Essays in Political Economy

by

Otis Russell Reid

B.A. Economics, B.A. Public Policy, Stanford University (2012)

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Economics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2018

Signature redacted

Author ....

Department of Economics
January 15, 2018

Signature redacted

Certified by ....

Abhijit Banerjee
Ford Foundation International Professor of Economics
Thesis Supervisor

Signature redacted

Certified by ....

Benjamin A. Olken
Professor of Economics
Thesis Supervisor

Signature redacted

Accepted by ....

Ricardo J. Caballero
Ford International Professor of Economics
Chairman, Department Committee on Graduate Theses

Essays in Political Economy

by

Otis Russell Reid

Submitted to the Department of Economics
on January 15, 2018, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Economics

Abstract

This thesis consists of three chapters on political economy. Each chapter explores the effects of a change to the equilibrium of a given market.

In the first chapter, Jon Weigel and I study a randomized controlled trial in the Democratic Republic of the Congo on corruption at tolls. We randomly vary incentives for drivers to comply with rules instead of engaging in corruption. These incentives affect the "supply" of corruption rather than the "demand" for corruption from bureaucrats. We find that sizable financial incentives produce a 7 to 10 percentage point increase in the probability that drivers get receipts, implying an elasticity of citizen supply of bribes ranging from -0.45 to -0.95. Social incentives have no effect. Similarly, providing information about other drivers' responses to treatment (to shift social norms) does not affect behavior. Drivers' appear remarkably inelastic in their supply of bribes. We argue this reflects the fact that bribe payment may increase the efficiency of transactions in the toll setting we examine and suggest that corruption may serve to "grease the wheels" in this context.

In the second chapter, Christopher Blattman, Horacio Larreguy, Benjamin Marx, and I study a large-scale randomized controlled trial designed to combat vote-buying in the 2016 election in Uganda. We find that the campaign did not reduce the extent to which voters accepted cash and gifts in exchange for their votes. In addition, we designed the study to take advantage of our large sample (covering 1.2 million voters) to examine both direct treatment and spillover effects. The spillover effects on vote-buying are also zero, but the campaign had large direct and indirect effects on vote-shares for candidates. Heavily treated areas had increases in visits from non-incumbent candidates and non-incumbent candidates improved their vote shares substantially in these parishes. Consistent with these effects, we find evidence that the campaign diminished the effectiveness of vote-buying transactions by shifting local social norms against vote-selling and by convincing some voters to vote their conscience, regardless of any gifts received.

In the third chapter, I examine the effect of the 26th Amendment, which lowered the voting age in the United States from 21 to 18. This change enfranchised a large population of new voters, expanding the electorate by almost 9%. However, I find that the Amendment had little effect on overall political outcomes in the United States. Although it did increase total turnout in areas with more young voters, it did not affect the partisan composition of the electorate and correspondingly did not lead to changes in representation or policy. These results stand in contrast to other well-studied expansions of the franchise and provide an important caveat to those findings: when the preferences of new voters are insufficiently distinct from those of existing voters, politicians have little reason to change their established positions.

Thesis Supervisor: Abhijit Banerjee
Title: Ford Foundation International Professor of Economics

Thesis Supervisor: Benjamin A. Olken
Title: Professor of Economics

Acknowledgments

This document marks the end of a long road. Over the last four and a half years at MIT, I have learned a tremendous amount both about economics and myself.

I am indebted to my advisors, Abhijit and Ben. I am hugely appreciative for their support, both in my research and in my choice to pursue a non-academic job after MIT. The papers included here are much better for having taken their advice at different forks along the way, though, as always, remaining mistakes are my own. In addition to my main academic advisors, I am grateful for the advice and support I received from Daron Acemoglu, Chris Blattman, Esther Duflo, Horacio Larreguy, Frank Schilbach, Tavneet Suri, and Heidi Williams. In particular, I want to thank Heidi not only for practical advice, but also for never failing to ask about myself or others in the department as people. I am also greatly indebted to the economics mentors that predated MIT, but who have continued to support me in graduate school: Larry Diamond, Avner Greif, Steve Stedman and Corinne Thomas, Jeremy Weinstein, and, most of all, Pete Klenow. Lastly, I want to thank the many other teachers who have impacted my journey along the way and, in particular, Marilyn Metzler and Keith Cooper of Chapel Hill High School.

I would not have made it – or at least would have had a much grimmer time – without my classmates. My office mates (Ben Marx, Ray Kluender, Dan Waldinger, Michael Stepner, Jon Petkun) have consistently made my days in the department brighter and the burdens of research lighter. I can scarcely recall a day where I haven't turned around to ask one of them a question about research or life. In particular, my research has been hugely strengthened by having a friend and co-author with such similar research interests in the form of Ben Marx. I am also thankful for the camaraderie of my entire year of students, who made the dark days of first year much brighter and more manageable, and in particular to Pooya Molavi for making TAing a much easier experience.

One of my regrets at MIT is that I did not get to know students in other years as well as I should have. However, I am thankful for the friendships that I was able to build with the Saturday spin group (Jane Choi, Maddie McKelway, Tamar Oostrom, and Carolyn Stein (honorary member by office proximity: Ryan Hill)), the development folks (Josh Dean, John Firth, Reshma Hussam, Matt Lowe, Gabriel Kreindler, Natalia Rigol, Mahvish Shaukat, Ashish Shenoy, Cory Smith), and other older students and mentors (Alex Bartik, Sally Hudson, and Brendan Price). I am also particularly grateful to Jon Weigel, not only for being a great co-author, but for making the many months in Kananga not just bearable, but even, at times, fun.

I was also lucky enough to live with excellent roommates while in Cambridge. Valentin Bolotnyy and Andrew Hillis made moving to Cambridge easy and living in 1130 Cambridge Street a blast. Colin Gray, Sam Gould, and Sam Moy all made 60 Bishop Allen feel much warmer than it had any right to feel (even when our heat broke). I also want to thank all of my other friends who, through trips and phone calls, have helped me through graduate school, most notably Dan Fuentes, Craig Gaffney, Alex Katz, Ishan Nath, Kelder Monar, and Zane Silver. Many other friends have lent a hand throughout my time in grad school and on my way here and my failure to list them all reflects time and word constraints on my part and not at all a failure to appreciate everything they have done for me.

Last, but most importantly, I want to thank my family, both current and future. I am blessed beyond words to have the support and love of my mom, Holly Russell, my dad, Donald Reid, and my sister, Hadley Reid. I have called them to laugh, cry, celebrate, and complain about every part of my life and graduate school was no exception. I could have written a thesis chapter on each of them and the ways they have enabled my successes at MIT and beyond, but I was told that they would not qualify for inclusion here, so those essays will have to wait. I also want to acknowledge my grandfather, Grampy, and my uncle Rusty for their support in graduate school and beyond.

I am most of all grateful to MIT for bringing me together with my fiancée, Christina Patterson. I could not have completed graduate school without her love and support and I am looking forward to creating a family with her over the many decades to come. I am deeply lucky to have found someone so beautiful, kind, and brilliant and I look forward to celebrating her many successes in the years to come. No one ever gives you the advice "go to econ grad school to find a wife," but based on my $N = 1$, they should. I also want to acknowledge Circe Patterson and use this occasion to formally adopt her – the saying in Washington, D.C. is that "if you want a friend, get a dog," but I think in Cambridge it's "if you want unconditional support for your work, get a cat." She has made the many set-backs in the research process more bearable, even if she has also cancelled my code a few times by walking on the keyboard.

My studies and research projects at MIT were made possible by the generous support of the MIT Presidential Fellowship, the Clyo F. Castle Graduate Fellowship Fund, the Shultz Fund, the MIT economics department, the Abdul Latif Jameel Poverty Action Lab, and the National Science Foundation. My work has been hugely aided by excellent work by our entire team of enumerators in the Democratic Republic of the Congo and Uganda, particularly Muana Kasongo and Elie Kabue, as well as excellent research assistance by Kelsey Barrera, Anemone Birkebaek, and Alex Nawar.

. . . .

This thesis is dedicated to the memory of my grandparents Nana, Grandad, and Gigi. It is also dedicated to Cara Nickolaus, whose loss is still felt in the halls of E52.

Contents

1 Citizen Participation in Corruption: Evidence from Roadway Tolls in the Democratic Republic of the Congo (with Jonathan Weigel) 13

1.1 Introduction . . . 13
1.2 Context . . . 19
1.3 Experimental Design . . . 20
1.4 Theory . . . 26
1.5 Data . . . 28
1.6 Empirical Specification . . . 32
1.7 Results . . . 36
1.8 Discussion . . . 39
1.9 Conclusion . . . 44
1.10 Tables and Figures . . . 45
1.11 Appendix . . . 60
1.12 Appendix Tables and Figures . . . 74

2 A Market Equilibrium Approach to Reduce the Incidence of Vote-Buying: Evidence from Uganda (with Christopher Blattman, Horacio Larreguy, and Benjamin Marx) 87

2.1 Introduction . . . 87
2.2 Background . . . 93
2.3 Experimental Design . . . 97
2.4 Data . . . 101
2.5 Empirical Framework . . . 102
2.6 Results . . . 105
2.7 Discussion . . . 112
2.8 Conclusion . . . 115
2.9 Tables and Figures . . . 119

3 A "Minor" Expansion: The 26th Amendment and Changes in U.S. Political Outcomes 147

3.1 Introduction . . . 147
3.2 Background . . . 150
3.3 Empirical Specification . . . 151
3.4 Data . . . 153

3.5 Results 156
3.6 Discussion 160
3.7 Conclusion 161
3.8 Tables and Figures 162

List of Figures

1-1 Effects of treatment on number of additional receipts brought each round (ITT) . . . 46
1-2 Illustrative timeline of the experiment . . . 46
1-3 Distribution of realized tax morale game winnings against the predicted distribution . . . 47
1-4 Bribes by round (each round is jittered by 10 FC for visibility) . . . 47

2-1 Main treatment effects . . . 121
2-2 Sample leaflet used in experiment . . . 121

3-1 Reported share of votes for Democrats in presidential elections, by age group 163
3-2 Proportion of the population between ages 18 and 20 as of 1970 (by county) 163
3-3 Event Studies of the Effect of the Amendment on Turnout in Presidential Races. . . . 164
3-4 Event Studies of the Effect of the Amendment on Turnout in House Races. . 165
3-5 Effects on votes for Democrats (left panel) and Republicans (right panel) in House races . . . 166
3-6 Effects on votes for Democrats (left panel) and Republicans (right panel) in presidential races . . . 167
3-7 Effects on the vote share for Democrats in House races . . . 168
3-8 Effects on turnout for House by college (left panel) and non-college (right panel) counties . . . 169
3-9 Effects on Democratic vote share for House by college (left panel) and non-college (right panel) counties . . . 170
3-10 Effects on Democratic vote share for president by college (left panel) and non-college (right panel) counties . . . 171
3-11 Effects of the 26th Amendment on total votes for House (left) and total votes for president (right) in placebo states . . . 172

THIS PAGE INTENTIONALLY LEFT BLANK

List of Tables

1.1 Summary statistics . . . 49
1.2 Choices over amounts . . . 49
1.3 Empirical distribution of choices over all rounds . . . 49
1.4 Test of balance across individual-level controls. . . . 50
1.5 Test of balance across individual-level controls (continued). . . . 51
1.6 Test of balance across trip-level controls. . . . 51
1.7 Treatment remembered versus treatment assigned . . . 52
1.8 IV first stage (each column is a regression with the outcome of recalling the treatment named at the top of the column) . . . 52
1.9 Main effects on bringing a receipt to the follow-up visit (unconditional on reporting haven taken a trip in the intervening period) . . . 53
1.10 Main effects on number of valid receipts brought to the follow-up visit (unconditional on reporting haven taken a trip in the intervening period) . . . 54
1.11 Main effects on receipts per trip. . . . 55
1.12 Total (selection and causal) estimates on equilibrium bribes. . . . 56
1.13 Tobit results on bribes. . . . 57
1.14 Covariates of amount paid when getting a receipt and not. . . . 58
1.15 Estimated elasticities of receipt-getting with respect to instantaneous monetary reward for various values of β. Bootstrapped 95% confidence interval shown in square brackets below each estimate. . . . 58
1.16 Receipts per reported trip under different conditions. . . . 59
1.17 Treatment remembered versus treatment assigned, as a percent of respondents assigned to that treatment group, in Follow-Up Visit 2 (top panel) and in the two other rounds (bottom panel) . . . 75
1.18 Treatment remembered versus treatment assigned in the previous round (not the current round), looking at people who were actually assigned the control in the current round (as a percent of those respondents), in Follow-Up Visit 2 (top panel) and in Follow-Up Visit 3 (bottom panel) . . . 76
1.19 Effects of treatment removing Follow-Up Visit 2 in whole or in part . . . 77
1.20 Effects of treatment when controlling for treatment history. . . . 78
1.21 Effects of treatment for toll regulars. . . . 79
1.22 Attrition by round. . . . 80
1.23 Treatment effects on reporting attempting to get a receipt. . . . 81
1.24 Self-reported reasons for missing a receipt from a trip. . . . 81
1.25 Treatment effects on likelihood of reporting a lost receipt. . . . 82

1.26 Overall trip reporting. 83
1.27 Trips recorded by validator, but not reported by drivers. 84
1.28 Estimated elasticities of receipt-getting with respect to instantaneous monetary reward for various values of β. Bootstrapped 95% confidence interval shown in square brackets below each estimate. 85
1.29 Heckman results on bribes. 86

2.1 Summary Statistics 123
2.2 Quality of Implementation 124
2.3 Treatment Effects on Vote-Buying (Z-Standardized) 125
2.4 Treatment Effects on Vote-Buying: Cash Received 126
2.5 Effects of the Campaign on Attitudes Towards Vote-Buying 128
2.6 Index of Treatment Effects on Vote Shares (Z-Standardized) 129
2.7 Campaign Effects on Turnout 130
2.8 Index of Treatment Effects on Campaigning (Z-Standardized) 131
2.9 Interactions on Key Outcomes 133
2.10 Balance Voter Respondent 140
2.11 Balance Key Informant 1 143
2.12 Balance on Pre-determined Electoral Data 145

3.1 Variation in voting ages prior to the 26th Amendment 174
3.2 Availability of age data by Census year 174
3.3 Effects of the 26th Amendment on total turnout by office 175
3.4 Effects of the 26th Amendment on votes for each party by office 175
3.5 Effects of the 26th Amendment on Democratic vote share 176
3.6 Effects of the 26th Amendment on vote shares for each party by office and college/non-college 176
3.7 Effects of the 26th Amendment on total votes in House and presidential races in the two placebo states, Georgia and Kentucky 177
3.8 Estimating population in intercensal years 178

Chapter 1

Citizen Participation in Corruption: Evidence from Roadway Tolls in the Democratic Republic of the Congo (with Jonathan Weigel)

1.1 Introduction

Corruption is often deemed detrimental to economic outcomes (Ferraz et al. (2012), Kaufmann and Wei (1999), Méon and Sekkat (2005)). It also limits the ability of governments to raise revenue. Although some interventions appear to reduce corruption, the magnitudes of such reductions are often small. For example, in Olken (2007) on a road-building project in Indonesia, the threat of government audits decreased estimated leakage by 8 percentage points, but even with a 100% probability of an audit, the level of missing expenditures remained at 19%. Indeed, perhaps the more striking result from this and other studies is

We are indebted to the guidance and support of our advisors Abhijit Banerjee, Nathan Nunn, Benjamin Olken, Rohini Pande, and James Robinson. Anemone Birkebaek provided excellent research assistance on this work. This paper has benefited from conversations with Daron Acemoglu, Esther Duflo, Matt Lowe, Benjamin Marx, Frank Schilbach, Michael Stepner, and the participants in the Political Economy lunch at MIT. This research would not have been possible without a grant from the Abdul Latif Jameel Poverty Action Lab. This work would not have been possible without the personal support of Christina Patterson, to whom Otis is eternally grateful. Any errors or omissions are the responsibility of the authors.

just how persistent corrupt transactions appear to be despite interventions targeted against them.

The persistence of corruption could be driven by a range of factors. In many corrupt interactions, there is both a bribe-taking bureaucrat and a bribe-paying citizen. Recent research has largely focused on the bureaucrat side. For example, monitoring technologies have proven effective in disciplining absenteeism (Duflo et al. (2012), Dhaliwal and Hanna (2013)) and monetary incentives have increased tax collector effort, albeit also increasing bribe levels (Khan et al. (2016)). The implicit view motivating these papers is that corruption is perpetuated by low-quality institutions that create weak or perverse incentives for bureaucrats. Less is known about the citizen side of corruption: in a literature review on corruption, Olken and Pande (2012) mention "bureaucrats" or "bureaucracy" 46 times; they mention 'citizen' 5 times. However, it is possible that citizens play a role in perpetuating corruption. In particular, an older theoretical literature notes how corruption can "grease the wheels" in settings of low institutional quality: paying a bribe might enable citizens to access public services faster than navigating the red tape in a bloated bureaucracy (Leys (1965), Lui (1985)). In short, citizens might supply bribes because they increase the efficiency of their transactions with the state.

This paper explores citizen supply of corruption in the context of roadway tolls in Kananga, Democratic Republic of Congo (DRC). Motorcycle taxi drivers were offered financial and social incentives to bring receipts proving that they paid the legal toll. The goal of the financial incentive treatments was to estimate the elasticity of citizen supply of bribes with respect to the price of complying with the toll. The financial incentives were either 1000 Congolese Francs (FC) – about \$1 (half the price of the toll) – or 2000 FC (the full price of the toll) for drivers with proper receipts. The goal of the social incentives was to estimate the extent to which corruption is affected by the perceived social value of tax compliance. Social incentives included (1) a pledge by the researchers to contribute 2000 FC to widows in Kananga, and (2) a pledge to contribute 2000 FC to the provincial government to subsequently transfer to widows in Kananga. The goal of these treatments is to test if citizens are willing to pay bribes because they perceive it to have low social cost in light of the pervasiveness of corruption in the government. If citizens trust our pledge to directly give widows 2000 FC for each valid receipt more than they trust the government to follow through on transferring this money, they should respond differentially to the former incentive relative to the latter. Finally, drivers in the control group were simply asked to bring receipts without any reward.

Participants who completed the baseline survey were randomized into different experimental conditions for their next trip outside of Kananga. After each of two three week periods, they returned for a follow-up visit and were subsequently re-assigned to a different treatment group, such that each participant was assigned to three different treatments over the course of the experiment. Non-attriting participants were always assigned once to a financial treatment, once to a social incentive, and once to the control group. Additionally, we cross randomized a social norms intervention seeking to alter drivers' beliefs about other drivers' propensities to pay bribes at the tolls. This intervention was intended to test the hypothesis that drivers justify participation in corruption due to their perceptions that bribery is widespread.

Figure 1-1 summarizes intent-to-treat (ITT) results. The financial incentive treatments caused participants to bring approximately .04 more receipts. IV estimates – in which driver treatment recall of their treatment status is instrumented with true treatment status – are larger, increasing the effect size to 0.11 receipts per round. The implied elasticity of citizen supply of bribes with respect to the effective price of the toll is negative, as expected, but is relatively small in absolute value, ranging from -0.45 to -0.95 depending on the driver discount factor used. Our estimated effects on the magnitude of the equilibrium bribe are mixed, though a Tobit model suggests negative effects consistent with the simple bargaining theory we introduce in Section 1.4.1. The social incentive groups are not statistically distinguishable from the control group. The social norms information also has no effect on rates of receipt bringing.

Corruption is notoriously persistent in many settings, and the treatment effects we find are comparable to those in many studies of anticorruption interventions, suggesting that citizen-side interventions have potential policy significance. However, perhaps the most striking result is that even when drivers could fully reimburse their toll payments by demanding a receipt, 87% of participants did *not* do so. Our preferred interpretation is that high citizen supply of bribes reflects the fact that bribes increase the efficiency of the toll transaction. In a collusive, Nash-bargaining setup, one would expect driver-side incentives to reduce bribery on the extensive margin and decrease the equilibrium bribe. However, if the toll officer faces lower time costs, he can strategically delay to increase the attractiveness of a quick bribe payment from time-constrained motorcycle taxi drivers. Issuing a receipt requires both parties to (i) enter a building a short distance from the road, (ii) complete paperwork, and (iii) complete an electronic form on a handheld receipt printer. In contrast, drivers who pay bribes do not even need to dismount their motorbikes. Drivers estimate that paying a bribe reduces the time spent at the toll by nearly 70%. To bolster this interpretation further, we exploit heterogeneity in driver time costs: drivers without passengers are more than twice as responsive to treatment as drivers with passengers. According to qualitative interviews, this finding reflects the fact that drivers carrying goods are less time pressed than drivers with impatient passengers. Knowing this, toll officers might strategically use delay tactics more often when they observe drivers with passengers to maximize the chances of a bribe payment, or drivers may simply have higher time costs (irrespective of toll officer behavior) in those situations.

This paper contributes to the growing field experimental literature on corruption. Prior work has demonstrated the effectiveness of monitoring technologies in decreasing absenteeism in schools (Duflo et al. (2012)) and public-sector health facilities (Dhaliwal and Hanna (2013)). Khan et al. (2016) find that incentive pay for tax collectors in Pakistan reduces the frequency of bribes and boosts government revenues. Most similar is Bertrand et al. (2007) who find that individuals in India who are promised financial rewards if they obtain a driver's license quickly are indeed more likely to obtain a license, but they are also more likely to pay bribes and they are worse drivers on average. This work predominantly focuses on the bureaucrat side, and ours is the first field experiment to incentivize citizens to forego paying bribes. The closest non-experimental study is Naritomi (2015), a study of a government policy in Brazil incentivizing consumers to obtain receipts for final products. In that context, she finds sizable decreases in evasion of the value-added tax in sectors with customers (not firms) as final consumers, consistent with other studies documenting the effectiveness of third-party information and VAT compliance (e.g. Pomeranz (2015)). A likely explanation for the different results we observe is that consumers are not trading off time costs and financial costs to the same degree as they are in the toll setting we analyze.[1] Our study suggests that efforts to intervene on the citizen side of a given bribe-taking transaction may be hampered when bribes function as an "efficient grease" to an otherwise

[1] An alternative explanation is that in Naritomi's setting, citizens are not publicly breaking a corrupt bargain, but instead are privately defecting by turning in their receipts.

slow bureaucratic process. This may be particularly true when toll officers can strategically increase the cost of compliance to elicit more bribes.

Our results therefore offer the first experimental evidence in support of the "greasing the wheels" hypothesis on corruption. This hypothesis comes from an older theoretical literature noting several channels by which corruption could improve efficiency in a non-Coasian, second-best world. Huntington (1968) memorably summarizes this view: "The only thing worse than a society with a rigid, over-centralized, dishonest bureaucracy is one with a rigid, over-centralized, honest bureaucracy." We focus on the "speed money" mechanism from this literature: corruption can speed up bureaucratic processes that are otherwise beset by red tape (Leys (1965); Lui (1985)).[2] If paying bribes enables individuals to obtain a needed government service faster than by navigating the bureaucracy, corruption can improve the efficiency of the transaction. Note that this argument takes institutional inefficiencies as exogenous. But, as Bardhan (1997) points out, "The distortions are not exogenous to the system and are instead often part of the built-in corrupt practices of a patron-client political system." Given these relationships, we expect endogenous growth of distortionary red tape – our study speaks to the effects of corruption in those situations, not to the overall welfare effects of corruption in general.

The observational literature on the effects of corruption on growth finds mixed results. Several analyses of the average effects of corruption on growth reject the "grease the wheels" hypothesis (Mauro (1995), Méon and Sekkat (2005), Fisman and Svensson (2007)), but other papers have found support for it in settings of low institutional quality. Indeed, Méon and Sekkat (2005) observed that average negative effects of corruption on growth could belie heterogeneity by institutional quality. Corruption could promote efficiency in settings with bad institutions – consistent with the "grease the wheels" hypothesis – even as it causes misallocation in settings with good institutions. Along these lines, Méon and Weill (2010) find cross-country evidence that corruption is associated with lower efficiency costs in settings of low institutional quality. Similarly, Dreher and Gassebner (2011) find

[2] The literature offers two other reasons why corruption could increase efficiency. First, corruption could enable individuals to dodge bad public policy (Leff (1964)). If a regulation does more harm than good, and firms can bypass it by paying bribes, again corruption could enhance the economy's efficiency. As Méon and Sekkat (2005) put it: "Graft may simply be a hedge against bad public policies." Second, firms making corrupt bids for government contracts could approximate a competitive auction and outperform other allocation rules – such as government favoritism – if these bids reflect the underlying efficiency of the bidding firms (Beck and Maher (1986), Lien (1986)). These two mechanisms are not relevant in our toll setting, so we focus on the "speed money" view of corruption.

that corruption is positively correlated with firm entry in heavily regulated countries. On a more micro level, Vial and Hanoteau (2010) find that Indonesian manufacturing plants that paid more bribes also grew more during the 1975-1995 period.[3] Kato and Sato (2015) reach a similar conclusion in a study of manufacturing in India. Our paper seeks to complement this literature by providing experimental, well-identified evidence on this topic in a relevant low-capacity context.

Understanding the nature of bribe-taking transactions at tolls in Congo is also of general interest because, in many developing countries, so-called "gatekeeper states" obtain a significant portion of their revenues from taxes on the movement of people and goods inside and outside of the country. In our setting, the Provincial Government of Kasai Central gets 28% of its revenues from tolls and other taxes on transportation in the province. Such revenue-generation strategies are common in countries with low state capacity, where other taxes might be harder to enforce (e.g. Sequeira and Djankov (2014)). Thus, characterizing the strategies of bribe payers and bribe receivers in literal and figurative gatekeeper relationships is key to understanding the persistence of corruption in developing countries. In addition, as we address further in our discussion, the evidence suggests that it is difficult *ex ante* to identify the mix of coercion and collusion present in bribe-taking relationships, as without exogenous variation to incentives, simply most observed equilibriums are consistent with a wide range of possibilities.

We begin by discussing the context of our study in Section 1.2. We then present the experimental design in Section 1.3 and a simple theory in Section 1.4. We discuss our data in Section 1.5 and we then present our results in Section 1.7. Finally, we discuss the interpretation of our results in Section 1.8 and offer concluding remarks.

[3] This finding contradicts evidence from Uganda that follows an identical firm-level estimation strategy (Fisman and Svensson (2007)). Vial and Hanoteau (2010) argue that this reflects the forward-looking nature of Indonesian officials during the Suharto period and the long-term deals they struck with firms to maximize rents overtime; they characterize firms and officials in Uganda as dealing with greater uncertainty and so more likely to extract more today at the expense of tomorrow.

1.2 Context

1.2.1 Setting

The Democratic Republic of Congo ranks 147th out of 168 countries in the Corruption Perceptions Index of 2015 according to Transparency International. Our experiment takes place in Kananga, a city of roughly 1 million (the fourth largest in the country) and the capital of the Kasai Central province. In a quasi-random sample of households, Lowes et al. (2016) find that self-reported median monthly household income is approximately \$70, or \$111 at PPP. The local currency is the Congolese Franc (FC) and during the period of our study, 1,000 FC was worth \$1.00-\$1.03.

Motorcycle taxis (known as "motos") are the most common form of transportation in Kananga and the surrounding areas. Moto trips out of town take an average of one full day to complete. All routes out of the city pass by a toll station, where motorcycles must stop and pay the tax of 2,000 FC, show a receipt that they have already paid at that toll within the last 48 hours, or bribe the toll officer. Toll officers are occasionally rotated to new posts, but this occurs only rarely – during our sample period, we do not observe any re-assignment of toll officers. Due to the extremely poor quality of roads in the area, it is very difficult to avoid passing by one of these tolls stations when leaving Kananga. Paying the full toll amounts to about 13% of the median pre-toll estimated trip profit (15,000 FC). On about 85% of trips, motorcycle taxis transport passengers; on the remainder of such trips, they transport goods.

Our experiment occurred from May to September, 2016, though 94% of data collection was completed by the end of August. May to August is the dry season, when travel is most common.[4]

1.2.2 Study Population

Motorcycle taxi drivers (known as "motards") are 100% male and generally in their mid twenties. Almost all of them have reached secondary school, speaking at least some French in addition to the local language of Tshiluba. The median driver has been driving for 4-5

[4] In mid August, there was a rare outbreak of violence near the city, when a prominent local sub-chief of the predominant tribe launched a rebellion against the government, culminating in an attack on the local airport on September 23-24. He is believed to have been killed in the ensuing fighting, but the violence and deployment of soldiers reduced travel in the direction of his home territory substantially. Fighting related to this rebellion is ongoing in the area.

years.

As we discuss further in Section 1.3.1, to participate, a driver first needed to complete an interview with an enumerator somewhere in the city of Kananga. The sampling process was not explicitly random, but we believe it approximated a (partial) census of motorcycle taxi drivers. In the initial interview, drivers were invited to come to a baseline visit if they had taken at least 1 trip outside the city in the prior 2 months. Importantly, however, we did not randomize drivers into treatment or control unless they completed the baseline interview. This reduced attrition during the study considerably.

In Table 1.1 we show observables gathered during the initial interview between drivers who were invited to the baseline survey and showed up compared to those who did not show up. Those who showed up do not substantially differ in observables from those who never came to the office. This suggests that our experimental estimates are generalizable to the broader population of drivers in Kananga.

1.3 Experimental Design

1.3.1 Timing

Our experiment was conducted over three phases, as follows:

1. **Recruitment:** we sent enumerators to intersections throughout Kananga where motorcycle taxis were known to gather or pass frequently. Our enumerators stopped individual motards and administered a brief survey to gather a small number of covariates and to determine eligibility for participation in our experiment. Drivers were paid 500 FC (about $0.50) for participation in this short survey and were eligible to earn an additional 200 FC in an accompanying game. Motards were eligible to participate if they had taken a trip outside the city of Kananga in the prior 2 months. Individuals who were eligible were invited to a baseline visit on a randomly selected day in the following three weeks.[5] 1,616 drivers were interviewed, of whom 1,219 were invited to a baseline visit.

[5] We used stickers, phone numbers, and a screening question to prevent the same individuals from joining the sample twice, as well as comparing photos of individuals reporting similar names to remove people from the sample who successfully entered twice.

2. **Baseline:** any driver who was invited for and attended a baseline visit was assigned to treatment and entered the experimental sample. Drivers were paid 3,000 FC for attending the baseline visit and were eligible to earn up to an additional 1,200 FC in an accompanying game. Of the 1,219 eligible drivers, 912 came for a baseline visit and became part of the experimental sample. 1.1 shows that motards who showed up for the baseline survey do not differ substantially from those who did not.

3. **Follow-up visits (3 total):** at the baseline, the driver was asked to return in 22 days (or 23 days, if the 22nd day was a Sunday) for a follow-up visit to collect outcomes from the treatment round. During this follow-up visit, the driver received an invitation to return to the office at the completion of the next treatment round in 3 weeks (21 days). During the second follow-up visit, the driver was again invited back in another 3 weeks at the conclusion of the last treatment round. In other words, drivers completed a maximum of 4 total visits: baseline, plus 3 follow-up visits conducted at the end of each treatment round). Drivers were paid 3,000 FC for attending each follow-up visit as a show-up fee to incentivize attendance.

An illustrative timeline is shown in Figure 1-2.[6]

1.3.2 Treatments

At each of the three follow-up visits, drivers were assigned to one of four treatments, or the control group. Specifically, each driver received exactly one assignment from each of the following treatment categories, given in a random order. Within each category, drivers were randomized across treatments. For example, within the financial incentive category, every driver received either the FC1000 or FC2000 treatment, but not both.

1. **Control:**

   - *Control*: the driver was asked to bring a receipt from his next trip through a toll, but no reward was offered.

2. **Financial Incentive:**

[6] Note that for a specific driver, since the date of baseline was random, this process could be as long as 13 weeks, if the baseline visit occurred at the end of the baseline period.

- *FC1000*: the driver was asked to bring a receipt from his next trip through a toll and told he would be paid 1000 FC (50% of the toll price) for each receipt that he brought up to a maximum of 2 receipts.

- *FC2000*: the driver was asked to bring a receipt from his next trip through a toll and told he would be paid 2000 FC (100% of the toll price) for each receipt that he brought up to a maximum of 2 receipts.

3. **Social Incentive:**

   - *Charity*: the driver was asked to bring a receipt from his next trip through a toll and was told that for each receipt he brought, we (the research group) would donate 2000 FC to a home for widows in the city of Kananga, up to a maximum of 2 receipts.

   - *Government*: the driver was asked to bring a receipt from his next trip through a toll and was told that for each receipt he brought, we (the research group) would give 2000 FC to the government of Kananga to, in turn, transmit to a home for widows in the city of Kananga, up to a maximum of 2 receipts.

We discuss the power advantages of this panel design, as well as some of the assumptions required to evaluate it in Section 1.11.2.

In addition to these main treatments, we cross-randomized a social norms intervention at the individual level. For this intervention, individuals' treatment status was constant across rounds. Selected participants were told by enumerators the proportion of drivers who reported paying the full amount at the toll during the baseline survey. The text read as follows: "Now, I'm going to give you an update about information we've learned speaking to motorcycle taxi drivers in Kananga over the past 3 weeks. In particular, did you know that 62 percent of motorcycle taxi drivers paid 2000 FC to the DGRKOC at the toll in their last trip? Is that 62 percent higher, lower, or the same as you would have expected?" Selected participants were given a chance to respond to these questions. This information was repeated at the first two follow-up visits.

The hypothesis we wish to test with this intervention is that individuals participate in petty corruption partly because they believe it is the status quo. Because 62% is likely to be construed as a high rate of tax compliance in this setting – where we estimate true tax compliance at 13% – this information should surprise individuals and potentially move their priors about the prevalence of bribe payment at the toll.[7] Indeed, 80.5% of individuals in the first follow-up survey admitted that this rate of full payment was equal to or higher than what they would have expected. We thus hypothesized that this treatment will increase drivers' efforts to obtain a receipt.

Simultaneously, there were two interventions at the toll-day level. Driver recall of their date of travel was poor, which means that we have trouble evaluating these treatments (since we cannot match a specific trip to whether or not treatment was active on that toll-day), but for completeness, we describe both of them here.

1. *Enumerator presence*: Beginning one month into the study, we assigned an enumerator to spend 4 days per week at the tolls around the city (as there are 4 main tolls and 7 days in a week, the probability of a given toll being monitored on a given day was $\frac{4}{7} * \frac{1}{4} = 14.3\%$).[8] The purpose of the enumerator was to validate that drivers were taking the trips that they claimed to have taken (and were reporting all trips) by doing very short surveys with study participants as they passed the tolls. The enumerator was *not* told to monitor the toll officer, but his presence at the toll was clear and, anecdotally, officers were suspicious that he might report on them.

2. *Financial incentive for toll officer*: Beginning at the same time as the enumerator monitoring, we began offering financial incentives to toll officers to issue more receipts. Specifically, on randomly selected days, we offered the toll officers 500 FC for each receipt they issued over a threshold set at the 25th percentile of *reported* receipts in the pre-period (this excludes a small number of inframarginal receipts). The randomization was set so that an average of every 5th day was treated, but the period between treatments was randomized.[9]

1.3.3 Experimental Measures

We use experimental measures of honesty, tax morale, and patience of drivers, which are described briefly in this section.

[7] This information was presented in terms of drivers reported payments, as reported to us and thus was truthful.

[8] The enumerator did not work on Sundays, so the probability of monitoring varied between Sunday and the other days, but this provides an approximation.

[9] The initial day of treatment was randomized. Then, the pause between each episode was randomized as follows: 3 days (65 percent of the time), 4 days (17 percent), 5 days (12 percent), 6 days (6 percent).

Honesty

We measured honesty using an experimental game administered in the field by the enumerator using a tablet during the recruitment stage. The participant received the following instructions (in French or Tshiluba):

"Now, I will ask you to play a short game to evaluate your chance to win a small amount of money. In one minute, a number will randomly appear on the screen. It will be a number between 1 and 10. Before the number appears on the screen, I will ask you to think of one of those numbers in your head, but not say it out loud. If the number you are thinking about appears on the screen, you will tell me so and I will give you 200 FC [$0.20]. If a different number appears, then tell me that it was different and you will not get a reward."

The enumerators then demonstrated two examples, one in which the participant won and one in which he lost. Participants were asked whether they understood the rules of the game or needed further explanation. If the participant understood the game, then he was asked to imagine a number. Once he had imagined a number, the enumerator would then pass to the next screen, where a randomly selected number would appear. The enumerator would then ask:

"Did you guess the number that appeared on the screen?"

If all participants were honest, we would expect an average of 10% of participants to guess correctly. However, participants could also lie about having guessed the number. In practice, 38.4% of participants claimed to have won the game, suggesting a significant rate of dishonesty.

Tax Morale

We measured tax morale, defined by Luttmer and Singhal (2014) as "nonpecuniary motivations for tax compliance" using an experimental game administered at our office during the baseline visit. This game is similar to the Random Allocation Game as described in Lowes et al. (2017). In the game, the enumerator presented the participant with a six-sided die and told the participant:

"Now, I would like to play a game using this dice. Each side has a different number of dots from 1 to 6. I am going to face away from you. While I am facing away, please roll the dice 4 times in the box, so that only you can see it. When you are done, I will ask you the total number of dots that were on the dice during those 4 rolls. At the end of the survey, I will give you 50 FC for each dot that you rolled during those four rolls and we will give the government 1200 FC (the maximum) minus the amount that we give you."

The enumerators then demonstrated several examples and administered test questions to make sure that the participant had understood the rules. The enumerator then faced away or left the room while the participant completed his rolls. If all participants were completely honest, we would expect the average amount received by a participant to be 700 FC (14 being the average over 4 rolls). In practice, the average was 779 FC, with excess mass in the right tail, as seen in Figure 1-3.

This game measures the experimentally relevant parameter of "willingness to deprive the government of revenue for personal gain," since participants were informed repeatedly that any money that they did not win from the game would go to the provincial government (also the recipient of toll revenue).[10]

Anticipated Discount Factor

We measured each participant's "anticipated discount factor" using choices over money today versus at the participant's next visit. At the baseline visit and the first two follow-up visits, the participant was given two choices between an amount of money at the current moment and a larger amount of money in the future. Those decisions are summarized in Table 1.2.

Drivers were told that one of the two decisions would be selected by the tablet to occur in real life, so they were incentivized to think seriously about their preferred option. If they received money at a future visit, that amount was clearly delineated so that they understood that we were, in fact, making good on our promise to deliver the payment.

We use the term "anticipated discount factor" to emphasize that this measure combines three elements: (1) the participant's true underlying time discount factor (2) the participant's perceived likelihood of returning for his follow-up visit and (3) the participant's trust that we would deliver on our promise in the future (relative to now). The anticipated discount factor is the product of these three factors. We want to emphasize that this measure is precisely the experimentally relevant parameter. Conceptually, there is no difference be-

[10] We find much less theft in this sample than in Lowes et al. (2017), working in the same context, but with a different sub-population. There are several reasons why this could be true, but we suspect that the most important difference was that the study population came to our office for this activity, where they may have felt more uncomfortable cheating than at home as in Lowes et al. (2017). This also explains why cheating in the "honesty game" described in Section 1.3.3 was significantly higher, as that game was played outside of our compound, in the field.

tween this decision and the decision of the driver to demand a receipt at the toll in exchange for a promised reward – in both cases he is trading off a short-term benefit against a future reward, which he may discount if either (1) he values the future little (2) he anticipates *not* coming to his next visit or (3) he expects the experimenters not to honor their word. Thus, from our perspective, the decision taken in this game is highly informative about the driver's valuation of our promises of a reward – even if it is true that this game does not measure patience alone.[11]

Many drivers selected the low return options. Specifically, if we combine the 3 times that drivers made these decisions (at Baseline, at Follow-Up 1, and at Follow-Up 2), we see the matrix of realized choices in Table 1.3, where the vertical axis is for choice 1 and the horizontal axis is for choice 2.

Many drivers change their choices across rounds, so only 21.4% of drivers always choose the immediate amount (indicating a stable weekly β, bounded above by 0.63) and only 15.5% of drivers always choose to wait (indicating a stable weekly β bounded below by 0.74). Overall, these results suggest that drivers are, on average, highly impatient or present biased. We will return to this fact when we interpret the results in Section 1.8 below.

1.4 Theory

1.4.1 Basic Bargaining

This section sketches a simple bargaining model applicable to this setting to elucidate the mechanism behind the financial incentive treatments. Section 1.11.1 in the Appendix explores a possible mechanism behind the social incentive treatments.

Consider a driver and a toll agent. The driver (subscript $_i$) receives an individual return from completing the trip π distributed $F(\cdot)$. This value is the driver's return net of all costs (e.g. fuel and motorcycle rental) besides paying at the toll.[12] The driver also has an individual specific tax morale, χ distributed $H(\cdot)$, which captures the driver's intrinsic valuation

[11] Importantly, we do not seek to distinguish between hyperbolic and non-hyperbolic preferences (unlike most recent work on time preferences, e.g. Andreoni et al. (2015)). This is because in our setting, all rewards (both in this experimental game and decisions about seeking a receipt) are in the future relative to an immediate pay-off or cost, so any hyperbolic factor would always be active. Thus, it is not conceptually important to separately estimate it.

[12] We assume that the value of $\pi \geq t$ so trip completion is not affected.

of completing a legitimate transaction and obtaining a receipt, absent other incentives and a trip-specific time cost, ξ distributed $G(\cdot)$ with mean 0. This time cost will only be paid if the driver decides to wait for a receipt, so it represents the driver's trip-specific shock to the value of getting a receipt.

The driver has two potential actions $\{t, b\}$. t means demanding a receipt and paying the full tax. In this case, the driver pays an amount t and the toll agent (subscript $_a$) receives w, a piece-rate value of reporting the traffic. b means paying a bribe. In this case, the driver pays b as a transfer to the toll agent. We assume that the toll agent can always refuse a bribe offer and instead issue a receipt, but cannot refuse to issue a receipt. However, in practice, he will never do so, as, conditional on the driver's action, he is always at least as well off accepting the bribe.[13]

The payoff of the driver if he demands a receipt is given below:

$$V_i(\pi, \xi, t) = \pi - t - \xi + \chi \tag{1.4.1}$$

If the driver does not demand a receipt, his payoff is instead:

$$W_i(\pi, \xi, b) = \pi - b \tag{1.4.2}$$

The agent has a corresponding set of payoffs. If he issues a receipt, his payoff is:

$$V_a(\pi, \xi, t) = w \tag{1.4.3}$$

If he takes a bribe, his payoff is:

$$V_a(\pi, \xi, t) = b \tag{1.4.4}$$

The amount b is determined by Nash bargaining with the toll agent, who has bargaining weight δ. This yields the following simple equation for the bribe, where the agent receives his outside option w plus a δ share of the surplus generated by collusion.

$$b^* = w + \delta \underbrace{(t - \chi - w + \xi)}_{\text{surplus from collusion}} \tag{1.4.5}$$

[13] Subscripts are largely suppressed in what follows for visual simplicity.

Note that any time that the surplus from collusion is negative, the driver will demand a receipt, but any time the surplus from collusion is positive, the driver will instead pay a bribe.

Treatment Effects

We can think of any of the treatments outlined in Section 1.3.2 as being a shock $k \geq 0$ to the driver's return to getting a receipt. Then the solution to the Nash bargaining problem is a slight adjustment to (1.4.5):

$$b^* = w + \delta \underbrace{(t - \chi - w - k + \xi)}_{\text{surplus from collusion}} \tag{1.4.6}$$

Again, any time the surplus from collusion is positive, the driver will still pay a bribe. However, an increase to k will cause the share of drivers who pay bribes to (weakly) fall. The share of drivers getting a receipt, $\bar{t}$ will change as follows in response to a change in k:

$$\frac{\partial \bar{t}}{\partial k} = \frac{\partial \bar{t}}{\partial w} = m(t - k - w) \text{ where } m(\cdot) \text{ is the PDF of } \chi_i + \xi_i \tag{1.4.7}$$

To evaluate the effect on the equilibrium average bribe, $\bar{b}$, we can do a similar exercise:

$$\frac{\partial \bar{b}}{\partial k} = -\delta(1 - \underbrace{(t - k - w)m(t - k - w)}_{\text{selection effect}}) \tag{1.4.8}$$

This equation makes clear that there are two off-setting forces: (1) the direct, "causal" effect of increasing k on bribe levels, which is, *ceteris paribus*, to decrease the bribe level by δk and (2) the selection effect of removing the marginal individuals from engaging in bribery. In this model, the marginal individuals are those for whom the value of collusion is already very low – i.e. those individuals already paying low bribes – so removing those individuals causes the average bribe to rise. In Section 1.6.3 below, we explain our strategy for separating these two effects.

1.5 Data

Our data are reported from drivers at follow-up visits to our office. At these visits, we ask them about trips they have taken in the period between their last visit and the current visit

– the period during which a given treatment applies.

1.5.1 Outcomes

We focus on two main outcomes. The first outcome is presence of a valid receipt corresponding to a trip taken by the driver. In our context, presence of a valid receipt is evidence of non-participation in corruption. In all but one of the toll stations around Kananga, the toll officers are equipped with electronic receipt issuing machines.[14] These machines record all receipts issued, and the toll officers are responsible for depositing an amount of cash equal to the receipts issued at the conclusion of each reporting period (generally, each week). As a result, if an officer issues a receipt, he cannot steal money associated with that receipt. If no receipt is issued, then there is no tracking mechanism for the money and, according to the office workers who took the reports, toll officers never turned in any money that was not backed by receipts. Only 8.4% of drivers in the control group produced any receipt proving payment at the toll and the average number of receipts (in the control group), conditional on producing any receipt, was 1.14.

Our second outcome is the amount paid as a bribe. There are several ways to measure this:

- *Self-reported amount paid*: we asked drivers how much they paid at the toll. We count this amount as a "bribe amount" if they do not have a receipt – this is important to avoid conflating effects on receipts issued with effects on equilibrium bribe paid, as those effects have different policy implications. Roughly 37% of drivers report paying less than the official rate. Conditional on underpaying, the average reported discount is 62%; the median discount is 50%.

- *Box amount*: for each of the same transactions, we asked drivers to privately record the amount that they had paid on a slip of paper and to put that paper in a sealed cardboard box. The paper slips were clearly marked with an individual-specific ID code, so the drivers knew that the information was identifiable, but it removes any direct embarrassment/social desirability associated with admitting to a bribe in the presence of the enumerator.[15]

[14] The final toll station uses specialized receipt pads, which we also accepted.

[15] One key problem with this measure is that there is a very large mass point at 200 FC, which we attribute to a lack of attention/innumeracy (we believe that the vast majority of these people meant to indicate 2000 FC). As a result, we recode all responses of 200 FC as 2000 FC if the individual stated in his self-report that

- *Self-reported "arrangement"*: for each trip past a toll, we also asked drivers to report whether they had an "arrangement" with the toll officer. The term "arrangement" is used locally to refer to an extralegal agreement with the toll officer, including underpayment of the toll or, potentially, an agreement to allow the driver to avoid paying on his return trip. This is thus a coarse measure of bribery that may be less sensitive than stating an amount paid.

All of the measures are highly correlated. Since the driver's self-report has the lowest measurement error (since unusual amounts could be discussed with the enumerator to check that they reflected our preferred definition of cost), it is our preferred measure. However, we recognize the possibility of social desirability bias in reporting potentially illegal activity. In Figure 1-4, we provide a chart of bribes paid in each round. Note that this intentionally excludes all payments associated with a receipt, since paying the full amount for a receipt would not be a bribe.

The figure suggests that there may have been some under-reporting of bribes. In particular, there is a clear shift in mass from people reporting paying 2000 FC (the legal amount) to lower amounts, starting in round 3, after they have encountered the research team several times and a stock of trust has been built. However, the amount of mass that shifts is small and there is no further shift in round 4. This provides some suggestive evidence that the effect of social desirability may be small: (a) many drivers are willing to report payments that are clearly illegal (any amount under 2000 cannot be explained by claiming a lost receipt or the like), (b) the shift in reporting after repeated contact with our research team is relatively small, and (c) there is no further shift during the 4th visit, suggesting that few individuals are marginal with respect to their willingness to reveal illegal behavior. This suggests (though does not prove) that most people are honestly revealing their amount paid.

1.5.2 Covariates

As we pre-specified in our Pre-Analysis Plan, we use the following individual covariates:

- $HONESTY_i$: a dummy that is 1 if the driver reported winning the "honesty game" (as described in 1.3.3 above) in our recruitment survey and 0 otherwise

- $INITRECEIPT_i$: a dummy that is 1 if the driver brings a receipt to the baseline

he paid 2000 FC.

visit and 0 otherwise

- $OWNBIKE_i$: a dummy that is 1 if the driver reports owning his own bike and 0 otherwise
- $TAXMORALE_i$: value from 4 to 24, as described in section 1.3.3 above
- $EDUCATION_i$: a set of dummies for different education levels (no schooling, primary completed, secondary completed, tertiary completed)
- $INCOME_i$: a measurement of income at baseline[16]
- $EXPERIENCE_i$: years of experience as moto driver
- AGE_i: age in years

We can see in Tables 1.4 and 1.5 that, by design, these covariates are balanced across treatments (a combined test of joint significance across all variables and treatments (shown at the bottom of all columns) shows that there is no systematic difference across treatment and control).

In addition, as pre-specified, we use the following trip-specific covariates:

- $EXEMPT_{ir}$: a dummy that is 1 if the driver reports carrying a document or passenger (e.g. senior government official) who exempted him from paying the toll and is 0 otherwise
- $AVOID_{ir}$: a dummy that is 1 if the driver reports avoiding the toll and is 0 otherwise
- $BOSSMONEY_{ir}$: a dummy that is 1 if the driver reports being given money from his boss to pay the toll and is 0 otherwise
- $ENUMPRESENCE_{ir}$: a dummy that is 1 if the driver passes a toll on a day during which we are validating traffic and 0 otherwise
- $TRIPVALUE_{ir}$: a continuous variable that is equal to the driver's self-reported profitability of the trip
- γ_{toll}: toll fixed effect

[16] We use a measure of consumption, amount spent on cellphone airtime, in the past week, as we think it is measured with more accuracy.

- γ_e: enumerator fixed effect

The trip-specific covariates may be problematic because they are post-treatment. As a result, if treatment induces drivers to change their behavior, such as whether or not they evade the toll, then including these covariates will bias our estimates of the effects of treatment (see Angrist & Pischke (2009) for a discussion of "bad controls"). A large majority of trips (81%) are passenger directed, which means drivers are unlikely to manipulate these trip features. However, in Table 1.6, we analyze whether any of the trip-specific covariates are unbalanced across treatments and find that they are balanced.

1.6 Empirical Specification

1.6.1 ITT

Our pre-specified specification is a standard intent-to-treat regression run at the individual-by-trip-level. For an individual i on round r taking trip t, we run:

$$\begin{aligned} Y_{irt} = \beta_0 + \beta_1 FC1000_{ir} + \beta_2 FC2000_{ir} + \beta_3 CHARITY_{ir} + \beta_4 GOVT_{ir} \\ + \beta_5 TOLLPAY_{irt} + \beta_6 ENUM_{irt} + \gamma_r + \varepsilon_{irt} \end{aligned} \tag{1.6.1}$$

The six coefficients measure the causal effects of each treatment, as outlined in Section 1.3.2. We cluster all standard errors at the level of the driver to allow for arbitrary serial correlation within individual. In some specifications, we add the aforementioned vector of driver controls, $X_i'\theta$ or a driver fixed effect, γ_i. We also, as discussed above, sometimes include a vector of trip-specific controls, $X_{irt}'\phi$.

However, audit results suggest that there was not insignificant misreporting of trips and, importantly, that treatment may have affected quantities of trips reported.[17] This

[17] We believe that the difference between the control and the four treatment groups is a false positive for three reasons. First, there are no systematic differences in misreporting when we compare across the treatment groups that *drivers themselves thought they were in.* Second, social desirability bias is an unconvincing explanation of the pattern of results we observe because the control was an "active" control: drivers were asked to bring receipts, even if there was no financial or social incentive. Finally, this increase in trip-taking relative to the control implies implausibly large back-of-the-envelope trip-cost elasticities of -2.6 to -5.2. Each of these points is examined in detail in Section 1.11.4.

finding motivates an alternative specification, using intent-to-treat at the individual-by-round level. Although under the plausible assumption that unreported trips lack receipts, trip mis-reporting would bias the coefficients from (1.6.1), there would be no such bias in a regression done at the individual-by-round level, since the new regression does not condition on trips taken. The new specification is very similar to (1.6.1), but it is not possible to look at the two toll-level treatments, $TOLLPAY_{irt}$ and $ENUM_{irt}$, without looking at individual trips.[18] For completeness, we show this individual-by-round regression below:

$$Y_{ir} = \beta_0 + \beta_1 FC1000_{ir} + \beta_2 FC2000_{ir} + \beta_3 CHARITY_{ir} + \beta_4 GOVT_{ir} + \gamma_r + \varepsilon_{ir} \quad (1.6.2)$$

1.6.2 IV

We also show results in an IV framework. Unlike many RCTs, we do not have any compliance problems in terms of receiving assigned treatment: conditional on showing up to a visit, every participant received the correct, randomly assigned treatment. However, since our treatment is a promise of a reward for a given action (bringing a receipt), we have the problem that participants may not be able to *recall* their treatment assignment correctly. Participants who forget their reward assignment (or forget that any reward is possible) are unlikely to respond to treatment. This motivates the following first stage, where $RECALL(T)$ is a dummy for recalling a given treatment, T. Note that since there are 4 treatments, there are four endogenous regressors and thus four first-stage equations.[19]

$$RECALL(T)_{irt} = \beta_0 + \beta_1 FC1000_{ir} + \beta_2 FC2000_{ir} + \beta_3 CHARITY_{ir} + \beta_4 GOVT_{ir} + \gamma_r + \varepsilon_{irt} \quad (1.6.3)$$

The exclusion restriction for this regression is that treatment assignment only affects

[18] Since the timing of those treatments was randomized, causal identification of the main treatments is unaffected.

[19] The four endogenous regressors are $RECALL(FC1000)$, $RECALL(FC2000)$, $RECALL(CHARITY)$, and $RECALL(GOVT)$. Note that there are two conditions under which all four of these endogenous regressors will have a value of 0 (i.e. the respondent does not recall any of them). One is if the respondent explicitly recalls that he is in the control. The other is if the respondent does not remember his treatment status at all. This means that we implicitly assume that the answer "don't know" is equivalent to recalling that one is part of the control, rather than representing some probability distribution over the treatments. Given the experimental design, this is the most natural assumption.

outcomes through its effect on what participants believe their treatment to be. *A priori* this is likely to hold: it is difficult to imagine that assignment to a certain treatment would have effects unless it was remembered by the participant. The main source of violations to this assumption would be if treatment assignment was remembered by the participant initially, thereby affecting his actions, but then forgotten prior to the follow-up interview. We believe that this is unlikely, as we think that participants who change their actions in response to treatment are likely to recall that treatment during the interview. We paid participants an additional 100 FC (about $0.10) if they correctly recalled their treatment status to attenuate cheap talk problems.

In Table 1.7 we show the matrix of recalled treatment against treatment assignment. Correct recollections lie in the main diagonal. From this matrix, it is easy to see that treatment recall was highly imperfect, with only 40.7% of participants correctly recalling treatment, but that treatment assignment is clearly predictive of recalled treatment. In Table 1.8, we show this more formally by running the first stage separately for each outcome and we see that there is a very strong first stage across all treatments.

One clear pattern of note is that many people in the control group "recall" having been assigned to a treatment category other than control. There are two potential reasons for this. First, some participants may have been attempting to "game" the system and fool the enumerator into believing that they are part of a different group. Based on enumerator reports, we suspect this did occur occasionally in spite of the small financial incentive to discourage this behavior. Second, we learned that there was confusion about the reassignment of people in a treatment condition in round 1 to the control in round 2. In particular, some participants did not understand that they were losing their reward offer and switching to a new condition in which they were asked to bring a receipt, but would not receive any financial compensation, nor would we donate to any social cause on their behalf.[20] In this case, we would have a set of individuals who are genuinely confused about their treatment status. Note that both of the problems should be fixed by the IV strategy: the IV identifies the LATE *precisely for the population who is induced into believing they are treated by the instrument*, that is to say, by treatment assignment.

[20] When such confusion became evident after starting the second treatment round of the study, we noted this in an addendum to our PAP uploaded during the study.

1.6.3 Estimating Causal Effects on Bribes Levels

Estimating the causal effect of treatment on bribes paid is challenging for two reasons. First, as discussed in Section 1.5.1, bribe levels may be misreported; however, as noted earlier, we do not believe that this is a major issue in our setting.

Second, since treatment affects whether or not a receipt is issued, then as outlined in Section 1.4.1, treatment will affect both the pool of bribe payers (selection effect) as well as having a direct effect on the amount that they pay (causal treatment effect). We can estimate the total effect with a naive regression using amount paid (conditional on not having a receipt) as the outcome, but to separate the effects, we need to do something slightly more complicated.

We approach this problem using the following methods,.

- Fixed effects regression
  - If we refer back to Section 1.4.1, there are two components that produce selection, individual tax morale χ_i and trip-specific time-cost ξ_{it}. Since tax morale is fixed within individuals, then *if there were no trip-specific shocks*, the use of individual fixed effects would eliminate the selection effect. In the absence of trip-specific shocks, then conditional on observing two bribery incidents for the same driver (under different treatments), we can interpret the difference between the bribes paid under each condition as being the causal effect of treatment, since any individual factor (e.g. tax morale) is netted out by the within-individual comparison.
  - In the presence of trip-specific shocks, this result will be biased upwards in proportion to the magnitude of $Var(\xi_{it})$, as discussed in Section 1.4.1.
- Tobit
  - Again, in Section 1.4.1, if we take the structure of the model seriously, then we observe that a driver choses to demand a receipt if and only if the latent surplus from bribery is negative.[21] Since the equilibrium bribe is w (the outside option of

[21] If we wish to take the theory less seriously, the ideal solution would be to use a Heckman selection correction. However, finding an instrument that predicts selection into bribery, but does not (separately) affect bribe levels is very difficult in a bargaining set up. We nonetheless explore available options in Appendix Section 1.11.5.

the toll officer – i.e. his value of issuing a receipt) plus a share of the surplus, this tells us that we can think of any case in which we observe a receipt as reflecting a latent bribe that is left-censored at w (since the minimum bribe is $w + \delta(0)$. We can thus use a Tobit to account for this censoring, where we set the "bribe" paid by the driver to be w for all cases in which the driver received a receipt.

– One issue is that w is *ex ante* unknown. However, we can estimate w by looking at the lowest bribes reported in the data, which, in the model, must be equal to w. In Section 1.5.1, we can see that roughly 15% of drivers report a bribe of 0 (but no receipt), so we set $w = 0$.

1.7 Results

1.7.1 Main Results

We now turn to the main results. First, we estimate (1.6.2), which is on the individual-by-round level. The regressions in Table 1.9 use as the dependent variable a dummy for bringing a valid receipt, while the regressions in Table 1.10 use a count variable for the total number of receipts reported by the driver in a given round.[22] Both tables suggest that the $FC1000$ and $FC2000$ treatments had a small but significant effects on the probability that drivers brought receipts. In other words, the financial incentives seemed to have induced a small subset of drivers to abstain from corruption at the toll and demand that a legitimate transaction take place. According to OLS estimates, drivers in the financial incentive treatments were about 4 percentage points more likely to have brought receipts to an interview. Specifications including controls (column 2) and individual fixed effects (column 3) look quite similar. According to IV estimates, which isolate the effect on drivers who remembered their treatments, the effect is larger: 10 percentage points for the $FC1000$ treatment and $FC2000$ treatment.

In contrast, the social incentive treatments have no effect on average. Promising donations directly or via the government does not motivate the average driver to negotiate for a receipt at Kananga's tolls. This null average result suggests that perceptions about the low social value of payments to the state do not explain citizen supply of bribes. Even

[22] As noted above, the data are pooled across rounds of the experiment.

when their money would go toward spending on public goods with higher probability, drivers were not more likely to request that a legitimate transaction take place. However, when we examine participants who are children of widows, a pre-specified sub-group for which we expect a larger effect of the donation treatments, these results change. Treatment effects are larger and marginally significant for this subgroup. As a result, when we estimate ϕ (the perception of government corruption) as in Section 1.11.1, we estimate a confidence interval that covers all values between 0 and 1. In addition, the social norms treatment had no effect. Drivers do not appear to participate in corruption simply because they think everyone else is doing the same. The null effect for the social norms treatment remains true in all estimations that follow, so we drop that coefficient for simplicity.

Tables 1.11 and 1.12 consider results on the individual-by-trip level. The number of observations decreases relative to the individual-by-round analysis because many drivers reported taking no trips in the roughly three week period between office visits. The dependent variable is whether the driver brought a receipt corresponding to his (reported) trip. As noted in Section 1.11.4, these results should be interpreted with caution, due to the issues with mis-reporting.

In Table 1.11, we can effectively sign the bias and so we should think of the results shown there as upper-bounds on the true effects. Having noted that caveat, the financial incentive treatments caused a 7 percentage point increase in receipt-bringing relative to the control. As in the individual-by-round analysis, the estimated effect is essentially the same using individual controls, trip controls, and individual fixed effects. The social incentive treatments do not have a consistent effect across specifications. As indicated in equation 1.6.1, these regressions also include (i) a dummy ($ENUM_{irt}$) that equals 1 if on a given trip there was an enumerator validating traffic at the toll, and (ii) a dummy ($TOLLPAY_{irt}$) that equals 1 if on a given trip the toll officer incentive was available at the relevant toll. The coefficient on $ENUM_{irt}$ is for the most part positive, but also never significant.[23]

Next we turn to the results on equilibrium bribes. In these regressions, the dependent variable is the reported amount paid at the toll. Results are essentially identical if we use the box measure.[24] The first three columns of Table 1.12 show estimations of (1.6.1) with

[23]Drivers often had difficulty recalling dates of travel. This makes it complicated to match the toll-level treatments to specific trips and thus biases the treatment effects on both of these treatments to zero.

[24] Note that these regressions are on the individual-by-trip level; it is not possible to recreate the bribery analysis on the individual-by-round analysis.

no covariates, with individual covariates, and with both individual and trip-level covariates. In the first two columns, the coefficients on the main treatments are negative, but never statistically significant.

Then, in Table 1.13, we attempt to establish the pure causal effect of treatment on equilibrium bribes, as described in Section 1.6.3. The results vary across methods. The fixed effect regression produces the peculiar result that the two different rewards have divergent effects on bribe amounts – to rationalize this result would likely require an extensive form bargaining game where toll officers can use a costly hassling technology to separate high and low time cost drivers.[25]

The Tobit methodology suggests large negative effects on bribe payment due to the assumption that non-payment of bribes implies a negative "latent" bribe amount. This largely result comes from the fact that in the Tobit, we assume all cases with receipts have a bribe of 0 or less (left-censored at 0) and, as seen in Table 1.11, we know that receipts-per-reported-trip rise significantly. As a result, this treatment effect on receipts shows up as a causal negative effect on latent bribe. This is consistent with the model sketched above.[26]

1.7.2 First-degree price discrimination?

This section probes the bribe results further by considering to what extent driver and trip covariates predict the magnitude of bribes paid. We find weak evidence that toll officers engage in first-degree price discrimination.

In Table 1.14, we show how the amount that a driver pays at the toll changes with respect to a proxy for driver consumption/wealth (amount spent on phone credit), the driver's estimate of the revenue he earned from the trip, and the two honesty/tax morale measures from Section 1.3.3. Column (1) shows the results for the sample of transactions where the driver does not have a receipt, showing that reported bribe levels do respond to these measures. This offers some evidence that toll officers engage in first-degree price

[25] The flavor of this model would be as follows. Toll officers are not able to observe the reward promised to the driver and must pay a cost if they wish impose a time hassle on the driver (in equilibrium they mix over hassling or not). Drivers have either low or high time cost (i.e. the penalty of being hassled is low or high). Conditional on being hassled, drivers can either pay their time cost or "surrender" and pay the maximum bribe (2000 FC). All high time cost drivers will surrender, conditional on being hassled. The divergent result for the low and high reward would come from the fact that for a small reward, only low cost drivers demand a receipt, whereas for a high reward, there is a pooling equilibrium where many drivers gamble on asking for a receipt, but the high cost drivers will surrender and pay a high bribe if they are hassled. This generates higher bribes for the 2000 FC reward group, but lower bribes in the 1000 FC reward group.

[26] In the Heckman model considered in Appendix Section 1.11.5, these effects disappear, and the data appear quite noisy.

discrimination. However, we should not exaggerate this claim. The coefficient on driver revenue is small and only marginally significant. Moreover, column (2) limits the sample to transactions in which the driver gets a receipt, showing that for drivers who "exit" bribery, they face a flat cost, even when they are wealthier or their trip is more valuable. Thus, although toll officers may have some ability to set different prices of bribes for different drivers, the results are far from perfect price discrimination by a monopolist.

In addition, as predicted by the theory in Section 1.4.1, tax morale weakly predicts bribe magnitude among drivers who do not have receipts. In particular, drivers who won more in the tax morale game (discussed in Section 1.3.3) and thus have *lower* tax morale pay higher bribes. Intuitively, individuals who intrinsically value paying official taxes to the government must be compensated with a lower bribe, or they will select out of the bargaining process, pay the full tax, and demand a receipt. Conversely, those with low tax morale pay relatively higher bribes. As expected, this effect disappears for those who get receipts.

1.8 Discussion

Overall, our results indicate that the elasticity of corruption with respect to monetary incentives is negative, but relatively small. We first show that this result is unchanged when we calculate the full elasticities, but that the elasticity is sensitive to the driver discount rate, which might have important implications for anti-corruption policy. Second, we argue that drivers are fairly inelastic in their supply of corruption in this setting because bribes increase the efficiency of toll transactions. In other words, bribes appear to function as an "efficient grease" in this setting.

1.8.1 Elasticity of Corruption

We can estimate the precise elasticity of supply for bribes per trip with respect to an incentive in the following way:

$$\epsilon = \frac{\Delta\% \text{ receipts-per-trip}}{\Delta\% \text{ monetary cost of compliance}} \tag{1.8.1}$$

The numerator can be written as follows. Let $_c$ index the result in the control and $_M$ index the result under monetary incentive M.

$$\Delta\% \text{ receipts per trip} = \frac{\frac{\text{Receipts}_M}{\text{Trips}_M} - \frac{\text{Receipts}_c}{\text{Trips}_c}}{\frac{\text{Receipts}_c}{\text{Trips}_c}} \tag{1.8.2}$$

Receipts can be calculated directly from the estimates of the total increase in receipts as seen in Figure 1.10, over the control average.[27] We use the estimates from column (1), the baseline ITT specification. However, to calculate the true number of trips, we need to account for the potential mis-reporting discussed in Section 1.11.4, which we show in equation (1.8.3). Let T_x represent *reported* total trips under condition x (either treatment or control) and let U_x be the under-reporting rate (i.e. the percent of trips recorded by the auditor, but not reported in interviews) under condition x.

$$\text{Trips}_x = \frac{T_x}{1 - U_x} \tag{1.8.3}$$

We can use estimates of T_c and T_M from Figure 1.26 (column (2)). For the estimates immediately below, we assume that U_c and U_M are the same, per our logic in Section 1.11.4, but we use estimates of U_c and U_M from Figure 1.27 (column (1)) in our additional analysis in Section 1.11.4 in the Appendix.

Finally, the denominator is calculated by discounting the potential reward faced by a given driver by the amount of time between his trip and his appointment.[28] We can see this formally in equation (1.8.4) below:

$$\Delta\% \text{ monetary cost of compliance} = -\frac{M * \beta^w}{2000} \tag{1.8.4}$$

To estimate this value, we need to know β, the driver discount rate, and w, the time between when the driver passes the toll (i.e. when he makes the decision to participate in corruption or not) and his interview.[29] w is known to the driver because at the time of his

[27] Technically, we will use the constant in the regression, which is the control average, net of round fixed effects. We will generally use the constant instead of the control mean throughout for this reason.

[28] Note that here we are using the "list price" of compliance (i.e. imagining changing the price of 2000 FC) as our baseline cost measure, not the "marginal financial cost" of compliance. The "marginal financial cost" of compliance is [List price - Bribe], where the bribe that a driver would have faced is unobserved in any instance where he takes up a receipt. We focus on the "list price" both for ease of exposition and because it is a policy-relevant: our results speak to the effect of lowering the list price, which is directly in the government's control. Since the marginal financial cost of compliance is bounded above by the list price (and can be as low as 0 for people currently paying a bribe of 2000 FC), looking at the marginal financial cost of compliance would cause us to estimate that drivers are *even more* inelastic than we already estimate them to be.

[29] M is the (known) value of the incentive (either 1000 FC or 2000 FC).

decision to seek a receipt, he knows the timing of his next appointment. w represents the number of weeks between a driver's trip and his next scheduled appointment, which we can calculate by subtracting the date of the driver's next appointment from his reported date of travel.[30] β is the driver's weekly discount factor, which we estimated using the methods described in Section 1.3.3. This discount factor includes both literal time preference, as well as any other factors that affect the driver's belief that he will receive a future payoff (e.g. trust that the experimenters will honor payments in the future), which is the empirically relevant discount factor.

One issue here is that we do not have a precise estimate of driver β, due to the small number of time preference decisions that we offered. As a result, if a driver was always impatient, then all we know is that his $\beta \leq 0.63$, while a driver who was always patient has $\beta \geq 0.74$. Meanwhile, drivers that sometimes were patient and sometimes were impatient have intermediate discount factors. The median driver took a patient decision 1 in 3 times (twice out of 6 possible opportunities). This suggests a median β of around 0.63 or slightly below, but we will show sensitivity to changing this parameter for our estimates in Table 1.28.[31] We then calculate the average β^w over the entire sample (w and β are uncorrelated with treatment) for the different values of β that we consider.

Table 1.15 shows our estimated elasticities for the two treatment dummies and separately for a specification where the $FC1000$ and $FC2000$ are combined linearly.[32] We bootstrap the standard errors using 5,000 draws. Our preferred estimates are in the last column, which combines the two monetary rewards into a single estimate. These estimates suggest that on average, the elasticity is negative, but close to 0. Our 95% confidence interval allows us to reject elasticities larger (more negative) than -0.95 for a low β or smaller (less negative) than -0.45 with no discounting. Citizen supply of bribes is thus relatively inelastic in this setting.

The table also shows the importance of time preferences in the elasticity of corruption with respect to financial incentives. Since driver discount rates are very high on average, the

[30] Some drivers missed their appointments by large margins, creating a negative w. However, this is more plausibly interpreted as a very large w, since most of those drivers did not come to the office until they were sought out by our enumerators and thus may not have been considering collecting the reward at all. We drop any w past -2 days (and replace those in $[-2, 0]$ with the empirical time until appointment), though we believe this still underestimates the "true" w.

[31] β elicitation was always done prior to treatment assignment, so, by design, there is no relationship between period-wise β and treatment assignment in that period.

[32] We construct a variable that equals 1000 in the $FC1000$ group, 2000 in the $FC2000$ group, and 0 for the control group, charity treatment, and government treatment.

present value of the reward they face for compliance with the toll is much lower than the face value of the reward. This finding is policy relevant. Even if the value of abstaining from corruption is high – but that value arrives in the future – high rates of discounting may perpetuate citizen supply of bribes. Liquidity constraints and other factors contributing to discounting in developing countries may thus indirectly fuel citizens' participation in corruption.

1.8.2 Bribes as an "efficient grease"

The high discount rates we observe among drivers suggest the most compelling explanation for the limited responsiveness of drivers to monetary incentives: bribe payment may increase the efficiency of transactions in the toll setting we study. We consider two types of evidence supporting this interpretation: (1) reported duration of toll transactions when drivers demand receipts versus when they do not demand receipts, and (2) differential treatment effects when drivers do and do not have passengers (a shifter of driver time costs).

The clearest evidence that bribes increase efficiency in this toll setting comes from drivers' reports about typical transactions when they do and do not demand a receipt. Obtaining a receipt is slow due to bureaucratic procedures and the extent to which toll officers deliberately delay the process to try to extract bribes. Although the toll officers wait on the road in front of the toll, they leave their handheld receipt printers in an office located a short distance from the road. To get a receipt, drivers must park their motorbikes and walk down to this office with the toll officer. The officer manually enters on the receipt printer the name of the motard, the type of vehicle, the chassis number, and the name of the agent. Then, he handwrites these same pieces of information in a large ledger. He might ask to see their driver's license and motorbike registration documents, too. In total, this process takes an estimated 15 minutes on average. It can take considerably longer if there are many other vehicles at the toll – whom we will likely attend to first if they do not require receipts – or if there is a problem with any of a driver's documents.

On the other hand, drivers can speed up the interaction considerably by paying the toll money and not asking for the receipt. They do not even half to dismount their motorbike in this case. "To pass through the toll rapidly," one enumerator recalled, "motards prepare their money in advance, and ... then they don't ask for the receipt."[33] Drivers often do

[33]Enumerator Interview, August 3, 2017, Kananga.

not ask for a receipt even after paying the full 2000 FC to speed up the interaction. To avoid confusion, drivers and officers have a code system that communicates their intent to pay a quick bribe without explicitly saying so. "Brother, here is your coffee," was how one motorbike lessor said motards frequently communicate their intent to pay a bribe.[34] According to enumerators' estimates, paying a bribe cuts the time of the toll transaction by nearly 70% (an average of 11.25 minutes). This large difference in the time cost of passing through the toll reinforces the notion that bribes may increase efficiency of toll transactions. Moreover, given that drivers have large observed discount rates, this time savings is likely very consequential in their decisions to ask for a receipt.

Drivers mentioned in particular that it is difficult to demand receipts when carrying passengers because they are under greater time pressure. When drivers are alone or transporting cargo, they may be more willing to park their bike, enter the toll office, and complete the necessary paperwork to obtain a receipt. However, drivers are likely less willing to pay this time cost when they have passengers who are also impatient and might factor a delay into the driver's final wage for the trip. Knowing this, toll officers might choose to delay more when they see drivers with passengers in order to maximize their chances of extracting a bribe. For these reasons, we would therefore predict more muted responses to incentives among drivers with passengers compared to drivers with cargo.

Table 1.16 shows results of the trip-level regression (1.6.1) of treatment on whether or not the driver has an associated receipt in different samples.[35] In column (1) we show our baseline results as reported in Table 1.11. In columns (2) and (3), we limit our sample to trips during which the driver reported having a passenger and not having a passenger, respectively.

Responsiveness to the financial incentives is more than twice as large when drivers do not have a passenger. This finding is consistent with drivers' reports that they are more time constrained when carrying passengers and thus less willing to wait for a receipt.

In sum, the evidence suggests that citizen supply of corruption may be inelastic in this setting because bribes substantially reduce the time costs associated with a toll transaction.

[34]Enumerator Interview, August 3, 2017, Kananga. Bardhan (1997) notes that it is common to have specific words for "speed money," such as "coffee" in Kananga, noting examples from the Philippines in particular.

[35] Due to the trip non-reporting issue discussed earlier, the results here are caveated by the aforementioned problems related to trip non-reporting. Nevertheless, since all of the columns use the same specification, they contain valuable information in reference to one another.

Bribes are quite literally "speed money" Bardhan (1997). As noted in the Introduction, speeding up bureaucratic procedures is one of the principal reasons identified in the literature why corruption can in certain circumstances "grease the wheels" and improve efficiency (Leys (1965), Lui (1985)). Instead of absorbing the time cost associated with bureaucratic red tape at the toll, a bribe enables drivers to pass quickly, thereby facilitating the flow of goods and people in and out of Kananga.

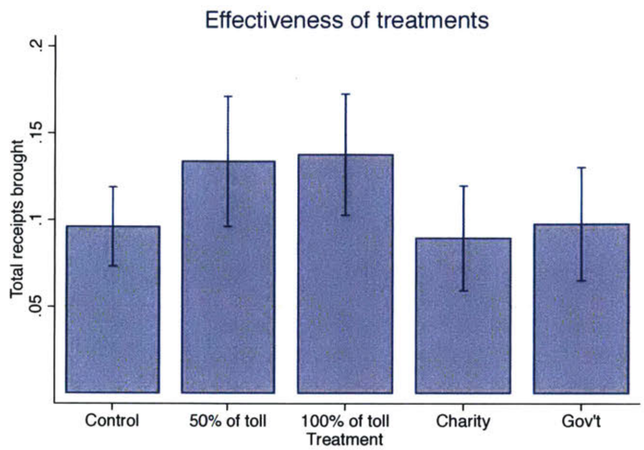
1.9 Conclusion

This experiment provides three key contributions to the corruption literature. First, it is one of the first studies to experimentally vary the returns to corruption in the field for citizens (as opposed to bureaucrats) and provides a template for future research in this area. Second, it credibly estimates the elasticity of tax compliance with respect to price. Third, it offers evidence that citizens can be inelastic in their supply of corruption in settings in which bribes increase the speed of official transactions. This is first experimental evidence in support of the "grease the wheels" hypothesis about corruption and economic efficiency.

1.10 Tables and Figures

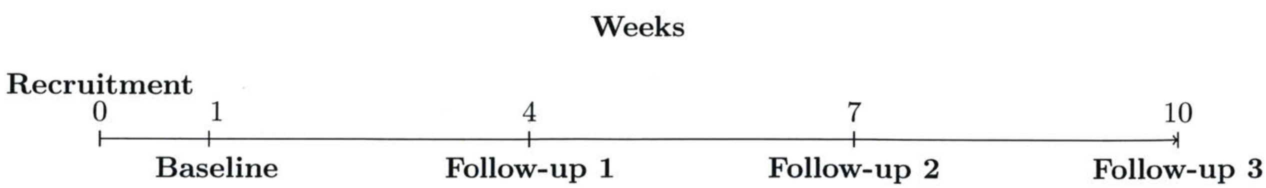
1.10.1 Figures

Figure 1-1: Effects of treatment on number of additional receipts brought each round (ITT)



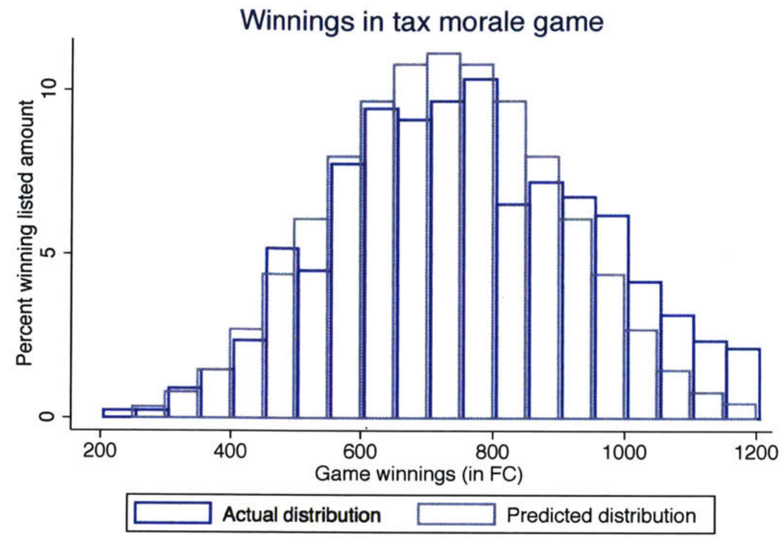
Notes: Treatment effects are estimated in a single regression using equation (1.6.2), so results include round fixed effects. The dependent variable is the number of valid (incentivized) receipts. The constant was excluded, so all results can be compared relative to the control group. Standard errors are clustered at the individual level. The confidence interval shown is a 95% confidence interval.

Figure 1-2: Illustrative timeline of the experiment



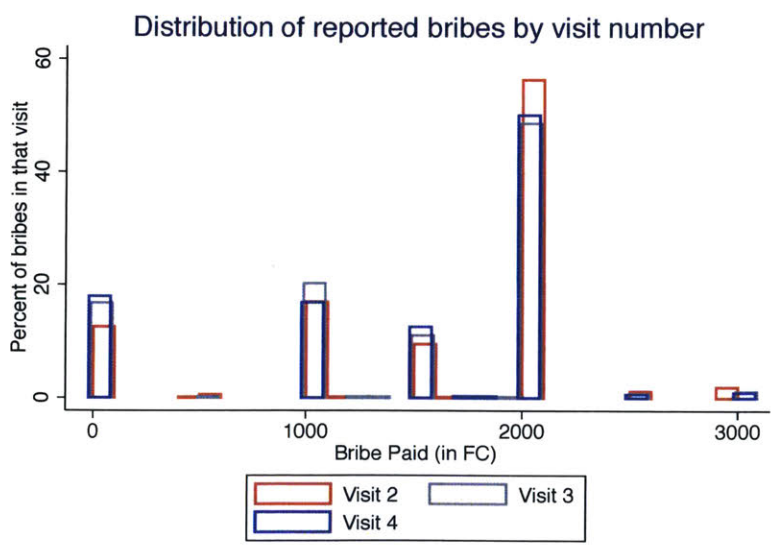
Notes: Illustrative timeline for a participant who was recruited on the first day of recruitment and requested to arrive for his baseline interview on the first day of the baseline interview period. Assigned timelines could be as much as three weeks longer than this timeline by assigning a baseline interview in week 4 instead of week 1. Gaps between follow-up interviews were always three weeks.

Figure 1-3: Distribution of realized tax morale game winnings against the predicted distribution



Notes: Dark blue indicates the empirical histogram of game winnings in the tax morale game described in Section 1.3.3. The pale blue indicates the predicted distribution that should arise by chance on average if there was no cheating.

Figure 1-4: Bribes by round (each round is jittered by 10 FC for visibility)



Notes: Empirical histogram of self-reported bribe payments (people with valid receipts are excluded from this graph). The histogram for Visit 3 is off center by 10 FC relative to Visit 2 and the histogram for Visit 4 is off center by 20 FC relative to Visit 2. Each bin is 100 FC wide.

1.10.2 Tables

Table 1.1: Summary statistics

VARIABLES	(1) In Study	(2) Out of Study
How old were you at your last birthday?	27.97	26.92
	(5.855)	(6.776)
Secondary or tertiary education	0.975	0.983
	(0.157)	(0.128)
Owns bike	0.420	0.380
	(0.494)	(0.486)
Weekly phone credit spend (in FC)	3,247	2,816
	(3,566)	(2,743)
Years as motard	5.097	4.713
	(3.029)	(2.933)
Trust in foreign researchers (1-4)	3.508	3.463
	(0.855)	(0.893)
Won dishonesty game	0.372	0.430
	(0.484)	(0.496)
Observations	866	300

Notes: This table shows summary statistics for our sample. Column (1) shows summary statistics for the population of drivers who did join the study. Column (2) shows summary statistics for drivers who were invited to join the study (based on the recruitment survey), but did not join.

Table 1.2: Choices over amounts

Choice	Immediate Amount	Amount at Next Visit	Weekly Discount Factor (β) for Indifference
1	200 FC	500 FC	0.74
2	100 FC	400 FC	0.63

Notes: This table shows the trade-offs faced faced by motards in the two different versions of the discount factor game. The first line shows one of the choices and the second line shows the other choice. The columns show, respectively, the amount received if demanded immediately, the amount received if he waited until the following visit, and minimum weekly discount factor that would justify choosing to wait for the higher amount.

Table 1.3: Empirical distribution of choices over all rounds

	Chose to Wait	Chose Immediate Amount
Chose to Wait	31.2%	10.5% FC
Chose Immediate Amount	13.5% FC	44.8% FC

Notes: This table shows the decisions made by the motards in the game. The unit of analysis is an interview round, so the results in each cell show the percentage of all visits that fell into that cell. The first row indicates that the motard chose to wait for 500 FC instead of taking 200 FC immediately. The opposite is true in the second row. The first column indicates that the motard chose to wait for 400 FC instead of taking 100 FC immediately instead of taking 100 FC immediately. The opposite is true in the second column. Since all participants faced both decisions at each visit, each visit can be allocated to one of the four resulting cells.

Table 1.4: Test of balance across individual-level controls.

VARIABLES	(1) Dishonesty game	(2) Receipt at baseline	(3) Owns bike	(4) Dice game score
Reward: 1000 FC	0.02	-0.02**	0.04**	0.16
	(0.02)	(0.01)	(0.02)	(0.14)
Reward: 2000 FC	-0.01	0.02**	-0.02	-0.11
	(0.02)	(0.01)	(0.02)	(0.14)
Reward: 2000 FC donation to charity	-0.00	0.00	0.03*	0.15
	(0.02)	(0.01)	(0.02)	(0.14)
Reward: Gov. donates 2000 FC to charity	0.00	-0.01	-0.03*	-0.22
	(0.02)	(0.01)	(0.02)	(0.14)
Constant	0.37***	0.04***	0.40***	15.78***
	(0.02)	(0.01)	(0.02)	(0.19)
Observations	2,414	2,484	2,481	2,475
R-squared	0.00	0.00	0.01	0.00
Ind. FE	N	N	N	N
Ind. Controls	N	N	N	N
Control avg.	0.374	0.0487	0.430	15.64
Joint p-value	0.205	0.205	0.205	0.205

Notes: This is a table testing balance across treatment categories. It is estimated using using equation (1.6.2), so all results include round fixed effects (not reported). The dependent variable is noted at the top of each column. The p-value from a test of joint significance of all regression coefficients across all variables tested is listed at the bottom of each column. Standard errors clustered by individual. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.

Table 1.5: Test of balance across individual-level controls (continued).

VARIABLES	(1) Education	(2) Weekly airtime spend (FC)	(3) Age	(4) Experience
Reward: 1000 FC	0.01	82.23	0.07	0.00
	(0.02)	(218.38)	(0.36)	(0.19)
Reward: 2000 FC	-0.01	-51.07	0.01	-0.03
	(0.02)	(216.27)	(0.36)	(0.19)
Reward: 2000 FC donation to charity	0.02	-88.38	0.08	0.04
	(0.02)	(215.74)	(0.36)	(0.19)
Reward: Gov. donates 2000 FC to charity	-0.03	135.98	0.02	0.04
	(0.02)	(218.68)	(0.36)	(0.19)
Constant	3.04***	3,265.73***	28.01***	5.19***
	(0.02)	(176.44)	(0.29)	(0.15)
Observations	2,481	2,481	2,481	2,481
R-squared	0.00	0.00	0.00	0.00
Ind. FE	N	N	N	N
Ind. Controls	N	N	N	N
Control avg.	3.061	3284	28.08	5.255
Joint p-value	0.205	0.205	0.205	0.205

Notes: This is a table testing balance across treatment categories. It is estimated using using equation (1.6.2), so all results include round fixed effects (not reported). The dependent variable is noted at the top of each column. The p-value from a test of joint significance of all regression coefficients across all variables tested is listed at the bottom of each column. Standard errors clustered by individual.
$^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$.

Table 1.6: Test of balance across trip-level controls.

Variable category	Test of joint significance p-value
Toll chosen	.55
Enumerator at interview	.78
Exempt from toll	.67
Trip revenue	.44
Toll allowance from boss	.18

Notes: This table tests joint signficance of all treatments within a variable or variable family (for categoric variables). Results are estimated using equation (1.6.1). Standard errors are clustered by individual.

Table 1.7: Treatment remembered versus treatment assigned

	Remembered treatment							
Assigned treatment	Control	1000FC	2000FC	Charity	Gov.	Other	DK	Total
Control	17	13	16	12	5	8	30	100
1000FC	3	58	7	6	2	8	17	100
2000FC	3	2	68	5	2	9	11	100
Charity	4	5	9	57	0	8	17	100
Gov.	6	7	14	9	30	10	24	100
Total	8	16	22	17	7	8	22	100

Notes: This table shows the treatments recalled by participants (columns) for each treatment assigned to participants (rows). All values are in percent terms, such that the total for each treatment assignment adds up to 100% when adding across all of the columns (i.e. each cell shows the percent of people who recall the treatment listed in that column, among those that are assigned the treatment listed in that row).

Table 1.8: IV first stage (each column is a regression with the outcome of recalling the treatment named at the top of the column)

VARIABLES	(1) 1000 FC	(2) 2000 FC	(3) Charity	(4) Gov't
Reward: 1000 FC	0.46***	-0.09***	-0.06***	-0.03***
	(0.03)	(0.02)	(0.02)	(0.01)
Reward: 2000 FC	-0.10***	0.52***	-0.07***	-0.03***
	(0.01)	(0.03)	(0.02)	(0.01)
Reward: 2000 FC donation to charity	-0.07***	-0.07***	0.44***	-0.04***
	(0.02)	(0.02)	(0.03)	(0.01)
Reward: Gov. donates 2000 FC to charity	-0.05***	-0.02	-0.03*	0.25***
	(0.02)	(0.02)	(0.02)	(0.02)
Observations	2,487	2,487	2,487	2,487
R-squared	0.28	0.27	0.25	0.17
Ind. FE	N	N	N	N
Ind. Controls	N	N	N	N
Control avg.	0.125	0.159	0.122	0.0462

Notes: This table shows the results from the first stage for each recalled treatment, as estimated using equation (1.6.3), so all results include round fixed effects (not reported). The dependent variable is noted at the top of each column (each is a dummy that is 1 if the driver recalled a given treatment and 0 otherwise). Standard errors clustered by individual. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.

Table 1.9: Main effects on bringing a receipt to the follow-up visit (unconditional on reporting haven taken a trip in the intervening period)

VARIABLES	(1) OLS	(2) OLS	(3) FE	(4) IV
Reward: 1000 FC	0.03*	0.04**	0.05**	0.10*
	(0.02)	(0.02)	(0.02)	(0.05)
Reward: 2000 FC	0.05***	0.05***	0.04**	0.12**
	(0.02)	(0.02)	(0.02)	(0.05)
Reward: 2000 FC donation to charity	-0.00	-0.00	0.01	0.03
	(0.02)	(0.02)	(0.02)	(0.05)
Reward: Gov. donates 2000 FC to charity	0.00	-0.00	-0.00	0.04
	(0.02)	(0.02)	(0.02)	(0.08)
Received norms information	-0.01	-0.01		-0.01
	(0.01)	(0.01)		(0.01)
Constant	0.10***	-0.16***		0.08***
	(0.01)	(0.05)		(0.02)
Observations	2,487	2,402	2,467	2,487
R-squared	0.01	0.03	0.42	0.00
Ind. FE	N	N	Y	N
Ind. Controls	N	Y	N	N
Control avg.	0.0839	0.0818	0.0833	0.0833

Notes: Columns (1)-(3) are estimated using equation (1.6.2), so all results include round fixed effects (not reported). The dependent variable is a dummy that is 1 if the driver had one or more valid receipts and 0 otherwise. Column (2) adds individual controls as described in the text. Column (3) adds individual fixed effects. Column (4) uses the instrumental variables strategy described in equation (1.6.3). Standard errors clustered by individual. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$.

Table 1.10: Main effects on number of valid receipts brought to the follow-up visit (unconditional on reporting haven taken a trip in the intervening period)

VARIABLES	(1) OLS	(2) OLS	(3) FE	(4) IV
Reward: 1000 FC	0.04*	0.04*	0.05**	0.11*
	(0.02)	(0.02)	(0.02)	(0.06)
Reward: 2000 FC	0.04**	0.04**	0.03	0.11**
	(0.02)	(0.02)	(0.02)	(0.05)
Reward: 2000 FC donation to charity	-0.01	-0.00	0.00	0.02
	(0.02)	(0.02)	(0.02)	(0.06)
Reward: Gov. donates 2000 FC to charity	0.00	-0.00	-0.00	0.04
	(0.02)	(0.02)	(0.02)	(0.09)
Received norms information	-0.02	-0.02		-0.02
	(0.02)	(0.02)		(0.02)
Constant	0.11***	-0.17***		0.09***
	(0.02)	(0.06)		(0.02)
Observations	2,487	2,402	2,467	2,487
R-squared	0.00	0.02	0.43	0.00
Ind. FE	N	N	Y	N
Ind. Controls	N	Y	N	N
Control avg.	0.0961	0.0931	0.0956	0.0956

Notes: Columns (1)-(3) are estimated using equation (1.6.2), so all results include round fixed effects (not reported). The dependent variable is a count of valid (incentivized) receipts brought by the driver. Column (2) adds individual controls as described in the text. Column (3) adds individual fixed effects. Column (4) uses the instrumental variables strategy described in equation (1.6.3). Standard errors clustered by individual. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$.

Table 1.11: Main effects on receipts per trip.

VARIABLES	(1) OLS	(2) OLS	(3) OLS	(4) FE	(5) FE
Reward: 1000 FC	0.07**	0.08***	0.07**	0.10***	0.08***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Reward: 2000 FC	0.07***	0.07***	0.08***	0.07**	0.07*
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
Reward: 2000 FC donation to charity	0.00	0.01	0.01	0.05*	0.06*
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Reward: Gov. donates 2000 FC to charity	0.03	0.02	0.02	0.05	0.02
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
Toll officer treatment in effect	-0.02	-0.03	-0.03	-0.03	-0.03
	(0.03)	(0.03)	(0.04)	(0.04)	(0.05)
Enumerator validating traffic at toll	0.06	0.05	0.04	0.02	0.00
	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)
Observations	1,723	1,659	1,418	1,510	1,262
R-squared	0.01	0.03	0.08	0.44	0.47
Ind. FE	N	N	N	Y	Y
Ind. Controls	N	Y	Y	N	N
Trip Controls	N	N	Y	N	Y
Control avg.	0.124	0.120	0.137	0.115	0.131

Notes: All columns are estimated using equation (1.6.1), so all results include round fixed effects (not reported). The dependent variable is the number of valid (incentivized) receipts per trip reported by the driver in his follow-up interview. Column (2) adds individual controls as described in the text. Column (3) has individual controls and trip-level controls, as described in the text. Column (4) adds individual fixed effects (no controls). Column (5) has individual fixed effects and trip-level controls. Standard errors clustered by individual. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.

Table 1.12: Total (selection and causal) estimates on equilibrium bribes.

VARIABLES	(1) OLS	(2) OLS	(3) OLS
Reward: 1000 FC	-17.51	-0.10	88.94
	(63.94)	(63.71)	(59.53)
Reward: 2000 FC	-94.80	-89.93	-55.08
	(66.32)	(65.95)	(61.69)
Reward: 2000 FC donation to charity	-22.42	-18.78	-40.95
	(68.32)	(68.70)	(66.92)
Reward: Gov. donates 2000 FC to charity	-38.03	-27.61	7.99
	(64.67)	(64.99)	(62.41)
Toll officer treatment in effect	8.63	16.72	-63.21
	(75.09)	(75.37)	(75.37)
Enumerator validating traffic at toll	-73.06	-72.31	-54.26
	(80.57)	(82.11)	(79.49)
Observations	1,355	1,305	1,168
R-squared	0.01	0.03	0.15
Ind. FE	N	N	N
Ind. Controls	N	Y	Y
Trip Controls	N	N	Y
Control avg.	1528	1534	1537

Notes: All columns are estimated using equation (1.6.1), so all results include round fixed effects (not reported). The dependent variable is amount of bribe (amount paid not backed by a receipt) reported by the driver at his follow-up interview. Column (2) adds individual controls as described in the text. Column (3) has individual controls and trip-level controls, as described in the text. Standard errors clustered by individual. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 1.13: Tobit results on bribes.

VARIABLES	(1) Tobit	(2) Fixed Effects
Reward: 1000 FC	-191.13**	-224.82***
	(95.29)	(61.52)
Reward: 2000 FC	-300.21***	-68.67
	(96.82)	(75.94)
Reward: 2000 FC donation to charity	-52.46	-89.26
	(95.09)	(69.90)
Reward: Gov. donates 2000 FC to charity	-126.03	-135.76*
	(97.86)	(74.70)
Toll officer treatment in effect	97.06	85.91
	(115.62)	(86.71)
Enumerator validating traffic at toll	-195.84	-53.98
	(124.83)	(97.53)
Observations	1,611	1,397
Ind. FE	N	Y
Ind. Controls	N	N
Trip Controls	N	N
Control avg.	1326	1357

Notes: The dependent variable is the bribe (amount paid not backed by a receipt) paid by the driver. In column (1), we show the results of a Tobit regression based on equation (1.6.1) where bribes are considered to be censored below at 0 and all drivers who receive a receipt (and do not overpay for said receipt) have a bribe of 0. In column (2), we show the result of a regression where we restrict the sample to drivers who are observed to not get receipts (but who report traveling) in two or more different rounds and the outcome is the amount paid as a bribe in each round – this regression includes individual fixed effects, but is otherwise the same as equation (1.6.1). Standard errors clustered by individual. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$

Table 1.14: Covariates of amount paid when getting a receipt and not.

VARIABLES	(1) No receipt	(2) Got receipt
Weekly phone credit expenditure (1000s of FC)	13.00**	0.78
	(6.16)	(1.73)
Driver estimated trip earnings (1000s of FC)	0.92*	-0.87
	(0.52)	(0.57)
Tax morale game winnings (in FC)	0.28*	0.06
	(0.16)	(0.11)
Won dishonesty game	8.16	14.25
	(61.87)	(23.40)
Observations	1,178	236
R-squared	0.03	0.06
Ind. FE	N	N
Ind. Controls	N	N
Trip Controls	N	N
Control avg.	1534	1956

Notes: All columns are estimated using equation (1.6.1), so all results include round fixed effects (not reported). The dependent variable is amount of money paid to the toll officer, as reported by the driver at his follow-up interview. Column (1) restricts the sample to trips where the driver did not have a valid receipt. Column (2) restricts the sample to trips where the driver did have a valid receipt. Standard errors clustered by individual. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$.

Table 1.15: Estimated elasticities of receipt-getting with respect to instantaneous monetary reward for various values of β. Bootstrapped 95% confidence interval shown in square brackets below each estimate.

Beta	FC1000	FC2000	Per FC
0.5	-1.71	-0.96	-0.95
	[-3.55, -0.12]	[-1.97, -0.18]	[-1.84, -0.25]
0.63	-1.37	-0.77	-0.77
	[-2.86, -0.10]	[-1.58, -0.15]	[-1.48, -0.20]
0.74	-1.16	-0.65	-0.65
	[-2.41, -0.08]	[-1.34, -0.12]	[-1.25, -0.17]
1	-0.81	-0.45	-0.45
	[-1.67, -0.06]	[-0.93, -0.09]	[-0.87, -0.12]

Notes: This table shows estimates of the elasticity of corruption (not getting a receipt) with respect to instantaneous monetary cost of compliance (i.e. cost of getting a receipt), using the calculations described in equation (1.8.1). Standard errors reflect the empirical 95% confidence interval from 5,000 bootstrap iterations. Bootstrapping was done using a block bootstrap, blocked by individual participant. The rows show the results using different assumed values of beta. Column (1) shows the results using only the effects of the 1000 FC treatment. Column (2) shows the results using only the effects of the 2000 FC treatment. Column (3) uses both treatments simultaneously (treating them linearly).

Table 1.16: Receipts per reported trip under different conditions.

VARIABLES	(1) Baseline	(2) Has passenger	(3) No passenger
Reward: 1000 FC	0.07**	0.05*	0.12*
	(0.03)	(0.03)	(0.07)
Reward: 2000 FC	0.07***	0.05*	0.14**
	(0.03)	(0.03)	(0.06)
Reward: 2000 FC donation to charity	-0.00	-0.01	0.07
	(0.02)	(0.03)	(0.06)
Reward: Gov. donates 2000 FC to charity	0.03	0.03	0.03
	(0.03)	(0.03)	(0.06)
Toll officer treatment in effect	0.00	0.00	0.06
	(0.03)	(0.03)	(0.07)
Observations	1,723	1,357	270
R-squared	0.01	0.01	0.04
Ind. FE	N	N	N
Ind. Controls	N	N	N
Trip Controls	N	N	N
Control avg.	0.124	0.138	0.0833

Notes: All columns are estimated using equation (1.6.1), so all results include round fixed effects (not reported). The dependent variable is the number of valid (incentivized) receipts per trip reported by the driver at his follow-up interview. Column (2) restricts the sample to trips where the driver reported having a passenger when he passed the toll. Column (3) restricts the sample to trips where the driver reported not having a passenger when he passed the toll. Standard errors clustered by individual.
$^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$.

1.11 Appendix

1.11.1 A Simple Theory of Social Incentives

The social incentives will not form the focus of this paper. Nonetheless, here we provide a simple theory of their intended impact. Consider an individual who places social value $\gamma \in [0, +\infty)$ on each dollar spent on a given social cause (here, donations to a home for widows). Assume that they place probability p on the research team actually making the promised payment to the widow's home. Then, the charity treatment, a promised payment of m dollars, is valued in expectation at $mp\gamma$.

Now, consider the government treatment, which has the exact same social cause, but involves transmission through the government. Now, there is an additional term, $\phi \in [0, 1]$, which represents the individual's perceived percentage of the money that actually will reach the social cause. In general, ϕ could be less than 1 either due to fraud or generic waste in government. In our setting, we believe that the parameter is best interpreted as perceived fraud, since the money is being transmitted directly as cash to the widow's home, not spent on goods or services (where waste might be a concern). This means that the government treatment, a promised payment of m dollars, is valued in expectation at $mp\phi\gamma$. Notice that the m, p, and γ parameters are all held constant, which means that comparing responses to the government and charity treatments allows us to recover perceived ϕ.

Note that this simple theory will break down if the framing of the government treatment changes the individual's utility function. For example, if the framing of the government treatment causes people to increase their belief that the experimenters are, in fact, working for the government, then that could affect the utility from the government treatment. If individuals increase their subjective view that the experimenters are working for the government by ε and place utility δ on obeying the wishes of the government, then the utility from the government treatment will be $mp\phi\gamma + \varepsilon\delta$ and comparing the two treatments will yield a biased estimate of ϕ. We believe that this is unlikely to have occurred – only 6.3% of people reported that they believed the research office was run by the government and this self-report was uncorrelated with treatment status – but we cannot explicitly rule it out.

1.11.2 Panel Structure

Discussion of the Panel Structure

As we laid out in Section 1.3.1 and Section 1.3.2, each participant in our study received different treatments in succession over the three rounds of the experiment. This design is novel: to our knowledge, ours is the first to randomize within individuals in a field setting in economics.[36] Obviously, this design is not possible for many experiments, which require a longer run period (e.g. an agricultural study that covers an entire growing period) or are built around a single event (e.g. an anti-vote buying intervention around a single election). However, a similar design may be possible for many shorter term interventions and we believe it is instructive to lay out and discuss some of the advantages and disadvantages of the panel design here.

The advantage of using a panel design is clear: increased power. Since each participant appears multiple times, total observations $T * N$ increase. Since individuals may have correlated errors over time, we cluster at the level of the participant, which means that *effective* observations do not rise at the same rate as total observations. However, the fact that we change the treatment status for each participant in each round mitigates this issue substantially. Even when clustering standard errors at the level of the individual, the fact that treatment changes across rounds for each participant creates within-participant (i.e. within cluster) variation, which means that the loss of power from clustering is small. If we were to have left participants in the same treatment repeatedly, then the panel structure would provide a smaller power advantage.[37] Note also that due to the randomized order of treatments, the experiment is equally valid with and without individual fixed effects. Fixed effects will improve power if the individual error component is large relative to the individual-by-round error and we observe multiple observations for a sufficient portion of

[36] Similar designs have been used in some lab experiments and in medical experiments (see for a review). The most similar field experiment design to our knowledge is Banerjee et al. (2007) on the Balsakhi program in India, but individuals in their study never experience multiple different treatments.

[37] In a simple three-period, two-treatment simulation where 30% of the variation is constant within person and 70% is randomly assigned each person-period, power is 8 percentage points higher when changing treatment each round (i.e. C-T1-T2) versus keeping each individual in the same treatment twice (i.e. C-T1-T1 or C-T2-T2) and 34 percentage points higher than a design in which individuals are in the same treatment/control category across all three rounds (i.e. C-C-C vs. T1-T1-T1 vs. T2-T2-T2). The power advantage of changing treatment each round is even larger (about 19 percentage points) when we add individual fixed effects (and individual fixed effects are not possible in the design where treatment status never changes).

the sample and they will worsen power if the opposite is true.

The disadvantages of using a panel design are potentially less clear, but we discuss them in more detail here. The main concern is inter-temporal spillovers. If treatment effects persist across periods, then this could create bias in our estimates. These spillovers could take several forms. First, particularly concerning for our design, participants could be confused about changes in their treatment status – since our treatment is a promise of a certain reward (or lack thereof), if participants do not realize that their treatment has changed, they will continue to respond as though in their prior treatment group, biasing our results. One sub-group in our experiment was particularly affected by this concern, which we discuss in Appendix 1.11.2 – our results are robust to removing this group. Second, participant behavior in one round (induced by treatment) could have persistent effects in future rounds, even if participants fully understand that their treatment status has changed. For example, if relationships between drivers and toll officers are important, and demanding a receipt affects those relationships, then a prior treatment that changes driver behavior could have persistent effects through its effect on the driver-toll officer relationship. Third, we could be concerned that there is a differential effect of treatment after a driver has previously been exposed to a different treatment (e.g. a disappointment or surprise effect). We explore these two effects in detail in Appendix 1.11.2 and find that they do not appear to create significant bias in our estimates of contemporaneous treatment effects.

Follow-Up Visit 2 Controls

As noted in Section 1.6.2, there was an issue in Follow-Up Visit 2 where participants who were assigned to the control did not understand that they were being removed from their prior treatment. The issue was that participants who were in the control in that round were asked to bring receipts (with no mention of a reward), but we did not emphasize that they would not receive any other reward, no matter what they had been told in the previous round. As a result, we believe that many participants in the control group believed that they were still in their old treatment category. We corrected this for Follow-Up Visit 3, but we will use this section to explore the implications of that implementation issue.[38]

Table 1.17 shows the difference in treatment recall for the affected round versus the other two rounds (pooled together here, though when we look at them separately, they look quite

[38] We pre-registered this issue once it was discovered part of the way through the experiment.

similar to one another). There are two important points to notice from this chart. The first is that the issue of people in the control incorrectly identifying themselves as being part of treatment is much bigger in Follow-Up Visit 2 than in the other rounds. In total, in Follow-Up Visit 2, a full *72 percent* of respondents in the control group identified themselves as being part of one of the four actual treatments versus only 32 percent in the other visits. The second is that in the treatment groups, this problem essentially disappears. Averaging across the 4 other treatments, in Follow-Up Visit 2, an average of 50 percent of people correctly identify their actual treatment. In the other two visits, an average of 54.75 percent of people correctly identify their actual treatment, an economically and statistically identical number.

In Table 1.18, we examine this issue in more detail by focusing on the control groups, identified earlier as the most likely place for confusion. In this table, we restrict the sample to people who were assigned to the control in Follow-Up Visit 2 (top panel) or in Follow-Up Visit 3 (bottom panel). Thus, if there were no confusion issues (or misremembered treatments), we should see all of the mass in the first column (i.e. everyone should remember that they are in the control). We then compare their recalled treatment to the treatment group they were assigned in the prior (not current) round. I.e. for Follow-Up Visit 2, we compare their recalled treatment at Follow-Up Visit 2 to the treatment they were assigned at Follow-Up Visit 1.[39] If it is in fact true that in Follow-Up Visit 2, people did not realize that they were no longer assigned to their prior treatment (and instead were in the control), then we should expect to see many more people match their remembered treatment to their former (but not current) treatment status. This is exactly what we see. For example, at Follow-Up Visit 2, 74% of people who were assigned to the FC1000 treatment at Follow-Up Visit 1 (but are in fact in the control) "remember" that they are in the FC1000 treatment – but that same quantity is only 38% at Follow-Up Visit 3. Averaging across treatments, 59 percent of the control group in Follow-Up Visit 2 incorrectly believe they are keeping their old treatment versus only 34 percent in Follow-Up Visit 3.

These facts suggest that our anecdotal information about the confusion with the Follow-Up Visit 2 control group is correct. We would expect this confusion to bias our results downwards, since if much of the control group believed that it was treated then they would

[39] Since no one was assigned the same treatment multiple times, and we are restricting to people who were actually assigned to the control in the current round, there is no one assigned to the control in the prior round.

behave as if they were treated, pushing the difference between control and treatment towards zero.

In Table 1.19, we show the results from running specification (1.6.2). In column (1), we show our baseline results, as seen in Table 1.11 column (1). In columns (2) and (3), we remove the entirety of Follow-Up Visit 2, without and with individual fixed effects, respectively. In columns (4) and (5), we just remove the control group for Follow-Up Visit 2, which we believe is the most reasonable way to deal with this issue. As we anticipated, the results are larger, though not statistically significantly different from the baseline results.

Path Dependence and Reputation

One area of additional substantive interest is the effect of the panel structure, with repeated exposure to different treatments over the course of the experiment. The treatment history of a given participant could have effects through three channels. First, the panel structure could result in respondent confusion. If respondents do not understand that they are changing treatment categories, then there will be a mechanical persistence of treatment effects.[40] Second, prior treatment exposure could affect participant behavior through an information or learning channel. If prior exposure leads participants to change their behavior at the tolls, then they may learn new information about the ease of getting receipts or the possibility of changing bribe levels that could be persistent. Third, and relatedly, prior treatment exposure may have persistent effects through a reputation channel. If treatment induces a change in driver behavior, even for a single period or interaction, then it may persistently affect their relationship with a given toll officer. The direction of this effect is not *a priori* clear. A toll officer could hassle them more in the future as punishment or become more acquiescent to their demands for a receipt (as he learns that they are a more honest type than he believed).

In Table 1.20, we estimate a modified version of (1.6.2), in which we regress the number of valid receipts brought in a given round on treatment in that round. In column (1), we replicate our baseline results, pooling across all rounds, as in column (1) of Table 1.10. In column (2), we use only the data from the first follow-up visit (where there cannot be any history effects, since it includes only the original treatment assignment). In column (3), we add a set of controls for the treatment history of the participant. In particular, we code up a

[40] We explore a particular case of this in Section 1.11.2.

set of dummy variables that have a value of 1 if the participant has *previously* been exposed to a given treatment and are 0 otherwise (in follow-up visit 1, all of these variables are 0, since the participant has not been exposed to any previous treatments). Since the order of treatments was randomly assigned for each individual, these dummies are valid, exogenous regressors to include.

In column (4), we add a control for whether the driver has ever brought a valid receipt to any previous interview. Note that this control combines the driver's type (honest or not) with any "treatment effect" of previously being induced into getting a receipt by treatment. In columns (5) and (6), we attempt to disentangle these two margins by using the same treatment history variables from column (3) as instruments for having previously gotten a receipt. Note that this will only be valid under the strong assumption that prior treatment affects current period behavior *only* through its effect on having past receipt-getting. This would require us to rule out the possibility of respondent confusion or any reputation effects that do not operate through successfully getting a receipt (e.g. if a participant demanded a receipt, but failed to get one, then that could affect future interactions with that toll officer, even though it would not show up in the instrument). We do not believe that this assumption is necessarily likely to hold, but we include it as an additional data point.

Overall, we believe that this exercise provides a number of interesting insights for the reader. First, the main (current period) treatment effects are relatively consistent across periods, no matter the specification. This suggests that our main effects are relatively unaffected by the panel structure of the experiment, which may be of interest to other experiments designing future interventions. One exception to this is that the results using only the data from the first follow-up are somewhat stronger than the results when we pool all rounds. Second, there is some evidence that prior treatments have a persistent effect on behavior, though the point estimates are generally insignificant.[41] Finally, there is some suggestive evidence for a persistence channel through past receipt-bringing, but this evaporates when individual fixed effects are added in column (6). This suggests to us that persistence is unlikely to be first-order in our context.

Next, we show several results designed to dig deeper into the issue of reputation. The earlier analysis explored the potential long-run effects of changing a participant's reputation

[41] Note that some of this effect is likely due to the confusion among members of the follow-up visit 2 control group as discussed in Section 1.11.2.

through prior treatment status. Here, we explore in more detail the potential effects that *forward-looking* reputation concerns might have on treatment effects. In particular, we might think that drivers with more exposure to a given set of toll officers would be less inclined to take up treatment if treatment affected their reputation. Since the treatment is temporary, but relationships presumably last for many periods, drivers might be unwilling to risk their reputations if they believe that they will continue to have significant exposure to a toll officer after the treatment period.[42] We do not have data on driver expectations about which tolls they are likely to frequent in the future, but we use data on self-reported past history of toll usage as a proxy.[43] We show the results using this proxy in Table 1.21.

In column (1), we repeat our analysis from column (1) of Table 1.11, using specification (1.6.1).[44] In column (2), we restrict our sample to the set of *trips* where the driver reported having visited that toll at least once in the two months before our study period began. In column (3), we use the same sample, but now we repeat the specification from column (5) of Table 1.20, in which we use prior treatment as an instrument for past receipt bringing. These results indicate that forward-looking reputation concerns are unlikely to be important for explaining the relatively low responsiveness of participants to treatment in our experiment. In particular, comparing columns (2) and (3), we see that, if anything, drivers visiting tolls that they have frequented in the past (and are likely to continue to frequent in the future) are *more* responsive to treatment than the population as a whole, though the difference is not statistically significant. Likewise, when we repeat the analysis looking at the effect of having been previously induced into bringing a receipt in column (3), we see that the point estimate on prior receipt is smaller and insignificant compared to its analogue in column (5) of Table 1.20. Again, this suggests that forward-looking receipt effects are small – if permanent relationships were heavily affected by participants demanding receipts, we would expect this effect to be larger.

[42] As discussed in Section 1.2.1, toll officer identities are constant during the period we study and tend to remain so for long periods.

[43] We validate this proxy by regressing the toll passed by the participant on his self-reported history. For each of the main tolls, prior history is a statistically significant and economically meaningful predictor of toll chosen, suggesting that prior history is a valuable proxy for the driver's expectations about his future interactions.

[44] We do not show the coefficients on enumerator presence and the toll officer treatment for visual simplicity.

1.11.3 Robustness and Additional Analysis

Attrition

In this section, we consider the issue of attrition. Table 1.22 contains the results of regressions in which the outcome was a dummy that equaled 1 if a driver failed to show up for the following visit. For example, in the first column, drivers receive a 1 if they attrited between rounds 1 and 2 (i.e. between the baseline visit and the first follow up visit). The results suggest that the charity and government treatments *decreased* the probability of attrition between baseline and follow-up visit 1 relative to the control. Likewise, the 1000 FC reward treatment, the 2000 FC reward treatment, and the government treatment *decreased* the probability of attrition between visits 2 and 3; and the government treatment *decreased* the probability of attrition between visits 3 and 4. Note that the sample changes slightly across columns, as, for example, in order to be observed in column (2), one cannot have attrited between rounds 1 and 2. Finally, in column (4), for the population of drivers who do not attrit, we regress the number of days between appointments on treatment status. If compliance with our protocol was perfect, we would have an average number of days between appointments of exactly 21. Instead, drivers take an average of 23 days, but there is no statistically or economically meaningful relationship between treatment and days between appointments, which suggests that we do not need to be worried about any effect of having more time to take trips on outcomes.

Overall, the most concerning of these findings is that in each of the three rounds, the government treatment appears to have had an effect. The direction of this finding is perhaps surprising, as it suggests that drivers in the government treatment were least likely to attrit (indeed, the coefficient is roughly 100% of the control mean, suggesting almost no attrition in this group). Ex ante, we believed that individuals in the government treatment would be most likely to attrit, over concerns that we (the experimenters) were involved with the government and might arrest drivers who lacked certain documents. However, the opposite appears to have been the case. One theory is that drivers in the government treatment may have believed that they would be found by the police or otherwise sought by the authorities if they did not return, but we cannot know for sure.

However, despite the slight evidence of differential attrition across treatment groups, overall attrition across rounds was not large: 3% between baseline and follow-up 1, 2%

between follow-ups 1 and 2, and about 5% between follow-ups 2 and 3. Overall, 87.6% of all baseline attendees completed all 4 visits. In our analysis, whether we add individual fixed effects, which means that all comparisons are within individual, or we can restrict to the set of individuals who attended all visits, there is neither a statistically or economically meaningful effect on the coefficients of interest.

Attempts to Get Receipt

In this section, we elaborate on the possible coercive power of the toll agents and the role of our treatments in encouraging "receipt-seeking effort" in addition to actual receipts gotten. In particular, we are interested in whether the elasticity of this effort may be greater than that of the actual success-rate, which would provide further evidence for the partial coercive power of the toll officers.

We do not have a perfect measure of receipt-seeking effort. However, for any trips in which people did not have a receipt, we asked people why.[45] Table 1.24 shows that 9 percent of respondents reported asking for a receipt, but having the agent refuse to grant them one.[46] Since this data is self-reported, we treat this as only suggestive evidence of driver effort. Drivers may feel social desirability bias to provide an excuse as to why they lack a receipt, even if the truth is that they did not pay the toll. However, we believe that there is likely still some signal in this measure, as we show in more detail later.

In Table 1.23, we use our trip-level regression specification, equation 1.6.1. Note that this means that we are conditioning on an individual reporting the trip to us. In column (1), we estimate our main effects, but restricting to a sample that includes only (a) trips with receipts and (b) the first trip without a receipt. Note that these estimates are very similar to those in Table 1.11, even though the sample is somewhat different. In column (2), we look only at the distribution of excuses (ignoring trips with receipts) and observe a positive, but insignificant effect of treatment on reporting trying (but failing) to get a receipt. Note, however, that this specification will treat as missing data any drivers who report only a sole trip, for which they have a receipt, and it is contaminated by selection bias. If treatment induces drivers to demand receipts more intensely, it could plausibly both induce drivers to

[45] Importantly for the estimation that follows, we only asked people once for their excuse, even if they had multiple trips without a receipt. Thus, in the estimation that follows, we restrict our sample to trips where the driver has a receipt *or* to the first trip for which he lacks one.

[46] We can also include the 2.5 percent of respondents who said that the machine was broken or out of power, both of which may be excuses used by agents who refuse to issue receipts, but the results are similar.

go from not demanding receipts to demanding them (but failing) *and* to go from demanding receipts (but failing) to demanding them more stridently and succeeding. This latter effect would cause us to underestimate the effect of treatment on effort in column (2). Thus, in column (3), we define a dummy that is 1 if either the participant brought a receipt or reported requesting one, but being denied.

Overall, these results again suggest that there may be coercive power for toll officers, but it is limited. The point estimates in column (2) are positive for the financial incentives, but, notably, we cannot reject that the coefficients in columns (1) and (3) are the same. If there was a substantial margin on which toll officers were able to resist demands from drivers for receipts, we would expect that incentivized drivers would be much more likely to report having requested receipts (due to their desire for the reward), but having been denied. These results are, however, only suggestive. It could also be the case that, in equilibrium, there are few unsuccessful demands in part because drivers know that toll officers are likely to resist.

Receipt Loss

One concern with interpreting our estimates as an elasticity of corruption with respect to incentives would be that we systematically mismeasured our outcome. If we measured our outcome equally accurately in both treatment and control (even with noise), this would not bias our estimator – but if our mismeasurement was not centered at zero and was more severe in treatment, this would be a serious concern. One obvious way this could occur would be if drivers commonly lost their receipts. Loss of receipts is not zero centered (no one "accidentally finds" receipts, so there are no positive shocks) and, importantly, one can only lose a receipt if it was requested in the first place. This suggests that lost receipts could be significantly more severe for our treatment groups, which are more likely to demand receipts (as we know from Section 1.7.1). Thus, if drivers commonly lose receipts, this could substantially bias our results towards zero.

Anecdotally, we do not believe that losing receipts is common. Drivers generally store the receipts with their money. Since there is no evidence that drivers are careless in potentially losing bills (even low quality Congolese francs close to disintegration), we believe that they would treat their receipts with the same care. Indeed, the receipts we received had often been kept for extended periods, as measured not only by the date of the receipt, but also by the wear-and-tear on the receipt itself. As a result, we believe that the vast majority of reported "loss" of receipts by participants is not truthful.

However, we also test directly for evidence that drivers are losing receipts. In Table 1.24, we show the reasons that drivers report for not having a receipt, *conditional on reporting having taken a trip*.[47] First, by far the most common reason for missing a receipt is self-reporting having underpaid the toll. The next most common reason is having left one's receipt at home, followed by reporting losing the receipt. Any driver who left their receipt at home could have fetched the receipt and brought it for whichever reward to which they were entitled. The fact that none did is already suggestive of the fact that most of these reports are likely not truthful.

Nevertheless, we also test more directly for evidence of lost receipts in Table 1.25. In column (1), we estimate our equation (1.6.1), restricting to the sample of (a) trips with receipts and (b) the first trip without a receipt, which shows that the financial incentives induced participants to bring more receipts. In column (2) we regress our treatments on a dummy variable for reporting having lost one's receipt. If losing receipts were an issue, we would expect that the treatments that show strong treatment effects in our main tables would also predict more people reporting having lost their receipts. In column (3), we construct a second dummy variable for either reporting losing one's receipt or bringing a valid receipt. As expected, the treatment effects in column (3) cannot be distinguished from the effects in column (1).[48] This does not prove that no receipts were lost. However, it tells us that the magnitude of any "lost receipt" effect must be small, since otherwise we would expect to see an identical pattern to treatment (since only receipts gotten in the first place could be subsequently lost, something that only treated individuals are likely to do).

1.11.4 Trip Misreporting

As discussed briefly, drivers do not report their trips with perfect accuracy. We explore possible consequences of trip reporting errors in this section. Table 1.26 shows the results with a dummy variable indicating the driver took at least one trip (column 1) and a count variable of the total trips taken (column 2) on the left-hand side, regressed on the treatment dummies. Negative coefficients on the government treatment suggests that either individuals

[47] Note that since this specification requires having reported a trip, then even if there is some under-reporting of trips, this likely overestimates the share of people reporting having lost their receipts.

[48] If anything, the effects are somewhat smaller for the financially incentivized groups, which is not surprising, if the receipts were more valuable to them and thus were better cared for.

in this treatment group went on slightly fewer trips or, perhaps more likely, they reported taking fewer trips. Although there is no reason that drivers in the government donation treatment group should have feared admitting to trips, it is possible that any mention of the government induced concerns among drivers, leading them to underreport trips. This measurement error would likely bias the estimated effect of the government donation treatment on receipts-per-trip upward, if we assume that unreported trips lacked receipts, as seems plausible. This does not meaningfully affect our results, since we did not detect a treatment effect for this treatment arm.

Another way to test for underreporting of trips is to consider the set of trips validated by enumerators at tolls. As discussed in Section 1.3.2, the validation was done by an enumerator stationed at the toll who flagged down participants (who had been warned to expect this possibility) and completed a very short survey with them. Because we started validating trips after a month and we only validated four toll-days per week, the set of validated trips is small: only 169 observations.[49] However, it is still possible to see whether the treatments appear to have affected misreporting among this subset. More specifically, Table 1.27 shows the results from a regression of a dummy variable for unreported trips on the treatments (restricting the sample to the 169 validated trips).[50]

The first column suggest that under-reporting is lower in the control group relative to *all four* treatment conditions. A joint F-test that the coefficients on the four treatment indicators are different from zero has a p-value of 0.096 using standard inference or of 0.13 using randomization inference.[51] We believe that the difference between the control and the treatment groups is likely a false positive for several reasons.

First, as shown in column 2 of Table 1.27, there are no systematic differences in misreporting when we compare across the treatment groups that *drivers themselves thought they were in.* Driver recall is obviously endogenous. However, if drivers were strategically underreporting, then using the endogenously recalled treatment should *strengthen* the "effect" of treatment on misreporting – precisely those drivers who believed themselves to be treated

[49] 203 driver-trips were observed, but only 169 could be matched to the sample on driver ID number, driver name, or phone number.

[50] Trips are considered unreported if, at his follow-up interview, the driver reported no trip within 7 days of the date of the validated trip (when the enumerator completed a short survey with a given motard at the toll). The 7 day window is used to account for the relatively poor recall of drivers about their exact date of travel.

[51] With a small sample (169 trips), the assumptions behind asymptotic normality may not hold, so randomization inference provides robust p-values in this case.

should be the ones under-reporting trips. However, we do not see that effect here – instead, there is no effect using recalled treatment instead of assigned treatment.[52] Related to this, in column 3, we restrict to the sample of drivers who specifically reported that they did not remember (at all) their treatment status. This group is small (only 31 trips), but the results are suggestive. Despite being the group of drivers who should be *least* likely to respond to treatment, by virtue of their total uncertainty about their treatment, the "treatment effects" in this population are very large and indeed are larger (significantly so, for some treatments) than those for the population as a whole.

Second, one might imagine that drivers were embarrassed about admitting to trips for which they did not get receipts. However, such embarrassment would exist in the control group, too, given that it was an "active" control: enumerators asked drivers in this group to bring receipts, using the same wording they used with drivers in other treatment groups (absent language about receiving a reward). Holding constant experimenter demand effects was precisely the objective of this active control. Moreover, it makes little sense why they would try to hide trips that had already been recorded by the research team.[53]

Finally, further evidence comes from economic theory. If the effects estimated in column 1 of Table 1.27 are not false positives, then this implies that each of the four treatment groups increased the total number of trips by around 20% (since reported trips in Table 1.26 are essentially the same in treatment and control). This is implausible for two reasons. First, if the treatments induced more trip taking, it must have been because the promised inducements were decreasing the effective cost of trips for drivers. But this can only be true if they anticipated bringing receipts and thus obtaining the incentives we offered.[54] Given the low rate of receipt-bringing, it is difficult to imagine that such a cost calculation could explain a 20% increase in trip taking. For example, drivers brought about .07 additional receipts per reported trip as seen in Table 1.11 when in either the FC1000 or FC2000 treatment group, on a base of .124 receipts. Even without any discounting of the reward (which we explore below), an expected return of 1000*.194 = 194 FC (or even 2000*.194 = 388 FC) is small. For reference, the cost of a trip from the center of town to one of the tolls

[52] This stands in stark contrast to the results of a regression of bringing a valid receipt on endogenously recalled treatment – recalled treatment strongly predicts receipt-bringing for all treatments – suggesting that this measure has valuable information content.

[53] Another interpretation is that drivers were confused about the validation survey, thinking that if they completed a survey at the toll they were not supposed to report it again at the office. However, it is difficult to explain why this would not have also applied for the control group.

[54] Or if there were large effects on average bribe paid, for which we do not find strong evidence.

usually costs about 5000 FC. Thus, for even a 388 FC cost difference to generate 20% more travel would imply an implausibly large back-of-the-envelope trip-cost elasticity of -2.6 (or -5.2 for the FC1000 treatment group).

Additional evidence for this theory comes from the social treatments (Charity and Government). The estimated misreporting rate is roughly the same across all four treatments. However, we find no effect from the Charity and Government treatments on receipt-bringing, so it is difficult to understand how these treatments would be causing drivers to take more trips, even if drivers did value the donations to the widow's group. That all four treatments groups have the same rate of estimated misreporting suggests that the lower rate of misreporting in the control group is more likely than not an artifact of noisy data. However, we wanted to flag the issue in the interest of maximal research disclosure. See Section 1.11.4 for more details and estimates of the elasticities (see Section 1.8.1 for calculations) when we assume that the misreporting is real.

Corruption Estimates Under Trip Misreporting

Even though we have reason to believe that the different rates of estimated trip misreporting in the treatment groups relative to the control is likely a false positive (for reasons explained above), in this section we consider its implications if true. If we assume that all unreported trips lacked receipts, then trip-level estimations will bias the effect of treatment upwards. This is because our results suggest that treated individuals were less likely to report their trips, which lowers the denominator of any regression with an outcome of receipts-per-trip. Thus, ignoring the unreported trips would lead us to conclude that the receipt-bringing rate in the treatment groups was higher than it actually was.

The best strategy to deal with this bias is to conduct our analysis on the individual-by-round level instead of the individual-by-trip level, as it is still well-defined to look at the effect of incentives on receipts-per-interview-period without controlling for or normalizing by the number of trips reported, as reported in Table 1.10 in the main text. We also then inflate the number of trips reported by our estimates of the rate at which the treatment groups tended to underreport trips, as reported in Table 1.27 in the main text. Together, these strategies allow us to account for trip non-reporting in the elasticity calculations from Section 1.8.1, which we summarize in Table 1.28.

As we might have expected from our discussion earlier, these results are much closer to zero (and indeed can neither statistically nor economically be distinguished from zero). This is because, by taking the under-reporting results at face value, we end up with a rate of receipts per "true" trip that is indistinguishable between the control and treatment groups. Since the rate of receipt bringing is then unchanged, it is unsurprising that we observe an elasticity of zero. These results suggest that the responsiveness of corruption to financial incentives is even more inelastic than we estimated earlier, which only further deepens the puzzle. Indeed, these results suggest that driver's high value of time, combined with the partial coercive power of toll officers means that corruption may be even more important as a way to "grease the wheels" of this bureaucracy.

1.11.5 Heckman selection correction

As noted in Section 1.6.3, the Tobit methodology we employ in the body of the paper takes the model very seriously. A more general solution would be to use a Heckman selection correction with an instrument that predicts selection into bribery, but does not (separately) affect bribe levels. However, it is difficult to imagine a predictor of bribery that would not also affect bribe levels because the decision to bribe rather than get a receipt is directly a function of expected returns to each action under any bargaining framework.

Notwithstanding this concern, in this section, we consider a Heckman selection correction using the presence of the enumerator at the toll on the day of travel as our instrument. The logic for this instrument is that bribery happens outside the view of the enumerator, in a small hut by the toll. The presence of the enumerator at the toll may provide some impetus on the part of the driver or toll officer to issue a receipt; but, conditional on agreeing to seek a corrupt bargain, the negotiation happens outside the view of the enumerator, who thus does not have a direct impact on the outcome. We present this result not as an ironclad solution to the selection issue, but rather another attempt –along with the fixed effects and tobit strategies discussed in section 1.6.3 –with the two other methods to provide a broader picture of the possible effects.

1.12 Appendix Tables and Figures

1.12.1 Appendix Tables

Table 1.17: Treatment remembered versus treatment assigned, as a percent of respondents assigned to that treatment group, in Follow-Up Visit 2 (top panel) and in the two other rounds (bottom panel)

	Remembered treatment							
Assigned treatment	Control	1000FC	2000FC	Charity	Gov.	Other	DK	Total
Control	4	20	26	18	8	8	16	100
1000FC	3	55	5	11	2	9	15	100
2000FC	3	2	68	5	4	8	10	100
Charity	5	7	14	52	0	8	14	100
Gov.	4	7	15	12	25	12	24	100
Total	4	19	25	20	8	9	16	100

	Remembered treatment							
Assigned treatment	Control	1000FC	2000FC	Charity	Gov.	Other	DK	Total
Control	23	9	11	9	3	7	37	100
1000FC	3	60	7	3	2	7	17	100
2000FC	2	3	68	6	1	9	12	100
Charity	4	4	6	59	0	8	19	100
Gov.	7	7	14	7	32	9	24	100
Total	10	15	20	16	7	8	24	100

Notes: This table shows the treatments recalled by participants (columns) for each treatment assigned to participants (rows). All values are in percent terms, such that the total for each treatment assignment adds up to 100% when adding across all of the columns (i.e. each cell shows the percent of people who recall the treatment listed in that column, among those that are assigned the treatment listed in that row). The top panel shows the results only for Follow-up Visit 2, while the bottom panel excludes Follow-Up Visit 2.

Table 1.18: Treatment remembered versus treatment assigned in the previous round (not the current round), looking at people who were actually assigned the control in the current round (as a percent of those respondents), in Follow-Up Visit 2 (top panel) and in Follow-Up Visit 3 (bottom panel)

	Remembered treatment							
Treatment from previous round	Control	1000FC	2000FC	Charity	Gov.	Other	DK	Total
1000FC	3	73	3	0	2	5	15	100
2000FC	7	3	77	0	0	10	4	100
Charity	4	4	10	53	1	9	19	100
Gov.	3	2	9	17	29	11	29	100
Total	4	20	26	18	8	8	16	100

	Remembered treatment							
Treatment from previous round	Control	1000FC	2000FC	Charity	Gov.	Other	DK	Total
1000FC	29	40	6	4	3	4	14	100
2000FC	37	3	40	8	0	2	10	100
Charity	20	6	12	40	0	8	14	100
Gov.	22	4	4	25	22	6	15	100
Total	27	14	15	19	6	5	13	100

Notes: This table shows the treatments recalled by participants (columns) for each treatment assigned to participants in the prior round (rows), restricted to the set of participants who were assigned to the control in the current round. All values are in percent terms, such that the total for each treatment assignment adds up to 100% when adding across all of the columns (i.e. each cell shows the percent of people who recall the treatment listed in that column, among those that were assigned the treatment listed in that row in the previous round). The top panel shows the results only for Follow-up Visit 2, while the bottom panel shows the results only for Follow-Up Visit 3.

Table 1.19: Effects of treatment removing Follow-Up Visit 2 in whole or in part

VARIABLES	(1) Original	(2) No R2	(3) No R2	(4) No R2 Control	(5) No R2 Control
Reward: 1000 FC	0.04*	0.06**	0.04	0.06**	0.06*
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Reward: 2000 FC	0.04**	0.07***	0.08***	0.07***	0.06**
	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)
Reward: 2000 FC donation to charity	-0.01	0.02	0.01	0.02	0.01
	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)
Reward: Gov. donates 2000 FC to charity	0.00	0.03	0.02	0.03	0.02
	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)
Observations	2,487	1,650	1,586	1,924	1,872
R-squared	0.00	0.01	0.57	0.01	0.51
Ind. FE	N	N	Y	N	Y
Control avg.	0.0961	0.0730	0.0716	0.0961	0.0957

Standard errors clustered by individual. $^{*}p<0.1,^{**}p<0.05,^{***}p<0.01$

Notes: All columns are estimated using equation (1.6.2), so all results include round fixed effects (not reported). The dependent variable is a count of valid (incentivized) receipts brought by the driver. Column (1) includes all rounds. Columns (2) and (3) remove the data from Follow-Up Visit 2. Columns (4) and (5) remove only the data from drivers assigned to the control group in Follow-Up Visit 2. Columns (3) and (5) add individual fixed effects. Standard errors clustered by individual. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$.

Table 1.20: Effects of treatment when controlling for treatment history.

VARIABLES	(1) Original	(2) R1 only	(3) History	(4) Prior Receipt	(5) IV	(6) IV FE
Reward: 1000 FC	0.04*	0.07*	0.06**	0.04**	0.06**	0.05*
	(0.02)	(0.04)	(0.03)	(0.02)	(0.03)	(0.03)
Reward: 2000 FC	0.04**	0.13***	0.06***	0.04**	0.05**	0.02
	(0.02)	(0.04)	(0.02)	(0.02)	(0.02)	(0.03)
Reward: 2000 FC donation to charity	-0.01	0.07*	0.01	-0.00	0.00	0.00
	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Reward: Gov. donates 2000 FC to charity	0.00	0.05	0.02	0.00	0.01	-0.01
	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Ever FC1000 in past			0.02			
			(0.03)			
Ever FC2000 in past			0.07**			
			(0.03)			
Ever charity in past			0.03			
			(0.03)			
Ever govt in past			0.05			
			(0.03)			
Ever brought receipt previously				0.14***	0.44**	-0.14
				(0.03)	(0.21)	(0.62)
Observations	2,487	857	2,487	2,484	2,484	2,464
R-squared	0.00	0.02	0.01	0.02	-0.05	0.49
Ind. FE	N	N	N	N	N	Y
Control avg.	0.0961	0.0638	0.0961	0.0962	0.0962	0.0957

Notes: All columns are estimated using equation (1.6.2), so all results include round fixed effects (not reported). The dependent variable is a count of valid (incentivized) receipts brought by the driver. Column (1) includes all rounds. Column (2) restricts to Follow-Up Visit 1 only. Column (3) adds controls for having ever had each of the treatments in the past. Column (4) controls for whether the driver ever brought a receipt to a prior round. Column (5) instruments for having ever brought a receipt to a prior round using prior treatment history as instruments. Column (6) adds individual fixed effects to the specification in column (5). Standard errors clustered by individual. $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.21: Effects of treatment for toll regulars.

VARIABLES	(1) Original	(2) Toll repeaters	(3) Toll repeaters - IV
Reward: 1000 FC	0.07**	0.10**	0.12***
	(0.03)	(0.04)	(0.05)
Reward: 2000 FC	0.07***	0.09**	0.10***
	(0.03)	(0.03)	(0.04)
Reward: 2000 FC donation to charity	0.00	-0.02	0.00
	(0.02)	(0.03)	(0.04)
Reward: Gov. donates 2000 FC to charity	0.03	0.03	0.04
	(0.03)	(0.04)	(0.04)
Ever brought receipt previously			0.19
			(0.18)
Observations	1,723	1,063	1,063
R-squared	0.01	0.03	0.05
Ind. FE	N	N	N
Control avg.	0.124	0.142	0.142

Standard errors clustered by individual. $^{*}p<0.1,^{**}p<0.05,^{***}p<0.01$

Notes: All columns are estimated using equation (1.6.1), so all results include round fixed effects (not reported). The dependent variable is the fraction of reported trips for which the driver brought a valid (incentivized) receipt. Column (1) includes all trips. Column (2) restricts to drivers who report traveling past a toll that they had reported visiting at least once in the two months prior to our study. Column (3) instruments for having ever brought a receipt to a prior round using prior treatment history as instruments, while still restricting to trips where drivers passed tolls they had reported visiting in the two months prior to the start of our study. Standard errors clustered by individual. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.

Table 1.22: Attrition by round.

VARIABLES	(1) R1-R2	(2) R2-R3	(3) R3-R4	(4) Days between appointments
Reward: 1000 FC	-0.0254	-0.0251**	-0.0305	0.0102
	(0.0187)	(0.0124)	(0.0231)	(0.399)
Reward: 2000 FC	-0.0275	-0.0242*	-0.0232	0.141
	(0.0180)	(0.0129)	(0.0231)	(0.382)
Reward: 2000 FC donation to charity	-0.0396**	0.0143	-0.0202	-0.0407
	(0.0159)	(0.0200)	(0.0239)	(0.362)
Reward: Gov. donates 2000 FC to charity	-0.0323*	-0.0246*	-0.0480**	-0.119
	(0.0175)	(0.0127)	(0.0198)	(0.358)
Constant	0.0528***	0.0318***	0.0702***	23.17***
	(0.0129)	(0.0105)	(0.0152)	(0.221)
Observations	905	856	835	2,486
R-squared	0.008	0.011	0.006	0.000
Avg. Dep. Variable	0.0319	0.0222	0.0503	23.17

Notes: This regression is estimated using equation (1.6.2), so all results include round fixed effects (not reported). For columns (1)-(3), the dependent variable is a dummy that is 1 if the driver stopped participating between the two survey rounds listed at the top of the column and 0 if the driver continued participating. The sample is restricted to the set of drivers who participated in the prior round of the survey. In column (4), the dependent variable is the number of days taken between survey visits, conditional on participating in the follow-up visit (averaged across all rounds). Standard errors clustered by individual. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.

Table 1.23: Treatment effects on reporting attempting to get a receipt.

VARIABLES	(1) Got a receipt	(2) Tried to get receipt	(3) Tried or got receipt
Reward: 1000 FC	0.08**	0.04	0.10***
	(0.04)	(0.03)	(0.04)
Reward: 2000 FC	0.09***	0.02	0.09**
	(0.03)	(0.03)	(0.03)
Reward: Charity	0.01	0.01	0.02
	(0.03)	(0.02)	(0.04)
Reward: Govt.	0.03	-0.00	0.03
	(0.03)	(0.02)	(0.04)
Observations	1,287	1,025	1,287
R-squared	0.01	0.01	0.01
Ind. FE	N	N	N
Control avg.	0.169	0.0842	0.239

Notes: All columns are estimated using equation (1.6.1), so all results include round fixed effects (not reported). In column (1), the dependent variable is the fraction of reported trips for which the driver brought a valid (incentivized) receipt. In column (2), the dependent variable is a dummy that is 1 if the driver reported demanding a receipt, but not receiving one, and 0 otherwise. In column (3), the dependent variable is 1 if the trip resulted in either a receipt or a demand for a receipt that was refused, and 0 otherwise. Columns (1) and (3) restrict the sample to the set of reported trips for which the driver reports receiving a receipt or the first trip for which he does not have a receipt. Column (2) restricts the sample to only the first trip for which a driver does not have a receipt. Standard errors clustered by individual. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$.

Table 1.24: Self-reported reasons for missing a receipt from a trip.

Underpaid toll	35.4
Left receipt at home	18.0
Lost receipt	16.5
Asked, but agent refused	9.3
Didn't ask for receipt	8.7
No toll on route	3.9
Exempt from toll	3.3
Machine was broken	1.2
Machine out of power	1.2
Other	0.9
No agents due to rebel group	0.7
Out of money	0.6
Boss or client took receipt	0.4
Total	100.0

Table 1.25: Treatment effects on likelihood of reporting a lost receipt.

VARIABLES	(1) Got a receipt	(2) Reported losing receipt	(3) Lost or got a receipt
Reward: 1000 FC	0.08**	-0.02	0.05
	(0.04)	(0.03)	(0.04)
Reward: 2000 FC	0.09***	-0.00	0.07*
	(0.03)	(0.03)	(0.04)
Reward: 2000 FC donation to charity	0.01	0.01	0.02
	(0.03)	(0.03)	(0.04)
Reward: Gov. donates 2000 FC to charity	0.03	0.03	0.05
	(0.03)	(0.04)	(0.04)
Observations	1,287	1,025	1,287
R-squared	0.01	0.00	0.01
Ind. FE	N	N	N
Control avg.	0.169	0.163	0.305

Notes: All columns are estimated using equation (1.6.1), so all results include round fixed effects (not reported). In column (1), the dependent variable is the fraction of reported trips for which the driver brought a valid (incentivized) receipt. In column (2), the dependent variable is a dummy that is 1 if the driver reported losing a receipt, and 0 otherwise. In column (3), the dependent variable is 1 if the trip resulted in either a receipt or a (reported) lost receipt, and 0 otherwise. Columns (1) and (3) restrict the sample to the set of reported trips for which the driver reports receiving a receipt or the first trip for which he does not have a receipt. Column (2) restricts the sample to only the first trip for which a driver does not have a receipt. Standard errors clustered by individual. $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.26: Overall trip reporting.

VARIABLES	(1) Any trip	(2) Total trips taken
Reward: 1000 FC	-0.0134	-0.0517
	(0.0266)	(0.0500)
Reward: 2000 FC	-0.0188	-0.0480
	(0.0273)	(0.0722)
Reward: 2000 FC donation to charity	-0.0382	-0.0434
	(0.0270)	(0.0550)
Reward: Gov. donates 2000 FC to charity	-0.0533*	-0.148***
	(0.0276)	(0.0544)
Constant	0.647***	1.084***
	(0.0241)	(0.0626)
Observations	2,487	2,487
R-squared	0.031	0.034
Control avg.	0.526	0.792

Notes: This regression is estimated using equation (1.6.2), so all results include round fixed effects (not reported). The dependent variable in column (1) is a dummy that is 1 if the driver reported one or more trips and 0 otherwise. The dependent variable in column (2) is a count of the number of trips reported by the driver. Standard errors clustered by individual. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.

Table 1.27: Trips recorded by validator, but not reported by drivers.

VARIABLES	(1) OLS	(2) Recalled	(3) If DK Treatment
Reward: 1000 FC	0.198*	-0.0475	0.703***
	(0.116)	(0.115)	(0.226)
Reward: 2000 FC	0.200*	-0.115	0.794***
	(0.104)	(0.0982)	(0.115)
Reward: 2000 FC donation to charity	0.205*	-0.0140	0.499*
	(0.121)	(0.111)	(0.283)
Reward: Gov. donates 2000 FC to charity	0.233**	0.0293	0.390
	(0.117)	(0.161)	(0.265)
Constant	0.282***	0.459***	
	(0.0961)	(0.107)	
Observations	169	169	31
R-squared	0.078	0.048	0.335
Trip non-reporting in control	0.404	0.404	0.404

Notes: This regression is estimated using equation (1.6.2), so all results include round fixed effects (not reported). The dependent variable is a dummy that is 1 if the driver did not report a given trip in his follow-up interview and 0 if he did report the trip. The sample is restricted to the set of trips that were validated by our enumerator at the toll. Column (1) shows the results using assigned treatment as the treatment. Column (2) shows the results using recalled treatment as the treatment. Column (3) restricts the sample to drivers who correctly recalled their treatment at their follow-up visit. Standard errors clustered by individual. $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 1.28: Estimated elasticities of receipt-getting with respect to instantaneous monetary reward for various values of β. Bootstrapped 95% confidence interval shown in square brackets below each estimate.

β	**FC1000**	**FC2000**	**Per FC**
0.5	-0.06	-0.08	-0.28
	[-2.41, 2.04]	[-1.23, 0.92]	[-1.29, 0.68]
0.63	-0.04	-0.06	-0.23
	[-1.94, 1.64]	[-0.99, 0.74]	[-1.04, 0.55]
0.74	-0.04	-0.05	-0.19
	[-1.64, 1.39]	[-0.84, 0.63]	[-0.88, 0.47]
1	-0.03	-0.04	-0.13
	[-1.14, 0.97]	[-0.58, 0.44]	[-0.61, 0.32]

Notes: This table shows estimates of the elasticity of corruption (not getting a receipt) with respect to instantaneous monetary cost of compliance (i.e. cost of getting a receipt), using the calculations described in equation (1.8.1). Standard errors reflect the empirical 95% confidence interval from 5,000 bootstrap iterations. Bootstrapping was done using a block bootstrap, blocked by individual participant. The rows show the results using different assumed values of beta. Column (1) shows the results using only the effects of the 1000 FC treatment. Column (2) shows the results using only the effects of the 2000 FC treatment. Column (3) uses both treatments simultaneously (treating them linearly).

Table 1.29: Heckman results on bribes.

VARIABLES	(1) Heckman
Reward: 1000 FC	99.19
	(70.33)
Reward: 2000 FC	-4.48
	(70.24)
Reward: 2000 FC donation to charity	-3.98
	(75.88)
Reward: Gov. donates 2000 FC to charity	0.84
	(72.65)
Toll officer treatment in effect	-40.76
	(78.49)
Reward: 1000 FC	-0.21**
	(0.10)
Reward: 2000 FC	-0.23**
	(0.11)
Reward: 2000 FC donation to charity	0.10
	(0.12)
Reward: Gov. donates 2000 FC to charity	-0.11
	(0.11)
Toll officer treatment in effect	0.13
	(0.14)
Enumerator validating traffic at toll	-0.20*
	(0.11)
Observations	1,519
Ind. FE	N
Ind. Controls	N
Trip Controls	N
Control avg.	1520

Notes: The dependent variable is the bribe (amount paid not backed by a receipt) paid by the driver. In column (1), we show the results of a Heckman selection correction based on equation (1.6.1) where bribes are considered to be missing for all drivers who receive a receipt. The top panel shows the second stage results on bribes paid and the bottom panel shows the selection equation. The presence of an enumerator at the toll on the date of travel is the excluded instrument. Standard errors clustered by individual.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Chapter 2

A Market Equilibrium Approach to Reduce the Incidence of Vote-Buying: Evidence from Uganda (with Christopher Blattman, Horacio Larreguy, and Benjamin Marx)

2.1 Introduction

Democracy in many countries is undermined by widespread vote-buying – the provision of gifts, in cash or in kind, in exchange for votes. Across political regimes, candidates use many tactics to buy votes (Gans-Morse et al., 2014), from giving likely supporters an incentive

For implementation we thank the Alliance for Election Campaign Finance Monitoring (ACFIM) and the National Democratic Institute (NDI) in Uganda; we are particularly indebted to Teresa Lezcano Cadwallader, Henry Muguzi, Simon Osborn, and Ivan Tibemanya. We are grateful to Kelsey Barrera, Alex Nawar, Harrison Pollock, for their outstanding research management and assistance in Uganda, to the entire executive staff at Innovations for Poverty Action Uganda for ensuring the completion of the survey work, and we thank Patryk Perkowski for excellent research assistance in the U.S. We also thank Pia Raffler and Melina Platas Izama for sharing data and providing advice on field operations. We benefited from helpful comments and suggestions from Aislinn Bohren, Berk Ozler, the EGAP peer response system (in particular, Katherine Casey, Nahomi Ichino and Macartan Humphreys), and participants at APSA 2016, the Harvard Development Retreat, the Institute for International Economic Studies (IIES) at Stockholm University, the Uppsala University Applied Micro Seminar, and the MIT development tea. We gratefully acknowledge financial support from the J-PAL Governance Initiative, and the International Growth Centre. Otis would also like to thank Christina Patterson for her unwavering support throughout this project.

to turn out (Nichter, 2008), to targeting the people who are most likely to reciprocate the gift with a vote (Finan and Schechter, 2012). Political intermediaries known as brokers target resources to voters and mobilize them around elections, while extracting significant resources for themselves (Camp and Szwarcberg, 2015; Larreguy, 2013; Larreguy et al., 2016; Stokes et al., 2013). Endemic vote-buying practices impede political and economic development by limiting the ability of citizens to hold elected officials accountable (Stokes, 2005), the emergence of credible political platforms (Keefer and Vlaicu, 2008), and public goods provision (Hicken and Simmons, 2008).

Policy experiments designed to combat vote-buying have found that legalistic appeals to resist vote-selling (Vicente, 2014) and behavioral interventions (Hicken et al., 2014) convince some voters to renounce selling their vote, and hurt the electoral performance of vote-buying candidates (Green and Vasudevan, 2016). But do these interventions reduce vote-buying, or merely displace it? For instance, candidates and their brokers could simply buy their votes elsewhere. Voters could also react by choosing to sell their votes to a different candidate, or to accept gifts from multiples candidates but still vote for their preferred candidate. Because the response to campaigns of this kind is complex and potentially involves spillovers, it is important to track the response of voters (the supply side) in both treated and untreated areas, as well as that of vote-buying parties and candidates (the demand side), to interventions designed to combat vote-buying.

This paper investigates these concerns through a large-scale policy experiment in Uganda. President Yoweri Museveni and his National Resistance Movement, or NRM, have held power since 1986 in a system that most analysts classify as a "hegemonic party system," not unlike that of the PRI in late 20th century Mexico (Tripp, 2010; Magaloni et al., 2013). Ahead of Uganda's 2016 general elections, we partnered with a Ugandan Civil Society Organization (CSO), the Alliance for Election Campaign Finance Monitoring (ACFIM), and an international NGO, the National Democratic Institute (NDI), to evaluate the causal effects of what is (to our knowledge) one the largest anti-vote-buying campaigns ever implemented. Historically, vote-buying has been endemic in Uganda (Conroy-Krutz, 2012; ACFIM, 2015). Starting with an experimental sample composed of 918 parishes (or collections of villages) where ACFIM was active, we randomized roughly two thirds to treatment and one third to pure control. Beyond being a local administrative unit, fieldwork prior to the intervention suggests that the parish is also the level at which local political brokers are known to operate.

Within treated parishes we randomized the fraction of villages targeted by the campaign, using a *randomized saturation design* where the level of local treatment intensity is itself randomly determined.[1] We denote villages in treated parishes as treated and spillover villages depending on whether they were assigned to receive the campaign or not, respectively. We denote all villages in control parishes as control villages.

ACFIM's campaign was conducted prior to the 2016 election and included five main elements: (i) a leaflet drop; (ii) three village meetings, organized by local ACFIM activists, to build awareness of and opposition to vote-buying; (iii) the organization of a public village-wide resolution against vote-buying; (iv) the posting of posters reminding voters that selling their vote would harm the community; and (v) an automated-call reminder on the eve of the election. ACFIM activists spread across the country ran their intervention in 1,427 villages. The villages in our experimental sample cover around 1.2 million people registered to vote in the 2016 Ugandan general election across 6% of the country's polling stations, and 12% of polling stations in the districts we study – unparalleled numbers for this type of intervention.

ACFIM aimed to shift local social norms against vote-selling – people's perceptions about how others will behave, what kinds of behavior are considered appropriate, and the social sanctions for violating norms. The leaflet and the first community meeting attempted to create common knowledge about the costs and inappropriateness of vote-buying in terms of future service delivery and politician corruption. The second and third meetings were designed to convince a critical mass of the community to take a coordinated action against vote-selling – the public, village-wide resolution. Posters and automated calls were intended to reinforce the new norm. All these actions sent a public signal about the new norm, including towards candidates and their brokers.

To assess impacts, we use a combination of administrative data, original survey data, and systematically-collected qualitative accounts from the implementation of the intervention. Shortly after the 2016 elections, we surveyed 28,454 villagers, collecting data on people's experience with vote-buying, as well as data on the local prices of goods commonly used for vote-buying purposes in Uganda. We surveyed all treatment and control villages, as well as 1,399 out-of-sample villages in order to increase our power to estimate spillovers of the ACFIM campaign and thus our ability to capture the reactions of voter and candidates or their brokers to the campaign. In addition, we obtained administrative data on electoral

[1]This design follows along the lines of Baird et al. (2014). See Section 2.3.3 for additional details.

results at the polling station level for the two most important ballots conducted in February 2016 (President and Members of Parliament (MP)).

Contrary to our expectations, as well as those of ACFIM and NDI, we see no evidence that the ACFIM campaign significantly reduced the extent to which voters were offered (and accepted) cash or gifts in exchange for their vote. We preregistered our central hypotheses: that cash and gift giving would fall in treated villages but likely increase in "spillover villages" – untreated villages in the same parish. However, survey respondents do not report a lower prevalence of vote-buying after the ACFIM campaign. For instance, an index of vote-buying offers received in cash and in kind increased by just 0.03 standard deviations in treated villages (not statistically significant). The fraction of survey respondents who reported receiving cash on behalf of any of the Presidential or MP candidates, which is 43% in control villages, increased by 2 percentage points (not statistically significant).

This null average result, however, masks considerable heterogeneity in the response of candidates and political machines to the ACFIM campaign. We see evidence that opposition candidates, but not incumbent candidates, increased their attempts to buy votes on average as a result of the campaign. For example, only 10% of people in control villages reported that representatives of challengers of either the Presidential or Parliamentary races offered them cash for their vote, and this rose, respectively, by 1.8 and 2.3 percentage points in treated and spillover villages. Incumbent politicians, meanwhile, marginally increased their attempts to buy votes only in spillover areas within parishes randomly assigned to a high level of ACFIM presence. Challenger candidates seem to have also increased their campaigning efforts in heavily treated parishes, though this result often falls short of statistical significance and must be interpreted with caution.

We also see some evidence that instead of refusing offers of cash, voters took the money offered by politicians but nonetheless voted for their preferred candidate. On average, the ACFIM campaign reduced the incumbent's vote share in both the Presidential and Parliamentary elections, though this decrease often falls short of statistical significance. In addition, incumbents received significantly less support in parishes where all villages received the campaign. For example, in these intensively treated parishes, incumbent MPs received a lower vote share by approximately 0.2 standard deviations (SDs). Villages in spillover villages within treated parishes experienced a similar decline in incumbent vote shares, reinforcing the effects of the ACFIM campaign on parish-level vote shares.

Consistent with these effects, we see evidence of a change in attitudes and perceived social norms around vote buying, especially in the most intensively treated parishes. For instance, 75% of respondents in control villages thought that other villagers would be angered over vote selling, and 57% said the consequence would be ostracization. These perceived social sanctions rose about 2 percentage points on average with treatment, with slightly larger shifts occurring in the most intensely treated parishes. Treatment also appears to have changed attitudes: respondents in treated villages were slightly more likely to say that vote buying has ill consequences for the village and is unacceptable, again with the largest effects found in intensely treated parishes.

While we anticipated impacts on incumbent vote shares, we did not anticipate the direction, magnitude, or importance of these effects (we preregistered the outcome as secondary, with the potential for the opposite effect). Thus we must take these vote share results with some caution. Nonetheless, the treatment effects and our qualitative data tell a plausible, coherent story, where candidates and their brokers responded strategically to ACFIM's campaign. When it reached enough villages in the parish, the ACFIM campaign weakened the enforceability and effectiveness of vote-buying by incumbents in particular, who rely more than challengers on vote-buying, and whom voters associated with poor service delivery and corruption in order to recover the money spent on vote-buying during their campaigns – a complaint voters emphasized during the ACFIM meetings. Moreover, challengers took advantage of the campaign by engaging in greater vote-buying and policy campaigning.

Our findings suggest some promising avenues for policy. To begin, anti-vote buying campaigns may do well to encourage voters to take the funds and vote their conscience, rather than refusing gifts. It is unclear, however, how parties will respond in the longer run. The incentives to buy votes (and the resulting corruption and crowding out of public services) may persist. Alternatively parties may find it optimal to shift away from vote buying to other tactics, if vote buying becomes sufficiently ineffective. Understanding the longer term effects of anti-vote buying campaigns is an important area of future research. Another implication of our study is that campaigning may need to be intense enough to be effective, which might mean either increased resources to anti-vote buying, or geographical targeting of these resources. More broadly, these results suggest that more experimentation is needed to identify cost-effective strategies to counter vote-buying.

This paper builds on two previous randomized evaluations of programs designed to combat vote-buying practices. Vicente (2014) finds that a voter education campaign in São Tomé and Príncipe reduced the influence of money offered on voting, decreased voter turnout, and favored the incumbent, in a context where (relative to Uganda) challengers rely more on vote-buying practices. Hicken et al. (2014)'s experiment tackles vote-selling as a time-inconsistency problem, using ex-ante promises to reject vote-buying offers, or to accept them but instead vote their preferred candidate, in the Philippines. Our experiment differs in its scale and visibility, so as to generate incentives for candidates or their brokers to react and voters to coordinate against vote buying.[2] Another experiment recently implemented in India shows that a radio campaign designed to reduce vote-buying decreases the vote share of candidates known to buy votes (Green and Vasudevan, 2016).

We also build on recent experimental work that studies whether campaigning on public goods provision (as opposed to using standard clientelistic strategies) reduces vote-buying. In Benin, Wantchekon (2003) randomly assigned clientelistic messages and broad-based messages (regarding nationwide issues) endorsed by candidates, and found that clientelism was more effective in generating electoral support. Contrary to this finding, Fujiwara and Wantchekon (2013) showed also in Benin that town hall meetings addressing specific policy platforms of broad-based public goods provision (as opposed to traditional vote-buying strategies) reduced self-reported measures of vote-buying, and lowered the vote shares for the candidate if the village represented a political stronghold.

Outside the experimental literature, various papers have studied how vote-buying works in practice.[3] Of particular relevance to this paper is the literature that highlights the role of brokers for the success of vote-buying (Camp and Szwarcberg, 2015; Larreguy, 2013; Larreguy et al., 2016; Stokes et al., 2013). Our intervention builds heavily on this recent work, both in terms of design and analysis. Accounts from interviews with candidates, brokers and voters from our focus groups suggest that vote-buying in Uganda is facilitated by pyramidal structures of brokers that mediate between candidates and voters. We conducted our treatment saturation at the lowest level at which these brokers are organized, monitored

[2] Vicente (2014) treats only 40 enumeration areas (out of 50 that composed the experimental sample). Hicken et al. (2014) treat 600 voters (out of 900 that composed the experimental sample) privately.

[3] For example, Dekel et al. (2008) provide a model of vote-buying in which vote prices remain low in equilibrium because only the winning party buys votes. Gans-Morse et al. (2014) study how political machines mix across 4 types of clientelist strategies: vote-buying, turnout buying, abstention buying, and double persuasion. Finan and Schechter (2012) show that politicians target their vote-buying offers towards reciprocity-minded individuals. Finan et al. (2016) further argue that brokers exploit their social networks to acquire information about partisanship and reciprocity, which they subsequently use to target voters.

and incentivized, namely at the level of parish.

Finally, our experiment addresses vote-buying as a market equilibrium problem. In doing so, it also builds on a new strand of empirical work designed to uncover spillover and general equilibrium effects in experimental settings. A comprehensive review of this literature is beyond the scope of this paper, but a good review can be found in Baird et al. (2014), who provided the conceptual framework for the design used in our experiment. Looking at voter education campaigns more specifically, Fafchamps et al. (2012) find that a voter education in Mozambique has 2 types of spillover effects: the treatment effect is reinforced when targeted individuals are surrounded with other targeted individuals; and non-targeted individuals are also affected when living in close proximity with targeted individuals. Ichino and Schündeln (2012) study the displacement of fraud due to the deployment of observers in the 2008 election in Ghana.

The rest of the paper is organized as follows. We provide relevant background on the 2016 Ugandan general election and vote-buying practices in Section 2.2. Section 2.3 describes our experimental design, and Section 2.4 our data. We present our empirical framework in Section 2.5. Section 2.6 presents our main results. Section 2.7 discusses potential mechanisms and section 2.8 concludes.

2.2 Background

2.2.1 The 2016 Ugandan general election

Uganda holds general elections in February, every five years. The President is elected in a two-round system, requiring at least 50% of the popular vote to be elected in the first round. Members of Parliament (MPs) are elected in single-member constituencies using first-past-the-post voting. In addition, voters also elect District Woman Representatives to sit in Parliament. 375 seats were contested during the February 2016 general elections.[4]

The 2016 general elections were held on February 18, in accordance with the electoral

[4]This includes 238 constituency seats, 112 District Woman Representatives, and 25 indirect (reserved) seats. At the same time, voters also elect local leaders. The country is divided into 111 districts, which are themselves divided into counties, subcounties, parishes, and villages. Voters elect a District (or "LC5") Chairman and Councilors, as well as a Subcounty ("LC3") Chairman and Councilors. Village leaders (or "LC1s") are elected through informal processes at the village level, and so are not included in our analysis, and there are no elected positions at the county or parish level (these are governed by LC3 and LC5 councils).

calendar. A total of 28,010 polling stations were set up for the election, 6% (1,603) of which were directly treated by the ACFIM campaign. Eight candidates contested the presidential election, among which two were considered the eventual frontrunners: the incumbent President, Yoweri Museveni, in office since 1986; and a long-time opposition leader running for the fourth time, Kizza Besigye.[5] Museveni's and Besigye's respective parties, the National Resistance Movement (NRM) and the Forum for Democratic Change (FDC), were also dominant in the campaigns for parliamentary seats and local positions, but these elections additionally involve a large number of independent candidates,[6] as well as candidates from several smaller parties. For the parliamentary election, a total of 1,743 candidates ran for MP positions across the country's 290 constituencies and 112 districts.

Though politics is fairly competitive at the local level, at the national level most analysts consider Uganda a "hegemonic party system" or a "multiparty autocracy" due to suppression of opposition parties and candidates, and the widespread use of patronage and vote-buying by incumbents (Tripp, 2010).[7] For example, several major incidents occurred throughout the 2016 electoral period. First, the leader of the opposition, K. Besigye, was arrested twice in the week leading to the election (Amnesty International, 2015), and subsequently kept under house arrest. Second, checkpoints were set up, and the presence of security forces massively increased throughout the country as the election unfolded (Amnesty International, 2016). Third, the government enforced a four-day social media blackout (The Guardian, 2016). Lastly, voting materials were delivered late to a number of polling stations where voters were expected to vote against Museveni. The alleged goal was to generate long lines in those polling stations in order to ultimately discourage voters from casting their vote (The Guardian, 2016).

On February 20, 2016, Museveni was declared the winner of the presidential election with 60.8% of the vote (against 35.4% for Besigye). Museveni's party, the NRM, also won 164 out of 238 constituency MP seats (69%), and 86 out of 112 (77%) District Women's Representative positions. Ugandan and international observation missions provided mixed opinions about the fairness and transparency of the election.[8] For example, the EU Observa-

[5] Museveni took power through military victory in 1986, under "no party rule". Elections began in 1996, but restricted party competition. Multiparty competition was first permitted in 2006, and 2016 represents the third multiparty election.

[6] Often, these are individuals who lost in the primaries to represent their favored party.

[7] The Ugandan political regime was classified by the Freedom House as "not free" in 2016 (with a score of 36%), and as a "closed anocracy" in 2015 by the Polity IV project (with a score of -1).

[8] We discuss allegations of vote fraud in Appendix 2.

tion Mission cited the lack of independence of the Electoral Commission, the excessive use of force against the opposition, the "intimidating atmosphere for both voters and candidates", and "the orchestrated use of state resources and personnel for campaign purposes" as major obstacles against a free and fair election (European Union Election Observation Mission, 2016).

2.2.2 Vote-Buying in Uganda

Uganda has some of the highest rates of vote-buying in the world. Out of 18 countries with Afrobarometer data, Uganda in 2006 had the second highest reported rate of vote-buying of any country in the sample (after Kenya), with 85% of respondents reporting that politicians "often" or "always" give gifts during political campaigns (Afrobarometer, 2006).[9] The culture of vote-buying in the country has been called "ubiquitous" (Democracy Monitoring Group, 2011), and previous studies have described sizeable payment amounts – one such study reported that the median vote price in 2011 was 5 times the daily average income (Conroy-Krutz, 2012).

Despite the magnitude of vote-buying in Uganda, little is known about how it is undertaken in practice. To fill this gap and to explore possible intervention designs, we conducted (prior to the launch of the ACFIM campaign) and through our partners (NDI and ACFIM) focus groups in 48 locations spread throughout our eventual experimental sampling frame. In addition, we interviewed several elected candidates and active brokers to gather information about their vote-buying operations, and how candidates fund these operations.

The focus groups highlighted the large extent of the vote-buying phenomenon and its importance in enabling candidates to win elections. While focus group participants agreed that some voters may choose to "eat widely but vote wisely," i.e., to take money for their vote but then vote for their preferred candidate, they also highlighted that a large share of voters reciprocate gifts with their actual vote since money "softens people's hearts." Participants also noted that vote-buying addresses short-term needs which are especially salient around elections, when inflation is high.[10]

[9] The average across all 18 countries in the sample was 70%. In the same survey, 35% of Ugandan respondents said they had themselves been offered incentives to vote in elections (the sample average was 18%).

[10] Participants also argued they often have very poor information to discern among the best candidates. In addition, elected officials reportedly argue they are not responsible for improving public service delivery. This discourse is often left undisputed since voters are also uninformed about whom they should hold accountable for service delivery.

All participants, candidates and brokers emphasized the importance of brokers for the success of vote buying. An NDI survey of 185 elected MPs after the intervention reflects that all respondents had brokers in the 2016 election – 96% in all villages and 4% only in selected villages. Brokers are not only responsible for handing over cash or gifts to voters – typically soap, sugar and other more idiosyncratic goods, mostly in the week preceding the election – but they also make sure people who received such gifts turn out on election day. The brokers' ability to mobilize voters is so important for the success of vote-buying that candidates admitted to us that they decide how much to invest in buying votes depending on such ability. To maximize the returns from vote-buying, candidates use sophisticated pyramidal structures, with chiefs at the constituency level, coordinators at the subcounty level (often LC3 chairpersons), and managers at the parish level (often LC3 councilor or LC1 chairpersons) who are the ones ultimately responsible for recruiting and managing village-level brokers.[11]

Brokers have both immediate and long-term financial incentives to deliver voters for the candidates they work for. They are first endowed with a budget to carry out voter mobilization in the village – a fraction of this budget, which comes in the form of cash or gifts, is often retained by the brokers themselves.[12] Importantly, brokers typically receive a bonus for their work based on an evaluation of their performance,[13] and they are able to build a connection with an elected official, as well as to receive the benefits that such a connection entails.

ACFIM conducted a separate survey of Ugandan MPs that indicated that a large majority of costs are borne by individual candidates, not parties. Candidates fund their campaigns using personal resources (savings, property) or sometimes take out loans explicitly for the purpose of campaigning. As a result, we anticipated that parties were unlikely to respond to the campaign strategically by moving resources across races and candidates (i.e. that they would not reallocate resources between the presidential and parliamentary races in response

[11] Interestingly, some brokers work for candidates that run in different races but belong to different parties. Higher level candidates often also coordinate with lower level candidates that they trust irrespective of their party since these have a fewer resources but much more local presence among both voters and brokers. The NDI survey indicates that 48% of the surveyed MPs shared brokers with other candidates.

[12] While this claim cannot be verified, some focus groups participants argue that brokers keep between two thirds and four fifths of what candidates given them to distribute.

[13] Candidates look closely at the election returns and brokers have to provide a report about the candidate's performance in their area on election day. Agents who did not do well do not even bother to provide such a report.The NDI survey shows that 97% of the surveyed MPs followed their brokers' performance, out of which 66% did so by looking at polling-station results and 21% by requesting brokers to submit reports after the election.

to the ACFIM campaign).

2.3 Experimental Design

2.3.1 Description of the intervention

ACFIM (along with its 13 local partner organizations) implemented the anti-vote buying campaign in January-February 2016 across 53 Ugandan districts, or about half the country. The design of the campaign was influenced by ACFIM and NDI's past interventions, by a survey of Ugandan MPs (collecting qualitative information on campaign financing), and by the focus groups described above. The campaign sought to reduce the incidence of vote-buying by fostering a change in local norms as well as collective commitments in the community to not sell any votes. The general goal was to convince participants that selling their vote was not only inappropriate but also costly, since it would undermine the accountability of elected officials and future delivery of public goods to the community. With the adoption of a community-wide resolution on the issue, the campaign sought to improve coordination by fostering a collective commitment at the community level to renounce vote-selling.

The campaign took place during the apex of the electoral period, when most vote-buying transactions take place (i.e., the final weeks leading up to the election), and involved several stages in each selected village. First, all households in treated villages received a leaflet explaining in simple terms the costs and risks of vote-buying to their communities. Leaflet recipients were also invited to participate in subsequent community meetings to discuss the vote-buying issue. The leaflets were delivered via door-to-door canvassing conducted by local ACFIM activists in January 2016. The content of the leaflets was approved by the Electoral Commission and entirely non-partisan. The leaflets contained a cartoon alongside the following message (in the language spoken by the community):[14]

> "You wouldn't sell your future, you wouldn't sell your village's future. So, why sell your vote? Stand together with your village, and don't sell your vote. It is your chance to demand a better future!"

[14] 18 languages were used as part of the campaign: Acholi, Alur, Aringa, Ateso, Kumam, Langi, Lubwisi, Luganda, Lugbara, Lusoga, Madi-Moyo, Ngakarimojong, Rufumbira, Rukhozo, Rukiga, Runyankole, Runyoro, and Rutoro.

A sample leaflet in English can be found in Figure 2-2 and shows individuals first receiving money from a candidate for their votes (in the left plot), and then seeing their request for a health center denied on the ground that the candidate had already bought them off (in the right plot). These plots and the caption embody the main messages behind the ACFIM campaign, which were later reemphasized during the complementing components of the intervention. First, individuals who sell their votes are unlikely to later be able to demand public service delivery from the candidates they sold their votes to. Second, community coordination is crucial to fight vote-buying and the associated lack of public service delivery.

Following the leaflet drop, three meetings were organized to discuss vote-buying in the village. The meetings were again facilitated by a local ACFIM activist. The first meeting focused on introducing the campaign, discussing the leaflet and gathering participants' thoughts and experience on vote-buying. The second meeting was designed to provide an avenue for a collective deliberation on vote-buying. Finally, during the third meeting, ACFIM activists invited the community to collectively commit to refuse offers of gifts or money in exchange for votes. ACFIM activists then placed posters through the village indicating the village is a "no vote-buying village."

Finally, on the eve of election day, individuals that attended the village meetings and provided their phone number on the attendance sheet received automatized phone calls reminding them about the harm caused by vote buying. The calls included the following message (in the appropriate local language):

> "Hello! This is an important message from ACFIM. We are calling you to ask you not to sell your vote. You might think it is harmless to accept some small money or gifts from politicians during election campaigns, but this will affect the future of your whole community. Do you not want good hospitals, good roads, good schools for your children? When you ask for these services after elections, the politician who wins through buying votes will tell you "I bought your vote, therefore do not bother me by asking me for more things." Don't let your community down. Don't let your country down. Don't sell your vote!"

2.3.2 Experimental sample

Our experimental sample included 2,796 eligible villages across 1,603 polling stations within 53 Ugandan districts. The sample villages were spread across 110 parliamentary constituencies and 918 parishes. Eligibility to receive the intervention was tied to the presence of a local ACFIM activist (i.e. one that resided in a nearby location to the village that comprised the local polling station).[15] Throughout the paper, we use "eligible village" to indicate that it was potentially treated. A parish generally consists of 3-10 villages. Parishes with eligible villages can, and usually do, also have ineligible villages, some of which we also sampled for our survey in order to maximize statistical power when looking at spillovers identify the reaction of candidates or their brokers to the intervention, as well as voter coordination around it. Randomization was done among eligible villages so as to preserve internal validity of the design. We provide additional details on sampling and external validity on this procedure in Appendix 1.

2.3.3 Randomization

The intervention used a randomized saturation design, along the lines of Baird et al. (2014), varying the level of saturation of treatment at the level of a parish. Among the 2,796 eligible villages in 918 parishes, we randomly selected 1,427 villages across 535 parishes for treatment. The remaining 383 parishes were allocated a pure control group with no villages treated. An additional 1,399 villages located in the same 918 parishes were added to the endline survey sample to look for spillovers of the intervention oversampling villages in parishes with a higher treatment saturation.

Because the campaign could only take place in areas where ACFIM activists had a local presence at baseline, the randomized saturation level is defined in terms of eligible villages, where eligible means that the partner had local activists who could work in those villages (note that all our specifications control for the baseline level of partner presence, as described in our pre-analysis plan). The fraction of eligible polling stations in a parish ranged from 3% to 100%, with an average of 48%. Accounting for the variation in the number of voters registered in each station, the fraction of eligible voters ranged from 1% to 100%, with an average of 54%.

In the first step of the randomization, parishes were allocated to one of three cells: a pure control cell (no treatment), a partial-saturation treatment cell (50% of *eligible* villages assigned to treatment), and a high-saturation cell (100% of eligible villages assigned to

[15] Due to cultural issues, it is very hard for an individual to conduct this type of intervention in villages where she is perceived as an "outsider". As ACFIM members explained it to us, activists had to be "sons of the soil" for villagers to listen to them.

treatment). To fix ideas, consider a parish with 8 equally sized villages, of which 4 have ACFIM activists. If assigned to 100% treatment, this would mean that all 4 of the *eligible* villages would be treated (equivalent to 50% "true" saturation). If assigned to partial (50%) treatment, then a randomly selected 2 of the 4 eligible villages would be treated (25% true saturation).

This randomization was stratified at the parish level along baseline measures of partner presence (defined in terms of the number of voters covered), parish-level voter population, and support for the incumbent political party in the 2011 presidential election. Specifically, a stratum was defined by the three-way interaction of quartile of partner presence, quartile of voter population, and quartile of district-level NRM support (63 strata in total).

In the second step, eligible villages were assigned to treatment within the partial-saturation parishes. Here, we randomized villages to treatment or control status at the polling station level. All eligible villages in treated polling stations were selected to receive the ACFIM campaign (up to a limit of 3 villages per activist). None of the villages falling under control polling stations were selected to receive the campaign. This creates an integer problem if all eligible villages fall under a single polling station. If only one polling station was eligible for treatment in a parish (i.e. a parish had only a local ACFIM activist), it was either fully treated (with 50% probability) or a full control (with 50% probability). No polling stations were split between treatment and control in order to maximize the usefulness of the official election outcomes.[16]

This design allows us to identify the spillover effect on the non-treated from (potentially) both the responses of those receiving the campaign (social norm coordination) and from changes in candidate or broker behavior, in addition to standard intent-to-treat estimates of direct treatment. Our design also allows us to recover precise estimates of how those estimates vary with treatment intensity. The spillover effects could differ substantially with local treatment intensity – only if a large number of villagers resist vote-buying, political candidates or their brokers may be forced to change vote-buying tactics, as well as other campaign strategies.

[16]To fix this concept clearly, we can return to our 8 village (4 with ACFIM presence) parish example from before. Imagine first that there are 4 polling stations, each with 2 villages. Then, if that parish was assigned to partial (50%) treatment, there would be no problem (1 eligible, treated polling station, 1 eligible, untreated polling station, and 2 ineligible, untreated polling stations). However, if there were only 2 polling stations (1 with all of the ACFIM villages, 1 with none), then this parish would either be assigned to have either its 1 eligible polling station treated (which is then equivalent to 100% treatment) or its 1 eligible polling station untreated (which is then equivalent to being in the control).

2.4 Data

2.4.1 Administrative Data

Overview

We use official electoral results obtained from the Ugandan Electoral Commission at the lowest possible level, the polling station. We use this data for two of the three of the ballots conducted in February 2016 (President and MP) for 1,585 out of the 1,603 (99%) polling stations in our experimental sample.[17] We also use data on turnout and vote share of the corresponding incumbents from the previous general election conducted in 2011, available for 98% of polling stations in our sample. We discuss the integrity of the electoral data in Appendix 2.

2.4.2 ACFIM administrative notes

We use data collected by the ACFIM partners during the implementation of the three village meetings. Two activists of each ACFIM partner took part in every meeting: one had the role of facilitator and the other one of note taker. The note taker had to fill in basic information about the meeting, which included the start time, end time, and location of each meeting, rough estimates of the number of participants from the village and from outside the village, the presence of influential individuals likely to engage in or mediate vote-buying activities (LC 1, 2, 3, or 5 officials, MPs, candidates or brokers), a range of questions addressing whether the facilitators conducted the meetings as specified during training and in the meeting scripts, the views of the community about the effect of vote buying and possible solutions against it, and activists' perceptions of how likely communities were to vote on a resolution against vote buying and whether they effectively did.

2.4.3 Survey Data

We conducted an endline survey of 28,454 Ugandan voters in the aftermath of the ACFIM campaign and the general election. The survey started on March 2, 2016, and ended on July 19, 2016, after some of our survey teams encountered administrative delays due to the sensitivity of the information collected. The survey involved three different questionnaires:

[17] Due to discrepancies in local names and spellings, we were unable to match 1% of polling stations in our sample with the official electoral data.

of registered voters, a "key informant" in each village, and a local market survey of the prices of goods commonly used for vote-buying, as well as the prices of goods not subject to vote-buying practices.

Survey respondents were randomly sampled from the official voter register in each village, stratifying into four categories by age (above or below the median for Ugandan voters) and gender.[18] All respondents were over 18, registered to vote, and living in the village. We also conducted one key informant survey in each sampled village.

2.5 Empirical Framework

2.5.1 Estimation

Our baseline equation is the following intent-to-treat (ITT) specification:

$$\begin{aligned} Y_{ivp} = \alpha_0 + \alpha_1 Treatment_{vp} + \alpha_2 Spillover_{vp} + \alpha_3 ACFIM\_Presence_p \\ + \alpha_4 ACFIM_{vp} + \Omega X_{ivp} + \varepsilon_{ivp} \quad (2.5.1) \end{aligned}$$

where $Treatment_{vp}$ is an indicator for assignment to the intervention in village v in parish p; $Spillover_{vp}$ is an indicator that village v is untreated but where there is a village in parish p that is treated; $ACFIM\_Presence_p$ is the baseline level of presence of the implementer in the parish; $ACFIM_{vp}$ is an indicator that village v is an eligible village (as opposed to the 1,399 spillover villages); and X_{ivp} is a vector of individual-level controls from the survey.[19] In addition, in the Appendix we report a modified version of equation (2.5.1) that includes strata fixed effects γ_s. We use the same specification for regressions conducted using the polling station-level data – in this case, observations are at the level of polling station j within parish p.

To estimate how the effects of the ACFIM campaign vary with the level of treatment

[18] The voter register for the 2016 election was available for all but two parishes in our sample. In those cases we used the voter register corresponding to the 2011 election. For villages with fewer than 40 individuals listed in the voter register, we included all individuals, irrespective of age and gender.

[19] These controls include, from the survey data, the age, years of education, and marital status of the respondent, whether the household owns any land, the number of adults and children in the household, an index of asset ownership (as defined in Appendix 3), and a set of occupation, ethnicity and religion dummies. From the electoral data, we include the 2011 turnout, the 2011 fraction of the vote received by the incumbent candidate in the corresponding election, and the number of registered voters in 2016.

saturation (at the level of the parish), in every table we report results from the following equation:

$$Y_{ivp} = \gamma_0 + \gamma_1 Saturation_p + \gamma_3 ACFIM\_Presence_p + \gamma_4 ACFIM_{vp} + \Omega X_{ivp} + \varepsilon_{ivp} \quad (2.5.2)$$

Where $Saturation_p$ is defined as the fraction of voters in parish p that are being treated (i.e the intensity of the treatment at the parish level). The main coefficient of interest in this equation is γ_1, which measures the average effect of random treatment saturation across treatment and spillover villages. Note that equation (2.5.2) was not specified in our pre-analysis plan. We present estimates from this equation mainly for ease of exposition, and because we consider the main effect of treatment saturation to also be of interest. Note that this regression specification assumes a constant effect of saturation on both treated and spillover villages – as our results make clear, this is empirically the case for many outcomes. We discuss later why this may be the case.

Finally, to estimate how treatment and spillover effects vary with saturation, we also run the following linear saturation model:

$$\begin{aligned} Y_{ivp} = \beta_0 &+ \beta_1 Treatment_{vp} + \beta_2 Spillover_{vp} + \beta_3 Treatment_{vp} \times Saturation_p \\ &+ \beta_4 Spillover_{vp} \times Saturation_p + \beta_5 ACFIM\_Presence_p + \beta_6 ACFIM_{vp} \\ &+ \beta_7 ACFIM\_Presence_p \times \beta_6 ACFIM_{vp} + \Omega X_{ivp} + \varepsilon_{ivp} \quad (2.5.3) \end{aligned}$$

Estimates from this specification are reported in Table 2.9 for our main outcomes of interest. The two main coefficients of interest here are β_3 and β_4, indicating how the treatment and spillovers effects, respectively, change with treatment saturation at the parish level. The $ACFIM\_Presence_p$ and $ACFIM_{vp}$ terms (in levels and interacted) purge the saturation model from the variation in saturation that comes from (non-randomly assigned) degree of ACFIM presence, giving us causal estimates for the slope effect of treatment and spillover status (i.e., for how treatment and spillover effects vary with the intensity of parish-level saturation). Note there is no main effect of $Saturation_p$ in this specification since all control parishes have zero saturation by design. In addition, the coefficients β_1 and β_2 in this specification are, respectively, simply intercepts for the treatment and spillover groups when parish saturation is zero, and thus do not have a meaningful interpretation.[20]

2.5.2 Dealing with multiple outcomes and comparisons

We sought to reduce the risks of false discovery or cherry picking results in a number of ways. First, we prespecified our hypotheses, estimation framework, and outcomes in a pre-analysis plan.[21]

Second, we singled out one family of outcomes as primary: survey-based reports that candidates gave cash and in-kind gifts to the respondent or other villagers, where we are interested in both the direct effect of treatment and the spillover effect of the ACFIM campaign. In addition, we pre-specified a number of secondary outcomes to shed light on mechanisms behind our primary results, including measures of the aggregate supply and demand for votes at the village level, policy campaigning, vote shares and turnout, as well as attitudinal outcomes.[22]

Finally, we reduced the number of primary hypotheses to test by combining them into mean effects indexes of all outcomes in that family.[23]

2.5.3 Randomization balance

Treatment is generally balanced along covariates. We present randomization checks in Tables 2.10, 2.11, and 2.12. We use a range of baseline or time-invariant variables from the voter survey, key informant survey, and official electoral data – these variables are described in Appendix 3. We regress these variables on our two main specifications, namely equations 2.5.1 and 2.5.2 from section 2.5, and report all the coefficients from these specifications. Of 99 coefficients (from 66 regressions), only 9 (9%) have a p-value less than 0.1 – almost exactly what should have occurred as a result of chance. Nonetheless, in the remainder of the analysis we show that our main results are robust to controlling for baseline covariates.

[20]Equation (3) includes a minor deviation from pre-specified equation (2) in our pre-analysis plan, which had two additional right-hand side terms ($ACFIM\_Presence_p \times Treatment_{vp}$ and $ACFIM\_Presence_p \times Spillover_{vp}$) but did not include the $ACFIM\_Presence_p \times \beta_6 ACFIM_{vp}$ interaction. The results obtained from both specifications are qualitatively similar, but equation (3) above is the correct specification since the previously included terms captured some of the relevant (exogenous) variation.

[21]See **https://www.socialscienceregistry.org/trials/377**, archived on December 18, 2015.

[22]We report experimental results on village inflation in a separate paper.

[23]We take averages of our outcome measures, coded to point in the same direction, akin to the approach by Kling et al. (2007)). Component variables are first standardized, then averaged, then standardized again to have mean zero and unit standard deviation. We do this first for all variables from the voter survey, and then for all the variables in the key informant survey, and then average the two. This gives the two sources of data equal weight.

2.6 Results

Figure 2-1 shows the main effects of the campaign, in terms of standard deviations of key indices, using the saturation specification in equation 2.5.2. The first two effects are the strongest and most robust: the campaign reduced the vote share of incumbents and increased the share of the vote accruing to challengers. The effects on vote-buying are more nuanced – there was no significant effect on incumbent vote-buying, but some evidence for increased challenger vote-buying activities in heavily treated parishes. In addition, there is noisy but sizable evidence for an increase in campaigning, particularly by challengers, in heavily saturated parishes.

Despite having not pre-specified it, we use the saturation specification (equation 2.5.2) as a baseline as we explore these effects in more detail below. As an expositional point, we believe that this specification is easily interpretable: it is the effect of total treatment saturation, which is randomly assigned, conditional on initial ACFIM presence (for which we control). However, this specification is only helpful *because* the effects on treated and spillover villages tend to almost always go in the same direction and be of similar magnitudes. If the effects were off-setting, as we had anticipated, then this specification would have masked important heterogeneity. However, this does not seem to be the case here. As we discuss in more detail later, we believe that this is likely true because candidates and brokers noticed the presence of the ACFIM intervention, but out of a lack of precise information or due to logistical returns to scale in vote buying and policy campaigning, tended to affect the entire parish when they changed activities. Likewise, the ensuing changes in perceptions about the candidates affected voters across the parish, not just voters in the targeted villages, but with an intensity that rose with treatment saturation in the parish.

2.6.1 Compliance and Quality of Implementation

Funding and logistical delays meant that ACFIM implemented the intervention later and more hastily than they originally anticipated, but qualitative data from ACFIM notetakers and our own survey data suggest a reasonably high level of treatment compliance and quality of implementation.

ACFIM estimated that the leaflet was received by 67,374 households across 1,427 targeted villages, or approximately 41% of the total population in these villages (there were

422,110 registered voters in total across all treatment villages).[24] Following the leaflet drop, an estimated 62,566 households participated in at least one meeting, which averaged 30 participants, and 21,390 posters (15 per village) were sent across all treatment villages. Finally, a total of 32,674 automated calls were made on the eve of the election (i.e. on February 17, 2016 between 5pm and 8pm) to individuals who provided their phone number to ACFIM in one of the previous meetings. 18,451 (56%) of these calls were answered according to administrative data provided by the implementing company.

In general, ACFIM administrative notes suggest that the activists implemented the meetings in accordance with their training and the meeting scripts.[25] The survey data tell a similar story. Table 2.2 reports control means and treatment effects on various implementation measures, including treatment and spillover effects from equation (2.5.1) in odd-numbered columns, and treatment effects from parish-treatment intensity from equation (2.5.2) in even-numbered columns. Respondents in treatment villages were 34 percentage points more likely to report an organization with anti-vote buying messages and the presence of leaflets, 29 percentage points more likely to have attended a meeting, and 3 percentage points more likely to have received a call.[26]

The control means in Table 2.2 are nonzero, suggesting that other civic or political organizations were active, as one would expect, but the absence of any statistically significant effects on spillover villages suggest that the villages were generally not experiencing ACFIM's campaign directly. All this is consistent with ACFIM administrative notes. The meetings were largely conducted with people from the village assigned to treatment – only an average of less that two in thirty individuals were outsiders in the sense that they belonged to another

[24]This percentage is estimated from a back-of-the-envelope calculation based on the following figures. Based on the 2014 Ugandan census, the average household had 4.7 members and the fraction of the population under 18 (thus ineligible to vote) was 55%. We validated this using our survey, which found that 37 percent of individuals in treatment villages said they received a leaflet.

[25] The note takers indicated that the facilitators followed the script in almost all of the meetings, and that the facilitators succeeded at conveying the purpose of each meeting. Consistent with the goal of the first meeting, when asked (not exclusively) about the goals of the first meeting that were conveyed to participants by facilitators, note takers indicated that in 73% of cases was the introduction of the campaign and discussion the leaflet content, and in 51% the sharing of the participants' views about on vote buying and selling. The second meeting was a transition meeting designed to provide an avenue for a collective deliberation on vote-buying. There is more variation in what note takers indicated but all meetings are consistent with the intended purpose. Similarly to the first meeting where its goal was clear, in the third meeting note takers indicated that in 61% of cases the conveyed goal was to deliberate on and arrive to a community resolution against vote buying.

[26] We expected a smaller effect on calls received, since calls were only made to individuals who voluntarily shared their cell phone numbers during the anti-vote-buying meetings organized by ACFIM in the village.

village.[27]

Based on data from ACFIM note-takers, in 70% of the village meetings there was at least one influential individual likely to engage in or mediate vote-buying activities, including local councilors, MPs, candidate, and brokers.[28] In 74% of the cases where at least one of such individuals was present, they reportedly tried to influence the meeting. Such participation rates could indicate that these individuals were well aware of the ACFIM campaign and potentially felt threatened by it.

2.6.2 Equilibrium outcomes of the ACFIM campaign

ACFIM's campaign urged villagers to refuse to sell their vote to brokers and candidates. This section looks at equilibrium outcomes of the campaign, with a particular focus on the primary prespecified outcome, the incidence of vote-buying transactions. We also examine a natural intermediary outcome, the impact on attitudes and social norms surrounding vote buying. We see no significant evidence that vote buying transactions decreased overall, although there is heterogeneity between types of candidates, with challengers being more likely to engage in vote-buying transactions both in treatment and spillover villages. Consistent with these effects, we see only a moderate effect of the campaign on social norms.

Vote buying transactions

Tables 2.3 and 2.4 report the effects on self-reported receipts of cash and other gifts in exchange for a vote, measured across all candidates and disaggregated across different types of candidates running for the Presidential and the Parliamentary election. We focus on these elections because these two offices are the ones that entail the largest access to public funds, and thus resources invested in vote-buying. We look at vote-buying measured across all candidates and across two types of candidates: incumbents,[29] and challenger (i.e., non-incumbent) candidates. For the purpose of this categorization, vote tallies are computed at the national level for the presidential election, and at the constituency level for

[27] Importantly, the share of outsiders across meetings was constant, which lessens the concern of cumulative effect characteristic of significant spillovers.

[28] If we use a less stringent definition of influential individual likely to engage in or mediate vote-buying activities, in almost all village meetings such an individual was present.

[29] For incumbent MPs, the identity of incumbents was obtained from a fuzzy match by district, constituency, name and party between the official election results (obtained from the Electoral Commission) and the list of MPs in the 2011-2016 Ugandan Parliament. When no incumbent candidate was found in a particular constituency, incumbent status was defined by party (i.e., if the seat of constituency A had been occupied by party B, the candidate running under the banner of party B was defined as the incumbent, except if party B is "Independent").

the parliamentary election.[30] We expected that vote buying would fall in treated villages, and potentially rise in spillover villages, with both effects increasing in saturation levels.

Table 2.3 reports the impacts of treatment on a standardized index of 4 variables capturing the prevalence of vote buying: whether the respondent received any gift in cash, the log of 1 plus the amount of cash received, whether the respondent received any gift in kind, and the log of 1 plus the value of gifts received in kind (disaggregated across the two types of candidates). Coefficient signs are generally in the opposite direction of what we expected, although the magnitudes are generally small and not statistically significant. When pooling all races and candidates, reports of vote buying transactions increase by 0.034 standard deviations in treated villages, and rise by 0.015 standard deviations in spillover villages. We can effectively rule out medium and large effects of the ACFIM campaign on these transactions. The coefficients on the parish saturation interaction term is positive – the opposite of what was expected – and non-significant.

To further explore heterogeneity across candidates, Table 2.5 illustrate campaign effects on the two main components of the index: a dummy variable indicating whether the survey respondent reported receiving cash, summed up across races and candidates (in part 1), and the log amount of cash (plus 1 Ush, to avoid dropping zeros) received in exchange for votes (in part 2).[31,32] The two tables tell a consistent story. First, across the board there is no evidence that the ACFIM campaign reduced the incidence of vote-buying for any type of candidate. Second, there is no evidence that the campaign affected vote-buying by incumbents (column (3) and (4) of both tables), as we cannot reject null average effects, nor that parish-level saturation had a null effect on vote-buying. Third, there is some evidence that the campaign *increased* vote-buying by challenger candidates (columns (5)-(6)). We find that these effects are positive on average across treatment and spillover villages, and increasing with parish-level saturation. The magnitude of these effects is not negligible, with challengers spending 25% more in fully treated parishes than control parishes.

It is noticeable that, when looking at vote-buying by challengers, the significant effect of

[30] These definitions make it unlikely that the coding of challenger candidates corresponding to each race is affected by our treatment.

[31] In our survey data, we collected data on all individuals (brokers) who approached the respondent to give her a gift in exchange for her vote, as well as the identity of the candidates these brokers were working for. A respondent is coded as having received a gift from a particular candidate if she mentioned this candidate among the individuals the brokers were working for.

[32] We do not condition on receiving a positive amount of money in these estimates, so they should not be interpreted as price effects, but rather as effects on average amount received (including both the intensive and extensive margins).

the campaign is similar in treatment and spillover villages. This suggests the possibility that, when challengers buy votes in a parish, they do so in all villages within the parish, possibly due to logistical returns to scale in vote buying. This would imply that most of the action takes place along the extensive margin, i.e., challengers entering treated parishes and the villages in those parishes. To explore this implication, Tables 4C and D, respectively, show as an outcome whether a candidate operates in a village or parish. These estimates sum across the presidential and MP races, so the variables in question are counts of candidates (note that there can be at most two incumbents (one president and one MP), but many more potential challengers). The strongest results are for challengers (about 0.13 more challengers operating in the parish), though they do not reach conventional levels of significance.

Attitudes and social norms

For a campaign such as ACFIM's to be effective, the campaign must succeed in changing perceived social norms and people's attitudes towards vote buying. Our survey measured a handful of perceived norms and attitudes. We report treatment effects of the campaign on these variables in Table 2.5. The estimates in this table suggest that a majority of respondents already held the belief that vote-buying had negative consequences and imposed a social cost on their villages, and that the ACFIM campaign increased these attitudes slightly. It also increased the expectation of social sanctions as a result of vote buying, a key way in which social norms are enforced by the community. However, across all of our measures, the changes in beliefs tend to be small, in part because many respondents in control villages already held anti-vote-buying beliefs. In addition, one caveat to this analysis is that our measures are self-reported. Whether individuals condemn or claim to have ostracized vote-sellers might be subject to social desirability bias, particularly in treated villages.

Column (1) in Table 2.5 reports effects on whether survey respondents thought exchanging money for votes had negative consequences for the village. 89% already agreed with this statement in control villages, and this increased by just 1.3 percentage points in treated villages (significant at the 10% level). This effect is also increasing with parish-level saturation (column (2)). Turning to perceived social sanctions, we asked whether people in the village would be understanding towards, or angry at, the respondent for selling his or her vote (columns (3)-(4)). Similarly, we asked whether a person selling his or her vote would be ostracized by the rest of the village (columns (5)-(6)). Across these columns, treatment increased perceived sanctions on average by roughly 2 percentage (columns (3) and (5)). The corresponding coefficients on parish-level saturation are positive but not statistically significant (columns (4) and (6)).

In addition, we conducted a vignette experiment collecting respondents' perceptions of a hypothetical hard-working man in financial distress selling his vote to provide for his household.[33] 73% of control villagers agreed that this was totally unacceptable. This does not rise significantly in the average treated village, nor with parish-level saturation (columns (7)-(8)). Finally, columns (9)-(10) test whether the campaign affected a measure of self-reported vote-buying – whether the respondent reported to vote for a candidate she also accepted a vote-buying offer from. The campaign had no significant effects on this variable.

2.6.3 Program impacts on electoral outcomes

Program impacts on candidates' vote shares

In Table 2.6, we report treatment effects on the electoral performance of incumbent and challenger candidates, using self-reports from the voter survey (in Table 6A) as well as administrative data (in Table 6B). The vote shares are z-standardized and results are pooled across the Presidential and Parliamentary races (the results for the individual races are similar). Regressions conducted using the survey data are run at the individual level, while regressions using the electoral data are run at the polling-station level. As above, we report the coefficients from equation (2.5.1) in odd-numbered columns and those from equation (2.5.2) in even-numbered columns. In both tables, we report outcomes computed across 2 types of candidates: incumbents and non-incumbents (labelled "All challengers").

As a cautionary note, whether one should put more weight on the self-reported data or the administrative data is a priori unclear. The self-reported data could be subject to social desirability bias. However, if anything, this bias should be directed towards incumbents and thus is not a source of major concern given our findings. Moreover, since there is some measurement error in the administrative data (coming from the fact that polling stations to which treated villages correspond had some voters from non-treated villages), self-reported

[33] The exact phrasing of the vignette experiment was: "Imagine a man who lives with his wife and children in this village. He works hard, but he frequently has trouble maintaining his family economically. During the electoral campaign, a member of a party offered him a certain amount of cash (with the actual amount randomly determined across respondents) so that he would vote for the party. The man accepted the money and voted as he was instructed. In your opinion, was the behavior of this man: completely acceptable, wrong but understandable, or totally unacceptable?"

data may yield more precise estimates of actual voting behavior.[34]

Overall, the estimates in Table 2.6 suggest that the ACFIM campaign led to a decrease in the electoral performance of incumbents in treated villages, which is statistically significant in the survey data (Table 6A, column (1)) but not in the administrative data (Table 6B, column (1)). Across both data sources, there is robust evidence that support for incumbents falls in heavily treated parishes (columns (2)). These estimates suggest two important takeaways. First, the campaign *negatively* affected the electoral performance of incumbents. These results are opposite to those in Vicente (2014), where a comparable anti-vote-buying campaign increased support for incumbent candidates on average. Second, this negative impact is driven by parishes with a high degree of treatment saturation. This could reflect several mechanism being at play, which we discuss later in Section 2.7.

In the remaining columns of Table 2.6, we show effects on the electoral performance of challenger candidates. These estimates are, as expected, symmetric to those observed for incumbents. The estimates obtained using the survey data (Table 6A) show a significant increase in support for challengers in both treated and spillover villages (column (3)) and in heavily saturated parishes (column (4). Results obtained using the administrative data are similar, but only statistically significant in column (4).

Program impacts on voter turnout

Table 2.7 presents treatment effects on voter turnout, measured using administrative data. Turnout is z-standardized and pooled across the Presidential and Parliamentary races in columns (1)-(2), and shown in levels for each race in columns (3)-(6). We do not report results on self-reported turnout given the implausibly high turnout in our survey data (95% for the Presidential election and 93% for the Parliamentary election). Turnout in the administrative data was 67% for the Presidential poll and 69% for the Parliamentary elections.

We see some evidence that the ACFIM campaign moderately increased voter turnout, when pooling across both races – the effect size for treated villages is approximately 0.07 SDs (not statistically significant). Looking at each race individually, the point estimates for the average treatment effect and the average spillover effect are less than one percentage point. Parish-level saturation has a positive effect on turnout, but this effect falls short of

[34]Lastly, there were allegations of vote fraud, on which the campaign could have an effect. However, we discuss and dismiss this possibility in Appendix 2.

statistical significance (columns (2), (4) and (6)).

2.6.4 Program impacts on campaigning

To explore the possibility that candidates substituted to other (non-vote-buying) strategies in response to the ACFIM intervention, in Table 2.8 we look at whether treatment status, spillover status, and parish-level saturation increased the occurrence of other forms of political campaigning, which was another of our pre-registered secondary outcomes. In our survey data, the three campaigning methods most cited by voters were displaying political posters in the village (cited by 87% of respondents across all ballots and candidates), village visits (62%), and campaigning through loudspeakers, SMS or phone calls (39%). The outcomes in this table are the sum of campaign activities (out of 5 possible activities: visiting the village, putting up posters, distributing leaflets, campaigning via loudspeakers/SMS, and distributing merchandise) conducted by each type of candidate in our sample villages, and averaged across Presidential and Parliamentary races.

Table 2.8 provides some evidence that some politicians (namely challengers) adapted their campaigning strategies in response to our intervention. For incumbent candidates, campaign effects in treatment villages are positive, but not significant, and smaller in magnitude, while spillover effects are negative and non-significant (column (3)). The coefficient on parish-level saturation is positive but also non-significant (column (4)). Challenger candidates (columns (5)-(6)) seem to increase their campaigning efforts in response to the ACFIM intervention: for example, full treatment saturation leads to a 0.5 SD increase in campaigning effort relative to zero-saturation parishes (column (6)). Overall these estimates suggest that, contrary to our expectations, vote-buying and policy-campaigning strategies are complements rather than substitutes. This is also supported by the positive correlation between the two variables in control parishes.

2.7 Discussion

Our main results show that the ACFIM campaign had a large effect on vote shares, and some effects on vote-buying and campaigning by candidates. In particular, challenger candidates appear to invest more in policy-campaigning and to engage more in vote-buying. In addition, total turnout rose slightly. Moreover, these results are largely driven by parishes with high treatment saturation, with little difference between the point estimates for treatment and spillover. So, any theoretical explanation behind these findings must account for effects that permeate the entire parish, rather than solely treated villages.

We believe that one of two complementary theoretical explanations can account these results when focusing on the supply side of votes – the voters. First, the treatment shifted voters to reject vote-buying offers or to no longer think about them as binding contracts, and vote for the candidate they deemed better suited for office. The latter mode of acting is described by the Ugandan adage, "Eat widely, vote wisely," which was adopted as the official resolution in 30% of villages that reported an official resolution. In any case, by convincing voters that they should be free to vote for their preferred candidate, as opposed to the highest-bidding candidate, the ACFIM campaign substantially reduced the advantage of incumbents, who generally have a stronger local presence and can afford to offer more money to voters.

The second alternative is that the campaign served, inadvertently, to coordinate voters on an anti-incumbent message. The framing of the campaign was about the pernicious effect of vote-buying (a practice more commonly used by incumbents) on public-service delivery (a responsibility at which most incumbents are perceived to fail). As a result, notwithstanding the non-partisan language of the campaign, the public meetings on this topic could have shifted local beliefs against incumbents.

Crucially, regardless of the driving force behind the voters' change of attitudes, subsequent to these changes (which favored challengers) candidates as well the brokers working for them responded by shifting their vote-buying and policy-campaigning efforts across parishes. In particular, we observe a larger increase in vote-buying and policy-campaigning by challengers. Importantly, this increase occurs *throughout* the treated parishes (in treated and spillover villages) and rises further with higher treatment saturation. We also see that these effects are largely driven by challenger or their brokers starting operation in villages and parishes where they would have not operated absent the campaign. These effect suggests that there are local returns to scale in vote-buying and policy-campaigning – due to fixed costs of operating in a parish, once a decision to enter the parish is made, candidates and brokers operate in all villages in the parish. Our qualitative accounts indicate that village-level brokers are recruited and managed by higher-level brokers who operate at the parish level, which supports the presence of such fixed costs. This vote-buying and policy-campaigning technology then accounts for the similar magnitude of the treatment effect in treatment and spillover villages: the comparable effects reflect changes in candidate behavior in the parish overall, rather than direct spillovers across villages.

Lastly, we rule out that several alternative explanations can fully account for these results. First, it is possible that the campaign did diminish vote-buying, but, contrary to most expectations about the effect of social desirability bias, induced people to more honestly report vote-buying in their villages, which yields a zero or positive effect on *reported* vote-buying. This does not seem to be the case. For instance, we find no significant effect of the campaign on self-reported vote-buying in the 2011 election (results available upon request). This is likely to be a good test for social desirability bias, since the 2011 election pre-dated the campaign and under random assignment there should be no relationship between treatment and 2011 vote-buying, except through a social desirability or salience channel. In addition, our results on attitudes about vote-buying suggest that the campaign intensified negative feelings about vote-buying (though only by a small amount).

Second, it is possible that the null results on total vote-buying are due in part to agency problems between candidates and their brokers. Interviews with elected candidates and focus groups indicate that brokers are subject to significant moral hazard and that the combination of imperfect monitoring ability and competition for brokers around elections allows them to extract large rents. Candidates often provide lump sums of cash or other gifts to brokers with which to secure the support of villages under their influence. This is part of a performance contract wherein these brokers are expected to deliver a certain electoral support to the candidates. Brokers who fail to reach these targets lose their position in future elections, as well as their connections with winning candidates. Brokers then do solve a cost-minimization problem to achieve that target and keep the remaining resources for themselves. If brokers responded to the campaign by increasing the fraction of the money that they spent on voters (reducing what they kept for themselves) in an effort to overcome a weakening of the traditional vote-buying arrangements, this would have undone some of the effects of the campaign on vote-buying. However, this should have been particularly true for vote-buying by incumbents (who were facing an actual loss of support), which we do not observe.[35] In addition, importantly, this story does not explain the increase in challenger

[35]Nor do we observe a bimodal distribution of gift-giving, where some brokers give up and others redouble their efforts, which could also occur.

vote-buying and policy-campaigning.

Third, it is possible that the campaign deterred electoral fraud that otherwise would have favored the incumbents by engaging citizens in the electoral process. As noted in Appendix 2, however, there is no evidence that the campaign was related (positively or negatively) to the presence of markers for electoral irregularities. Although these tests are imperfect, their outcome is also supported by the results that instead focuses on self-reported data as an outcome. Overall, we take this as strong evidence that this cannot be a primary driver of our results.

2.8 Conclusion

This paper documents the effect of the largest anti-vote-buying campaign ever evaluated – with almost half a million voters treated across nearly 1,500 treated villages in Uganda – on vote- buying, policy-campaigning, and electoral outcomes. We found that the campaign, in spite of its relatively heavy footprint – leaflets, three village meetings, and a village-wide resolution – was not effective at preventing vote-buying, but we provide suggestive evidence that it did free people from traditional vote-buying relationships. As a result, voters were less likely to vote for incumbent candidates and more likely to vote for the main challengers. These effects were large, especially in heavily treated parishes, enough to reverse the position of the average incumbent and challenger in parishes with high saturation.

Our results on the prevalence of vote-buying runs counter to previous experimental evidence on such campaigns, as in Hicken et al. (2014) and Vicente (2014), both of whom find that comparatively less intensive interventions had sizable impacts on votes sold. We believe that the differences between our findings and those previously reported in the literature are best explained by the difference in scope and, inadvertently, in message, between our experiment and those cited above. The large scale and high degree of publicity of the ACFIM campaign, as well as the fact that local political brokers attended the community meetings intended to coordinate citizens' efforts against vote-buying, prompted candidates to respond to the ACFIM campaign. In addition, the common local decision to decide to "eat widely, vote wisely" (the public context reduced the ability of the campaign to control the precise message) meant that the effects may have shifted away from changes in vote-buying levels and towards changes in voting decisions *conditional* on the vote-buying offer they accepted.

However, our results indicate that it may be possible to disrupt its effectiveness by unmooring the relationship between vote-buying and voter behavior. In a dynamic game, where candidates seek to use the most cost-effective methods of gaining voter support, this breakdown in voter willingness to honor the vote-buying "contract" should induce candidates to shift towards other methods in future elections.

Thus, in future elections, and during the current tenure of the newly elected officials, we might expect this result to change candidate behavior. In particular, we believe that these results may induce candidates to emphasize and keep promises of future public goods rather than vote-buying, which could have substantial impacts on governance in Uganda. Future research should continue to examine this aspect of candidate message-optimization and its implications for electoral and economic outcomes.

In terms of welfare, as with any intervention around elections, the effects are difficult to estimate and a full accounting is beyond the scope of this paper. Previous research (e.g. Besley et al. (2010)) suggests that increasing the competitiveness of local elections improves the quality of governance. In this sense, since the campaign appeared to relatively advantage challengers, we might expect it to have positive effects. In addition, since total amount received by voters did not fall, it does not appear as though voters had short-term costs in foregone vote-buying offers.

In addition, future work should continue to explore how to break down the vote-buying equilibrium. Our results highlight that one-sided interventions of large scale and visibility are likely to fail to eradicate vote-buying if candidates respond to them. Future work would ideally then target both candidates and voters for treatment. In particular, in addition to tackling vote-selling, as we did in our intervention, there is the need to convince candidates to pledge not to buy votes to then undermine the demand for vote-buying. These efforts are politically sensitive and thus would need to be taken by a local organization with strong connections to multiple political parties, but could yield important insights about the relative merits of intervening on the demand, as opposed to simply the supply side of the votes' market.

We believe that this paper opens new avenues of research on both vote-buying and on campaigning in low development countries more broadly. This remains a fruitful area for more work, with important policy implications and potential for contributions to our knowledge about voter behavior and governance.

Appendix

Appendix 1: Discussion of external validity

The presence of a local ACFIM activist is clearly non-random. Our treatment randomization was within the sample of parishes/villages with local ACFIM activists, so this is *not* a problem for internal validity, but it does require a brief discussion on external validity. The first note on external validity is that, from the perspective of civil society organizations considering similar campaigns, the villages/parishes with pre-existing civil society presence may, in fact, be the policy relevant sample. The strength of CSOs often lies in their local credibility, built over multiple years and sustained through the presence of local members of the larger national CSO, so few CSOs would launch a campaign in villages (or parishes) to which they had never been. However, it is still worth noting the differences.

First, to be in our sample, a parish must contain at least 1 village where a local ACFIM activist works or lives. Since we do not survey any parishes with zero ACFIM presence, we cannot compare our sample directly to other parishes. However, we can correlate the degree of ACFIM presence (i.e. the percent of voters in a given parish who live in villages with ACFIM presence) with covariates to explore this selection indirectly. As we might expect, ACFIM presence is correlated with lower vote-share in 2011 for the incumbent president – a parish with 100% ACFIM presence voted for the incumbent by 7 percentage points fewer, on average, than one with 0% ACFIM presence. In our survey, we asked voters whether they received a gift for their vote in 2011 (the prior election). Again, as we might expect, ACFIM presence is correlated with less prior vote-buying: using the same 100% to 0% comparison, full ACFIM presence is correlated with a 5 percentage lower share of respondents reporting receiving a gift in 2011.

Second, within each parish, we sample every village where an ACFIM activist had the potential to work (whether in the treatment or control). However, in addition, we sampled 1,399 additional villages in the same parishes that were ineligible for treatment, but where we could look at spillovers. Throughout the analysis, we control for a dummy indicating that a village was not part of the experimental sampling frame. As can be seen in the results later, this dummy is usually insignificant, indicating that these villages do not generally differ from the untreated villages that were part of the experimental sample, though in some specifications a small difference appears.

Appendix 2: Electoral data integrity

Opposition leaders in Uganda and international observers challenged the integrity of the voting data in the aftermath of the election (All Africa, 2016; Daily Mail, 2016; Newsweek, 2016). Analysts noted several potentially suspicious patterns. We acknowledge these issues, but believe that the electoral data can still be useful for our analysis for several reasons. First, we generally obtain similiar results using self-reported voting outcomes from our voter survey and using the official election data. Second, we show in the Appendix that our treatment is uncorrelated with traditional markers of electoral malfeasance (Beber and Scacco, 2012).

Appendix 3: Variables used for Randomization Checks

From the voter survey, we use the age, years of education, marital status (a dummy variable for married individuals), land ownership (a dummy for households that own any land), the number of adults and children in the household, an index of asset ownership,[36] four measures of occupational status (dummy variables for individuals working in farming, trade/retail, any high-skill activity, or not actively working),[37] dummy variables indicating the individual belongs to one of Uganda's three largest ethnic groups (Ganda, Nkole and Soga), and three dummy variables for being a Catholic, a Protestant, or a Muslim. From the key informant survey, we use the years of education and marital status of the respondent, as well as the same four measures of occupational status, ethnicity and religion as above (note that age, land ownership, number of members in the household and assets were not collected in the key informant survey), as well as four dummy variables for whether the key informant is a local chief or elder, a member of a civil society group (a religious, youth, or women's group), a village committee member or a local council member. Finally, from the official electoral data we use the number of valid votes cast in 2011, the voter turnout in 2011, the vote shares of the NRM and of the FDC in 2011, and the number of registered voters in 2016.

[36] To construct this index, we simply add up dummy variables indicating ownership of a TV, radio, motor vehicle, and cell phone

[37] High-skill individuals include artisans or skilled manual workers, clerks and secretaries, supervisors, managers, security providers, mid-level professionals such as teachers, and upper-level professionals. Individuals not actively working include students as well as unemployed, retired, and disabled individuals.

2.9 Tables and Figures

2.9.1 Figures

Figure 2-1: Main treatment effects

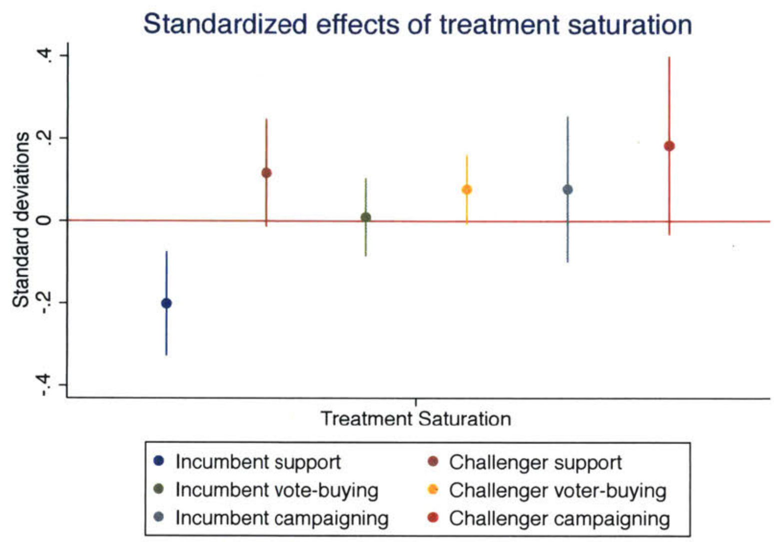
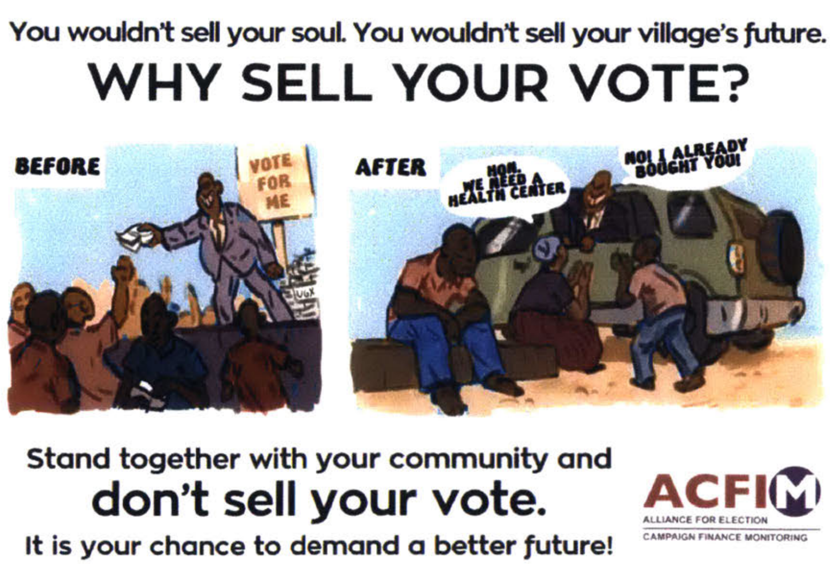


Figure 2-2: Sample leaflet used in experiment



2.9.2 Tables

Table 2.1: Summary Statistics

	Mean	SD	N
Survey data			
Recalls NGO visit in village	.324	.468	27807
Received a leaflet	.172	.377	28060
Recalls meetings took place	.129	.335	27755
Attended meeting	.207	.651	27745
Received a robo-call	.053	.224	28507
Recalls posters	.129	.335	28133
Negative consequences	.895	.306	28507
People angry	.756	.43	28507
Vote sellers ostracized	.579	.494	27732
Vote-buying unacceptable	.744	.437	28501
Any cash received, any candidate	.4	.49	28507
Any cash - incumbents	.331	.578	28507
Any cash - challengers	.111	.321	28507
Cash amount received (USh)	1526.1	4269.3	28507
Cash amount - incumbents	1004.0	2864.7	28507
Cash amount - challengers	697.8	2668.5	28507
Reported vote for incumbent	.657	.349	27112
Campaign activities, all candidates	5.901	4.246	28507
Campaign activities - incumbents	3.504	2.536	28507
Campaign activities - challengers	2.397	2.25	28507
Administrative data			
Registered Voters	574.0	202.9	3659
Turnout 2016 - President	.675	.09	3659
Turnout 2016 - MP	.689	.086	3112
Incumbent vote share 2016 - President	.614	.184	3654
Challengers vote 2016 - President	.386	.184	3654
Incumbent vote share 2016 - MP	.441	.246	3104
Challengers vote share 2016 - MP	.559	.246	3104
Turnout 2011 - President	.601	.103	3641
Incumbent vote share 2011 - President	.678	.186	3641

Table 2.2: Quality of Implementation

	NGO visit		Received leaflet		Meetings Attended		Received call		Posters	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment village	0.335***		0.338***		0.290***		0.029***		0.189***	
	[0.011]		[0.009]		[0.016]		[0.004]		[0.009]	
Spillover	0.018		0.007		0.002		-0.004		0.003	
	[0.011]		[0.006]		[0.013]		[0.004]		[0.006]	
Treatment Saturation		0.416***		0.433***		0.357***		0.035***		0.250***
		[0.021]		[0.017]		[0.026]		[0.007]		[0.016]
Outside Sampling Frame	-0.018	-0.216***	-0.006	-0.216***	-0.023*	-0.204***	0.005	-0.016***	0.008	-0.110***
	[0.011]	[0.011]	[0.006]	[0.009]	[0.013]	[0.012]	[0.005]	[0.003]	[0.007]	[0.007]
ACFIM Presence	0.022	-0.217***	-0.001	-0.250***	-0.014	-0.221***	0.007	-0.014*	0.025*	-0.119***
	[0.018]	[0.024]	[0.015]	[0.019]	[0.026]	[0.031]	[0.007]	[0.008]	[0.015]	[0.016]
R^2	0.14	0.10	0.20	0.14	0.06	0.04	0.04	0.04	0.09	0.07
Control Mean	0.198	0.198	0.052	0.052	0.113	0.113	0.040	0.040	0.062	0.062
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	27756	27756	28007	28007	27693	27693	28454	28454	28081	28081

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by parish in brackets. All regressions control for an ACFIM dummy (in-sample villages) and the parish-level ACFIM presence. dependent variables in this table are self-reported indicators of program implementation: whether the NGO visited, distributed leaflets, held meetings, conducted robocalls, or posted signs.

Table 2.3: Treatment Effects on Vote-Buying (Z-Standardized)

	All Candidates		Incumbents		All Challengers	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment village	0.034		0.005		0.059**	
	[0.024]		[0.026]		[0.023]	
Spillover	0.015		-0.011		0.045*	
	[0.025]		[0.029]		[0.027]	
Treatment Saturation		0.052		0.012		0.084*
		[0.044]		[0.047]		[0.043]
Outside Sampling Frame	-0.011	-0.022	-0.000	-0.012	-0.017	-0.022
	[0.025]	[0.018]	[0.028]	[0.019]	[0.026]	[0.016]
ACFIM Presence	-0.006	-0.034	-0.060	-0.067	0.070	0.025
	[0.043]	[0.046]	[0.047]	[0.051]	[0.044]	[0.046]
R^2	0.06	0.06	0.06	0.06	0.04	0.04
Control Mean	-0.008	-0.008	0.006	0.006	-0.024	-0.024
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	28454	28454	28454	28454	28454	28454

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by parish in brackets. All regressions control for an ACFIM dummy (in-sample villages) and the parish-level ACFIM presence. These dependent variables are standardized index of the following variables: any cash received, natural log of the amount of cash received, any gift received, and log of the value of any gift received. These dependent variables are restricted to the Presidential and Parliamentary (MP) races.

Table 2.4: Treatment Effects on Vote-Buying: Cash Received

4A: Any cash received (individual level)

	All Candidates		Incumbents		All Challengers	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment village	0.020		0.002		0.018*	
	[0.019]		[0.015]		[0.009]	
Spillover	0.026		0.003		0.023**	
	[0.021]		[0.017]		[0.010]	
Treatment Saturation		0.049		0.019		0.030*
		[0.034]		[0.026]		[0.017]
Outside Sampling Frame	-0.012	-0.007	0.001	0.001	-0.013	-0.008
	[0.020]	[0.014]	[0.015]	[0.011]	[0.009]	[0.006]
ACFIM Presence	-0.052	-0.077**	-0.051**	-0.061**	-0.001	-0.016
	[0.033]	[0.036]	[0.026]	[0.029]	[0.017]	[0.017]
R^2	0.13	0.13	0.11	0.11	0.06	0.06
Control Mean	0.430	0.430	0.327	0.327	0.102	0.102
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	28454	28454	28454	28454	28454	28454

4B: Log cash received (individual level)

	All Candidates		Incumbents		All Challengers	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment village	0.013		-0.008		0.129*	
	[0.096]		[0.089]		[0.076]	
Spillover	0.094		0.019		0.214**	
	[0.109]		[0.101]		[0.087]	
Treatment Saturation		0.104		0.061		0.247*
		[0.162]		[0.149]		[0.140]
Outside Sampling Frame	-0.056	-0.002	0.032	0.045	-0.115	-0.045
	[0.100]	[0.071]	[0.094]	[0.067]	[0.078]	[0.047]
ACFIM Presence	-0.424**	-0.472***	-0.402***	-0.433**	0.020	-0.101
	[0.166]	[0.179]	[0.152]	[0.168]	[0.140]	[0.142]
R^2	0.12	0.12	0.10	0.10	0.07	0.07
Control Mean	2.580	2.580	2.136	2.136	1.117	1.117
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	28454	28454	28454	28454	28454	28454

Treatment Effects on Vote-Buying: Cash Received (continued)

4C: Any cash received (village level)

	All Candidates		Incumbent		All Challengers	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment village	0.023		-0.002		0.042	
	[0.058]		[0.044]		[0.030]	
Spillover	0.086		0.002		0.091***	
	[0.065]		[0.047]		[0.034]	
Treatment Saturation		-0.002		-0.031		0.048
		[0.112]		[0.083]		[0.056]
Outside Sampling Frame	-0.130**	-0.078**	-0.052	-0.046*	-0.082***	-0.041**
	[0.057]	[0.031]	[0.040]	[0.025]	[0.030]	[0.016]
ACFIM Presence	-0.351***	-0.342***	-0.330***	-0.313***	-0.082	-0.101*
	[0.100]	[0.105]	[0.073]	[0.079]	[0.051]	[0.054]
R^2	0.04	0.04	0.04	0.04	0.04	0.03
Control Mean	1.216	1.216	0.919	0.919	0.372	0.372
Observations	4111	4111	4111	4111	4111	4111

4D: Any cash received (parish level)

	All Candidates	Incumbent	All Challengers
	(1)	(2)	(3)
Treatment Saturation	0.159	0.103	0.127
	[0.144]	[0.100]	[0.078]
ACFIM Presence	-0.181	-0.233**	0.077
	[0.136]	[0.097]	[0.076]
R^2	0.03	0.04	0.04
Control Mean	1.627	1.208	0.545
Observations	909	909	909

Table 2.5: Effects of the Campaign on Attitudes Towards Vote-Buying

	Neg Consequences		Social Cost		Ostracize		Vignette		Vote Choice	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment village	0.013*		0.023**		0.022*		0.004		-0.004	
	[0.007]		[0.011]		[0.012]		[0.011]		[0.011]	
Spillover	0.010		0.015		-0.005		0.018		0.007	
	[0.008]		[0.013]		[0.015]		[0.014]		[0.013]	
Treatment Saturation		0.030***		0.029		0.014		0.025		0.006
		[0.011]		[0.018]		[0.020]		[0.020]		[0.018]
Outside Sampling Frame	-0.012	-0.014**	-0.008	-0.011	-0.004	-0.021**	-0.002	0.007	-0.005	0.002
	[0.008]	[0.006]	[0.013]	[0.009]	[0.014]	[0.010]	[0.014]	[0.009]	[0.011]	[0.008]
ACFIM Presence	-0.019*	-0.035***	0.005	-0.010	-0.003	-0.012	-0.053***	-0.065***	-0.053***	-0.056***
	[0.011]	[0.013]	[0.017]	[0.020]	[0.020]	[0.023]	[0.020]	[0.021]	[0.018]	[0.020]
R^2	0.04	0.04	0.04	0.04	0.07	0.07	0.02	0.02	0.10	0.10
Control Mean	0.888	0.888	0.745	0.745	0.567	0.567	0.733	0.733	0.242	0.242
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	28454	28454	28454	28454	27680	27680	28448	28448	28454	28454

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by parish in brackets. All regressions control for an ACFIM dummy (in-sample villages) and the parish-level ACFIM presence. Dependent variabless in this table include (1) an indicator of any negative consequences to participating in vote-buying, (2) an indicator of any social cost, (3) whether others would ostracize vote-buying participants, (4) whether the respondent approved of vote-buying behavior in a vignette survey experiment , and (5) whether the respondent accepted a vote-buying offer.

Table 2.6: Index of Treatment Effects on Vote Shares (Z-Standardized)

6A: Electoral Support, Survey Data

	Self-Report (Incumbent)		Self-Report (All Challengers)	
	(1)	(2)	(3)	(4)
Treatment village	-0.087**		0.087**	
	[0.038]		[0.038]	
Spillover	-0.083**		0.083**	
	[0.041]		[0.041]	
Treatment Saturation		-0.236***		0.236***
		[0.071]		[0.071]
Outside Sampling Frame	0.012	0.018	-0.012	-0.018
	[0.033]	[0.019]	[0.033]	[0.019]
ACFIM Presence	-0.234***	-0.108	0.234***	0.108
	[0.067]	[0.073]	[0.067]	[0.073]
R^2	0.01	0.01	0.01	0.01
Control Mean	0.052	0.052	-0.052	-0.052
Controls	Yes	Yes	Yes	Yes
Observations	27065	27065	27065	27065

6B: Electoral Support, Administrative Data

	Electoral (Incumbent)		Electoral (All Challengers)	
	(1)	(2)	(3)	(4)
Treatment Polling Station	-0.112		0.112	
	[0.069]		[0.069]	
Spillover Polling Station	-0.003		0.003	
	[0.082]		[0.082]	
Saturation		-0.271**		0.271**
		[0.132]		[0.132]
Outside Sampling Frame	-0.116***	-0.062**	0.116***	0.062**
	[0.041]	[0.029]	[0.041]	[0.029]
ACFIM Presence	-0.700***	-0.571***	0.700***	0.571***
	[0.130]	[0.145]	[0.130]	[0.145]
R^2	0.03	0.03	0.03	0.03
Control Mean	0.041	0.041	-0.041	-0.041
Controls	Yes	Yes	Yes	Yes
Observations	3657	3657	3657	3657

Table 2.7: Campaign Effects on Turnout

	Turnout (z-index)		Presidential Election		Parliamentary Election (MP)	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment Polling Station	0.065		0.005		0.004	
	[0.045]		[0.004]		[0.004]	
Spillover Polling Station	0.017		0.002		-0.000	
	[0.050]		[0.005]		[0.005]	
Treatment Saturation		0.146		0.013		0.011
		[0.097]		[0.008]		[0.009]
Outside Sampling Frame	-0.103***	-0.127***	-0.010***	-0.012***	-0.008**	-0.010***
	[0.038]	[0.029]	[0.004]	[0.003]	[0.004]	[0.003]
ACFIM Presence	-0.076	-0.144	-0.008	-0.014	-0.009	-0.014
	[0.091]	[0.103]	[0.008]	[0.009]	[0.009]	[0.010]
R^2	0.33	0.33	0.31	0.31	0.30	0.30
Control Mean	-0.008	-0.008	0.674	0.674	0.690	0.690
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3659	3659	3659	3659	3112	3112

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by parish in brackets. All regressions control for an ACFIM dummy (in-sample villages) and the parish-level ACFIM presence. The first dependent variable is a standardized index of the following dependent variables: presidential and parliamentary turnout. Turnout is defined as valid votes divided by registered voters.

Table 2.8: Index of Treatment Effects on Campaigning (Z-Standardized)

	All Candidates		Incumbents		Primary Challenger	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment village	0.228		0.117		0.112	
	[0.194]		[0.104]		[0.112]	
Spillover	-0.186		-0.056		-0.130	
	[0.215]		[0.115]		[0.124]	
Treatment Saturation		0.776**		0.282		0.494**
		[0.386]		[0.200]		[0.227]
Outside Sampling Frame	0.249	-0.080	0.113	-0.015	0.136	-0.065
	[0.177]	[0.100]	[0.100]	[0.060]	[0.097]	[0.050]
ACFIM Presence	-0.682*	-1.130***	-0.371*	-0.535***	-0.311	-0.596***
	[0.368]	[0.369]	[0.191]	[0.195]	[0.211]	[0.214]
R^2	0.11	0.11	0.12	0.12	0.09	0.09
Control Mean	5.759	5.759	3.400	3.400	2.359	2.359
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	28454	28454	28454	28454	28454	28454

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by parish in brackets. All regressions control for an ACFIM dummy (in-sample villages) and the parish-level ACFIM presence. All variables in this table include only Presidential and Parliamentary (MP) races. These dependent variables are the sums of indicators of campaiging activities: visit to the village, posters, leaflets, advertising over loudspeakers, and merchandise.

Interactions on Key Outcomes (continued)

	Any Cash Received (Table 4A)			Log Cash Amount (Table 4B)			Extensive Margin, Villages (Table 4C)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All Cand.	Incumbents	Challengers	All Cand.	Incumbents	Challengers	All Cand.	Incumbents	Challengers
Treatment village	0.015	-0.003	0.018	0.006	0.009	0.037	0.115	0.061	0.091*
	[0.033]	[0.027]	[0.015]	[0.173]	[0.160]	[0.133]	[0.099]	[0.074]	[0.053]
Spillover	-0.022	-0.043	0.021	-0.106	-0.246	0.289*	0.106	-0.014	0.106*
	[0.040]	[0.030]	[0.019]	[0.206]	[0.182]	[0.172]	[0.114]	[0.077]	[0.064]
Treatment*Saturation	0.010	0.010	0.000	0.011	-0.040	0.196	-0.194	-0.132	-0.103
	[0.061]	[0.047]	[0.029]	[0.300]	[0.273]	[0.252]	[0.208]	[0.151]	[0.103]
Spillover*Saturation	0.164	0.156**	0.009	0.676	0.885*	-0.233	-0.080	0.044	-0.056
	[0.102]	[0.079]	[0.047]	[0.518]	[0.465]	[0.418]	[0.270]	[0.185]	[0.149]
Outside Sampling Frame	-0.002	0.027	-0.029*	-0.003	0.181	-0.286*	-0.126	-0.001	-0.105*
	[0.040]	[0.031]	[0.018]	[0.209]	[0.190]	[0.159]	[0.109]	[0.078]	[0.059]
ACFIM Presence	-0.124	-0.150**	0.027	-0.729	-0.924**	0.313	-0.240	-0.374**	0.027
	[0.087]	[0.068]	[0.036]	[0.455]	[0.413]	[0.332]	[0.227]	[0.163]	[0.117]
ACFIM Village*Presence	0.049	0.088	-0.038	0.229	0.486	-0.416	-0.037	0.107	-0.083
	[0.086]	[0.069]	[0.036]	[0.453]	[0.415]	[0.331]	[0.229]	[0.164]	[0.119]
R^2	0.13	0.11	0.06	0.12	0.10	0.07	0.04	0.04	0.04
Control Mean	0.430	0.327	0.102	2.580	2.136	1.117	1.216	0.919	0.372
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	28454	28454	28454	28454	28454	28454	4111	4111	4111

Table 2.9: Interactions on Key Outcomes

	Campaign Implementation (Table 2)					Vote-Buyi	
	(1) NGO	(2) Leaflet	(3) Meetings	(4) Call	(5) Posters	(6) All Cand.	Inc
Treatment village	0.374***	0.348***	0.312***	0.032***	0.182***	0.050	
	[0.021]	[0.019]	[0.031]	[0.008]	[0.017]	[0.045]	
Spillover	-0.003	-0.018**	-0.005	-0.003	-0.009	-0.026	
	[0.018]	[0.009]	[0.023]	[0.007]	[0.010]	[0.047]	
Treatment*Saturation	-0.082**	-0.021	-0.048	-0.005	0.015	-0.034	
	[0.040]	[0.037]	[0.056]	[0.015]	[0.033]	[0.085]	
Spillover*Saturation	0.064	0.081***	0.019	-0.003	0.043	0.137	
	[0.045]	[0.024]	[0.058]	[0.019]	[0.028]	[0.120]	
Outside Sampling Frame	-0.007	-0.006	-0.010	0.002	-0.008	0.005	
	[0.022]	[0.013]	[0.024]	[0.009]	[0.012]	[0.046]	
ACFIM Presence	0.023	-0.014	-0.020	0.017	0.036	-0.059	
	[0.037]	[0.021]	[0.048]	[0.019]	[0.024]	[0.098]	
ACFIM Village*ACFIM Presence	0.020	0.007	0.025	-0.009	-0.032	0.051	
	[0.043]	[0.026]	[0.050]	[0.019]	[0.027]	[0.098]	
R^2	0.14	0.20	0.06	0.04	0.09	0.06	
Control Mean	0.198	0.052	0.113	0.040	0.062	-0.008	
Controls	Yes	Yes	Yes	Yes	Yes	Yes	
Observations	27756	28007	27693	28454	28081	28454	

Interactions on Key Outcomes (continued)

	Attitudes Toward Vote-Buying (Table 5)				
	(1)	(2)	(3)	(4)	(5)
	Neg Consequences	Social Cost	Ostracize	Vignette	Vote Choice
Treatment village	-0.009	0.036*	0.049**	-0.019	-0.008
	[0.013]	[0.019]	[0.022]	[0.020]	[0.018]
Spillover	0.008	0.007	-0.002	0.014	-0.020
	[0.013]	[0.021]	[0.024]	[0.023]	[0.023]
Treatment*Saturation	0.045**	-0.029	-0.056	0.050	0.007
	[0.022]	[0.033]	[0.039]	[0.037]	[0.031]
Spillover*Saturation	0.009	0.026	-0.014	0.017	0.090
	[0.033]	[0.050]	[0.059]	[0.058]	[0.056]
Outside Sampling Frame	-0.015	-0.014	-0.014	-0.037	-0.001
	[0.014]	[0.024]	[0.028]	[0.025]	[0.024]
ACFIM Presence	-0.038	0.023	0.048	-0.017	-0.092*
	[0.030]	[0.045]	[0.053]	[0.051]	[0.050]
ACFIM Village*ACFIM Presence	0.002	-0.017	-0.037	-0.075	0.025
	[0.031]	[0.048]	[0.057]	[0.052]	[0.050]
R^2	0.04	0.04	0.07	0.02	0.10
Control Mean	0.888	0.745	0.567	0.733	0.242
Controls	Yes	Yes	Yes	Yes	Yes
Observations	28454	28454	27680	28448	28454

Interactions on Key Outcomes (continued)

	Official Electoral Support (Table 6)		Turnout (Table 7)		
	(1)	(2)	(3)	(4)	(5)
	Incumbent	Challengers	Turnout (z-index)	Presidential Turnout	MP Turnout
Treatment Polling Station	0.148*	-0.148*	0.005	-0.001	-0.003
	[0.085]	[0.085]	[0.082]	[0.007]	[0.008]
Spillover Polling Station	0.030	-0.030	-0.061	-0.003	-0.008
	[0.093]	[0.093]	[0.084]	[0.008]	[0.008]
Treatment*Saturation	-0.462**	0.462**	0.139	0.016	0.016
	[0.181]	[0.181]	[0.170]	[0.015]	[0.017]
Spillover*Saturation	-0.086	0.086	0.292	0.021	0.028
	[0.227]	[0.227]	[0.237]	[0.021]	[0.022]
Outside Sampling Frame	0.147**	-0.147**	-0.061	-0.010	0.000
	[0.062]	[0.062]	[0.073]	[0.007]	[0.007]
ACFIM Presence	0.127	-0.127	-0.101	-0.015	-0.007
	[0.121]	[0.121]	[0.119]	[0.011]	[0.011]
ACFIM Village*ACFIM Presence	-0.337**	0.337**	-0.150	-0.003	-0.026
	[0.151]	[0.151]	[0.170]	[0.016]	[0.016]
R^2	0.47	0.47	0.33	0.31	0.30
Control Mean	0.043	-0.043	-0.007	0.674	0.690
Controls	Yes	Yes	Yes	Yes	Yes
Observations	3657	3657	3659	3659	3112

Interactions on Key Outcomes (continued)

	Index of Campaigning (Table 8)		
	(1)	(2)	(3)
	All Candidates	Incumbents	Challengers
Treatment village	-0.636*	-0.202	-0.434**
	[0.327]	[0.172]	[0.187]
Spillover	-0.441	-0.006	-0.434**
	[0.364]	[0.196]	[0.216]
Treatment*Saturation	1.816**	0.673*	1.144***
	[0.704]	[0.356]	[0.410]
Spillover*Saturation	0.985	-0.123	1.107**
	[0.875]	[0.462]	[0.519]
Outside Sampling Frame	-0.122	-0.025	-0.097
	[0.332]	[0.189]	[0.188]
ACFIM Presence	-1.175*	-0.407	-0.768*
	[0.700]	[0.375]	[0.413]
ACFIM Village*ACFIM Presence	-0.432	-0.233	-0.200
	[0.695]	[0.378]	[0.403]
R^2	0.12	0.12	0.09
Control Mean	5.759	3.400	2.359
Controls	Yes	Yes	Yes
Observations	28454	28454	28454

Table 2.10: Balance Voter Respondent

	Age		Years Education		Married		Own Land		Adults		Children	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treatment village	-0.247		0.012		-0.011		-0.002		-0.030		-0.096	
	[0.302]		[0.117]		[0.010]		[0.010]		[0.056]		[0.075]	
Spillover	0.123		-0.120		-0.006		0.002		-0.056		-0.248***	
	[0.338]		[0.146]		[0.011]		[0.011]		[0.062]		[0.083]	
Treatment Saturation		-0.079		-0.004		-0.011		0.008		-0.051		-0.240*
		[0.494]		[0.213]		[0.018]		[0.020]		[0.094]		[0.143]
Outside Sampling Frame	-0.843**	-0.650***	0.170	0.083	-0.014	-0.012	-0.001	0.001	-0.005	-0.027	0.102	-0.009
	[0.342]	[0.245]	[0.139]	[0.097]	[0.011]	[0.008]	[0.010]	[0.007]	[0.059]	[0.041]	[0.073]	[0.050]
ACFIM Presence	-1.085**	-1.062**	-0.176	-0.185	-0.022	-0.018	-0.044***	-0.048**	0.317***	0.335***	0.739***	0.829***
	[0.467]	[0.518]	[0.209]	[0.223]	[0.016]	[0.018]	[0.017]	[0.020]	[0.096]	[0.111]	[0.127]	[0.153]
R^2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01
Control Mean	40.088	40.088	5.487	5.487	0.741	0.741	0.872	0.872	3.213	3.213	3.605	3.605
Observations	27375	27375	28452	28452	28454	28454	28454	28454	28454	28454	28451	28451

Balance Voter Respondent (continued)

	Assets		Farmer		Trade		High Skill		Not Working	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment village	-0.006		0.027		-0.010		-0.005		-0.006	
	[0.034]		[0.017]		[0.007]		[0.006]		[0.004]	
Spillover	-0.004		0.010		-0.010		-0.009		0.003	
	[0.039]		[0.020]		[0.008]		[0.007]		[0.006]	
Treatment Saturation		-0.009		0.053		-0.022		-0.013		-0.006
		[0.069]		[0.033]		[0.013]		[0.010]		[0.008]
Outside Sampling Frame	-0.009	-0.008	-0.008	-0.019**	-0.003	-0.004	0.009	0.006	-0.004	0.001
	[0.032]	[0.018]	[0.015]	[0.009]	[0.007]	[0.004]	[0.006]	[0.004]	[0.005]	[0.003]
ACFIM Presence	-0.055	-0.050	-0.011	-0.040	-0.007	0.005	0.014	0.021*	0.010	0.013
	[0.063]	[0.071]	[0.031]	[0.036]	[0.011]	[0.014]	[0.011]	[0.011]	[0.008]	[0.008]
R^2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Control Mean	1.638	1.638	0.687	0.687	0.088	0.088	0.078	0.078	0.053	0.053
Observations	28454	28454	28453	28453	28453	28453	28453	28453	28453	28453

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by parish in brackets. All regressions control for the parish-level ACFIM presence (interacted with Treatment and Spillover status in even columns), and an ACFIM dummy.

Balance Voter Respondent (continued)

	Ganda		Nkole		Soga		Catholic		Protestant		Muslim	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treatment village	0.021		-0.007		-0.017		0.034*		-0.017		-0.019	
	[0.021]		[0.014]		[0.015]		[0.019]		[0.019]		[0.013]	
Spillover	0.033		-0.012		-0.030**		0.064***		-0.022		-0.040***	
	[0.025]		[0.014]		[0.014]		[0.021]		[0.021]		[0.013]	
Treatment Saturation		0.022		-0.012		-0.020		0.065*		-0.027		-0.038
		[0.048]		[0.024]		[0.034]		[0.036]		[0.038]		[0.027]
Outside Sampling Frame	-0.013	-0.001	0.016*	0.012*	0.026***	0.014**	-0.032*	-0.008	0.003	-0.003	0.024**	0.007
	[0.019]	[0.006]	[0.009]	[0.006]	[0.010]	[0.006]	[0.018]	[0.011]	[0.017]	[0.010]	[0.009]	[0.005]
ACFIM Presence	0.158***	0.148***	-0.116***	-0.110***	0.055**	0.064*	-0.001	-0.032	-0.064*	-0.051	0.074***	0.092***
	[0.043]	[0.042]	[0.025]	[0.028]	[0.027]	[0.034]	[0.033]	[0.038]	[0.033]	[0.036]	[0.019]	[0.027]
R^2	0.02	0.02	0.02	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.01
Control Mean	0.075	0.075	0.061	0.061	0.060	0.060	0.423	0.423	0.429	0.429	0.087	0.087
Observations	28451	28451	28451	28451	28451	28451	28454	28454	28454	28454	28454	28454

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by parish in brackets. All regressions control for the parish-level ACFIM presence (interacted with Treatment and Spillover status in even columns), and an ACFIM dummy.

Table 2.11: Balance Key Informant 1

	Chief or Elder		Civil Society		Village Committee		Local Council	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment village	-0.019		0.009		-0.001		0.026	
	[0.017]		[0.008]		[0.026]		[0.021]	
Spillover	0.023		-0.003		-0.045		0.012	
	[0.023]		[0.008]		[0.031]		[0.025]	
Treatment Saturation		-0.038		0.014		-0.015		0.072*
		[0.031]		[0.011]		[0.047]		[0.039]
Outside Sampling Frame	-0.028	0.002	-0.002	-0.010	0.055**	0.024	-0.005	-0.016
	[0.021]	[0.013]	[0.008]	[0.007]	[0.027]	[0.016]	[0.025]	[0.016]
ACFIM Presence	0.143***	0.166***	-0.028**	-0.036***	-0.213***	-0.208***	0.152***	0.113***
	[0.028]	[0.035]	[0.011]	[0.013]	[0.042]	[0.049]	[0.034]	[0.039]
R^2	0.01	0.01	0.00	0.00	0.02	0.02	0.01	0.01
Control Mean	0.187	0.187	0.031	0.031	0.430	0.430	0.247	0.247
Observations	4090	4090	4090	4090	4090	4090	4090	4090

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by parish in brackets. All regressions control for the parish-level ACFIM presence (interacted with Treatment and Spillover status in even columns), and an ACFIM dummy.

Balance Key Informant (continued)

	Ganda		Nkole		Soga		Catholic		Protestant		Muslim	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treatment village	0.006		-0.004		-0.015		0.005		-0.013		-0.018	
	[0.024]		[0.016]		[0.016]		[0.025]		[0.025]		[0.015]	
Spillover	0.031		-0.005		-0.029*		0.029		-0.001		-0.026	
	[0.030]		[0.017]		[0.016]		[0.029]		[0.028]		[0.016]	
Treatment Saturation		-0.003		0.003		-0.021		0.033		-0.023		-0.043
		[0.055]		[0.029]		[0.035]		[0.046]		[0.045]		[0.030]
Outside Sampling Frame	-0.019	0.002	0.011	0.009	0.027**	0.015**	-0.021	-0.005	0.006	0.014	0.004	-0.002
	[0.022]	[0.008]	[0.012]	[0.008]	[0.011]	[0.007]	[0.026]	[0.017]	[0.025]	[0.017]	[0.013]	[0.008]
ACFIM Presence	0.177***	0.181***	-0.115***	-0.117***	0.051*	0.061*	0.011	-0.004	-0.034	-0.021	0.066***	0.088***
	[0.048]	[0.051]	[0.029]	[0.030]	[0.028]	[0.036]	[0.041]	[0.048]	[0.041]	[0.046]	[0.023]	[0.031]
R^2	0.03	0.02	0.02	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.01
Control Mean	0.095	0.095	0.063	0.063	0.063	0.063	0.449	0.449	0.421	0.421	0.091	0.091
Observations	4090	4090	4090	4090	4090	4090	4090	4090	4090	4090	4090	4090

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by parish in brackets. All regressions control for the parish-level ACFIM presence (interacted with Treatment and Spillover status in even columns), and an ACFIM dummy.

Table 2.12: Balance on Pre-determined Electoral Data

	Reg'd Voters 2011		Turnout 2011		NRM Vote 2011		FDC Vote 2011		MP Incumbent Vote 2011		Reg'd Voters 2016	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treatment Polling Station	-2.555		0.012		-0.011		0.012		-0.019		-3.216	
	[14.153]		[0.009]		[0.015]		[0.013]		[0.015]		[10.404]	
Spillover Polling Station	-6.448		0.005		-0.007		-0.008		-0.001		-14.253	
	[13.730]		[0.009]		[0.017]		[0.015]		[0.016]		[9.097]	
Treatment Saturation		-18.682		0.014		-0.026		0.015		-0.034		4.105
		[26.344]		[0.018]		[0.031]		[0.028]		[0.029]		[17.272]
Outside Sampling Frame	-18.833	-20.776**	-0.017***	-0.020***	-0.021**	-0.018***	0.016*	0.007	-0.020**	-0.011**	-81.538***	-86.651***
	[11.664]	[9.489]	[0.006]	[0.005]	[0.010]	[0.007]	[0.008]	[0.006]	[0.008]	[0.005]	[9.383]	[7.341]
ACFIM Presence	1.120	8.776	-0.082***	-0.089***	-0.149***	-0.138***	0.039*	0.031	-0.077***	-0.061**	-44.869***	-48.285***
	[23.369]	[26.539]	[0.016]	[0.019]	[0.027]	[0.032]	[0.023]	[0.027]	[0.025]	[0.029]	[15.573]	[17.935]
R^2	0.00	0.00	0.03	0.03	0.03	0.04	0.00	0.00	0.01	0.01	0.04	0.04
Control Mean	615.658	615.658	0.605	0.605	0.689	0.689	0.258	0.258	0.554	0.554	575.342	575.342
Observations	2820	2820	2820	2820	2814	2814	2814	2814	3217	3217	3665	3665

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by parish in brackets. All regressions control for the parish-level ACFIM presence (interacted with Treatment and Spillover status in even columns), and an ACFIM dummy.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 3

A "Minor" Expansion: The 26th Amendment and Changes in U.S. Political Outcomes

3.1 Introduction

Voting is at the core of democratic engagement and leader or policy selection. As a result, the study of voting has long held a foremost position in the field of political economy. One of the most important elements determining outcomes in an electoral democracy is the nature of *who* is allowed to participate in the franchise. Changes to the franchise have important implications not only substantively, but also because these changes allow us to examine the validity of different models of voting, which often respond to changes in the composition of the franchise in different ways. These different models of voting, in turn, have important implications for the functioning of democracy itself and for predicting the effects of other changes to electoral rules, such as reserving elected seats for women, as in Chattopadhay and Duflo (2004).

I am grateful to advice on this project from Abhijit Banerjee, Ben Olken, Thomas Fujiwara, and Daron Acemoglu. This project was born out of work done for 14.770. Christina Patterson provided a tremendous amount of both economic insight and emotional support during this work, which could not have been completed without her. This project would not have been possible without the data provided by the Inter-University Consortium on Political and Social Research. All mistakes remain, of course, with me.

Two of the most important workhorse models of voting are the median voter model (Black (1948)) and the citizen candidate model (Osborne and Slivinski (1996), Besley and Coate (1997)). In the former model, politicians are able to commit before each election to a set of policies – in a unidimensional policy space with moderate entry costs, this results in a single Nash equilibrium, with two candidates clustered at the midpoint of the policy space (pleasing the infamous median voter). In the latter model, candidate beliefs are known, but commitment to a set of policies is impossible and thus all commitments are a form of cheap talk. These models have a much richer set of equilibria, often diverging widely from the median voter's views. A crucial empirical distinction between these models relates to the flexibility of an elected politician's views in response to a change in the franchise –in a median voter model, policies immediately shift towards the new median point, while in a citizen candidate framework, policies do not shift at all and *will not* shift in the future, unless the politician herself is replaced.

I study one important franchise expansion that, to my knowledge, has not been examined in the literature. This paper examines the effects of the 26th Amendment, which lowered the voting age in the United States from 21 to 18 – expanding the pool of *eligible* voters by about 8.8%. This franchise expansion is interesting for several reasons. First, this is a quantitatively large expansion – roughly the same size as the change in the potential electorate after the end of Jim Crow. Second, young voters are believed to have different policy preferences than older voters. As demonstrated in Figure 3-1 below, 18-29 year-olds favor Democrats by larger margins than all other age groups in six of the 11 elections and just behind 30-44 year-olds in the five others. Third, this expansion affected every region of the United States, as young people are present in every electoral jurisdiction.

In the United States, two of the most important extensions of the franchise, one *de jure* and one *de facto*, that have already been studied are the granting of the vote to women in the late 1910s (fully ratified in 1920) and the forcible end to anti-black voting measures in the U.S. South in the 1950s and 1960s. Looking at the former of these franchise extensions, Miller (2008) uses differential adoption across states of the right of women to vote to examine the effects of women's suffrage on public health spending and child mortality. He finds large and essentially immediate effects of women's suffrage on health, sanitation, and other social spending and large and economically meaningful reductions in the deaths of children from preventible causes, such as diarrheal diseases or diphtheria. His finding of an immediate shift accords with the median voter logic found in Black (1948), as politicians respond to the existence of women in their future electorates by changing policy to meet their concerns in the current period.

Cascio and Washington (2014) look at the distribution of state funds after the passage of the Voting Rights Act (VRA), which generated large increases in black voter registration in the South. They find that this *de facto* enfranchisement (which eliminated discriminatory poll taxes and "literacy" tests) significantly increased state transfers to counties with higher black population shares, in states where the VRA was effected. They find that the increase in black turnout following the passage of the VRA had an elasticity of about 1 with respect to new transfers: each one percent increase in new turnout yielded a one percent increase in transfers. Another related paper, which also looked at the elimination of poll taxes and literacy tests, is Husted and Kenny (1997), which finds an increase in the total size of welfare payments. These papers also find similar support for the median voter theorem, with Cascio and Washington (2014) in particular pointing out that even after the passage of the VRA, the proportion of state legislators who were black was very small, so white legislators (including many who pre-dated and opposed the VRA) must have been responding as well.

Using a difference-in-difference approach based on the estimated share of a county's population that gained the right to vote, I find that the 26th Amendment caused a significant increase in total turnout, especially in House races, in areas with more new potential voters. Looking at states that had already lowered their voting ages prior to the Amendment, I find that its passage also increased voting in those areas, potentially due in part to general equilibrium shifts in campaigning behavior induced by the Amendment. Despite these effects on turnout, I find little aggregate effect on the share of votes for Democrats in House races and heterogenous effects in Presidential races that appear to be candidate specific. These small aggregate effects mask suggestive evidence of heterogeneity across counties with colleges (which see flat effects or increases in vote shares for Democrats) and counties without them (which see reductions in Democrat vote shares).

These results represent an important null finding in the literature on the effects of changing the electorate on political outcomes. Although the 26th Amendment did induce young people to vote, it did not have a meaningful impact on the politicians who were elected and correspondingly did not affect policy in a meaningful way. These results speak to the importance of considering within-group heterogeneity, as well as the differences between the group and the rest of the electorate when considering the effects of expanding the franchise on political outcomes.

In the remainder of this paper, I first briefly discuss the background behind the 26th Amendment. I then discuss my data sources, my empirical strategy, my results, and conclude with a brief discussion of my findings.

3.2 Background

The question of the voting age in the United States was tied to fighting in major U.S. conflicts from World War II until the ratification of the 26th Amendment during the Vietnam War. Prior to World War II, the voting age was set in each state at 21, which was also the age of eligibility for a wartime draft. In November 1942, President Franklin Delano Roosevelt lowered the age at which a man was eligible for the draft from 21 to 18. In response, Congress considered an amendment to lower the voting age to 18, sponsored in the Senate by Arthur Vandenberg (R-MI) and in the House by Jennings Randolph (D-WV). However, Congress failed to move the amendment beyond committee hearings and the push failed (Saldin, 2010).

However, under the slogan "old enough to fight, old enough to vote", a campaign was launched in Georgia to lower the voting age to 18. In 1943, Georgians voted to add the measure to their state constitution by an overwhelming margin of 79% to 21% (Hays (1951)). An additional 29 states considered the same change, but the cessation of hostilities in 1945 ended the impetus for lowering the voting age.

A second push to lower the voting age came during the Korean War from 1950 to 1953. President Dwight Eisenhower even proposed lowering the voting age in his 1954 State of the Union – Congress considered the issue, but the amendment failed to receive the necessary two-thirds majority in the Senate by five votes, failing 34-24 (Gale 2012). As memory of the war faded, the impetus for lowering the voting age also fell, though a single state moved forward on its own, with Kentucky lowering the voting age to 18 in a 1955 referendum (Neale (1983)). An additional two states, Oklahoma and South Dakota both *rejected* amendments to lower the voting age during the Korean War in 1952.

The issue finally came to a head during the Vietnam War. As the Vietnam War dragged on, youth protests over the draft grew larger and larger. One element of these protests was again the idea of "old enough to fight, old enough to vote." This time, the effort to lower the voting age was successful. In 1970, championed by Senators Edward Kennedy (D-MA) and Mike Mansfeld (D-MT), an amendment lowering the voting age to 18 was added to the Voting Rights Act renewal bill, which sailed through both houses by large margins. However, the change in voting age was immediately challenged by the state of Oregon as an infringement on the right of states to set their own conditions for voting. In December 1970, the Supreme Court ruled in *Oregon v. Mitchell 1970* by a 5-4 margin that: "(1) The 18-year-old minimum-age requirement of the Voting Rights Act Amendments is valid for national elections. (2) That requirement is not valid for state and local elections."

To resolve the potential chaos of needing a federal ballot (accessible to 18 year-olds) and a state ballot (only accessible to those meeting the state eligibility criteria), Congress took up the 26th Amendment, which was passed by overwhelming margins: 94-0 in the Senate and 401-19 in the House. These overwhelming margins did not reflect popular opinion – Gallup polling found support for changing the voting age reaching a high of 66% in 1968, but falling to 57% by 1970 (Lyons (2004)). In addition, between 1966 and 1971, there were 27 state referenda on lowering the voting age – of which 19 failed. In 1970, the same year that the Voting Rights Act was amended, 17 states held referenda, of which 11 failed (Saldin (2010)).

3.3 Empirical Specification

To examine the effect of this national change in voter eligibility, I use two different difference-in-difference strategies that leverage variation in the proportion of newly eligible voters at the time of the passage of the 26th Amendment. The first strategy is to identify the effect of the franchise expansion off of "permanent" populations of 18-20 year-olds – essentially identifying off of "college counties." To implement this strategy, I compare places that had more newly-eligible voters in 1970 (the closest census to the reform) with other places in the same state-year that had fewer newly eligible voters. With that strategy, the "new voters" value is frozen at the 1970 level for all years in the sample, so it assumes that (as with a college area), the 18-20 year-old population is rejuvenated at a fairly constant rate, so their proportion of the total population is constant over time.[1]

The second strategy is to identify the effect off of "differential baby booms" – using the aging of different cohorts over time and assuming that there is minimal migration. For example, a 17 year-old in 1970 stays in the county and becomes a 19 year-old in 1972. Differences in the relative size of cohorts across counties within a state-year create the variation that I exploit in this version.

As with other difference-in-difference designs, the identifying assumptions is parallel trends in the outcome variable between the "treated" and "untreated" counties. Specifically, it must be the case that counties with a younger population are not increasing their turnout faster than counties with fewer young people. This may be a strong assumption, as there may be reasons to think that either college towns (strategy 1) or baby-boom areas (strategy 2) have electorates that respond differently to various political movements of the time. This would contaminate the identification of the effect of the amendment, as it would conflate the effect of the enfranchisement of young voters with other differential sways in the voting behavior of the electorate. I will show event-study graphs throughout testing this assumption, and will show that despite the differences across these areas, the assumptions of parallel trends seems mostly reasonable.

The main specification that I use throughout the analysis is given b Equation 3.3.1

$$\ln(Y_{cst}) = \beta_0 + \beta_1 \ln(POP21+_{cst}) + \beta_2 \frac{NEWVOTER_{cst}}{POP_{cst}} + \beta_3 POST_t * \frac{NEWVOTER_{cst}}{POP_{cst}} + \gamma_{st} + \alpha_c + \varepsilon_{cst} \tag{3.3.1}$$

c (county), s (state), t (year), and p (party). Y is defined as an electoral outcome for one political office (e.g. House or president). $LN(POP21+_{cst})$ is the natural log of total (pre-26th Amendment) voting-eligible population in the county, to account for higher total turnout in higher population areas.[2] $POST_t$ is 0 for all elections before 1972 and 1 for all years thereafter. $NEWVOTER_{cst}$ is a count of the number of people aged 18-20 in a county in a given year t. γ_{st} is a state-year fixed effect and α_c is a county fixed effect.

This regression identifies β_3 using variation between counties within a given state-year

[1] Enrollment-by-college data only begins in 1980, so it is not possible to use that information to adjust the college population for each year.

[2] An important note is that I use proportional measures such as natural log in order to allow for state-year fixed effects to have the interpretation of being a proportional change to the size of the electorate.

in the share of 18-20 year-olds, the population that is granted the right to vote in 1972. The county fixed effect removes any effect of time invariant characteristics of places that have more young people, so the threats to identification would need to come from trends in voting that are correlated with the trends in the share of young people. I will test for the existence of such trends by showing event study plots throughout my main results.

Standard errors are clustered at the level of the county, though they are robust to being clustered at the level of the congressional district (for House races) and at the level of the state (for Presidential elections). In both cases, I should have well over the number of clusters needed to avoid problems from insufficient clusters (see for example the concerns raised by Bertrand et al. (2004)).[3]

In this work, I show only results on elections for House and for President. House races are contested in all counties every two years, while presidential elections are the highest profile races. In addition, total turnout and votes for Democratic candidates are highly correlated across elections, so in my examination of other races, there was little evidence of any meaningful heterogeneity by type of election. For House races, I restrict my sample, unless otherwise noted, to include only races that were contested by members of both major parties.

3.4 Data

The data for this project come from several sources: (1) historical sources on voting ages across states, (2) county-level population, including by age category, from the National Historical Geographic Information System (NHGIS) at the University of Minnesota, (3) county-level presidential, state-wide, and Congressional vote totals for each party from the Inter-University Consortium for Political and Social Research (ICPSR), and (4) data on undergraduate enrollment at colleges and universities in 1980 from the National Center for Education Statistics (through the Integrated Postsecondary Education Data System (IPEDS)), as well as data on date of accreditation by university from the Department of Education.

[3] I cluster at the level of the county in most regressions because counties may (a) contain multiple districts, (b) be split across multiple districts, and (c) change districts over time, so it is not uniformly more conservative to cluster at the district level throughout and it is generally less transparent to do so.

3.4.1 Voting Ages

The voting age was originally lowered to 18 via an amendment to the extension of the Voting Rights Act in 1970 (effective January 1, 1971). However, it was challenged in court and ruled as unconstitutional for non-federal elections by the Supreme Court. To avoid having two voting ages for different types of elections, Congress passed the 26th Amendment on March 23rd, 1971 and it was ratified by the necessary three-fourths of the states on July 1st, 1971, the fastest Amendment ratification process in U.S. history. Note that although the 26th Amendment was needed to allow younger voters to vote in non-federal elections, they would have been allowed to do so for federal elections even without it and thus it should not be seen as a problem that there is not a large amount of time between July and November, since registration would have started as early as January.

Even prior to the 26th Amendment, several states had different voting ages, as seen in Table 3.1. These different voting age changes make a traditional difference in difference harder to implement, since the exact change differs across state and time. In addition, since I lack exact ages at the county level, I have difficulty with small age bins such as only 20 year-olds. As a result, I drop these states (Georgia, Kentucky, Alaska, and Hawaii) from my analysis, using only those states with exogenous variation imposed by the 26th Amendment.[4] However, I will leverage the fact that Georgia and Kentucky had reduced their voting ages to 18 prior to the ample period to examine general equilibrium effects of the amendment.

3.4.2 Age Data

In order to estimate the effect of the 26th Amendment, my empirical strategy requires accurate estimates of the number of newly eligible voters, both before the Amendment comes into effect (to test for pre-trends) and afterwards, to measure the effects. However, county-level age distributions are not available except at each census. In addition, before the 1970 census, data was not available at the exact age level (i.e. number of 19 year-olds, number of 20 year-olds, etc.). This limitation means that more interpolation is needed to estimate the county level age distribution in each year. In Table 3.2, I show the Census data available for each year.

To calculate the fraction of the population between the age of 18 and 20 (inclusive)

[4] The states that lowered their voting ages in 1970 did so using referenda, so they never actually saw any independent effect of lowering their voting ages "before" the 26th Amendment (besides any local elections in 1971, which I do not study).

for the years before 1970, I assume a uniform distribution over all ages in any age group (i.e. I assume that one fifth of the population that is listed as being aged 20-24 is 20). If I have specific data for an age category, that value supersedes any estimate and the true value is subtracted from the grouped data and then I assume a uniform distribution over the remaining years (i.e. if there are X 20 year-olds and Y 20-24 year-olds, then I assume that there are $\frac{Y-X}{4}$ 21 year-olds).

To interpolate data between two census years, I use linear weights, so that the expected number of 18 year olds in intercensal year $t+s$ (where t is the prior census) would be $\frac{(10-s)*AGE_{18,t}}{10} + \frac{(s)*AGE_{18,t+10}}{10}$. In Appendix 3.8.3, I discuss this method in more detail in the context of estimating intercensal populations.

In Figure 3-2, I show a map with the distribution of the fraction of the population that was between the age of 18 and 20 in 1970 (which is known directly). There are clear geographic concentrations, but also substantial variation within each state, which is the source of the identification in my analysis.[5]

3.4.3 College Data

The correlation of youth population and college locations is high and an important potential confounder of the outcomes, especially vis à vis vote shares between the different parties.[6] As a result, I classify counties based on whether or not they had an accredited four-year college in 1972 (the year of the 26th Amendment), using data from the Department of Education on the accreditation history of all accredited colleges and universities in the United States. I merge this data with data on college location and type from IPEDS at the National Center for Education Statistics. The IPEDS data only exists beginning in 1980, so any colleges that closed between 1972 and 1980 will be missing from my sample. Approximately 25% of my counties have at least one four-year college in them.

3.4.4 Election Results

The Inter-university Consortium for Political and Social Research (ICPSR) provides data at the county level on presidential, House, Senate, and gubernatorial races from 1950-1990, plus very limited data on a non-gubernatorial state-wide office for some election years. I

[5] The missing data reflects changes to county codes between 1990 and earlier years in some states, not actual missing data.

[6] Democrats are generally believed to do better in college towns in modern elections, due to more liberal beliefs among students as well as faculty.

use the sample from 1950-1980. The data include total votes, votes for the Democratic candidate, votes for the Republican candidate, and votes for any independent candidate in each race. The data are linked to congressional districts, though if a county is split between multiple districts, all votes cast in that county (across multiple Congressional races) are aggregated and the county is assigned an error code as its district.[7] The result of this sample restriction is that the counties included in the electoral analysis will be slightly less urban, as urban counties are more likely than rural counties to be split across congressional districts.

3.5 Results

3.5.1 Turnout

The main results of the effect of the amendment on voter turnout are shown in Figures 3-4 and 3-3. I find that turnout to elect members of Congress increases significantly in counties with more new potential voters after the passage of the 26th Amendment. By comparing the left and right panels, we can see that this is true whether the measure of new potential voters (people aged 18-20) is based on differential baby booms or based on static differences fixed in 1970 (the census directly before the 26th Amendment was passed). Importantly, prior to the passage of the law, there is no trend in turnout in the counties with more young voters, though they do exhibit a larger difference between presidential and non-presidential years than other counties (likely because counties with more 18-20 year olds also have more other young people, who are known to be less likely to vote in off-year elections). Figure 3-3 shows that a result similar in magnitude holds for presidential elections. This is reassuring, since since we would expect anyone who cast a vote for the House of Representatives to also vote for president.

In Table 3.3, I show these results in difference-in-difference form. As can be seen from the figures, for each additional percentage point of the population that is aged 18-20, the number of votes in the county increases by approximately 1.2-1.4 percent. These magnitudes are plausible. In 1970, the median county was 60.1 percent aged 21 or older. Therefore,

[7] For example, if a county is split between District 1 and District 2 and Democrats receive 1000 votes for their candidate in District 1 and 500 votes for their candidate in District 2, the Democrat total would be 1,500 for the county and the district coding would be "98" – an error code.

at the median, each 1 percentage point of the total population that is newly voting eligible therefore represents a 1.66 percent increase in the size of the voting eligible population. A 1.2 percent increase in total votes would therefore represent an increase in votes per 1 percent increase in voting eligible population of 0.72%, as compared to a coefficient of slightly more than 1% on the percent increase in votes for each 1% increase in population over the age of 21. Thus, this suggests that new voters turned out at a rate of about 72% that of existing voters. According to a report from the Census (U.S. Bureau of the Census (1973)), the ratio of turnout between new and existing voters nationally was 75%, almost identical to this estimate.

Since the results are quite similar across the two specifications (using time-varying or fixed population of new potential voters), for the remainder of the results, I will use the fixed population measure.

3.5.2 Votes by Party

Figures 3-5 and 3-7 show that the effects on the composition of votes for the two major U.S. parties are much more mixed. In the House, Democrats and Republicans both appear to have increased their total votes after the passage of the 26th Amendment (unsurprising, given the results on turnout above). There are also some slight pre-trends in the number of votes for Republicans, which ticked up slightly in 1970 in areas with more new potential voters, even prior to the passage of the 26th Amendment. The trend break is much clearer for Democratic candidates. Indeed, Table 3.4, which pools the post-treatment years, shows that the point estimate on the new voter share is statistically indistinguishable for the two parties.

Since the votes are logged, these proportional increases cannot be directly compared, since Democrats generally outperformed Republicans in this era.[8] However, when we combine the estimates to look at vote shares for each party in contested House elections in Figure 3-7 , we find that these aggregate effects on the democratic vote share is approximately 0 – new voters seem to have roughly split their votes across the two parties. The lone possible exception to this was in 1972, shortly after the passage of the 26th Amendment. Even so, Table 3.5 shows that on average, there was no statistically significant effect of the amendment on the share of votes going to Democrats.

[8] On average, Democrats got 10% more votes than Republicans across all contested races in 1970. In areas with above median new potential voters, this was 15%.

Figure 3-6 shows that at the presidential level, Democrats appear to have benefitted slightly, but the estimates are more variable over time. This is unsurprising when we consider that presidential races are shaped by the candidates, who may specifically appeal to different demographic (e.g. age) groups. House races also have candidate specific effects, but by looking across all (contested) races, those group-specific appeals that are *not* directly linked to the overall party platform average out.

We can partially validate these somewhat surprising null results by comparing them to exit polls done at the time of the election – these exit polls qualitatively match the findings here. In 1972, the New York times found that 18-29 year-olds supported McGovern (the Democrat) at a rate of 10 points higher than his average support in the population. In the 1976 election, CBS News found that 18-21 year-olds supported Jimmy Carter (the Democratic candidate) at a rate that was slightly lower than his national average – but in the 1980 election, this relationship flipped, with Carter out-performing with young voters relative to his national total.

3.5.3 Heterogeneity by College Status

One key potential source of heterogeneity in the effects of the 26th Amendment, to which I alluded earlier, is the existence of a four-year college within a given county. In the late 1960s and early 1970s, college campuses tended to be at the forefront of the anti-war sentiment, with almost one-third of all surveyed campuses reporting organized protest activities (President's Commission on Campus Unrest, 1970). For this reason, in Figure 3-8, I show the same results on turnout, but this time dividing the sample into counties with a college and counties without a college. The results show a similar magnitude in both types of counties, albeit slightly larger for non-college counties. One reason for this result could be absentee voting – if some college students vote absentee (but are counted in the census as living at their college), then this will bias the effect of the 26th Amendment downward in places where most young people are college students.

The results on vote shares for Democrats shown in Figure 3-9 show larger differences. In non-college counties, vote shares for Democrats appear to decline slightly, while they are flat in counties that contain colleges. The magnitudes (seen in Table 3.6, with college estimates in even columns and non-college estimates in odd columns) are plausible, but large (for non-college counties) – a 1 percentage point increase in the share of new potential voters in the population is associated with a decrease in Democratic vote margins of 0.3 percentage points in non-college counties. This would correspond to a split of approximately 65-35 in favor of Republicans in non-college counties among new voters. However, since there are fewer new potential voters in those counties on average, even winning those voters by a large margin does not result in overall gains for the Republican party relative to the Democratic party.

The results for president seen in Figure 3-10 exhibit some pre-trends that make them harder to interpret, but they largely match the results seen earlier in Figure 3-6 on votes for Democratic and Republican presidential candidates – even in college counties, where one may have expected the largest democratic swing, there is no detectable shift in the democratic vote share.

3.5.4 Quasi-Placebo Test

I now examine early switching states as a quasi-placebo test. In the absence of spillovers or general equilibrium effects, states that had already lowered their voting ages to 18 prior to the passage of the 26th Amendment should not have any change in turnout or vote shares in response to its passage. I focus on Georgia and Kentucky, which had a voting age of 18 well before the rest of the country (Georgia as of 1943, Kentucky as of 1955). In Figure 3-11, I show "placebo" tests looking at the effect of $FRACAGE18 - 20_{i,t}$ using the same specification from (3.3.1) for Congressional races. Here, the series starts with 1956 as the omitted category (the first year that youth voting was allowed in Kentucky).[9]

In the left panel of 3-11, I show a null effect of the 26th Amendment on votes for House candidates. However, in the right panel, I show that turnout for President does appear to increase after the passage of the 26th Amendment. One explanation for this difference could be that the passage of the 26th Amendment had general equilibrium effects wherein campaigns devoted much more effort nationally to turning out 18-20 year-olds. In that case, House elections, which are decided locally, would not necessarily see an effect (since their strategies would remain unaffected), but the presidential races could. House turnout does show upticks in presidential years (but not midterm elections), which would accord with a spillover from presidential organizing to House elections.

In Table 3.7, I show the same results in difference-in-difference form. The magnitude of

[9] Here I use the time-varying measure of the fraction of eligible voters since the relevant population is already able to vote earlier in the period and I want to absorb that effect directly.

the shift is approximately 1% more total votes for president per 1 percentage point more people between the ages of 18-20. This is about 70% of the size of the total effect from column (2) in Table 3.3. The effects are noisy, but they suggest that changes to presidential campaign strategy and organizing accounted for a relatively large share of the total change in turnout.

3.6 Discussion

The 26th Amendment was not a small change to the potential electorate. Adding people aged 18 to 20 to the ranks of eligible voters expanded the electorate by approximately 8.8%. For comparison, when the Voting Rights Act was passed in 1965, African-Americans made up approximately 10.5%-11% of the population (and the portion of African-Americans directly impacted by the legislation was smaller).[10] Yet, while the Voting Rights Act has had large impacts on the identities of politicians who represent African American areas, as well as the policies adopted by those politicians (Cascio and Washington (2014)), I find that the 26th Amendment had minimal effects on vote shares and, in results not shown here, I find suggestive evidence that politicians representing areas with more new voters were less likely to be replaced between 1970 and 1972 than those representing fewer new voters.

The lack of impact found here is not because the change to the voting age did not in fact cause any new young voters to turn out. I find that young voters turned out at approximately 75% of the rate of older voters – increasing total votes cast in House races by approximately 6.6%, a quantitatively large effect. However, examining the partisan split of those votes suggests why the aggregate impacts of the 26th Amendment were small. New voters appear to have split their votes for the national legislature (the House) evenly between the two major parties. Given the strong importance of parties in the modern American political system, a change to voting rules that empowers each of them equally produces little overall effect.

Even on the issue of the Vietnam War, which was the stated reason for allowing 18 year-olds to vote, young voters were divided. Even as late as 1973, Gallup found that young voters (under the age of 30) were almost perfectly evenly divided on whether sending troops to war was a mistake – 53% viewed it as a mistake to 47% opposed. In this context,

[10] Hispanics made up an additional 3-4% of the population and were also impacted.

it is little surprise that allowing young people to vote failed to change either the identity of the politicians in office or their policy choices. In any political economy model, the driving force explaining why changes in the electorate affect policy is that the new voters differ systematically from the incumbent voters in their policy preferences. The results here suggest that in the early 1970s, policy preferences were not sufficiently different between new and old voters for the 26th Amendment to have real impacts. These findings could differ substantially today. Democrats now win a much larger share of youth votes than they did in the 1970s, which suggests that young voters now hold systematically different views than older voters.

3.7 Conclusion

This project is the first to empirically investigate the effects of the 26th Amendment on voter turnout, votes for specific political parties, and politician behavior. The headline results of this project are that the 26th Amendment is associated with a substantial rise in total voter turnout that is in line with aggregate calculations done at the time. The effects on party vote share tend to be small and not clearly signed, though they do not show any substantial shift in favor of Democrats, which was a prevailing opinion at the time. There is some heterogeneity between college and non-college counties, with Republicans performing better in non-college counties than in counties with colleges after the lowering of the voting age.

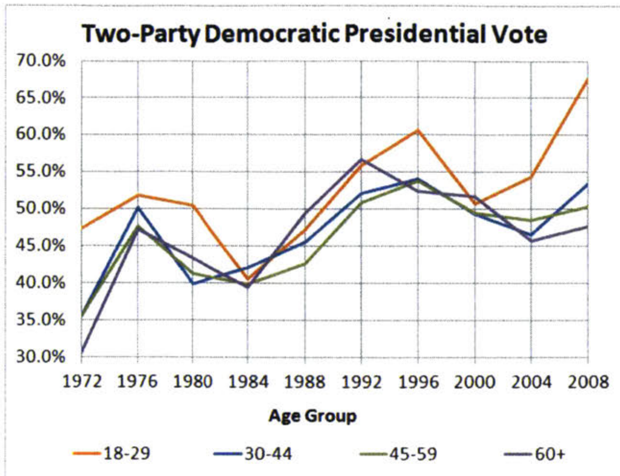
Looking at "placebo" states that had already reduced their voting ages prior to the 26th Amendment suggest that changes to campaigning to target new voters in particular appear to have been important. Particularly for presidential elections, it appears as though there was a substantial uptick in political participation among all young people around 1972, even in states where the voting age had previously been lowered.

Taken together, this work suggests a surprisingly small role for the reduction in the voting age in modern political outcomes. In spite of a sizable increase in the total population of potential voters, neither the turnout numbers nor the effects on vote shares or legislator behavior suggest any major shifts in outcomes. This stands in contrast to the work on suffrage for women (Miller, 2008) and blacks in the American South (Cascio and Washington, 2014), which found important shifts in spending and outcomes in the wake of those reforms. This comparative lack of results suggests that youth may be insufficiently focal as an identity, limiting the ability of a specific interest group to develop. Without a specific platform, it is then difficult to elect new officials or to put pressure on existing officials to implement relevant policies.

3.8 Tables and Figures

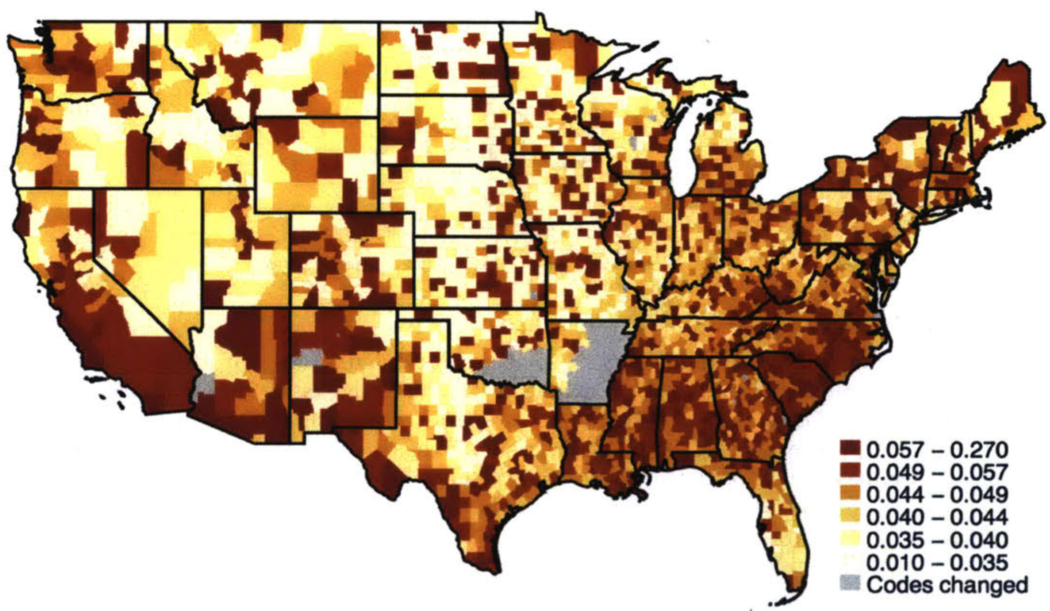
3.8.1 Figures

Figure 3-1: Reported share of votes for Democrats in presidential elections, by age group



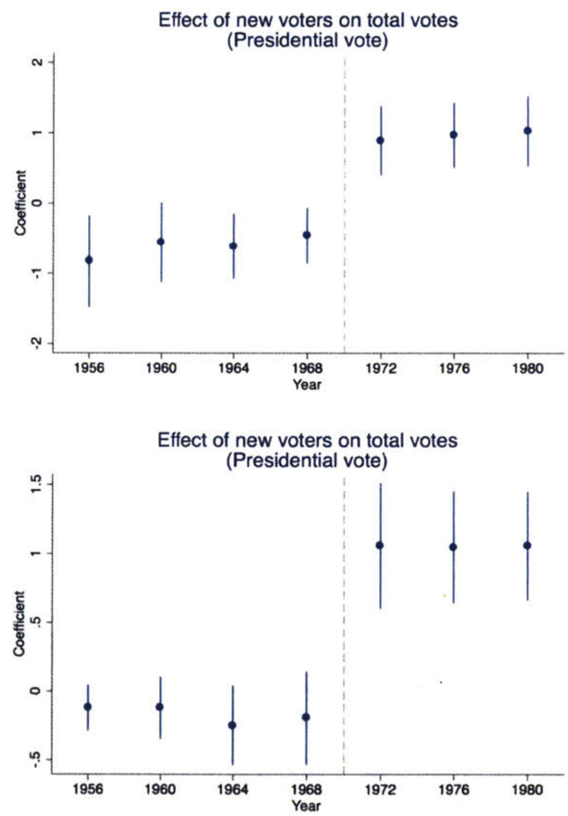
Notes: Exit poll results by age group for Democratic share of the presidential vote, from *Exit Polls: Surveying the American Electorate, 1927-2010* via Sabato's Crystal Ball.

Figure 3-2: Proportion of the population between ages 18 and 20 as of 1970 (by county)



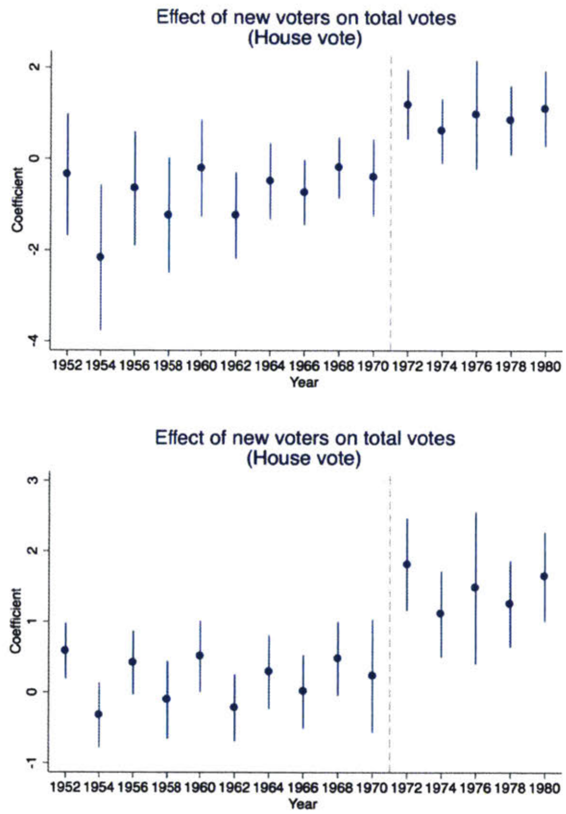
Notes: Map shows the proportion of the population in each county that is between the ages of 18 and 20 in the 1970 census. Dark lines indicate state boundaries. Gray counties had changes to their county codes between 1970 and 1990 (the mapping year).

Figure 3-3: Event Studies of the Effect of the Amendment on Turnout in Presidential Races.



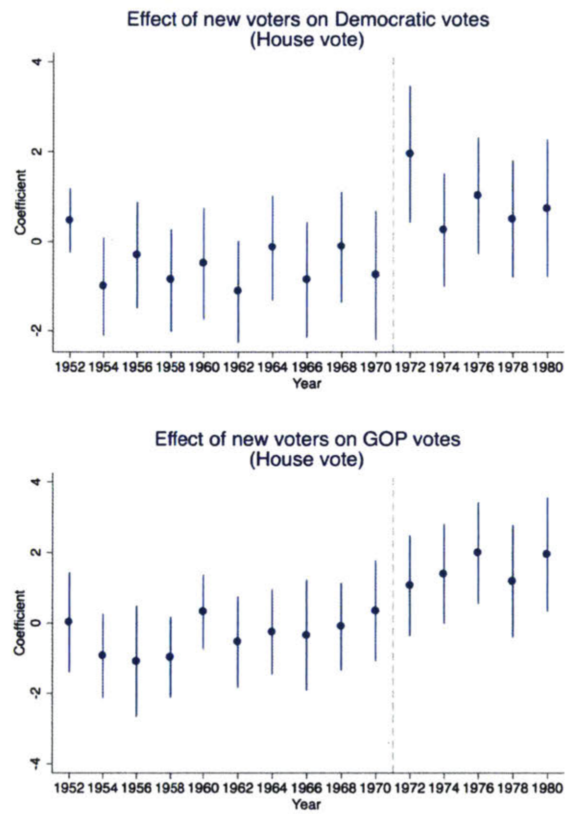
Notes: Results estimated using equation (3.3.1), so all results include state-year fixed effects and a control for the log of total population (not shown). Dependent variable is the log of total votes cast in presidential elections. The left panel defines the fraction of new eligible voters using the time-varying fraction of the population aged 18-20. The right panel shows the results fixing the fraction aged 18-20 in 1970. The gray line indicates the passage of the 26th Amendment. Confidence intervals are 95% confidence intervals. Standard errors are clustered by county.

Figure 3-4: Event Studies of the Effect of the Amendment on Turnout in House Races.



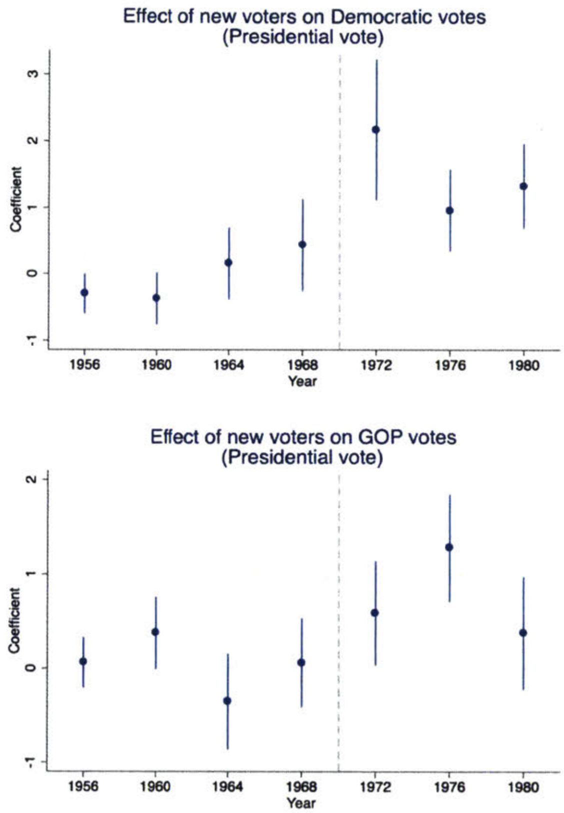
Notes: Results estimated using equation (3.3.1), so all results include state-year fixed effects and a control for the log of total population (not shown). Dependent variable is the log of total votes cast in House elections, excluding non-contested elections. The left panel defines the fraction of new eligible voters using the time-varying fraction of the population aged 18-20. The right panel shows the results fixing the fraction aged 18-20 in 1970. The gray line indicates the passage of the 26th Amendment. Confidence intervals are 95% confidence intervals. Standard errors are clustered by county.

Figure 3-5: Effects on votes for Democrats (left panel) and Republicans (right panel) in House races



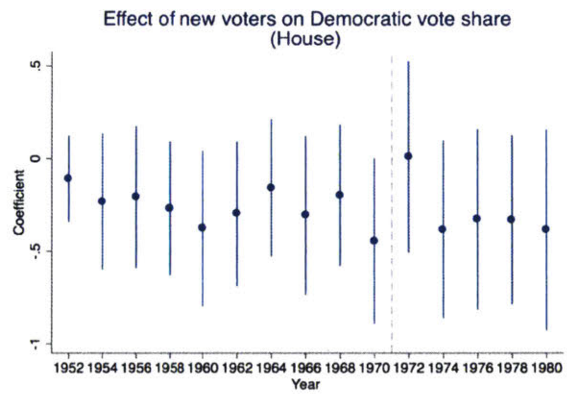
Notes: Results estimated using equation (3.3.1), so all results include state-year fixed effects and a control for the log of total population (not shown). In the left panel, the dependent variable is the log of total votes cast for Democratic candidates in contested House elections. In the right panel, the dependent variable is the log of total votes cast for Republican candidates in contested House elections. In both panels, the measure of new voters is the fraction aged 18-20 in 1970. The gray line indicates the passage of the 26th Amendment. Confidence intervals are 95% confidence intervals. Standard errors are clustered by county.

Figure 3-6: Effects on votes for Democrats (left panel) and Republicans (right panel) in presidential races



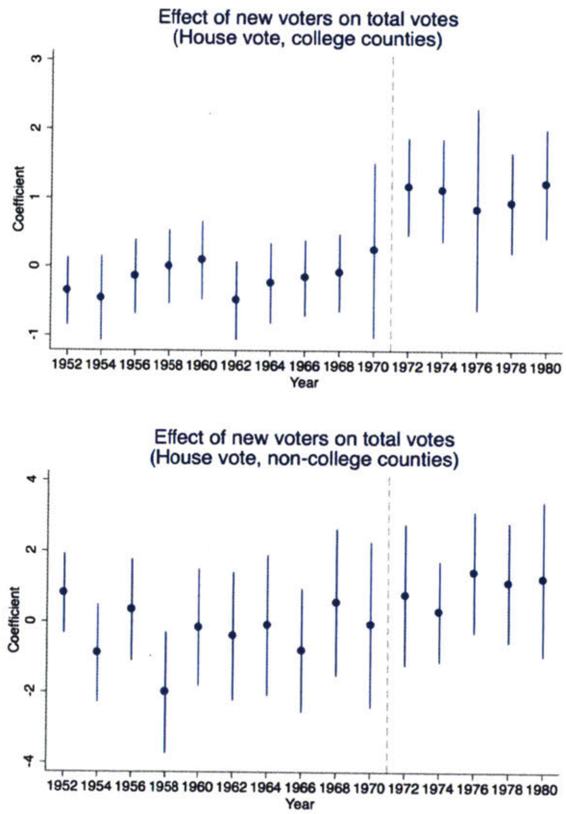
Notes: Results estimated using equation (3.3.1), so all results include state-year fixed effects and a control for the log of total population (not shown). In the left panel, the dependent variable is the log of total votes cast for Democratic candidates in presidential elections. In the right panel, the dependent variable is the log of total votes cast for Republican candidates in presidential elections. In both panels, the measure of new voters is the fraction aged 18-20 in 1970. The gray line indicates the passage of the 26th Amendment. Confidence intervals are 95% confidence intervals. Standard errors are clustered by county.

Figure 3-7: Effects on the vote share for Democrats in House races



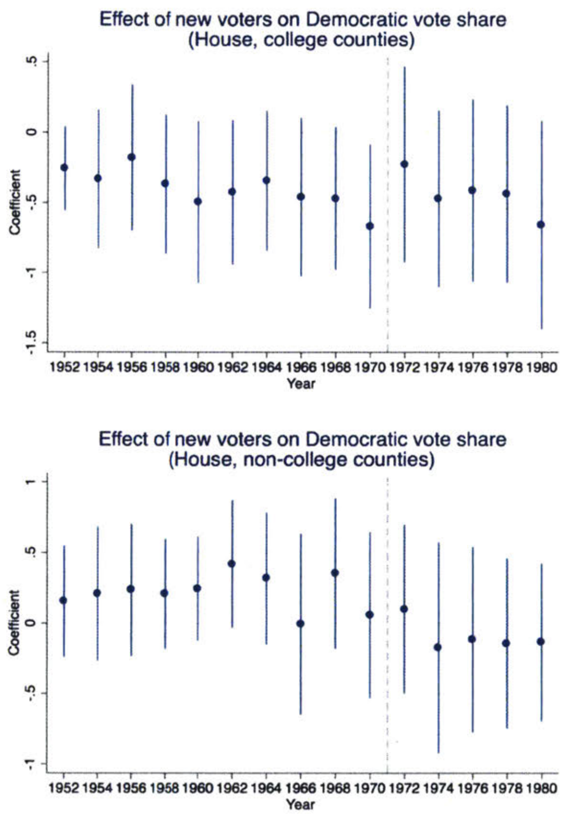
Notes: Results estimated using equation (3.3.1), so all results include state-year fixed effects and a control for the log of total population (not shown). The dependent variable is the share of total votes cast for Democratic candidates in contested House elections. The measure of new voters is the fraction aged 18-20 in 1970. The gray line indicates the passage of the 26th Amendment. Confidence intervals are 95% confidence intervals. Standard errors are clustered by county.

Figure 3-8: Effects on turnout for House by college (left panel) and non-college (right panel) counties



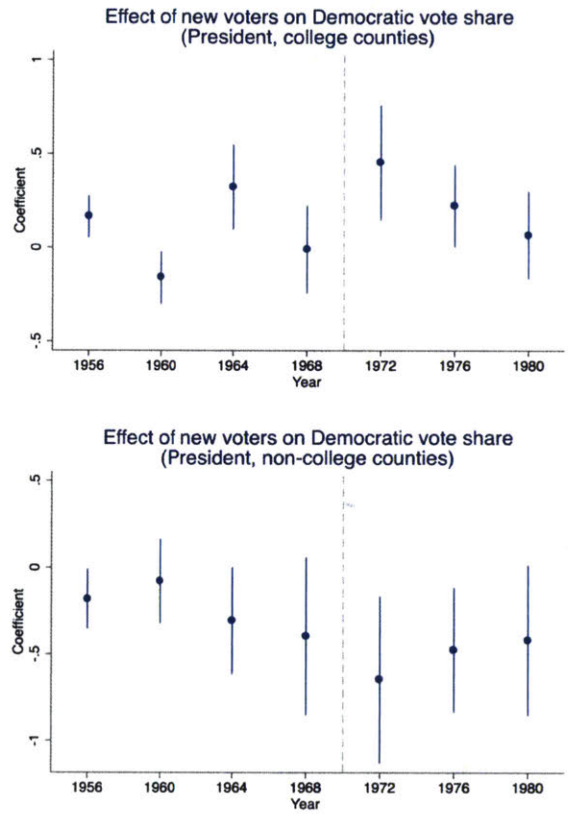
Notes: Results estimated using equation (3.3.1), so all results include state-year fixed effects and a control for the log of total population (not shown). In both panels, the measure of new voters is the fraction aged 18-20 in 1970 and the dependent variable is total votes cast in contested House elections. In the left panel, the sample is counties that have one or more colleges, while in the right panel, the sample is counties without a college. The gray line indicates the passage of the 26th Amendment. Confidence intervals are 95% confidence intervals. Standard errors are clustered by county.

Figure 3-9: Effects on Democratic vote share for House by college (left panel) and non-college (right panel) counties



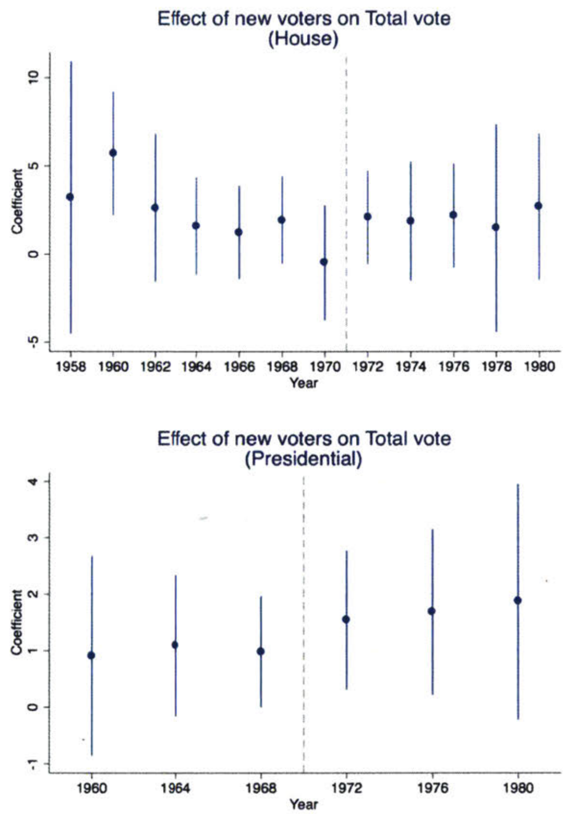
Notes: Results estimated using equation (3.3.1), so all results include state-year fixed effects and a control for the log of total population (not shown). In both panels, the measure of new voters is the fraction aged 18-20 in 1970 and the dependent variable is the share of total votes cast for Democratic candidates in contested House elections. In the left panel, the sample is counties that have one or more colleges, while in the right panel, the sample is counties without a college. The gray line indicates the passage of the 26th Amendment. Confidence intervals are 95% confidence intervals. Standard errors are clustered by county.

Figure 3-10: Effects on Democratic vote share for president by college (left panel) and non-college (right panel) counties



Notes: Results estimated using equation (3.3.1), so all results include state-year fixed effects and a control for the log of total population (not shown). In both panels, the measure of new voters is the fraction aged 18-20 in 1970 and the dependent variable is the share of total votes cast for Democratic candidates in presidential elections. In the left panel, the sample is counties that have one or more colleges, while in the right panel, the sample is counties without a college. The gray line indicates the passage of the 26th Amendment. Confidence intervals are 95% confidence intervals. Standard errors are clustered by county.

Figure 3-11: Effects of the 26th Amendment on total votes for House (left) and total votes for president (right) in placebo states



Notes: Results estimated using equation (3.3.1), so all results include state-year fixed effects and a control for the log of total population (not shown). In both panels, the measure of new voters is the fraction aged 18-20 in each year and the sample is restricted to the states of Georgia and Kentucky, years 1956-1980. In the left panel, the dependent variable is the log of total votes cast in contested House elections, while in the right panel, the dependent variable is the log of total votes cast in presidential elections. The gray line indicates the passage of the 26th Amendment. Confidence intervals are 95% confidence intervals. Standard errors are clustered by county.

3.8.2 Tables

Table 3.1: Variation in voting ages prior to the 26th Amendment

State	Voting Age	Year of Change
Georgia	18	1943
Kentucky	18	1955
Alaska	19	1959 (year of entry as a state)
Hawaii	20	1959 (year of entry as a state)
Alaska	18	1970
Maine	20	1970
Massachusetts	19	1970
Minnesota	19	1970
Montana	19	1970
Nebraska	20	1970

Notes: Data is from Saldin (2010). States may appear more than once if they had more than one voting age change below age 21. All changes were approved in the year listed and only applied to elections occurring *after* that year (not in the same year).

Table 3.2: Availability of age data by Census year

Year	Grouped Age Data	Detailed Age Data (20 and under)	Age-By-Year (all years)
1950	X	X	
1960	X	X	
1970			X
1980			X

Notes: An X indicates that the form of data listed in the column header is available for the census year listed in that row. A given census year may have more than one type of data available.

Table 3.3: Effects of the 26th Amendment on total turnout by office

VARIABLES	(1) House	(2) House	(3) Pres.	(4) Pres.
Log total pop. aged 21+	1.078***	1.077***	1.045***	1.041***
	(0.0285)	(0.0284)	(0.0240)	(0.0239)
Frac. new voter in 1970*Post	1.289***		1.197***	
	(0.194)		(0.147)	
Frac. new voter*Post		1.440***		1.335***
		(0.208)		(0.149)
Observations	35,926	36,269	21,491	21,965
R-squared	0.994	0.994	0.999	0.999

Notes: Results estimated using equation (3.3.1), so all results include state-year fixed effects (not shown). In columns (1) and (2), the dependent variable is the log of total votes cast in contested House elections. In columns (3) and (4), the dependent variable is the log of total votes cast in presidential elections. Columns (1) and (3) use the fraction of voters between the ages of 18-20 in 1970 as the measure of new voters post-1970. Columns (2) and (4) use the time-varying fraction of voters between the ages of 18-20 as the measure of new voters post-1970. Standard errors are clustered by county. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Table 3.4: Effects of the 26th Amendment on votes for each party by office

VARIABLES	(1) Dem. House	(2) Dem. Pres.	(3) GOP House	(4) GOP Pres.
Log total pop. aged 21+	0.933***	0.978***	1.237***	1.092***
	(0.0528)	(0.0334)	(0.0659)	(0.0282)
Frac. new voter in 1970*Post	1.386***	1.463***	1.835***	0.712***
	(0.336)	(0.239)	(0.469)	(0.219)
Observations	34,860	20,681	34,860	20,681
R-squared	0.988	0.996	0.978	0.996

Notes: Results estimated using equation (3.3.1), so all results include state-year fixed effects (not shown). In columns (1) and (2), the dependent variable is the log of total votes cast for Democratic candidates. In columns (3) and (4), the dependent variable is the log of total votes cast for Republican candidates. Columns (1) and (3) limit the sample to contested House elections. Columns (2) and (4) limit the sample to presidential elections. Standard errors are clustered by county. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Table 3.5: Effects of the 26th Amendment on Democratic vote share

VARIABLES	(1) Dem. share House	(2) Dem. share Pres.
Log total pop. aged 21+	-0.0831***	-0.0314***
	(0.0173)	(0.00858)
Frac. new voter in 1970*Post	-0.0402	0.0970
	(0.135)	(0.0725)
Observations	34,860	20,927
R-squared	0.769	0.909

Notes: Results estimated using equation (3.3.1), so all results include state-year fixed effects (not shown). The dependent variable is the share of votes cast for Democratic candidates. Column (1) limits the sample to contested House elections. Column (2) limits the sample to presidential elections. Standard errors are clustered by county. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Table 3.6: Effects of the 26th Amendment on vote shares for each party by office and college/non-college

VARIABLES	(1) Dem. % House	(2) Dem. % House	(3) Dem. % Pres.	(4) Dem. % Pres.
Log total pop. aged 21+	-0.0281	-0.104***	-0.0510***	-0.0313***
	(0.0195)	(0.0223)	(0.00988)	(0.0112)
Frac. new voter in 1970*Post	-0.312	-0.0628	-0.295**	0.172**
	(0.210)	(0.191)	(0.141)	(0.0846)
Observations	25,356	9,429	15,477	5,416
R-squared	0.804	0.773	0.896	0.929
Sample	Non-College	College	Non-College	College

Notes: Results estimated using equation (3.3.1), so all results include state-year fixed effects (not shown). The dependent variable is the share of votes cast for Democratic candidates. Columns (1) and (2) limit the sample to contested House elections. Columns (3) and (4) limit the sample to presidential elections. Odd-numbered columns ((1) and (3)) restrict the set of counties to counties with no four-year colleges. Even-numbered columns ((2) and (4)) restrict the set of counties to counties with one or more four-year colleges. Standard errors are clustered by county. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Table 3.7: Effects of the 26th Amendment on total votes in House and presidential races in the two placebo states, Georgia and Kentucky

VARIABLES	(1) House turnout	(2) Presidential turnout
Log total pop. aged 21+	1.026***	0.882***
	(0.0952)	(0.0515)
Frac. new voter*Post	0.768	1.001**
	(0.608)	(0.419)
Observations	1,895	1,945
R-squared	0.993	0.996

Notes: Results estimated using equation (3.3.1), so all results include state-year fixed effects (not shown). The sample is restricted to the states of Georgia and Kentucky, years 1956-1980. In column (1), the dependent variable is the log of votes cast in contested House elections. In column (2), the dependent variable is the log of votes case in presidential elections. The measure of new voters is the fraction aged 18-20 in each year. Standard errors are clustered by county. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Appendix

3.8.3 Population

The Census does not provide measurements of total population by county for intercensal years until 1970 (i.e. there are only population measures in 1950, 1960, and 1970, and not intervening years, such as 1956). As a result, I estimated this population using a linear interpolation between the starting and ending population for a given county. For example, consider a county with population X at census year t and population Y at $t+10$. Then the estimated population for the intervening years can be found in Table 3.8.

This method is similar to that used by the Census to provide some intercensal estimates.

Table 3.8: Estimating population in intercensal years

Year	Population
t	X
$t+2$	$\frac{2}{10}(Y-X)+X$
$t+4$	$\frac{4}{10}(Y-X)+X$
$t+6$	$\frac{6}{10}(Y-X)+X$
$t+8$	$\frac{8}{10}(Y-X)+X$
$t+10$	Y

Notes: This table shows the method used for estimating total population in each county for intercensal years, where t is the year of one census and $t+10$ is the year of the following census (and thus X and Y are known).

Bibliography

ACFIM (2015): "Who Pays the Piper? MP Survey on the Commercialization of Politics," Tech. rep.

AFROBAROMETER (2006): "Afrobarometer Round 3," `http://www.afrobarometer.org`.

ALL AFRICA (2016): "Uganda: Opposition Right to Contest Poll Outcome - Rights Watchdog," Tech. rep.

AMNESTY INTERNATIONAL (2015): "Uganda: Arbitrary arrests and excessive use of force hindering debate in run-up to elections," .

——— (2016): "Amnesty International International Report 2015/2016," Tech. rep.

ANDREONI, J., M. A. KUHN, AND C. SPRENGER (2015): "Measuring time preferences: A comparison of experimental methods âňİ," *Journal of Economic Behavior and Organization*, 116, 451–464.

BAIRD, S., J. A. BOHREN, C. MCINTOSH, AND B. OZLER (2014): "Designing Experiments to Measure Spillover Effects," *PIER Working Paper No. 14-006.*

BANERJEE, A. V., S. COLE, E. DUFLO, AND L. LINDEN (2007): "Remedying Education: Evidence From Two Randomized Experiments in India," *Quarterly Journal of Economics*, 122, 1235–1264.

BARDHAN, P. (1997): "Corruption and Development : A Review of Issues," *Journal of Economic Literature*, 35, 1320–1346.

BEBER, B. AND A. SCACCO (2012): "What the Numbers Say: A Digit-Based Test for Election Fraud," *Political Analysis*, 20, 235–247.

BECK, P. J. AND M. W. MAHER (1986): "A comparison of bribery and bidding in thin markets," *Economics Letters*, 20, 1–5.

BERTRAND, M., S. DJANKOV, R. HANNA, AND S. MULLAINATHAN (2007): "Obtaining a Driver's License in India: An Experimental Approach to Studying Corruption," *Quarterly Journal of Economics*, 1639–1676.

BERTRAND, M., E. DUFLO, AND S. MULLAINATHAN (2004): "How Much Should We Trust Differences-In-Differences Estimates?" *Quarterly Journal of Economics*, 119, 249–275.

Besley, T. and S. Coate (1997): "An Economic Model of Representative Democracy," *Quarterly Journal of Economics*, 112, 85–114.

Besley, T., T. Persson, and D. M. Sturm (2010): "Political Competition, Policy and Growth: Theory and Evidence from the US," *Review of Economic Studies*, 77, 1329–1352.

Black, D. (1948): "On the Rationale of Group Decision-Making," *Journal of Political Economy*, 56, 23–34.

Camp, E. and M. Szwarcberg (2015): "Stealing for the Party: Brokers and Competitive Distribution in Argentina," .

Cascio, E. U. and E. Washington (2014): "Valuing the Vote: The Redistribution of Voting Rights and State Funds Following the Voting Rights Act of 1965," *Quarterly Journal of Economics*, 129, 379–433.

Chattopadhay, R. and E. Duflo (2004): "Women as Policy Makers: Evidence From a Randomized Policy Experiment in India," *Econometrica*, 72, 1409–1443.

Conroy-Krutz, J. (2012): "What Determines the Price of a Vote? Motive, Means, and Opportunity in Vote-Buying Practices," .

Daily Mail (2016): "Ugandan election commission lacks "independence": EU observers," Tech. rep.

Dekel, E., M. O. Jackson, and A. Wolinsky (2008): "Vote Buying: General Elections," *Journal of Political Economy*, 116, 351–380.

Democracy Monitoring Group (2011): "Report on Money in Politics: Pervasive Vote Buying in Ugandan Elections," Tech. Rep. January.

Dhaliwal, I. and R. Hanna (2013): "Deal with the Devil: The Successes and Limitations of Bureaucratic Reform in India," .

Dreher, A. and M. Gassebner (2011): "Greasing the wheels? The impact of regulations and corruption on firm entry," *Public Choice*, 155, 413–432.

Duflo, E., R. Hanna, and S. P. Ryan (2012): "Incentives work: Getting teachers to come to school," *American Economic Review*, 102, 1241–1278.

European Union Election Observation Mission (2016): "Uganda, Presidential, Parliamentary and Local Council Elections," Tech. rep.

Fafchamps, M., A. Vaz, and P. C. Vicente (2012): "Voting and Peer Effects: Experimental Evidence from Mozambique," .

Ferraz, C., F. Finan, and D. B. Moreira (2012): "Corrupting learning Evidence from missing federal education funds in Brazil," *Journal of Public Economics*, 96, 712–726.

Finan, F., H. Larreguy, and L. Schechter (2016): "Vote Buying and Networks: Information, Enforcement or Both?" .

Finan, F. and L. Schechter (2012): "Vote-Buying and Reciprocity," *Econometrica*, 80, 863–881.

FISMAN, R. AND J. SVENSSON (2007): "Are corruption and taxation really harmful to growth? Firm level evidence," *Journal of Development Economics*, 83, 63–75.

FUJIWARA, T. AND L. WANTCHEKON (2013): "Can Informed Public Deliberation Overcome Clientelism? Experimental Evidence from Benin," *American Economic Journal: Applied Economics*, 5, 241–255.

GANS-MORSE, J., S. MAZZUCA, AND S. NICHTER (2014): "Varieties of Clientelism: Machine Politics during Elections," *American Journal of Political Science*, 58, 415–432.

GREEN, D. P. AND S. VASUDEVAN (2016): "Diminishing the Effectiveness of Vote Buying: Experimental Evidence from a Persuasive Radio Campaign in India," *Working paper.*

HAYS, J. (1951): "Georgia's Official Register 1945-1950," Tech. rep., State of Georgia, Department of Archives and History, Atlanta.

HICKEN, A., S. LEIDER, N. RAVANILLA, AND D. YANG (2014): "Temptation in Vote-Selling: Evidence from a Field Experiment in the Philippines," .

HICKEN, A. AND J. W. SIMMONS (2008): "The Personal Vote and the Efficacy of Education Spending," *American Journal of Political Science*, 52, 109–124.

HUNTINGTON, S. (1968): *Political Order in Changing Societies*, New Haven, CT: Yale University Press.

HUSTED, T. A. AND L. W. KENNY (1997): "The Effect of the Expansion of the Voting Franchise on the Size of Government," *Journal of Political Economy*, 105, 54–82.

ICHINO, N. AND M. SCHÜNDELN (2012): "Deterring or Displacing Electoral Irregularities? Spillover Effects of Observers in a Randomized Field Experiment in Ghana," *The Journal of Politics*, 74, 292–307.

KATO, A. AND T. SATO (2015): "Greasing the wheels? The effect of corruption in regulated manufacturing sectors of India," *Canadian Journal of Development Studies/Revue canadienne d'études du développement*, 36, 459–483.

KAUFMANN, D. AND S. WEI (1999): "Does "grease money" speed up the wheels of commerce?" .

KEEFER, P. AND R. VLAICU (2008): "Democracy, Credibility, and Clientelism," *Journal of Law, Economics, and Organization*, 24, 371–406.

KHAN, A., A. KHWAJA, AND B. OLKEN (2016): "Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors," *The Quarterly Journal of Economics.*

KLING, J. R., J. B. LIEBMAN, AND L. F. KATZ (2007): "Experimental Analysis of Neighborhood Effects," *Econometrica*, 75, 83–119.

LARREGUY, H. (2013): "Monitoring Political Brokers: Evidence from Clientelistic Networks in Mexico," .

LARREGUY, H., J. MARSHALL, AND P. QUERUBIN (2016): "Parties, Brokers and Voter Mobilization: How Turnout Buying Depends Upon the Party's Capacity to Monitor Brokers," *American Political Science Review*, 110, 160–179.

LEFF, N. H. (1964): "Economic development through bureaucratic corruption," *American behavioral scientist*, 8, 8–14.

LEYS, C. (1965): "What is the Problem about Corruption?" *The Journal of Modern African Studies*, 3, 215–230.

LIEN, D.-H. D. (1986): "A note on competitive bribery games," *Economics Letters*, 22, 337–341.

LOWES, S., N. NUNN, J. A. ROBINSON, AND J. L. WEIGEL (2017): "The evolution of culture and institutions: Evidence from the Kuba Kingdom," *Econometrica*, 85, 1065–1091.

LUI, F. T. (1985): "An equilibrium queuing model of bribery," *Journal of political economy*, 93, 760–781.

LUTTMER, E. F. P. AND M. SINGHAL (2014): "Tax Morale," *Journal of Economic Perspectives*, 28, 149–168.

LYONS, L. (2004): "Gallup Brain: History of the Youth Vote," *Gallup Poll Tuesday Briefing*, 1–4.

MAGALONI, B., E. MIN, AND J. CHU (2013): "Autocracies of the World, 1950-2012," *Version 1.0). Dataset, Stanford University.*

MAURO, P. (1995): "Corruption and Growth," *Quarterly Journal of Economics*, 110, 681–712.

MÉON, P. G. AND K. SEKKAT (2005): "Does corruption grease or sand the wheels of growth?" *Public Choice*, 122, 69–97.

MÉON, P.-G. AND L. WEILL (2010): "Is Corruption an Efficient Grease?" *World Development*, 38, 244–259.

MILLER, G. (2008): "Women's Suffrage, Political Responsiveness, and Child Survival in American History," *Quarterly Journal of Economics*, 123, 1287–1327.

NARITOMI, J. (2015): "Consumers as Tax Auditors," .

NEALE, T. H. (1983): "The Eighteen Year Old Vote: The Twenty-Sixth Amendment and Subsequent Voting Rates of Newly Enfranchised Age Groups," Tech. rep., Congressional Research Service, Washington, DC.

NEWSWEEK (2016): "Uganda 2016: Mbabazi mounts challenge to election results," Tech. rep.

NICHTER, S. (2008): "Vote Buying or Turnout Buying? Machine Politics and the Secret Ballot," *American Political Science Review*, 102, 19–31.

OLKEN, B. A. (2007): "Monitoring Corruption: Evidence from a Field Experiment in Indonesia," *Journal of Political Economy*, 115, 200–249.

OLKEN, B. A. AND R. PANDE (2012): "Corruption in Developing Countries," *Annual Review of Economics*, 4, 479–509.

Osborne, M. J. and A. Slivinski (1996): "A Model of Political Competition with Citizen-Candidates," *Quarterly Journal of Economics*, 111, 65–96.

Pomeranz, D. (2015): "No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax," *American Economic Review*, 105, 2539–2569.

President's Commission on Campus Unrest (1970): "The Report of the President's Commission on Campus Unrest," Tech. rep., Washington, DC.

Saldin, R. P. (2010): *War, The American State, and Politics Since 1898*, Cambridge, UK: Cambridge University Press.

Sequeira, S. and S. Djankov (2014): "Corruption and Firm Behavior: Evidence from African Ports," *Journal of International Economics*.

Stokes, S. C. (2005): "Perverse accountability: A formal model of machine politics with evidence from Argentina," *American Political Science Review*, 99, 315–325.

Stokes, S. C., T. Dunning, M. Nazareno, and V. Brusco (2013): *Brokers, Voters, and Clientelism: The Puzzle of Distributive Politics*, Cambridge University Press.

The Guardian (2016): "Ugandans cast votes in presidential elections after lengthy delays," Tech. rep.

Tripp, A. M. (2010): *Museveni's Uganda: Paradoxes of Power in a Hybrid Regime*, Boulder, CO: Lynne Rienner Publishers.

U.S. Bureau of the Census (1973): "Voting and Registration in the Election of November 1972," Tech. rep., U.S. Bureau of the Census, Washington, DC.

Vial, V. and J. Hanoteau (2010): "Corruption, manufacturing plant growth, and the Asian paradox: Indonesian evidence," *World Development*, 38, 693–705.

Vicente, P. C. (2014): "Is Vote Buying Effective? Evidence from a Field Experiment in West Africa," *The Economic Journal*, 124, F356–F387.

Wantchekon, L. (2003): "Clientelism and Voting Behavior: Evidence from a Field Experiment in Benin," *World Politics*, 55, 399–422.