questions before we start the Trial?

Task B Summary

1. You will randomly be Player X or Player Z. You will keep the same role for the entire experiment.

2. At the beginning of the first part of the period, a coin is flipped.

3. Player X knows the result of the coin flip and knows his alternatives. If the coin toss is "HEADS" there are THREE alternatives; if the coin toss is "TAILS" there are TWO alternatives.

4. Player X makes his first-part choice. For each possible choice, Player X knows the first-part payoffs for himself AND for Player Z.

5. Player Z does NOT know the result of the coin toss and does not know the number of alternatives faced by Player X.

6. Player Z knows his first-part payoff for sure, but does not know Player X's first-part payoff. The following table–which is also displayed on the screen at the front of the room–shows the possible first-part payoffs for Player X depending on Player Z's first-part payoff.

	If Player Z has...	Then, Player X has...
Case a	1.50	1.50
Case b	0.75	1.50 OR 0.75
Case c	0.16	2.35 OR 1.35

7. Player Z "accepts" or "rejects." This decision determines whether or not Player X makes another choice in the second part of the period. It does NOT determine or affect first-part payoffs for either player. First-part payoffs are whatever Player X chose them to be.

8. If he "accepts," both players get the first-part payoffs chosen by Player X, but in addition Player X sees another coin toss and makes a choice that determines payoffs in the second part. Player Z does NOT "accept" or "reject" for a second time.

9. If he "rejects," both players get the first-part payoffs chosen by Player X, but Player X's second-part payoff is zero. Player Z's second-part payoff is determined randomly. The possibilities are shown in the table below.

If Player Z "rejects," Player Z's Second-Part Payoff is:
1.50 with 25% chance
0.75 with 25% chance
0.16 with 50% chance

10. The above describes the two parts within a period. This task consists of ten of these periods.

Task Y

1. There are ten periods in this task. Each period has two parts, a first and a second part.

2. Each period you will be paired with a DIFFERENT person in the room. So, you will NOT interact with the same person for more than one period.

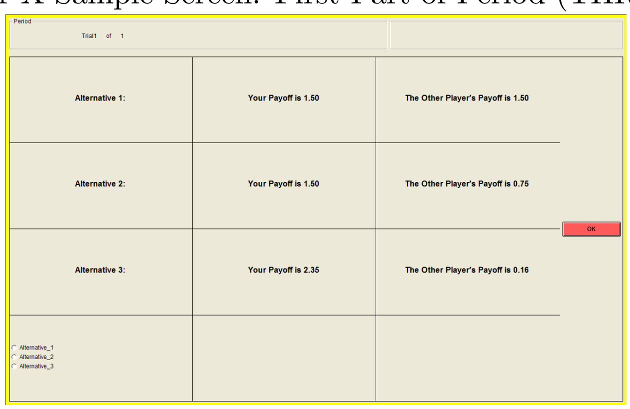
3. At the beginning of the first part of the period, a coin is flipped. Player X knows the result of the coin flip. Player Z does not. Player X should NOT reveal the result of the coin flip to Player Z. The coin flip takes place in front of the first Player X. He verifies the coin flip and I input it into the computer.

4. If the coin flip is "HEADS," Player X chooses between THREE alternatives; if it is "TAILS," TWO alternatives.

5. Each alternative is a payoff pair: The first is Player X's; the second is Player Z's.

6. Suppose the coin lands "HEADS." Player X sees the screen in figure 17. Please look at the screen and the payoffs. The first column is the respective alternative. The second and third columns are Player X's and Player Z's payoffs. If Player X chooses alternative 1, Player X and Player Z get 1.50. If Player X chooses alternative 2, Player X gets 1.50 and Player Z gets 0.75. Finally, if Player X chooses alternative 3, Player X gets 2.35 and Player Z gets 0.16.

Figure 17: Player X Sample Screen: First Part of Period (THREE alternatives)



7. Suppose the coin lands "TAILS." Player X sees the screen in figure 18. The screen is similar to the previous; but, Player X chooses between TWO rather than THREE alternatives. Also, the payoffs are different.

Figure 18: Player X Sample Screen: First Part of Period (TWO alternatives)

8. Player X selects his choice and clicks OK to confirm. This choice determines payoffs for the first part of the period.

9. Player Z does NOT see the coin flip. So, Player Z does NOT see whether Player X chose between the THREE alternatives in figure 15 or the TWO alternatives in figure 16.

10. Player Z sees a screen like the one in figure 19. Figure 19 is a sample screen. Please look at the screen in figure 19.

    - The first cell in the first row indicates Player Z's payoff. In the sample screen, the payoff is indicated as "BLANK." This number can be either 1.50, 0.75 or 0.16 depending on the alternative chosen by Player X.
    - The second cell in the first row indicates Player X's possible payoffs. In the sample screen, the possible payoffs are indicated as "BLANK." In the actual task, Player X's possible payoffs depend on Player Z's payoffs as shown in the following table.

	If Player Z has...	Then, Player X has...
Case a	1.50	1.50
Case b	0.75	1.50 OR 0.75
Case c	0.16	2.35 OR 1.35

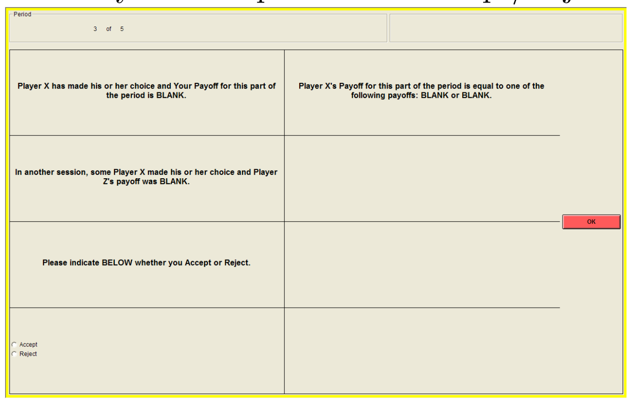
    - So, if Player Z has a first-part payoff of 1.50, then Player X has a first-part payoff of 1.50 also. If Player Z has a first-part payoff of 0.75, then

Player X could also have a first-part payoff of 0.75, but Player X could have a first-part payoff of 1.50 instead. This depends on the number of alternatives that Player X chose from. Finally, if Player Z has a first-part payoff of 0.16, then Player X either has a first-part payoff of 2.35 or 1.35. This depends on the number of alternatives that Player X chose from.

- The first cell in the second row shows information from a session similar to the one that you are in today. This session was called session B. The cell shows a particular Player Z's first-part payoff in session B. This was the result of a choice made by the Player X that he was paired with. Session B was the same as today's session in terms of procedures: A coin was flipped and verified by the first Player X and Player Z did NOT get to see the coin flip. Furthermore, if the coin landed "HEADS," Player X chose from the screen in figure 17. If the coin landed "TAILS," Player X chose from the screen in figure 18. However, session B was potentially different from today's session in terms of the result of the coin flip: In other words, while you see another Player Z's first-part payoff coming from session B, you do NOT know whether the Player X that made this choice saw the same coin flip as the Player X you are paired with. So, your Player X could have observed "HEADS" while the other Player X observed "TAILS" or vice versa. In the sample screen below, the other Player Z's first-part payoff is indicated as "BLANK." During the task, this number is either 1.50, 0.75 or 0.16. The first-part payoff shown in the cell has been chosen at random from all Player Z first-part payoffs in session B in any given period.

11. Player Z "accepts" or "rejects" at the end of the first part of the period. Player Z selects his choice at the bottom of the screen and clicks OK to confirm. The

Figure 19: Player Z Sample Screen: Accept/Reject Screen



decision to "accept" or "reject" determines whether or not Player X makes another choice in the second part of the period. It does NOT determine or affect first-part payoffs for either player. First-part payoffs are whatever Player X chose them to be.

12. So, if Player Z "rejects," both players still get the payoffs chosen by Player X in the first part of the period. However, Player X does NOT make a choice in the second part of the period. Instead, Player X's payoff for the second part of the period is zero and Player Z's second-part payoff is determined randomly as indicated in the following table.

If Player Z "rejects," Player Z's Second-Part Payoff is:
1.50 with 25% chance
0.75 with 25% chance
0.16 with 50% chance

13. On the other hand, if Player Z "accepts," both players still get the payoffs chosen by Player X in the first part of the period. But, in addition, Player X makes a choice in the second part of the period and this choice determines second-part payoffs for both players. The computer will pause and I will flip the coin again. This second coin toss affects the alternatives that Player X chooses from in the second part of the period. This second coin flip will be verified in the same way as the first coin flip. Players Z will NOT see the result of this coin flip.

14. If the coin lands "HEADS," Player X sees the screen in figure 20. Please look at the screen and the payoffs.

Figure 20: Player X Sample Screen: Second Part of Period

15. If the coin lands "TAILS," Player X sees the screen in figure 21. Please look at the screen and the payoffs.

Figure 21: Player X Sample Screen: Second Part of Period

16. Player X's second-part choice determines Player X's and Player Z's payoffs for the second part. Player Z will NOT get to "accept" or "reject" at the end of the second part of the period.

17. The above steps describe the two parts within a period. This task consists of ten of these periods.

Task Payoffs

Payoffs in each period are the sum of the payoff in the first and the second part of the period. At the end of each period, you will see how much you earned during that period and any previous periods. You will be paid your total earnings for all periods. All payoffs shown are in experimental dollars. Each experimental dollar is a quarter of a real U.S. dollar. So, for example, if you make 8 experimental dollars in this task, you will get paid 2 real U.S. dollars for this task at the end of the experiment.

Waiting Involved...

Note that each decision round is complete when everyone participating in the experiment (i.e., all Players X and all Players Z) has submitted a decision. This means that the delay of a single participant's decision can prevent the next round from starting. Please take time to make your decision, but please also keep the above in mind.

Questions

Are there any questions before we discuss the Trial?

Trial & Quiz

Now, we will have five minutes of what is called a trial. During this trial, you will NOT be assigned a role. This means that you will get to make choices as if you were Player X. Then, you will get to observe those choices as if you were Player Z. During this process, you will get to see the screens that both Players X and Z will get to see once Task Y starts. During this trial, you will NOT be interacting with any other player. So, the trial is a practice opportunity for you to familiarize yourself with the software and the instructions.

The trial will last five minutes. If you have any questions during the trial period, please raise your hand. We will come to your cubicle and answer your questions.

After the trial is completed, you will be given a quiz on the computer to make sure you understand the procedures and how your earnings are determined. The questions in the quiz are based on the trial and the instructions.

You will be paid 3.00 REAL U.S. dollars for participating in this trial and completing the quiz REGARDLESS of how many trial periods you complete or how many quiz questions you answer correctly. So, please take your time and pay attention to the screens you see in the trial and the questions you are asked in the quiz.

Are there any questions before we start the Trial?

Task Y Summary

1. You will randomly be Player X or Player Z. You will keep the same role for the entire experiment.

2. At the beginning of the first part of the period, a coin is flipped.

3. Player X knows the result of the coin flip and knows his alternatives. If the coin toss is "HEADS" there are THREE alternatives; if the coin toss is "TAILS" there are TWO alternatives.

4. Player X makes his first-part choice. For each possible choice, Player X knows the first-part payoffs for himself AND for Player Z.

5. Player Z does NOT know the result of the coin toss and does not know the number of alternatives faced by Player X.

6. Player Z knows his first-part payoff for sure, but does not know Player X's first-part payoff. The following table–which is also displayed on the screen at

the front of the room–shows the possible first-part payoffs for Player X depending on Player Z's first-part payoff.

	If Player Z has...	Then, Player X has...
Case a	1.50	1.50
Case b	0.75	1.50 OR 0.75
Case c	0.16	2.35 OR 1.35

7. Player Z also knows the first-part payoff of some Player Z in another session. The procedures in this session were the same as in the experiment today. While Player Z sees this other Player Z's first-part payoff, he does NOT know whether the Player X that chose it saw the same coin flip as the Player X that he is paired with.

8. Player Z "accepts" or "rejects." This decision determines whether or not Player X makes another choice in the second part of the period. It does NOT determine or affect first-part payoffs for either player. First-part payoffs are whatever Player X chose them to be.

9. If he "accepts," both players get the first-part payoffs chosen by Player X, but in addition Player X sees another coin toss and makes a choice that determines payoffs in the second part. Player Z does NOT "accept" or "reject" for a second time.

10. If he "rejects," both players get the first-part payoffs chosen by Player X, but Player X's second-part payoff is zero. Player Z's second-part payoff is determined randomly. The possibilities are shown in the table below.

If Player Z "rejects," Player Z's Second-Part Payoff is:
1.50 with 25% chance
0.75 with 25% chance
0.16 with 50% chance

11. The above describes the two parts within a period. This task consists of ten of these periods.

APPENDIX D: EMAIL TO HOUSEHOLD RECIPIENTS

From: labgsu@gmail.com

To: First Name, Last Name

Subject: Research Study–Mail Sent (Confirmation Necessary)

This email is to inform you that you have been selected to participate in our research study.

As mentioned in the original invitation, we will send you a total of four (4) envelopes in the mail, each of which has different content.

Every time we send you an envelope, we will notify you via email. The _________ envelope was sent on _________ and you should receive it soon.

All envelopes contain a card. In addition, some contain two U.S. dollars and some contain lottery numbers. You should keep the contents of the envelopes. In particular, you should hold onto the lottery numbers, since they will be used to award further prizes at the end of the research study. The two dollars are yours to keep.

The research study requires that you inform us no later than _________ whether or not you have received the envelope. Furthermore, it requires that you answer the following questions:

1) The sender's name and sender's address (not yours.)

2) The date of receipt.

3) The envelope's condition–i.e., whether or not the mail arrived in good condition, open, wrinkled.

4) The contents of the envelope–i.e., whether or not the envelope contained the card and/or they money and/or the lottery number.

5) If the envelope contained money, how much was it? If the envelope contained a lottery, what was the number?

Due to the confidential nature of the research study, it is extremely important that neither you (nor any member of your household) inquire about arrival of the envelope to either the mail carrier or the postal office. If you do, you will jeopardize the purpose of the research. All you need to do is inform us whether or not the envelope has arrived and answer the above questions.

To reward you for your time and cooperation, all participants will be entered into a lottery that contains five prizes of 50 U.S. dollars, five prizes of 25 U.S. dollars and 10 prizes of 10 U.S. dollars.

The lottery numbers that we are sending you will be used for this purpose and you will only be part of the lottery drawing if you send responses to our questions before _________.

Those of you who do not receive mail will be entered into a separate lottery consisting of five prizes of 25 U.S. dollars. The winning lottery numbers will be announced via email.

Many thanks for your collaboration,

Angelino Viceisza

Georgia State University

REFERENCES

Abbink, Klaus. 2005b. Laboratory experiments on corruption. In *Handbook of economic corruption (forthcoming)*, ed. Susan Rose-Ackerman. Northhampton, U.S., and Cheltenham, U.K.: Edward Elgar Publishers.

Abbink, Klaus, and Matthew Ellman. 2005. The donor problem: University of Nottingham and Universitat Pompeu Fabra.

Aguirregabiria, Victor, and Pedro Mira. 2007. Sequential estimation of dynamic discrete games. *Econometrica* 75, no. 1: 1-53.

Aidt, Toke S. 2003. Economic analysis of corruption: A survey. *The Economic Journal* 113, no. 491: F632-F652.

Arellano, Manuel, and Stephen Bond. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58, no. 194: 277.

Banerjee, Abhijit, Esther Duflo, Rachel Glennerster, and Michael Kremer. 2006. Randomization in development economics: A toolkit. In *Handbook of development economics*, 4.

Banks, Jeffrey, Colin Camerer, and David Porter. 1994. An experimental analysis of Nash refinements in signaling games. *Games and Economic Behavior* 6, no. 1: 1-31.

Bardhan, Pranab. 1997. Corruption and development: A review of issues. *Journal of Economic Literature* 35, no. 3: 1320.

Bardhan, Pranab, and Dilip Mookherjee. 2005. Decentralization, corruption and government accountability: An overview. In *Handbook of economic corruption (forthcoming)*, ed. Susan Rose-Ackerman. Northhampton, U.S. and Cheltenham, U.K.: Edward Elgar Publishers.

Bayer, Patrick, and Stephen L. Ross. 2006. Identifying individual and group effects in the presence of sorting: A neigborhood effects application: Duke University and University of Connecticut.

Becker, Gary S. 1968. Crime and punishment: An economic approach. *Journal of Political Economy* 76, no. 2: 169.

Belleflamme, Paul, and Jean Hindriks. 2005. Yardstick competition and political agency problems. *Social Choice & Welfare* 24, no. 1: 155-169.

Bertrand, Marianne, Simeon Djankov, Rema Hanna, and Sendhil Mullainathan. 2006. Obtaining a driving license in India: An experimental approach to studying corruption *Quarterly Journal of Economics (forthcoming).*

Besley, Timothy, and Anne C. Case. 1995a. Incumbent behavior: Vote-seeking, tax-setting, and yardstick competition. *The American Economic Review* 85, no. 1: 25-45.

Besley, Timothy, and Michael Smart. 2007. Fiscal restraints and voter welfare. *Journal of Public Economics* 91, no. 3-4: 755-773.

Bodenstein, Martin, and Heinrich Ursprung. 2004. Political yardstick competition, economic integration, and constitutional choice in a federation: Board of Governors of the Federal Reserve System and University of Konstanz.

Bordignon, Massimo, Floriana Cerniglia, and Frederico Revelli. 2004. In search for yardstick competition: Property tax rates and electoral behavior in Italian cities: Universita Cattolica , Catholic University of the Sacred Heart of Milan–General and University of Turin.

Brandts, Jordi, and Charles A. Holt. 1992. An experimental test of equilibrium dominance in signaling games. *American Economic Review* 82, no. 5: 1350.

Brandts, Jordi, and Charles A. Holt. 2005. Naive Bayesian learning and adjustment to equilibrium in signaling games: Instituto de Análisis Económico and University of Virginia.

Brown, George W. 1951. Iterative solutions of games by fictitious play. In *Activity analysis of production and allocation*, ed. Tjalling C. Koopmans. New York: Wiley.

Browning, Edgar K. 1976. The marginal cost of public funds. *Journal of Political Economy* 84, no. 2: 283.

Camerer, Colin, and Teck Hua Ho. 1999. Experience-weighted attraction learning in normal form games. *Econometrica* 67, no. 4: 827-874.

Chakrabarti, Rajesh, and Ajay Subramanian. 2006. Power, compensation and corruption: Georgia Institute of Technology and Georgia State University.

Cox, James C. 2004. How to identify trust and reciprocity. *Games and Economic Behavior* 46, no. 2: 260-281.

Cox, James C., and Cary A. Deck. 2006. When are women more generous than men? *Economic Inquiry* 44, no. 4: 587-598.

Cho, In-Koo, and David M. Kreps. 1987. Signaling games and stable equilibria. *Quarterly Journal of Economics* 102, no. 2: 179-221.

Dahlby, Bev. 1998. Progressive taxation and the social marginal cost of public funds. *Journal of Public Economics* 67, no. 1: 105-122.

Dincer, Oguzhan, Christopher Ellis, and Glen Waddell. 2006. Corruption, decentralization and yardstick competition: Massey University and University of Oregon.

Djankov, Simeon, Rafael La Porta, Florencio Lopez-De-Silanes, and Andrei Shleifer. 2002. The regulation of entry. *Quarterly Journal of Economics* 117, no. 1: 1-37.

Dusek, Libor, Andreas Ortmann, and Lubomir Lizal. 2005. Understanding corruption and corruptibility through experiments. *Prague Economic Papers* 14, no. 2.

Duflo, Esther. forthcoming. Field experiments in development economics. In *Advances in economic theory and econometrics*, ed. Richard Blundell, Whitney Newey and Thorsten Persson, 2. Cambridge, U.K.: Cambridge University Press.

Fischbacher, Urs. 2007. Z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics (forthcoming)* 10, no. 2: 171-178.

Frank, Bjorn, and Gunther G. Schulze. 2000. Does economics make citizens corrupt? *Journal of Economic Behavior & Organization* 43, no. 1: 101-113.

Fudenberg, Drew, and Jean Tirole. 1991. Perfect bayesian equilibrium and sequential equilibrium. *Journal of Economic Theory* 53, no. 2: 236-260.

Fudenberg, Drew, and Jean Tirole. 1992. *Game theory*. Cambridge, MA: The MIT Press.

Fukuyama, Francis. 1995. *Trust: The social virtues and the creation of prosperity*: Free Press.

Gillette, Ann B., Thomas H. Noe, and Michael J. Rebello. 2003. Corporate board composition, protocols, and voting behavior: Experimental evidence. *Journal of Finance* 58, no. 5: 1997-2032.

Grether, David M. 1980. Bayes rule as a descriptive model: The representativeness heuristic. *Quarterly Journal of Economics* 95, no. 3: 537-557.

Grosskopf, Brit, and Rajiv Sarin. 2006. Is reputation good or bad? An experiment: Texas A&M University.

Gupta, Sanjeev, Hamid R. Davoodi, and Rosa Alonso-Teme. 1998. Does corruption affect income inequality and poverty? IMF Working Paper 98/76.

Harrison, Glenn W., and John A. List. 2004. Field experiments. *Journal of Economic Literature* 42, no. 4: 1009-1055.

Harsanyi, John C. 1967. Games with incomplete information played by "Bayesian" Players, i-iii. Part i. The basic model. *Management Science* 14, no. 3.

________. 1968. Games with incomplete information played by "Bayesian" Players, i-iii. Part ii. Bayesian equilibrium points. *Management Science* 14, no. 5.

________. 1968. Games with incomplete information played by "Bayesian" Players, i-iii. Part iii. The basic probability distribution of the game. *Management Science* 14, no. 7.

Harsanyi, John C. 1973. Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points. *International Journal of Game Theory* 2: 1-23.

Holmstrom, Bengt R. 1982. Moral hazard in teams. *Bell Journal of Economics and Management Science* 13, no. 2: 324-340.

Huntington, S. P. 1968. Modernization and corruption. In *Political order in changing societies*: 59-71. New Haven: Yale University Press.

Jain, Arvind K. 2001. Corruption: A review. *Journal of Economic Surveys* 15, no. 1:71.

Jonhson, Lynn C. 2006. Yardstick competition: An empirical investigation using state taxes and media markets: Massachussetts Institute of Technology.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2006. Measuring corruption: Myths and realities: The World Bank.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. forthcoming. Measuring governance using cross-country perceptions data. In *Handbook of economic corruption*, ed. Susan Rose-Ackerman. Northhampton, U.S. and Cheltenham, U.K.: Edward Elgar Publishers.

Kaufmann, Daniel, and Paul Siegelbaum. 1997. Privatization and corruption in transition economies. *Journal of International Affairs* 50, no. 2: 419.

Klitgaard, Robert E. 1988. *Controlling corruption*. Berkeley, California: University of California Press.

Klitgaard, Robert E. 1991. Gifts and bribes. In *Strategy and choice*, ed. Richard Zeckhauser. Cambridge, MA: MIT Press.

Leff, Nathaniel H. 1964. Economic development through bureaucratic corruption. *American Behavioral Scientist* 8, no. 3: 8-14.

Lockwood, Benjamin. 2005. A note on the hybrid equilibrium in the Besley-Smart model: University of Warwick.

Lui, Francis T. 1985. An equilibrium queuing model of bribery. *Journal of Political Economy* 93, no. 4: 760.

Mauro, Paolo. 1995. Corruption and growth. *Quarterly Journal of Economics* 110, no. 3: 681.

Martinez-Vazquez, Jorge, Javier Arze del Granado, and Jameson Boex. 2007. *Fighting corruption in the public sector. Contributions to economic analysis*: Elsevier.

Miller, Ross M. and Charles R. Plott. 1985. Product quality signaling in experimental markets. *Econometrica* 53, no. 4: 837-872.

Olken, Benjamin. 2006. Corruption perceptions versus corruption reality: NBER.

Olken, Benjamin A. 2007. Monitoring corruption: Evidence from a field experiment in Indonesia. *The Journal of Political Economy* 115, no. 2: 200.

Potters, Jan, Martin Sefton, and Lise Vesterlund. Leading-by-example and signaling in voluntary contribution games: An experimental study. *Economic Theory (forthcoming)*.

Potters, Jan, Bettina Rockenbach, Abdolkarim Sadrieh, and Eric van Damme. 2004. Collusion under yardstick competition: An experimental study. *International Journal of Industrial Organization* 22, no. 7: 1017-1038.

Robinson, Joan. 1951. An iterative method of solving a game. *Annals of Mathematics* 54: 296-301.

Rose-Ackerman, Susan. 1978. *Corruption: A study in political economy.* New York: Academic Press.

Roth, Alvin E. and Ido Erev. 1995. Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior* 8, no. 1: 164-212.

Samuelson, Paul. 1947. *Foundations of economic analysis.* Cambridge, MA: Harvard University Press.

Shleifer, Andrei. 1985. A theory of yardstick competition. *RAND Journal of Economics* 16, no. 3: 319-327.

Shleifer, Andrei, and Robert W. Vishny. 1993. Corruption. *Quarterly Journal of Economics* 108, no. 3: 599.

Smith, Vernon L. 1989. Theory, experiment and economics. *The Journal of Economic Perspectives* 3, no. 1: 151-169.

Spence, Michael. 1973. Job market signaling. *Quarterly Journal of Economics*, 87:3, pp. 355-74.

Stigler, George J. 1961. The economics of information. *The Journal of Political Economy* 69, no. 3: 213-225.

Tanzi, Vito. 1998. Corruption around the world: Causes, consequences, scope, and cures. International Monetary Fund. *Staff Papers - International Monetary Fund* 45, no. 4: 559.

The World Bank, PREM. 1997. *Helping countries combat corruption: The role of the World Bank.* Washington, D.C.: The World Bank.

Viceisza, Angelino C. G. 2004. Annals of government competition: Georgia State University.

________. 2007a. An experimental inquiry into the effect of yardstick competition on corruption: Georgia State University.

________. 2007b. Corruption in the mail sector: Evidence from a randomized field experiment in Lima, Perú: Georgia State University.

________. 2007c. Determinants and control of corruption: Efficiency, transparency and yardstick competition: Georgia State University.

Warlters, Michael, and Emmanuelle Auriol. 2005. The marginal cost of public funds in developing countries: An application to 38 African countries: World Bank and ARQADE, IDEI (Toulouse).

Weibull, Jörgen. 2004. Testing game theory. In *Advances in understanding strategic behaviour: Game theory, experiments and bounded rationality*. Essay in honour of Werner Güth., ed. Steffen Huck: 85-104: Palgrave MacMillan.

Wilson, John Douglas. 1999. Theories of tax competition. *National Tax Journal*, 52:269.

Wooldridge, Jeffrey M. 2002. *Econometric analysis of cross section and panel data*. Cambridge, MA: The MIT Press.

VITA

Angelino Casio Giovanni Viceisza was born on April 27, 1979 on Curaçao, Netherlands Antilles. He holds a Bachelor of Science in Business Administration (Major: Accounting) from the University of the Netherlands Antilles (Curaçao), a Master of Business Administration (Major: International Business) from Temple University (Philadelphia, PA), a Master of Arts in economics from Boston University (Boston, MA) and a Master of Arts in economics with a policy track from Georgia State University (Atlanta, GA).

In the fall of 2003, Angelino joined Georgia State University (Georgia State) to pursue a doctoral degree in economics. Throughout his time at Georgia State, he has obtained numerous internal and external awards. Internally, Angelino has won (1) the Harold Ball Economics Award for distinct academic achievement in quantitative economics, (2) the George Malanos Scholarship for fostering a community of scholars and (3) the Theodore C. Boyden Award for excellence in teaching. Angelino also obtained a dissertation grant from Georgia State University.

Externally, Angelino has received support from several organizations. In 2007, he received a travel grant from the Public Choice Society to present the paper *"An Experimental Inquiry into the Effect of Yardstick Competition on Corruption"* at the First World Meeting of the Public Choice Societies in Amsterdam. The paper received an honorable mention as one of the three best papers submitted for the Wicksell Prize–a prize that is awarded annually by the European Public Choice Society to the best paper by a scholar under the age of thirty. He also received a travel grant from the organizing committee of the the 2007 joint Latin American and Caribbean Economic Association (LACEA) and Latin American Meetings of the Econometric Society (LAMES) meetings in Bogotá to present his ongoing work on corruption, yardstick competition and efficiency.

In 2006 Angelino received the Joseph K. Heyman Scholarship from the Atlanta Economics Club for outstanding academic performance in business economics. In former years, he also obtained support from institutions in his home country including the Gregory Elias Foundation, United Telecommunication Services (Talent Management Program) and the University of the Netherlands Antilles (Faculty of Social and Economic Studies).

While at Georgia State, Angelino was involved with many internal and external activities. Internally, he served as a graduate research assistant to both the Experimental Economics Center and the Fiscal Research Center. He also served as a graduate teaching assistant for two courses in the Ph.D. in economics core, viz. microeconomic theory and statistics. In 2007, Angelino was selected to be the student speaker (on behalf of the economics department) at the Eleventh Annual Andrew Young School of Policy Studies Honor's Dinner. He also presented in the Georgia State Economics Department brown bag seminar series.

Externally, Angelino presented at several conferences and seminars including but not limited to (1) the joint 2007 LACEA-LAMES meetings (Bogotá), (2) the First World Meeting of the Public Choice Societies (Amsterdam), (3) the International Food Policy Research Institute's seminar series (Washington, D.C.) and (4) the Federal Reserve Bank of Atlanta's Research Department brown bag seminar series (Atlanta, GA).

Angelino received his doctorate in economics from the Andrew Young School of Policy Studies at Georgia State University in December 2007. He has since accepted the position of post-doctoral fellow in experimental economics with the markets, trade and institutions division of the International Food Policy Research Institute in Washington, D.C. Angelino plans to pursue a long-term career in research and academia and can be reached by email at a.viceisza@cgiar.org or via his homepage, http://www.ifpri.org/srstaff/ViceiszaA.