Title

Building Fiscal Capacity in Weak States: Experimental and Qualitative Evidence from Sierra Leone

Author

Grieco, Kevin

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Building Fiscal Capacity in Weak States:
Experimental and Qualitative Evidence from Sierra Leone

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Political Science

by

Kevin Michael Grieco

2024

ABSTRACT OF THE DISSERTATION

Building Fiscal Capacity in Weak States:
Experimental and Qualitative Evidence from Sierra Leone

by

Kevin Michael Grieco
Doctor of Philosophy in Political Science
University of California, Los Angeles, 2024
Professor Daniel N. Posner, Co-Chair
Professor Darin Eugene Christensen, Co-Chair

The three essays in this dissertation explore how weak states can build fiscal capacity, which is essential for economic development and political stability. In Chapter 2, I argue that governments in weak states can build fiscal capacity by collaborating with non-state, traditional political institutions (TPIs). Using an experiment I embedded within the local government's campaign to collect property taxes, I show that this collaboration increases citizens' tax compliance because TPIs possess coercive capacity and legitimacy. Chapter 3 further explores the legitimacy of TPIs. I argue that traditional leaders are popular because they enforce local laws that increase community welfare. I support this argument with qualitative data from 300 interviews conducted in Kono district, Sierra Leone. Chapter 4 investigates whether governments in weak states can enhance their legitimacy and foster tax compliance by increasing public participation in political affairs. We experimentally evaluate a phone-based participatory budgeting intervention in Freetown, Sierra Leone and find that the intervention increases citizens' perceptions of government legitimacy. Effects on tax compliance, however, are conditional on political affiliation and policy preferences about taxation.

The dissertation of Kevin Michael Grieco is approved.

Araceli Cruz

Kate Baldwin

Daniel N. Posner, Committee Co-Chair

Darin Eugene Christensen, Committee Co-Chair

University of California, Los Angeles

2024

Table of Contents

List of Figures

List of Tables

Acknowledgments

Daniel Posner and Darin Christensen supported my graduate education from the moment I stepped on to campus at UCLA. I could not have asked for a better tandem to co-chair my committee. I learned more from watching them work than in any course I could have taken. Our weekly zoom meetings during COVID-19 prevented my dissertation from being yet another victim of the global pandemic. I am extremely grateful to Kate Baldwin and Cesi Cruz for rounding out my committee. Kate's book on chiefs is the first book of political science I ever read and is a major source of inspiration for my own work. Cesi's positive energy is infectious. Many academics are smart. I feel very lucky to have had the fortune of working together with a group of scholars who are also fun to be around.

At UCLA, I benefited intellectually and spiritually from an array of coffee-based meet ups. Each week I carried my coffee mug to the sculpture garden behind Bunche Hall to where Darin convened a small group of students. In these meetings Daniel Carnahan, Aurora Echavarria, and Frank Wyer listened patiently and attentively as I detailed my (lack of) progress from the previous week. These meetings kept me on track and were a source of energy. Jihae Hong was always ready to walk with me to get coffee after lunch, preferably at Pete's. During our first year, Julian Michel and I met religiously on Sundays at Cafe Profetta; we also spent countless afternoons on the benches in front of Bunche, talking shop over his jumbo latte. Zep Kalb shared research ideas alongside a mocka on many mornings in the apartment we shared on Kelton Avenue with Julian and Derek Holliday. I had the honor of hosting the Bunche Hall Coffee Group in the office I shared with Jenny Hamilton, Ciara Sterbenz, and Carolyn Steinle. Julian and Derek, as well as Brigid Morris, Elayne Stecher, and Chase Privett were core members and regular attendees. When it was too late to drink coffee, this group, accompanied by others, carpooled down to Broxton for beers. For better or for worse, most graduate students spent most of their free time with other graduate students. Certainly this was the case for me. And I could not have been luckier with the friends I made at UCLA.

It is just over eight years since my first trip to Sierra Leone as a master's student at Wageningen

University. I am grateful to Maarten Voors for giving me that opportunity, and for the many opportunities that followed. It was on this first trip that I met Niccolò Meriggi at a conference at Njala University in Mokonde. I visited Nicco the next week in Freetown and promptly moved in. It was my first taste of research and Sierra Leone; I became hooked on both. Unfortunately, the two people who taught me the most about Sierra Leone, and to whom I am most indebted to for this dissertation, are no longer with us. Mr. Jalloh was a mentor and friend whose dedication to his work and his country kept my cynicism and pessimism at bay. Mr. Atta opened his home to me for sixth months in 2021 and shared with me the beauty of Lebanese cooking, the calming properties of whiskey on ice, and his vision of the quiet life ahead. When Yasmina came to stay with me at his house, he let it be know that I must marry her at once. I miss both these men.

It would have taken me twice as long to finish this dissertation if not for the hard work of the front office. I would like to extend a particular thanks to Esther Blair. If one were to dive into UCLA's bureaucratic underbelly (god save them), the administrative records would undoubtedly show that I have filed more late drop petitions and leave of absence requests than any graduate student in the university's recorded history. Esther had my back each and every time. I would also like to thank Belinda Sunnu her energy and chattiness around the microwave.

Lastly, I would like to thank my family and friends. I have not been available for as many phone calls as I would have liked. I promise to be better.

Chapters 2 and 3 received generous funding for data collection from the International Centre for Tax and Development. The International Growth Centre provided funding for scoping research. I am grateful for feedback from Kate Baldwin, Graeme Blair, Darin Christensen, Cesi Cruz, Barbara Geddes, Ellen Lust, Colette Nyirakamana, Daniel Posner, Wilson Prichard, Sridhar Telidevara, Frank Wyer, and Kerem Yildirim. I would like to thank participants at the Politics of Order and Development lab and COMPASS workshop at UCLA. Seminar participants at the 6^{th} GLD Conference, the Sierra Leone Tax for Development Conference, the GLD Work-in-Progress session, and WGAPE 2024 (Accra) provided helpful comments and suggestions. I gratefully acknowl-

edge a hardworking enumeration team for data collection. Tamba Moitenga, Henry Mondeh, and Aiah Morsay provided excellent research assistance throughout this project. Of course, this project would not be possible without the support and collaboration of the Kono District Council, the Kono Council of Paramount Chiefs, and Knowledge for Community Empowerment Organization. The preanalysis plan for this project is registered at osf.io/g7xwv.

Chapter 4 is co-authored with Abou Bakarr Kamara, Niccolò Meriggi, Julian Michel, and Wilson Prichard and is currently in preparation for publication. This chapter was supported by generous funding from the International Growth Center and the International Center for Tax and Development. We received helpful comments from Graeme Blair, Katherine Casey, Darin Christensen, Cesi Cruz, Adrienne LeBas, Giulia Mascagni, Nara Monkam, Oyebola Okunogbe, Daniel Posner, Soledad Prillaman, and Jon Weigel. Chapter 4 also benefited greatly from participants at the CaliWEPSs IV workshop, APSA 2022, the Global Development Conference 2022, and the 2024 World Bank Land Conference. We are grateful to Jacques Courbe, Emile Eleveld, Yojin Higashibaba, Xenia Rak, Michael Rozelle, and Ella Tyler for excellent research assistance. Thanks also to our wonderful team of supervisors and enumerators. Of course, this project would not have been possible without the contributions and cooperation of the Freetown City Council. The preanalysis plan associated with Chapter 4 is registered at: https://osf.io/dhkfe.

VITA

MS, International Development, Wageningen University, 2016

BA, Economics & Philosophy, University of New Hampshire, 2012

Chapter 1

Capacity Building Experiments in Sierra Leone

In *States and Power in Africa*, Jeffrey Herbst (2014) argues that the fundamental problem facing state-builders in Africa is "project[ing] authority over inhospitable territories that contain relatively low densities of people." While Herbst wrote these words 24 years ago, they would ring true today for any visitor to Kono District in eastern Sierra Leone, where the median village contains 17 buildings and the only paved road is the Chinese-built highway that links to Freetown, the country's capital. The dirt roads connecting the district's sparsely populated villages create headaches for administrators. In the dry season, travel is expensive—the transportation cost (by motorbike) from the median village to the district capital is greater than the country's minimum wage (Grieco, 2024)—and these costs sky-rocket in the rainy season.[1] when roads turn to mud and many become impassible.

The Kono District Council (KDC), the local government with jurisdiction over the rural parts of

[1]Roads are central to Herbst's analysis of state capacity, as he uses road density to proxy colonial administrators' "ability to broadcast power." I calculate median number of buildings from the 2015 National Census. The 75th and 25th percentile villages have 33 and 8 buildings, respectively.

the district, has certainly struggled to project its authority. Outside of the district capital, state-provided public services are minimal: there is no connection to the electricity grid or piped water, and while schools and health facilities exist, they are chronically understaffed and under-equipped. Rule of law is largely left to traditional leaders in the country's 190 chiefdoms, where traditional chiefs enact and enforce laws and resolve disputes.

Perhaps unsurprisingly the KDC collects very little taxes. In-person collection is both logistically challenging and resource intensive and given that the KDC has no track record of service delivery, citizens are likely to be reluctant to pay. Of course, without tax revenue the KDC cannot invest in the physical or administrative infrastructure to efficiently demand taxes; cannot invest in the enforcement infrastructure to compel payments; and cannot improve service delivery to engender citizens' compliance. In short, the KDC seems stuck in a pernicious equilibrium: they are too weak to compel or induce tax compliance and, thus, lack the tax revenue to build capacity.

This situation is not unique to Sierra Leone. Many governments in sub-Saharan Africa are weak, a product of colonialism (Michalopoulos and Papaioannou, 2020) and, ironically, the stability of the state system they were born into (Herbst, 1990).[2] A consequence of this legacy of state weakness is that many governments in sub-Saharan Africa struggle to effectively raise taxes (Lee and Gordon, 2005; Fjeldstad et al., 2014), which, in turn, is a cause of contemporary state weakness.

The essays in this dissertation explore the political economy of weak states by addressing two questions. First, how can weak states build fiscal capacity? For students of politics, understanding how states build fiscal capacity is important because governments need resources to project power and accomplish their goals. Understanding how they do this is a fundamental question in political science. For scholars of development, studying the state's ability to collect taxes is important because fiscal capacity, and state capacity more generally, are key driver of economic development (Besley and Persson, 2011). With tax revenues, governments can make growth-stimulating

[2] According to Hanson and Sigman (2021), state capacity is lower in sub-Saharan Africa than any other region in the world and has been since at least the 1960's.

investments in human capital and public infrastructure.

Second, if the state is too weak to govern, how do non-state institutions provide social order? Across much of the world, weak states coexist alongside traditional political institutions (TPIs) (Holzinger et al., 2016), especially in sub-Saharan Africa where Baldwin and Holzinger (2019) estimate that 83% of the continent's population is governed, at least in part, by TPIs. Since at least Kate Baldwin's (2016) landmark *The Paradox of Traditional Chiefs in Democratic Africa*, political scientists have known that these institutions influence their constituents' behavior in ways that matter for development. And while there is increasing scholarly interest in these institutions, we still know little about how these institutions exert social control. Given the prevalence of these institutions and their political and economic significance, understanding the mechanisms through which TPIs project their authority is a crucial question for political scientists.

Chapter 2 addresses both questions by investigating the KDC's recent tax reform. In 2018, the KDC initiated a tax reform to systematically collect taxes in rural villages for the first time since the post-war reintroduction of district councils in 2004. To overcome capacity constraints, the KDC partnered with the district's 14 Paramount Chiefs: the chiefs would help the KDC collect a newly introduced property tax and, in return, half of the collected revenue would be earmarked for spending in the chiefdom where it was collected.[3] As the Chief Administrator told me, "If you want to succeed in anything in the chiefdoms, it is very important for you to involve the chiefs."[4] By collaborating with TPIs, the KDC was following a well-trodden path. Colonial governments in Africa often collaborated with TPIs to accomplish their goals—such as the suppression of the slave trade, security of trade routes, and tax collection (Crowder, 1968, 1964; Migdal, 1988)—and the post-colonial state has followed in its footsteps (see Buur and Kyed, 2007). Using a field experiment, Chapter 2 tests whether the KDC's collaboration with TPIs was an effective strategy for increasing citizens' tax compliance and therefore building state capacity. I also attempt to

[3]By April 2018, the KDC had signed a memorandum of understanding with all the district's paramount chiefs that outlined this revenue sharing agreement between the KDC and the respective chiefdom council.

[4]Interview with Kono District Council Chief Administrator, November 30, 2022.

tease out the mechanisms that make this collaboration effective and in doing so shed light on how non-state TPIs exert authority in rural areas where the state is weak.

While politicians and government officials in Kono sometimes clashed with traditional leaders, they were well aware of the respect traditional leaders commanded from constituents. "In Africa, and especially in this country," the Chief Administrator of the KDC told me, ". . . chiefs are the symbol of respect." He conceded than many chiefs were "more respected than even the councilors and the [Council] chairman," the elected officials in local government. The Chief Administrator's impression that traditional leaders have broad public support that outstrips elected government leaders is in keeping with Afrobarometer survey data, which consistently finds that traditional leaders are more trusted than elected politicians. Moreover, there is little indication that Africans want to see the governance of their communities handed over to the state at the expense of TPIs: a striking 80% of respondents think that traditional leaders' influence in governing local communities should either increase or stay the same (Logan and Amakoh, 2022).

Chapter 3 explores TPIs' popularity and digs deeper into how they govern areas where the state is relatively absent. In rural Sierra Leone, the government lacks capacity to effectively communicate its laws. "If you ask anyone about laws from parliament, they won't tell you anything," a Paramount Chief told me, "the people are more accustomed to the [chiefdom] bylaws than the [government] laws."[5] Recent research finds that "a core task of the chieftaincy is meeting bottom-up demand for the provision of order by resolving disputes and addressing crime" (Wilfahrt and Letsa, 2023). Whereas the state lacks the coercive capacity to enforce its laws in rural areas, chiefs possess the coercive capacity to deter law-breaking. "If you violate [the law], there are some penalties that you have to pay for it," said a Paramount Chief I interviewed. "And people fear that."[6] In Chapter 3, I argue that traditional leaders popularity is tied to their role in maintaining the rule of law. That is, chiefs are popular because they enforce laws that increase community welfare. I make this argument using qualitative data from 300 semi-structured interviews conducted in 29 villages.

[5]Interview with PC Foryoh, July 23, 2020.
[6]Interview with PC Foryoh, July 23, 2020.

While Herbst (2014) is primarily concerned with the challenge to state leaders of projecting power in *rural* areas, in many countries resource-constrained leaders struggle to project power in growing urban centers. For example, while living in Freetown during COVID-19, it was clear that the government could not enforce social distancing measures in crowded places, such as markets and public transportation; could not effectively track suspected cases; and could not organize widespread vaccinations.[7] Aware of these capacity constraints, Mayor Yvonne Aki-Sawyerr made tax reform central to her pledge to transform Freetown when she was elected in 2018. With technical support from members of our research team, the Freetown City Council (FCC) overhauled the existing property register by introducing a simplified and objective assessment valuation scheme (Grieco et al., 2019) that registered over 60,000 new properties to receive property bills.[8] For this reform to generate revenue to enable state-building, the FCC needed to do more than send out thousands of new property tax bills—they needed to get people to pay.

While the local governments in rural Kono and urban Freetown shared the state-building goal of raising more tax revenue, the potential strategies available to each differed markedly. On the one hand, some options available in Kono were not on the table in Freetown: chieftaincy institutions do not exist in Freetown.[9] On the other hand, the stronger mobile phone networks, broader internet coverage, and higher phone ownership in Freetown presented the FCC with options to leverage technology for public administration (Brockmeyer and Sáenz Somarriba, 2022; Okunogbe and Tourek, 2024) in ways that would not be possible in Kono. We worked with the FCC to develop a phone-based participatory budgeting intervention, where participants shared policy preferences with neighbors and local politicians and then voted for local public services that were subsequently implemented. The aim of this intervention was to strengthening citizens' perception of government legitimacy (Levi, 1988, 1997), and thereby increase tax compliance. Chapter 4 report results from

[7] See Grieco et al. (2020) for a report on the government's initial response to COVID-19 in Sierra Leone.

[8] A article in *The Economist* in June 2020 referred to the existing property register as a "bundle of papers covered in ink scrawls" (The Economist, 2020).

[9] In Sierra Leone, chiefdoms are the administrative legacy of the colonial strategy for governing the provinces; these institutions were never set up in Freetown. On the "bifurcated" colonial state in Africa and its legacy, see Mamdani (1996).

this policy experiment.

In what follows, I briefly summarize the main arguments and findings from each chapter.

Chapter 2: *Building Fiscal Capacity with Traditional Political Institutions*

To test whether the KDC's collaboration with tradition leaders increased tax compliance, I embedded an experiment in the KDC's property tax collection campaign. Potential taxpayers were shown awareness videos that varied in their content, particularly in terms of whether and how their local paramount chief characterized his involvement in tax collection. Using survey data from 1,752 property owners across 118 villages, I estimate the effects of notifying property owners about the state's collaboration with their TPI. Alerting property owners to the collaboration significantly increases tax compliance and increases property owners' belief that they should pay taxes.

I then probe why citizens are more likely to comply when the state collaborates with TPIs. Existing accounts of chiefs' authority can be divided into two camps. First, TPIs are "decentralized despots" a la Mamdani (1996), whose authority is fundamentally coercive. Collaboration boosts compliance by invoking the chiefs' capacities to detect and punish tax evasion. Second, TPIs are legitimate political institutions that faithfully advance their constituents' interests. Collaboration increases compliance because taxpayers do not believe the chief would align themselves with an effort that did not promote collective well-being.

To assess whether TPIs' authority stems from coercion or legitimacy, I created additional versions of the tax awareness video that manipulated the paramount chiefs' statements to emphasize their coercive powers or their accountability to constituents. I find experimental evidence that TPIs' authority stems from both their legitimacy and coercive capacity, though the evidence for the coercion mechanism is stronger.

Chapter 3: *Enforcing the Public Will: Participation and Coercion in TPIs*

Motivated by my finding that TPIs can effectively secure citizens' compliance with state policy (Chapter 2), I returned to my research site to understand why some chiefs are more effective collaborators than others. In doing so, I inductively developed the argument I present in this chapter: traditional leaders' legitimacy stems from their role in enforcing laws that improve community welfare.

In the first two empirical sections I present evidence that laws (1) reflect citizens' interests and (2) are fairly and effectively enforced. In the first section, I focus on law-making and enforcement, generally. I first provide a rich description of citizens' participation in law-making. I argue that local laws are likely to reflect citizens' preferences because citizens are heavily involved in local policy-making. As evidence that these laws are effectively and fairly enforced, I show that citizens are generally satisfied with the way chiefs enforce laws. In the second section I dig deeper into a widespread law that allows traditional leaders to demand mandatory labor from community members for community-oriented projects. I present two pieces of evidence that this communal labor law is in citizens' interests. First, I show that mandatory communal labor is directed at projects with broad social benefits (e.g., road maintenance). Second, I show that respondents are much more likely to report that projects undertaken with communal labor benefit of the whole community, rather than benefiting only a few. As evidence that this law is effectively enforced, I provide a granular description of TPIs' capacity to monitor participation in communal labor and to punish non-compliers.

The final empirical section provides some inferential leverage for my argument by showing that citizens' approval of traditional leaders is higher where TPIs fairly and effectively enforce laws. First, using chiefdom-level case studies, I document that the fairness of law enforcement varies across chiefdoms. Second, I have research assistants code 261 interviews and show that traditional

leaders are more popular in chiefdoms where they enforce laws fairly.

Chapter 4: *Participation, Legitimacy and Tax Compliance*

Working in collaboration with the FCC, we rolled out the participatory budgeting intervention as tax bills were being delivered in early 2021. To identify causal effects, we use a matched-pair design to randomize half of 3,618 property owners into treatment. We observed individual-level tax compliance through administrative records and surveyed the treatment and control groups at three stages: before the process, after services were selected but before they were delivered, and after services were delivered.

We find that participating in the DTHs increases perceptions of government legitimacy: treated citizens report they have greater influence over policy; that government is delivering better services and is more responsive; that governing institutions are more competent; and that elected representatives are performing better. However, despite the intervention's positive and durable effect on perceptions of government legitimacy, we find the intervention has no average effect on property tax compliance behavior.

This null average effect masks substantial treatment effect heterogeneity. The property tax reform in Freetown was a highly politicized affair, where the mayor publicly battled with central government over the City Council's authority to adjust property tax rates. We find that the intervention's effect on compliance is moderated by political affiliation and tax policy preferences. In our context, these countervailing treatment effects wash out any average effect.

The essays in this dissertation test strategies governments in weak states are using to build fiscal capacity. Canonical accounts about the development of fiscal capacity mostly tell us why leaders in modern Africa may have had less incentive to invest in fiscal capacity, relative to other historical leaders. The stability of the state system has reduced the pressure to raise revenue by minimizing the threat of external war. The advent of international credit markets and development aid have

reduced leaders' incentives to bargain with domestic elites in exchange for taxes. These arguments tell us little about what leaders in weak states *should* do if they want to build fiscal capacity. The two interventions I study address this question.

One way to strengthen state capacity is to improve the effectiveness of the state institutions that carry out government policy. An exciting body of work in political science and economics studies recruitment, monitoring, and incentive strategies to improve the performance of state agents (for reviews, see Finan et al., 2017; Brierley et al., 2023). My research suggests two alternative strategies for improving state capacity that look beyond improving the effectiveness of state agents. First, governments can collaborate with existing, non-state institutions that already have effective mechanisms of social control. This may lead to revenue gains that allow for investments in state capacity. Second, governments can invite greater citizen participation in public affairs, which may increase state capacity by building government legitimacy. Indeed, altering beliefs is often an essential part of institutional change (Callen et al., 2023). With greater legitimacy, governments may find it politically easier to invest in fiscal capacity or take enforcement actions that generate revenue.

This research also contributes to our understanding of authority in traditional political institutions. Recent research has found little evidence that traditional leaders' influence constituents' behavior through coercive mechanisms. Brierley and Ofosu (2023) find no evidence that chiefs in Ghana influence vote choice through coercive mechanisms. Similarly, in their study of "domain congruence" in Kenya, Malawi, and Zambia, Baldwin et al. (2023) find no evidence that coercion driver traditional leaders' influence over presidential vote choice or voluntary contributions to burial or education funds. In contrast to these studies, I find experimental evidence that citizens will pay state taxes because they fear punishment for noncompliance and qualitative evidence that chiefs have a strong enforcement infrastructure. One potential explanation for this difference in our findings is the nature of the outcomes we study: while I study laws where noncompliance *should* be punished by political authorities, Brierley and Ofosu (2023) and Baldwin et al. (2023) study out-

comes (vote choice and voluntary contributions) where noncompliance is not *illegal* and therefore should not be punished by political authorities. Finally, while I find evidence of that TPIs exert social control through coercion, my findings that TPIs usually use coercion to fairly enforce laws is in contrast with accounts that paint TPIs as "decentralized despots" (Mamdani, 1996; Richards, 1996).

Chapter 2

Building Fiscal Capacity with Traditional Political Institutions:

Experimental Evidence From Sierra Leone

While it is widely acknowledged that fiscal capacity is crucial for economic development (Besley and Persson, 2011), many poor countries are trapped in a low-capacity equilibrium: governments lack the capacity to collect taxes and, therefore, lack the tax revenues to invest in more capacity. Existing literature in economics and political science offers one pathway out of this trap: improving the effectiveness of state institutions. However, little attention has been given to the traditional political institutions (TPIs) that coexist alongside governments throughout the developing world (Holzinger et al., 2016), which govern important parts of day-to-day life (Baldwin, 2016). This paper explores whether weak states can increase their fiscal capacity by collaborating with TPIs.

One way that state collaboration with TPIs might increase fiscal capacity is by increasing citizens' compliance. On the one hand, if TPIs possess coercive capacity or are seen by local populations as legitimate (Logan, 2013), collaboration may boost compliance with the government's tax demands.

On the other hand, if traditional leaders possess little independent authority, or they wield their authority despotically (Mamdani, 1996), collaboration may be ineffective or even backfire.[1]

This article examines the impact of collaboration between state officials and TPIs on citizens' compliance with a newly introduced property tax in rural Sierra Leone. Collaborating with the local government in Kono District, I embedded an experiment in their campaign to collect property taxes. Potential taxpayers were shown awareness videos that varied in content, particularly in terms of whether and how their local paramount chief characterized his involvement in tax collection.[2] Using survey data from 1,752 property owners across 118 villages and 5 chiefdoms, I estimated the effects of notifying property owners about the state's collaboration with their TPI. I find that alerting property owners about the collaboration significantly increases a preregistered index of tax compliance measures, which comprised two survey-based indicators and the outcome of a behavioral game. I also find that collaboration increases property owners' belief that they *should* pay taxes.[3]

To assess whether TPIs' authority stems from legitimacy or coercion, I created additional versions of the tax awareness video that manipulated the paramount chiefs' statements to emphasize their legitimacy (e.g., accountability to their constituents) or their coercive powers. I find experimental evidence that TPIs' authority stems from both their legitimacy and coercive capacity, though the evidence from primary experimental tests is stronger for the coercion mechanism. Priming property owners to TPIs' legitimacy increases my measure of tax compliance above and beyond the effect of collaboration; however, while the size of this effect is only 18% smaller than the size of the collaboration effect, this increase is not statistically significant (t-statistic = 1.33). In line with arguments that emphasize coercive capacity as a source of TPIs' authority, I find that priming re-

[1] I use the term "traditional leaders" to refer generically to the political leaders of TPIs. In Sierra Leone, as in many other countries, these leaders are called "chiefs."

[2] The use of gendered language in this paper is intentional, to convey that chiefs in Kono, and Sierra Leone generally, are overwhelmingly men. While women chiefs are not impossible, they are rare. In Kono, where the study takes place, the 14 highest level chiefs (paramount chiefs) are all men. There are hundreds of village level chiefs, so while it is likely that some of them are women, this is not common.

[3] In the tax compliance literature, this belief is often referred to as tax morale.

spondents to think about punishment for noncompliance increases the measure of tax compliance, above and beyond the effect of collaboration. The size of this effect is roughly the size of the collaboration effect and is statistically significant.

This research contributes to the literature on fiscal capacity. Influential accounts of the historical development of fiscal capacity can be split into two bins. First, war making, and the external threat it poses, creates incentives for leaders to invest in extractive capacity, winnows out weak states, and generates a common interest (i.e., military victory or defense) that society is willing to contribute to (Tilly, 1990; Doner et al., 2005; Besley and Persson, 2008). However, as Herbst (1990) notes, these forces have been largely absent in modern Africa due to the stability of its state system.[4] Second, in elite bargaining theories, leaders trade political rights for tax revenues (e.g., North and Weingast, 1989; Bates and Lien, 1985). Yet, in contemporary democracies, where citizens already possess (*de jure*) property and political rights, it is unclear if political leaders have sufficient leverage to initiate these bargains. What strategies can state leaders in contemporary democracies employ to build capacity? To date, the literature has attempted to explain what makes bureaucracies, and the state agents they comprise, more effective. One set of scholars has focused on improving the performance of state agents by solving principal-agent problems through recruitment, monitoring, and incentive strategies (for reviews, see Finan et al., 2017; Brierley et al., 2023). Others have investigated how bureaucrats' relationship with society (i.e., their "embeddedness") impacts their effectiveness (e.g., Evans, 1989; Bhavnani and Lee, 2018).[5]

I contribute to the literature on fiscal capacity by showing that governments can better achieve their goals—in this case tax collection—by collaborating with non-state, traditional political institutions. Most similar to this project is Balán et al. (2022), who employ neighborhood chiefs (*chefs d'avenue*) to collect taxes in urban DRC. They find that these local elites collect more revenues, because they know the people in their neighborhood and can better target households with a higher

[4]Relative to the tumultuous state system of medieval Europe, which is the context for most studies in the bellicist tradition.

[5]See Pepinsky et al. (2017) for a review focused on the social embeddedness of "street-level" state agents. Besley et al. (2022) reviews these literatures together.

propensity to pay. This aligns with my findings that non-state actors can be effectively engaged in tax collection. Our work differs in that the neighborhood chiefs under study in Balán et al. (2022) are not leaders within non-state *political* institutions: for example, they cannot make laws, impose fines, or collect taxes independent of the state.[6] Gottlieb et al. (2021) also study the role of non-state actors in tax collection, but again, the local intermediaries they study are not leaders of non-state political institutions.[7] To my knowledge, this paper is the first experimental investigation of state collaboration with TPIs during tax collection.

More generally, this paper adds to research exploring the opportunities for collaboration between state actors (or funder agencies) and non-state actors (Buur and Kyed, 2007). Scholars have explored these types of collaborative relationships for targeting beneficiaries for social assistance programs (Basurto et al., 2020; Alatas et al., 2019), distributing development aid (Carlson and Seim, 2020), and implementing and coordinating development projects (Casey et al., 2018; Voors et al., 2018).

This research also contributes to the literature on traditional political institutions. Recent research has demonstrated the central role played by contemporary TPIs in many countries in sub-Saharan Africa, where they impact local development outcomes (Acemoglu et al., 2014; Baldwin, 2019), mobilize votes during elections (De Kadt and Larreguy, 2018; Nathan, 2019; Brierley and Ofosu, 2023), and influence local governance (Baldwin, 2016). Indeed, Baldwin and Holzinger (2019) estimate that 83% of the population of sub-Saharan Africa is governed, at least in part, by TPIs. Despite their importance, we have little understanding of the source of TPIs' authority—why do people consent to demands and rulings issued by their traditional leaders?[8] As I elaborate below,

[6] On the inability of *chefs d'avenue* to enforce tax compliance, the authors note that it is "unlikely that chiefs would have more credibly threatened official sanction [i.e., fines and legal consequences]" than state agents (Balán et al., 2022, section 7.3).

[7] These authors investigate the impact of business formalization appeals made by representatives from local market associations (whom they refer to as "local intermediaries") vis-à-vis state agents.

[8] A notable exception is Brierley and Ofosu (2023), who investigate the mechanisms behind paramount chiefs' influence over vote choice in Ghana. They argue this influence works through a signaling mechanism where voters exposed to chiefs' candidate endorsements update positively about candidate characteristics and expected performance. This aligns with my finding that legitimacy is a source of TPIs' authority. In contrast with my findings, they find no evidence that chiefs' endorsements influence vote choice through coercive channels. This difference is plausibly

the literature offers two diametrically opposed models of TPIs, painting them as either unaccountable despots that govern through coercion or legitimate institutions whose authority is rooted in constituents' consent. To my knowledge, this paper is the first to test and horse race these arguments experimentally.[9]

2.1 The Sources of TPIs' Authority and the Impact of Collaboration

In this project, I explore whether the state can increase citizens' compliance with tax collection efforts by collaborating with TPIs. Important work in political science and adjacent fields discusses the authority and capacity of TPIs. This research generates divergent predictions about the impact of collaboration on tax compliance.

There are at least two good reasons to believe that collaboration will increase citizens' tax compliance. First, individuals might voluntarily comply with TPIs' directives because these institutions are perceived as legitimate (Logan, 2013). Scholars have long argued that legitimacy—the belief that political actors or institutions have the "right to rule"—influences compliance with rules, regulations, or directives put forward by political authorities (Rousseau, 1762; Beetham, 2013). Support for this argument is found in empirical work that links legitimacy (or related concepts such as political trust) to citizens' compliance with the police and courts (Tyler, 2006), military service demands (Levi, 1997), taxes (Levi, 1988), and health regulations (Bargain and Aminjonov, 2020).[10]

One reason why citizens may perceive TPIs as legitimate is because they allow citizens input in policy-making (sometimes called "process legitimacy"). TPIs often contain elements of direct democracy (Skalník, 1996) and have inclusive and transparent decision-making processes (Baldwin and Holzinger, 2019). Other scholars argue that TPIs' legitimacy stems from their perfor-

explained by differences in the nature of the outcomes we study: chiefs (whether in Ghana or in Sierra Leone) do not observe vote choice, whereas in Sierra Leone chiefs have mechanisms in place to monitor compliance with tax demands.

[9] Note that the test of mechanisms in Brierley and Ofosu (2023) is not experimental.

[10] On political trust see Levi and Stoker (2000).

mance.[11] For example, TPIs maintain public order and political stability (Krämer, 2016), preside over systems of customary law that many people perceive as fair (Sawyer, 2008), help their communities obtain important services from government (Williams, 2010), and organize local input (e.g., labor and materials) to produce local public goods (e.g., road maintenance) (Baldwin, 2019). Consistent with the idea that TPIs are perceived as legitimate, survey results from 31 African countries (Afrobarometer, round 8) show that traditional leaders are more trusted, seen as less corrupt, have higher approval ratings, and are perceived as better listeners than elected officials. Further, 51% of the respondents said they would like to see these institution have more influence in local governance, while only 13% of respondents said they would like to see them have less influence (Logan and Amakoh, 2022).[12]

Second, collaboration with TPIs may increase tax compliance because they possess the coercive capacity to deter evasion (Allingham and Sandmo, 1972; Kleven et al., 2011). Indeed, coercive capacity is often included within the definition of TPIs, where the assumption is that they administer and enforce a system of customary law (Holzinger et al., 2016), typically through a network of sub-chiefs (Manning, 2009).[13] While TPIs' use of their coercive capacity may often be benign, leaders' abuse of their coercive capacity is a dominant theme in the literature. In Mamdani's (1996) influential account, colonial governments undermined existing accountability mechanisms within indigenous political institutions and enabled traditional leaders to become "decentralized despots." For many scholars, TPIs' authority stems from their willingness to abuse their central position in systems of local governance (Acemoglu et al., 2014; Fanthorpe, 2004). Most notably, scholars have documented traditional leaders' interference and bias in both local dispute resolution mechanisms (Mokuwa et al., 2011; Maru, 2006) and decisions regarding land allocation (Ntsebeza, 2005; Goldstein and Udry, 2008; Koter, 2013; Acemoglu et al., 2014). Following these accounts, state

[11] See Keele (2007) on the link between governing performance and citizens' trust in government, which is often linked to citizens' willingness to comply with government directives. As Hetherington and Rudolph (2008) point out, *perceptions* of government performance also matter.

[12] The trust gap is widening. While, trust in elected officials has declined since 2008/2009 survey rounds, trust in traditional leaders has held steady.

[13] Holzinger et al. (2016) note that their concept of traditional governance includes, following Fukuyama (2013), the "ability to make and enforce rules".

collaboration with TPIs may increase tax compliance because individuals fear that noncompliance will be punished with bias in future decisions regarding law or land. This discussion motivates the following hypotheses:

H1: Collaboration between state leaders and TPIs increases citizens' compliance with state tax demands.

H2: TPIs' legitimacy enables them to obtain citizens' compliance with state tax demands.

H3: TPIs use (the threat of) coercion to obtain citizens' compliance with state tax demands.

We have little evidence to evaluate whether state collaboration with TPIs increase citizens' compliance with state policy. If governments could increase tax compliance by collaborating with TPIs, we might expect to see more instances of such collaboration across the continent. Indeed, there are reasons to doubt that government collaboration with TPIs increases tax compliance. By collaborating with TPIs, the state is attempting to harness TPIs' authority over local populations. However, it is not clear that TPIs' authority can be leveraged to secure compliance with state policy. First, if citizens' compliance with TPIs' directives stems from direct involvement in policy making, citizens may be less willing to comply with TPIs' directives to adhere to a government policy that citizens had no part in crafting. Second, TPIs may possess limited coercive capacity independent of the state. In this paper, I investigate collaboration in a setting where the state's own agents (i.e., local government officials) are already making tax demands. If TPIs' coercive capacity ultimately rests with the state, collaboration may not affect citizens' perceptions regarding sanctions for noncompliance.

Moreover, collaboration may send signals to citizens that cause them to update their perceptions of the state's capacity. By informing citizens that they are collaborating with TPIs, the state may be signaling their low capacity to citizens, undercutting citizens' compliance. Finally, in settings where trust in state leaders is relatively low, collaboration may also send signals about the quality of state actors and their willingness to work in citizens' interest. On the one hand, if TPIs are despotic

and abusive towards local populations, collaboration may cause citizens to update negatively about the quality of state actors, undermining compliance. On the other hand, if TPIs are legitimate and well respected, collaboration may send a positive signal about the quality of state actors, boosting compliance. In summary, while the existing literature establishes TPIs' local authority, it is unclear whether the state can leverage this authority to secure compliance with state policy.

2.2 Context: TPIs and Property Taxation in Sierra Leone

Most of Sierra Leone (but for the peninsula that includes the capital city), is divided into 190 chiefdoms. Each chiefdom—while also under the jurisdiction of the state (both central and local governments) is governed by a chiefdom council and a group of hierarchically organized chiefs. The top traditional political authority in each chiefdom is the paramount chief, who is elected for life by the chiefdom's elite, with candidates drawn from a restricted set of ruling families (see Reed and Robinson, 2013). Chiefdoms are further divided into sections that contain a number of villages. Sections and villages are headed by section chiefs and village chiefs, respectively.

TPIs in Sierra Leone are relevant for local governance, an important scope condition for my argument. In a recent Afrobarometer survey (Round 8, conducted 2020) most rural respondents in Sierra Leone reported that TPIs have "some" or "a lot" of influence over local governance (92%), dispute resolution (95%) and land allocation (78%). Sierra Leone is a good context to test my hypotheses because existing evidence is mixed about the source of TPIs' authority, which also generates an ambiguous prediction about the effect of collaboration. On the one hand, influential accounts about the causes of the 11-year internal war point to the coercive and authoritarian chieftaincy system and its control over local resources (e.g., land, labor, women) (Richards, 1996, 2005). According to commentators, this narrative was a primary motivator for the World Bank/DFID/EU funded decentralization reform, where District Councils were reintroduced to counterbalance the authority of chiefs (Jackson, 2007; Fanthorpe et al., 2011). On the other hand, scholars have pointed to the legitimacy of TPIs and the public's trust in these institutions to explain citizens'

compliance with disease control measures during the Ebola outbreak (Wilkinson and Fairhead, 2017; Richards, 2016). Survey data is consistent with this ambiguity. Rural Sierra Leoneans' attitudes toward TPIs (e.g., trust, job approval, and perceived corruption) lie near the middle of the distribution for the 31 countries surveyed in 2019-2021 Afrobarmometer round. This also gives us confidence that the findings from this study can be generalized to other countries where TPIs are politically relevant. Appendix Table A.1 presents summary statistics for the perceived influence of and attitudes towards TPIs in Sierra Leone, compared to other surveyed countries.

In 2018 the local government in Kono (called the District Council) began a tax reform with the goal of systematically collecting taxes in rural areas for the first time since their post-war reintroduction in 2004. At the center of this reform was a property tax, which levied a rate on all residential and commercial building structures. Building structures were divided into tax rate bands based on their size and material and all structures within a given band were taxed at the same rate.[14] Tax collectors were assigned mutually exclusive areas of the district and were compensated with a share of the revenue they collected.

Revenues were very low ($\approx 2\%$ of potential revenue in 2019), in large part due to low compliance. Based on data from 2019, tax compliance rates were less than 10% in the villages visited by tax collectors.[15] While the District's TPIs were involved in the tax effort—for example, recruiting tax collectors and supervising local bank accounts—their participation was not widely communicated to tax payers.[16] Chiefs and local government officials were cognizant of this lack of communication: plans had been made for chiefdom meetings where chiefs could discuss the tax reform, but these plans fell through because neither the chiefs nor the district council could marshal the funding to hold them.[17]

[14] In 2019, 95% of the building structures in the district had an assessed rate of $1.50 or less.

[15] Note that in 2020, there was no attempt to collect the property tax, due to COVID-19.

[16] A memorandum of understanding (MoU) between the Kono District Council and the District's 14 Paramount chiefs was signed in early 2018. The MoU outlined the nature of the revenue sharing agreement. Half of the collected revenue was earmarked for development projects in the chiefdom where it was collected; five percent of the collected revenue was to be allocated to the chief, to support supervision responsibilities.

[17] This failure caused one paramount chief to complain to government officials that they had not followed through on their promise to support chiefs to communicate with their people about the tax. I was in attendance at the July 2020

Tax collectors offered in interviews that the lack of awareness about this collaboration was a cause of low compliance. One tax collector speculated that his collection efforts were successful in some villages "because the authorities passed the message [of TPI involvement] to the people, and the people have respect for the authorities" and less successful in other villages because "maybe the message [of collaboration] didn't reach them soon enough."[18] Another collector suggested that compliance would increase the following year if "the paramount chief calls a meeting. When the chiefs are more strongly backing [the reform], that's going to make people pay."[19]

In an effort to remedy this lack of awareness before the 2021 tax collection season, and therefore boost compliance, I worked with the Kono District Council (KDC) and the district's paramount chiefs to design a tax awareness campaign. The tax awareness video at the center of this campaign is the intervention under investigation in this study.

2.3 Interventions

Working with the local government and the district's paramount chiefs, I designed and recorded a tax awareness video intended to provide property owners with information about the property tax (e.g. valuation and rates).[20] We recorded four video segments:

1. **District council chairman provides information about tax collection:**
   - First, he introduces himself: *"Greetings my people! Good morning, good afternoon and good evening. This is your son Solomon Sahr Gbondo who is heading the Kono District Council*."
   - Second, he provides information about the tax rates: *". . . Stick house, with local roof. That is, palm trees leaves. You pay 20,000 Leones . . ."*
   - He concludes with an appeal to pay: *"Please, let us pay our taxes in order for us to able to carry out development projects in the district like roads rehabilitation, digging of boreholes,*

stakeholder meeting where this complaint was made.

[18]Interview with tax collector from Lei Chiefdom.

[19]Interview with tax collector from Tankoro Chiefdom.

[20]In Appendix A.6, I describe the process through which the tax awareness video was developed. A translation of the full text for all Kono videos can be found in Appendix A.9.

building of schools, and other things."

2. **Paramount chief mentions collaboration with local government:**
   - First, the paramount chief introduces himself: *"My Gbane people, I greet you all. This is your paramount chief Aiah Bindi Faefankongor the* 2^{nd}*."*[21]
   - Second, he explains that the Chiefdom Council is working with the Kono District Council to collect property tax on all the houses in the chiefdom: *"Gbane Chiefdom Council and Kono District Council are working in unity to collect property taxes, which is a tax for houses, which we should pay."*

3. **Paramount chief primes legitimacy:**
   - First, the paramount chief says that he will convene a meeting of subchiefs after taxes are collected to discuss how tax revenue will be spent: *"After we have finished collecting the tax payment, I will summon a meeting. In this meeting, I shall request the presence of other subordinate chiefs in the chiefdom for us to discuss and map out ways of how the collected money is going to be utilized."*
   - Second, the paramount chief acknowledges that the people of the chiefdom will not be happy if the tax revenue is not used for development: *"I am of the belief that if we do not utilize the funds collected in the best way for the development of the chiefdom, you the chiefdom people, will be annoyed."*

4. **Paramount chief primes coercion:**
   - First, the paramount chiefs says that he will convene a meeting to discuss how noncompliance will be punished: *"After the collection of these taxes, I will hold a meeting with the chiefs to brainstorm what to do with those that have refused to pay taxes for their houses."*
   - Second, the paramount chief says that he and other chiefdom authorities will not be happy with those who do not pay taxes: *"Let me emphasize that I and the rest of the chiefs will not be merciful on anyone who has refused to pay the tax."*

[21] Each paramount chief used slightly different words to deliver each message. Here I provide an example from Gbane Chiefdom.

These segments were combined into different awareness videos. Working in conjunction with the KDC and the district's paramount chiefs, I led a door-to-door tax awareness campaign in the summer of 2021, in which a team hired through a local civil society organization met with property owners to share different versions of the tax awareness video. I embedded an experiment in this campaign. Treatment conditions are different tax awareness videos that combine different video segments and are designed to test different hypotheses (see Table 2.1). Property owners assigned to the control condition see only the first video segment where the local government official provides information about tax collection. For property owners assigned to the first collaboration treatment condition (T1), the government information segment is followed by the second segment where the chief mentions their collaboration with local government. I expect T1 to increase property owners' perception that TPIs are collaborating with the local government on the property tax, relative to control. I test my first hypothesis by comparing tax compliance outcomes between T1 and C.

In a second legitimacy treatment condition (T2) the paramount chief's statement is expanded to include the third video segment where the chief primes legitimacy. The goal of T2 is to prime aspects of TPI's legitimacy that are highlighted in the literature—namely that important decisions will have the input of additional actors beyond the paramount chief and his close inner circle, that revenue spending decisions must be justified in public, and that poor governance will anger constituents. I can address my second hypothesis—TPIs' legitimacy enables them to obtain citizens' compliance with state tax demands—by comparing tax compliance outcomes between T1 and T2.[22] In a coercion third treatment condition (T3), the additional messaging from the paramount chief focuses on punishment for non-compliers, rather than legitimacy. The goal is to prime punitive actions that can be taken against non-compliers. I assess my third hypothesis by comparing tax compliance outcomes between T1 and T3. Table 2.1 summarizes the video components that make up each treatment condition and the comparisons that I will make to test each hypothesis.

[22]This is a plausible policy promise. Under the MoU governing the property tax, 50% of the collected revenue is earmarked for development in the chiefdom where it is collected and is to be allocated to the chiefdom council for that purpose. As the head of the chiefdom council, the Paramount Chief is on solid ground promising more inclusion from citizens.

Table 2.1: Summary of Treatment Conditions

Treatment Condition	Video Segment	Comparison	Hypothesis Tested
C: Tax information (n=428)	1		
T1: TPI collaboration (n=454)	1 + 2	T1 - C	H1
T2: Legitimacy (n=437)	1 + 2 + 3	T2 - T1	H2
T3: Coercion (n=433)	1 + 2 + 4	T3 - T1	H3

I randomly assigned property owners to treatment conditions with equal probability using simple randomization.[23] Table 2.2 presents balance. Columns 6-8 present differences between each treatment group's mean and the control group mean, standardized by the control group standard deviation. For metrics to gauge the magnitude of these differences, I provide two test statistics from a model that regresses a given covariate on the three treatment indicators. First, where a treatment group mean for a given covariate is statistically different than the control mean ($\alpha < 0.1$), I star the corresponding standardized difference in columns 6-8. Second, Column 9 presents the *F*-statistic for the joint null hypotheses—a significant result here implies that the treatment indicators collectively have predicative power (i.e., treatment group means are different than the control group mean).

Observed differences between treatment groups for these immutable covariates are no more than we might expect. Given the 33 tests I run in Columns 6-8, under the null hypothesis of no differences between groups, we would expect 3.3 tests to appear significant at the 90% confidence level; I find only one significant difference on the *education* covariate. In Column 9, I run 11 tests and therefore expect 1.1 to appear significant at the 90% confidence level; I find one significant difference on the *gender* variable. The observed imbalance on gender and education would only

[23]Every property owner was assigned to a version with a probability of 0.25. As respondents watched the tax awareness video on the tablets or phones that enumerators used to conduct the survey, I programed the treatment randomization into the tablet-based survey.

be a concern if these covariates predicts our primary outcome of interest, the compliance index (described in Section 2.5). The bivariate the relationship between gender and the tax compliance index is small and statistically insignificant.[24] Education, however, is positively associated with the primary compliance outcome. As the control group is more educated than any of the treatment groups, this imbalance could introduce downward bias into the estimates, though the magnitude of that bias is likely to be small.[25] I account for this when estimating treatment effects, as I preregistered the education variable as a control variable in my main specification.

Table 2.2: Balance Table

	Mean				**SD**	**Std. difference**			***F*-stat**
	C	T1	T2	T3	C	T1-C	T2-C	T3-C	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Demographics									
Age	46.47	46.96	46.41	47.11	14.73	0.03	0.00	0.04	0.24
Educated (received any schooling)	0.39	0.36	0.32	0.34	0.49	-0.06	-0.14*	-0.10	1.58
Kono speaking	0.81	0.81	0.81	0.82	0.39	-0.01	-0.02	0.02	0.12
Gender (female = 1)	0.28	0.24	0.31	0.30	0.45	-0.08	0.07	0.05	2.17*
Married	0.78	0.80	0.80	0.78	0.41	0.05	0.04	-0.01	0.39
Community social / political position	0.27	0.25	0.26	0.24	0.44	-0.04	-0.02	-0.07	0.32
Value to animal stock (100's USD)	2.25	2.36	2.40	2.41	4.62	0.02	0.03	0.04	0.11
Owns multiple properties	0.23	0.18	0.18	0.19	0.42	-0.11	-0.10	-0.08	1.04
Employment									
Has farm & no outside employment	0.57	0.56	0.56	0.56	0.50	-0.02	-0.02	-0.02	0.04
Farm & non-farming employment	0.33	0.32	0.35	0.32	0.47	-0.01	0.05	-0.03	0.47
Non-farming employment only	0.10	0.12	0.09	0.12	0.30	0.06	-0.04	0.08	1.31

Table 2.2 reports balance across immutable covariates. Columns 1-4 report treatment group means; Column 5 reports the control group standard deviation; Columns 6-8 report differences standardized relative to the control group standard deviation; Column 9 reports the *F*-statistic for the joint null hypothesis.
Significance: * $p < 0.10$

Table 2.2 also allows us to characterize the sample. Columns 1-4 display group means for each covariate and column 5 presents the control group standard deviation. The average respondent is

[24]The *p*-value on the regression coefficient of this estimated bivariate relationship is 0.81 (using only control group data).

[25]According to a bivariate regression (using only control data), moving from no education to some education increases the tax compliance index by 0.18 standard deviations. Respondents in control are seven percentage points more likely to have received some form of education than respondents in T2. Therefore, if left unadjusted, we should expect bias in the order of 0.0125 standard deviations.

about 46 years old, uneducated (≈ 65%), kono speaking (≈ 81%), male (≈ 72%), and married (≈ 79%). Property owners in my sample do not appear to be wealthy. In rural Kono District wealth is largely held in animal stocks, and the average respondent has animal stocks with a market value of $225.[26] The sample also captures a mix of elite and non-elite respondents, with roughly a quarter of the sample holding a community position of social or political importance (e.g., chief, mammy queen, religious leader, youth leader). Finally, respondents are primarily engaged in small-scale agriculture. Roughly 56% of the sample works exclusively on their own farm, while an additional third of the sample mixes work on their personal farm with outside employment.[27]

2.4 Data Collection

Five out of 14 chiefdoms were included in this study. Chiefdoms were excluded if (i) the paramount chief was unavailable to record a video; (ii) the recorded video deviated too much from the agreed script; or (iii) data collection costs were prohibitively high. I used geographic cluster sampling to select 123 villages for the study from a set of 434 eligible villages in the five eligible chiefdoms. As road infrastructure in Kono district is poor, making traveling between villages time and resource intensive, geographic cluster sampling helped minimize transportation costs. Figure 2.1 visualizes this sampling: blue triangles represent villages sampled for surveying.[28] More details on survey sampling can be found in Appendix A.4.

Working with a team of 33 enumerators between May and June 2021, we completed 1,752 surveys across 118 villages, selecting households to interview through a random walk procedure.[29] All

[26] Animal stock value calculated based on market value in the district headquarter town at the time of data collection. The mean household owns 0.86 goats, 0.30 sheep, 3.22 chickens, 0.17 ducks, 0.10 pigs, and 0.05 head of cattle

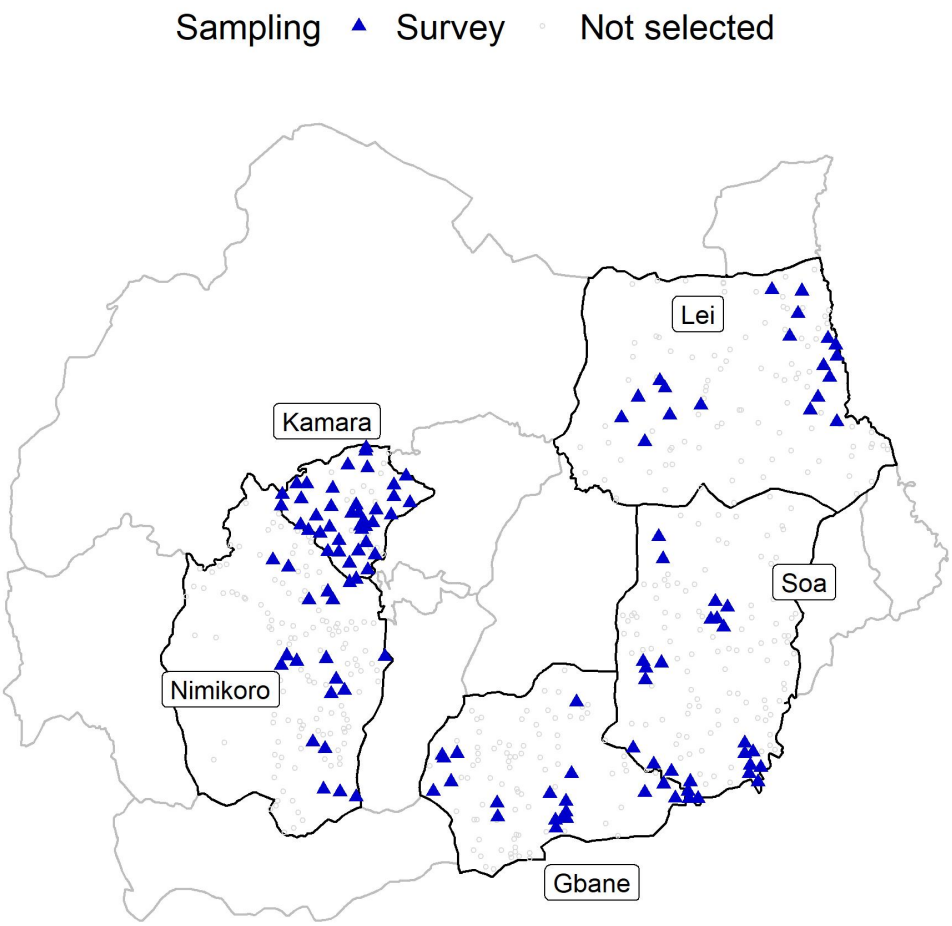
[27] The most commonly named forms of non-farm employment are: trader, miner, and wage laborer.

[28] In the five study chiefdoms, I excluded villages based on three criteria. The research design relies on the ability to manipulate respondents' beliefs about the involvement of TPIs in the new property tax collection. I reasoned that this belief would be more malleable in villages where property tax had not been previously collected. Therefore, I excluded all villages visited by tax collectors in 2019. Second, to increase enumeration efficiency, I excluded villages that are listed in the 2015 census as having fewer than three building structures. Third, I excluded chiefdom headquarters towns from the sample. This left me with a sample frame of 434 eligible villages.

[29] Enumerators explained their role to respondents in the following way: "I work for an organization that is between the people and the government, which is called KoCEPO. This organization is doing some research to find out ways to improve conditions in Sierra Leone." Note that the enumeration team was unable to locate five villages. For a sampling

survey respondents received three Maggi spice cubes upon completing the survey as a token of thanks. In addition, respondents kept their proceeds from a modified dictator game.[30]

Figure 2.1: Sampling Map



For the majority of questions in this survey, respondents were asked to gauge their expectations or perceptions on a 10-point scale. To make this scale more concrete to survey respondents, all enumerators were given 10 beans and a plastic plate, which served as a visual aid regarding the 10-point scale. Respondents were asked to allocate some, none, or all of the 10 beans to the plastic plate to represent their perceptions and expectations. Enumerators were trained on how

frame I used a list of villages from Sierra Leone's 2015 national census. This list of villages may not perfectly describe the set of villages in Kono today—there may have been errors during the census, new villages may have been created since 2015, and others may have been abandoned.

[30]For the control group, the mean amount kept by the respondent was 1,670 SLL (aprox. US$0.15).

to discuss the concept of probability with respondents in familiar terms and how probabilistic expectations could be expressed using the beans. Measurement validity is discussed in greater detail in Appendix A.2.

2.5 Results

I estimated treatment effects using the centered covariate-treatment interaction specification proposed by Lin (2013):

$$Y_i = \alpha + \beta_1 T1_i + \beta_2 T2_i + \beta_3 T3_i + \theta \mathbf{X}_i + \gamma C_c + \delta N_k + \epsilon_i \tag{2.1}$$

where Y_i is the outcome of property owner *i* and T1, T2, and T3 are dummy variables for each treatment condition. Following Lin (2013), $\mathbf{X}_i$ is a set of preregistered control variables, centered and interacted with with each treatment condition. Prespecified control variables include: (i) educational level (a dummy indicating whether the respondent received any schooling); (ii) a set of dummy variables for community positions of social or political importance; (iii) expected travel likelihood to the district and country capital; (iv) expected occurrence of an unlikely event; (v) the village level literacy rate; (vi) the percentage of households in the village that own a radio; and (vii) percentage of village residents born in the chiefdom.[31] I also include chiefdom and enumerator fixed effects as C_c and N_k, respectively. ϵ_i is the idiosyncratic error term.[32]

[31] Community position dummies include: Chief or deputy chief (village or section), women's leader (village or section), youth leader (village or section), religious leader, and other, which comprises societal heads, tribal chiefs (i.e., leaders of non-Kono ethnic groups), and chiefdom councilors. I include the respondent's perceived likelihood of an unlikely event—the president visiting the respondent's village on the following day—because it tells us how the respondent is using the 10-point scale and is therefore prognostic. I selected prognostic variables for covariate adjustment using a LASSO model that predicted my outcomes of interest (Details in Section 8 of the PAP).

[32] As randomization occurs at the level of the observation (respondent), I do not cluster standard errors.

2.5 Does State Collaboration with TPIs Increase Compliance?

Before reporting the impact of the TPI collaboration treatment (T1) on compliance, I first present a set of manipulation and attention checks. T1 attempts to manipulate respondents' perceptions about the collaboration between state actors and TPIs. To measure perceptions of involvement, enumerators presented respondents with a laminated paper divided into four squares, where each square represented one of the four actors: (i) Kono District Council, (ii) TPIs, (iii) the central government, and (iv) NGOs and civil society organizations. Each respondent was then given 10 beans and asked to allocate the beans across the four squares, placing more beans on actors they thought were more involved in and responsible for the property tax. The blue columns in Figure 2.2 displays baseline (i.e., control) perceptions of the involvement of the four actors and the red column shows the treatment group (T1) level. T1 increases the perceived involvement of TPIs by 0.27 standard deviation (p-value < 0.001), equivalent to 0.56 beans, or 24% of the baseline mean.[33]

Table 2.3 presents the effect of the TPI Collaboration treatment (T1) on compliance outcomes, where the preregistered primary outcome is an additive index (compliance index) that comprises two survey questions that attempt to measure propensity to pay and a behavioral game that attempts to capture voluntary compliance.[34] These indicators were measured using a survey administered to respondents directly following the tax awareness video.[35] The first survey question (*self-reported propensity to pay*) directly asked the respondents to state the likelihood of them paying the full tax rate if a tax collector came to their house the next day. The second survey question (*perceived neighbors' propensity to pay*) asked respondents what proportion of other property owners in the

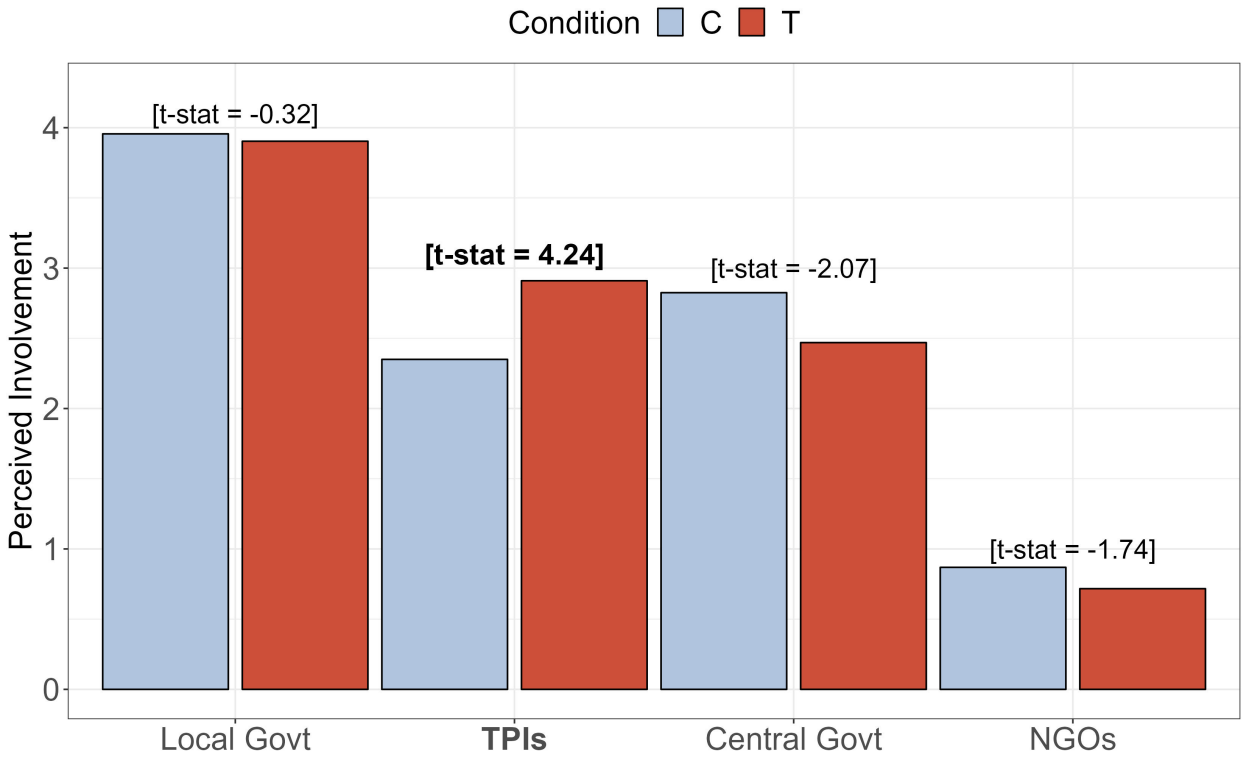
[33] At baseline, respondents allocated 68% more beans to local government than to TPIs. The increase in perceived involvement of chiefs due to T1 led to a corresponding perceived decrease in involvement from the central government and NGOs but not from local government. Note that the measurement strategy forces respondents to allocate a finite number of beans. Therefore, an increase for one actor must lead to a decrease for one or more of the other actors. Attention checks are presented and discussed in greater detail in Appendix A.5.

[34] To construct the summary index of the three compliance measures, I follow Kling et al. (2007) and standardize each sub-indicator relative to the control group and combine them in an equally weighted index that averages across standardized sub-indicators. I impute missing sub-indicators using the group mean.

[35] The three outcome measures were placed at the beginning of the survey, preceded only by two treatment comprehension questions.

village they thought would pay the new property tax, thereby attempting to measure propensity to pay indirectly.

Figure 2.2: T1 Increases Perceived Involvement of TPIs



After answering the above tax compliance survey questions, the respondents played a behavioral game (dictator game), where they were asked to distribute a small sum of money between themselves and the local government's property tax fund (*coins given to KDC's house fund*). Respondents were told that the money collected from property taxes went into a government bank account, which we referred to as the "house money fund." The enumerator then handed the respondents five 500 Leone coins (each valued about $0.05) and told the respondents that they should distribute these coins between themselves and the house fund. It was made clear that the respondents could keep any coins they allocated to themselves and that they could split the coins anyway they like between themselves and the house fund. Enumerators then recorded the number of coins the re-

spondents allocated to the house fund.[36] This indicator attempts to capture voluntary aspects of property tax compliance, as these voluntary contributions go to the same bank account as property tax revenue, and political authorities have no way of knowing how much a given respondent contributed.

Table 2.3: Effect of Collaboration (T1) on Compliance

	Mean	**Estimate**	**N**
Outcome	(Control)	(T1-C)	
Compliance Index	**0.000**	**0.068***	**1,752**
	(0.668)	(0.040)	
Self-reported propensity to pay tax	6.729	0.090	1,751
	(3.000)	(0.058)	
Perceived neighbors' propensity to pay tax	5.965	0.113	1,657
	(2.322)	(0.070)	
Coins given to KDC's house fund	1.664	0.001	1,752
	(1.438)	(0.056)	
Tax morale (secondary outcome)	7.357	0.155**	1,751
	(2.817)	(0.063)	

Table 2.3 reports the effect of the collaboration treatment (T1) on manipulation check outcomes (Panel A), the compliance index (in bold), and its subcomponents (Panel B). Column 1 reports the control group mean for each indicator, with the standard deviation in parentheses; Column 2 presents treatment effects estimates, with standard errors in parentheses. Reported effects are standardized effects. Models are estimated using OLS with preregistered specifications. Column 3 reports the number of non-missing observations. The *Tax morale* measure was not included in the compliance index (as per the PAP).
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Treatment effects are reported in standard deviation units. The TPI Collaboration treatment (T1) increases the compliance index by 0.068 standard deviations, relative to the control, a difference which is statistically significant. Looking at the effect of the T1 on the sub-indicators, we can see that this effect is driven by the measures of self-reported propensity to pay and perceptions that neighbors' will pay. To obtain a concrete understanding of the effect size, consider the point estimate for the effect on the indicator *self-reported propensity to pay*. The effect size is 0.09

[36] As stated, coins allocated to the house fund were handed over to the KDC to be deposited in their property tax revenue bank account.

standard deviations, which corresponds to 0.27 beans on the 10-bean scale. Given that each bean represents 10 percentage points, we can interpret this effect as a 3 percentage point increase of the self-reported likelihood to pay property tax.

I prespecified a lone secondary compliance outcome: a survey-based measure of the respondents' belief that they *ought* to pay property tax, often refereed to as "tax morale" in the tax compliance literature. Respondents were asked to imagine a situation in which they would not be fined or penalized if they did not pay their property tax and were then asked if they thought it was (morally) right to pay their tax. In the last row of Table 2.3, a large and statistically significant impact of the tax morale measure can be seen.[37] Considered together, these findings can be interpreted as strong evidence that individuals are more willing to comply with the newly introduced property tax when they know that their local government is collaborating with leaders of TPIs.

2.5 Sources of TPIs' Authority: Primary Mechanism Results

Why does government collaboration with TPIs increase citizens' compliance with property tax? This section examines two arguments for why TPIs are able to generate citizen compliance: *legitimacy* and *coercion*.

Table 2.4 reports the effect of the Legitimacy treatment (T2) (Columns 2-3) and Coercion treatment (T3) (Columns 4-5) on the compliance index. Column 2 shows that the Legitimacy treatment (T2) increases the compliance index over and above the effect of the TPI collaboration treatment (T1), but this difference is not statistically significant (p-value = 0.18). While statistically insignificant, it is worth noting that (i) all three sub-indicators move in the expected direction and (ii) the estimated effect of 0.056 standard deviations is only 18% smaller than the effect of the collaboration treatment (T1). Further, the positive T2 point estimate is driven by increased *coins given to KDC's development fund*, in accordance with theoretical expectations that legitimacy should lead

[37] Survey-based tax morale measures are often used in the tax compliance literature as proxies for tax compliance (e.g., Besley, 2020). Given the analytical distinction between a belief about paying taxes and behavior related to paying taxes, I decided to not to include this outcome in the compliance index. Doing so would increase the t-statistic on the compliance index to 2.45.

to consent-based compliance. Column 3 reports the effect of the Legitimacy treatment relative to the control, which we can interpret as the joint effect of the TPI collaboration treatment (T1) and the Legitimacy treatment (T2). While this comparison bundles theoretical mechanisms, it has important policy relevance. Legitimacy-based appeals made by traditional leaders work: They are more effective at generating tax compliance (p-value = 0.004) and increasing tax morale (p-value = 0.011) relative to generic appeals to pay made by government officials.

Column 4 shows that the Coercion treatment (T3) increases the compliance index above and beyond the effect of T1, and this effect is statistically significant at the 90% confidence level (p-value = 0.09). The point estimate on this increase is 0.066 standard deviations, almost exactly the size of the T1 effect. This positive effect on the compliance index is primarily driven by an increase in the direct measure of tax compliance, *self-reported propensity to pay tax*. In contrast to the Legitimacy treatment, the Coercion treatment has no impact on the voluntary compliance measure (*coins give to KDC's house fund*) suggesting that T3 impacts compliance through non-voluntary channels. Column 5 shows that the joint effect of the TPI collaboration treatment (T1) and the Coercion treatment (T3) and provides evidence that coercion-based appeals to pay are also effect. Relative to generic appeal made by government officials, coercion-based appeals from traditional leaders increase tax compliance (p-value < 0.001) and tax morale (p-value < 0.001).[38]

In summary, the main set of experimental results provide strong evidence that the TPI Collaboration treatment (T1) increases the measures of tax compliance used in this study. There is also evidence that the effect of collaboration is driven by both legitimacy and coercion, though the evidence for the coercion mechanism is stronger.

[38]In Table 2.4, I found evidence in favor of the coercion hypothesis: general statements made by the paramount chief about punishing noncompliance (T3) increased the tax compliance index, relative to T1. But what type of potential punishment are respondents reacting to? If TPIs' coercive capacity stems from chiefs willingness to abuse their control of local governance institutions, citizens may fear that noncompliance will be punished with biased future treatment; alternatively, if citizens may fear that noncompliance will be met with fines. In Appendix Table A.7, I attempt to tease out the specific tools of TPIs' coercive power by examining the impact of the Coercion treatment (T3) on six intermediate experimental outcomes. While the results in Table A.7 are murky, there is more evidence that the Coercion treatment (T3) is priming beliefs about fine-based punishment, rather than governance bias-based punishment.

Table 2.4: Effects of Mechanism Treatments (T2/T3) on Compliance Outcomes

	Mean	**Legitimacy**		**Coercion**		**N**
Outcome	(T1) (1)	(T2-T1) (2)	(T2-C) (3)	(T3-T1) (4)	(T3-C) (5)	(6)
Compliance Index	**0.063**	**0.056**	**0.124*** **	**0.066***	**0.134*** **	**1,752**
	(0.649)	(0.042)	(0.043)	(0.039)	(0.040)	
Self-reported propensity to pay tax	6.874	0.034	0.124**	0.133**	0.223***	1,751
	(2.920)	(0.061)	(0.060)	(0.058)	(0.058)	
Perceived neighbors' propensity to pay tax	6.226	0.058	0.170**	0.038	0.150**	1,657
	(2.381)	(0.071)	(0.071)	(0.070)	(0.069)	
Coins given to KDC's development fund	1.703	0.076	0.077	0.019	0.021	1,752
	(1.446)	(0.056)	(0.058)	(0.057)	(0.058)	
Tax morale (secondary outcome)	7.720	0.010	0.165**	0.064	0.219***	1,751
	(2.598)	(0.060)	(0.065)	(0.059)	(0.064)	

Table 2.4 reports the effect of the Legitimacy treatment (T2) and the Coercion treatment (T3) on the compliance outcomes. Column 1 reports the control group mean for each indicator, with the standard deviation in parentheses. Columns 2-3 present treatment effects for T2, relative to T1 and Control, respectively. Columns 4-5 present treatment effects for T3 relative to T1 and Control, respectively. Reported effects are standardized effects. Models are estimated using OLS with preregistered specifications. Column 6 reports the number of non-missing observations. The *Tax morale* measure was not included in the compliance index (as per the PAP).
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.5 Additional Legitimacy Results

To further investigate the legitimacy hypothesis, I collected both additional intermediate experimental outcomes. Table 2.5 presents the impact of the Legitimacy treatment (T2) on four intermediate experimental outcomes. The first two outcomes measure expected benefits of taxation, aimed at capturing performance-based aspects of legitimacy. The first indicator captures respondents' perception that their own village will benefit from taxation; the second, that other villages in their chiefdom will benefit. The third and fourth indicators capture perceptions that tax revenue will be spent transparently and efficiently, respectively; These outcomes are indicators of process-based legitimacy. Absent priors regarding whether TPIs' legitimacy stems from performance or process, the preregistered outcome for hypothesis testing is an index comprising all four indicators (legitimacy index). For these outcomes, the appropriate comparison group is the pure control (C; Tax

information). Using T1 (TPI collaboration treatment) as the comparison group for these outcomes is problematic because questions about legitimacy may prime respondents to TPIs' legitimacy. For example, having just seen their paramount chief in a video, T1 respondents may think of their chief's governance performance (a performance legitimacy prime) when asked if they expect to benefit from taxation, which would undermine the impact of the Legitimacy treatment (T2).

Table 2.5: Effects of Legitimacy Treatment (T2) on Intermediate Outcomes

Secondary Outcome	**Mean** (1)	**T2-C** (2)	**N** (3)
Legitimacy Index	**0.000**	**0.060**	**1,752**
	(0.698)	(0.042)	
Own village will benefit from tax	7.114	0.088	1,722
	(2.853)	(0.061)	
Other villages will benefit from tax	6.710	0.023	1,646
	(2.838)	(0.068)	
Ease of discovering how tax revenue was spent	4.460	0.103	1,742
	(3.094)	(0.064)	
Proportion of revenue towards development	6.180	0.021	1,709
	(2.614)	(0.066)	

Table 2.5 reports the effect of Legitimacy treatment (T2) on the secondary legitimacy outcomes. Column 1 reports the control group mean for each indicator, with the standard deviation in parentheses. Column 2 present standardized treatment effects for T2, relative to control. Models are estimated using OLS with preregistered specifications. Column 3 reports the number of non-missing observations. **Significance**: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Column 2 shows that the Legitimacy treatment (T2) increases the legitimacy index, relative to the control group, but this effect is not statistically significant (p-value = 0.15). However, the point estimate on all four sub-indicators is positive and of similar magnitudes across performance and process measures. These additional pieces of evidence complement the suggestive findings from my primary mechanism analysis (Table 2.4), where I found that the Legitimacy treatment (T2) increased the main compliance index above and beyond the effect of the Collaboration treatment (T1), but that this difference was not statistically significant (p-value = 0.18). Taken together, I interpret this collection of results as evidence in favor of the hypothesis that legitimacy is a source

of TPIs' authority.

2.6 Conclusions and Implications

Against the predictions of modernization theorists (e.g., Huntington, 1968), traditional political institutions are "resurgent" across contemporary sub-Saharan Africa (e.g., Englebert, 2005), with increasing legal protection (Holzinger et al., 2020; Baldwin, 2016, chpt. 3) and high levels of public support (Logan and Amakoh, 2022).[39] Using a field experiment in Sierra Leone, I find evidence that governments can increases citizens' tax compliance by collaborating with TPIs. This finding appears robust: using a preregistered specification, all three treatments that mention collaboration increase both a preregistered tax compliance index and respondents' belief that they ought to pay taxes. While the literature on state capacity in weak states has tended to focus on strategies for improving the effectiveness of *government* institutions, my results show that collaborating with traditional, non-state institutions can help governments to achieve their policy goals. Moreover, wile TPIs are often dichotomously cast as either coercive and despotic or legitimate, I find support for both coercion and legitimacy-based channels, suggesting that TPIs govern through a combination of these two tools.

Should weak states collaborate with TPIs to achieve their policy goals? First, governments thinking of such collaborations should recognize that chiefs likely vary in the extent to which they are effective collaborating partners.[40] Figure A.1 presents the treatment effect of collaboration conditional on chiefdom. The bolded black lines display the joint treatment effect, pooling across all treatment conditions (T1+T2+T3). Pooling buys more statistical power, as T1, T2, and T3 all feature the relevant paramount chief and contain messaging about collaboration. While this strategy bundles the effects of collaboration with the effects of legitimacy appeals and coercion appeals, this is appropriate because the objective is to explore variation in the authority of TPIs across chiefdom, and

[39] Legal protection of TPIs today is stronger than in independence era constitutions (Baldwin, 2016) and the constitutions of 1990s (Holzinger et al., 2020).

[40] A distinct, but related question is whether rural property tax can raise meaningful revenue. See Grieco (2024).

not necessarily to tease out the mechanism behind that authority. For transparency, the fainter lines alongside these pooled estimates present the treatment effect of each treatment condition individually, relative to control. Figure A.1 provides evidence that magnitude of the collaboration effect varies across chiefdoms. Focusing on the pooled estimates, we see that the effect of collaboration in Gbane Chiefdom is 0.284, more than four times the ATE of 0.068. By contrast, the effect of collaboration in Lei Chiefdom is 0.039, 43% smaller than the ATE. The difference between these effects is statistically significant at a 90% confidence level. Second, state leaders must consider the potential impacts of collaboration on long-term statebuilding. Does collaboration with TPIs undermine state leaders' attempts to build effective bureaucracies? While recent work suggests that citizens may view TPIs and government as complementary (Van der Windt et al., 2019), we have little evidence about how the state's collaboration with TPIs impacts citizens' attitudes and behaviors towards state officials.[41] This is one potentially fruitful path for future research.

[41] Also relevant here is Henn (2023), who argues that whether TPIs and the state are complements or substitutes depends on whether TPIs are integrated into a country's constitutional framework.

Chapter 3

Chiefs in Weak States: Enforcers of the Public Will

Traditional leaders have broad public support across much of sub-Saharan Africa.[1] While it has been argued that public support for traditional leaders is the foundation of their "resurgence" in contemporary, democratic Africa (Logan, 2013), we know little about what drives popular support.[2] Some scholars argue that "culture and identity," are the sources of traditional leaders' popularity, rather than their governing performance: "support for chieftaincy is not based on high satisfaction with the way chiefs perform their tasks" (Ubink, 2007). Indeed, Acemoglu et al. (2014) find that traditional leaders have greater support where they are *less* accountable, and therefore produce worse development outcomes.

In contrast, this article argues that traditional leaders' governing performance is central to understanding the public's support for these institutions. Specifically, this article argues that traditional

[1] Survey results from 31 African countries (Afrobarometer, Round 8, conducted 2020) show that 51% of the respondents would like to see these traditional leaders have more influence in local governance, while only 13% of respondents said they would like to see them have less influence. Compared to elected officials, traditional leaders are more trusted, seen as less corrupt, have higher approval ratings, and are perceived as better listeners (Logan and Amakoh, 2022).

[2] See Englebert (2002) for the resurgence of traditional chiefs in Africa.

leaders are popular because they help communities maintain the rule of law by enforcing local laws that people want and that increase community welfare. Traditional leaders facilitate public participation in law-making, such that local laws reflect community preferences. However, in the absence of enforcement mechanisms, citizens are unlikely to follow even pro-social laws because doing so entails individual costs (Olson, 1971). Traditional leaders facilitate the rule of law by detecting and punishing law-breakers.

I make this argument with qualitative data from 300 semi-structured interviews conducted in 29 villages in rural Sierra Leone. Motivated by the goal of explaining variation in chiefs' effectiveness as collaborators (Chapter 2), I probed interview data for variation in chiefs' governance practices. In doing so, I documented significant variation in citizens' perceptions of traditional leaders' enforcement of local laws: some respondents complained that chiefs enforced rules unfairly. By contrast, I found little variation and high overall levels of (1) public participation in law-making or (2) TPIs' enforcement capacity. My argument arises inductively from these documented patterns, which I validate by having research assistance code 261 interviews.

The empirical section of the paper proceeds in three parts. The first two parts show that (i) local laws are likely to benefit the public broadly and (ii) TPIs enforce those laws. Part one focuses on law-making and enforcement, generally. I investigate how local laws are made and citizens' general satisfaction with law enforcement. If local laws are to generate broad public benefits they should reflect citizens' preferences. I find high levels of citizen engagement in local policy-making, where village level laws are enacted with direct citizen input, and section and chiefdom level byelaws are enacted with the input from village representatives (e.g., village chiefs). Given these high levels of public participation, local laws likely reflect citizens' preferences. But for these laws to produce a public good (i.e., rule of law), they must be effectively and fairly enforced. I find that citizens are generally satisfied with the way chiefs enforce laws: respondents are six times more likely to mention law enforcement as a positive aspects of their chief's performance than a negative aspect of their chief's performance.

Part two takes a deeper look at a wide-spread local law that allows traditional leaders to demand mandatory labor from community members for community-oriented projects. I first show that citizens benefit from this communal labor law. Mandatory communal labor is directed at projects with broad social benefits—such as road maintenance and the construction of public buildings. Moreover, respondents are much more likely to report that projects undertaken with communal labor have broad, rather than narrow, social benefits.[3] I show how traditional leaders facilitate these communal labor projects by providing a granular description of TPIs' enforcement capacity, documenting their capacity to monitor participation in communal labor and to punish non-compliers. Taken together, I interpret the first two parts of the empirical section as evidence that TPIs facilitate a public good (i.e., the rule of law) by enforcing local laws.

In part three I take up whether traditional leaders role in maintaining the rule of law explains their popularity. I present evidence that traditional leaders are more popular and have more public support where they use their coercive capacity to fairly and effectively enforce laws and less popular where they fail to do so. First, I use chiefdom-level case studies to document substantial variation in the extent to which chiefs enforce laws fairly. Second, I have research assistants code 261 interviews and show that traditional leaders are more popular where they enforce laws fairly. Finally, I show that research assistants' coding validates my interpretation of chiefdom level variation in the fairness of law enforcement.

This article makes several contributions to the our understanding to TPIs. First, despite the central role TPIs play in dispute resolution and law enforcement (Baldwin and Raffler, 2019), "there is almost no literature on the topic [of chiefs' judicial role] in contemporary political science" (Wilfahrt and Letsa, 2023). We know little about how TPIs enforce laws, how law enforcement is perceived by local communities, and how it impacts citizens' perceptions of traditional leaders. This article attempts to fill this gap in the literature. Using qualitative evidence from 300 interviews, I argue that citizens perceive TPIs as legitimate because they enforce pro-social laws. This article builds

[3]Respondents are over eleven times more likely to report broad, rather than narrow, benefits for village level communal labor.

on prior research that suggests TPIs' enforcement capacity as a mechanisms enabling traditional governance systems to effectively deliver public goods in Mexico (Magaloni et al., 2019). My argument is also in keeping with lab experiments that find punishing noncompliers is important for cooperation (e.g., Fehr and Gächter, 2002; Ostrom et al., 1992).[4]

Second, I contribute to our understanding of public participation in TPIs. These institutions often have inclusive and transparent decision-making processes (Baldwin and Holzinger, 2019) and are often described in the literature contain elements of direct or participatory democracy (Skalník, 1996; Díaz-Cayeros et al., 2014; Magaloni et al., 2019). I contribute to the literature by, first, providing a rich description of public participation in law-making and second, distinguishing between public participation at different levels of policy-making. While public participation in enacting village-level laws is direct—mirroring the accounts in the literature—at higher levels of policy-making (i.e., section and chiefdom level laws) participation is law-making *representative*: section and chiefdom level laws are enacted at public meetings attended by village and section leaders, respectively.

Third, the extant literature tells us little about the *level* of bias in within traditional systems of governance. Qualitative accounts that richly document governance bias in TPIs (e.g., Mamdani, 1996; Richards, 1996)—for example, in dispute resolution (Maru, 2006) or land allocation (Ntsebeza, 2005)—do not systematically document instances of unbiased governance, making it impossible to establish the magnitude of governance bias in TPIs. Moreover, quantitative scholarship that investigates the causes (Baldwin et al., 2022) or consequences (Goldstein and Udry, 2008; Acemoglu et al., 2014; Mokuwa et al., 2011) of governance bias in TPIs does not measure this bias directly, but infers bias from behavioral patterns or institutional arrangements. For example, Goldstein and Udry (2008) demonstrate bias exists by showing that individuals who hold an office of social or political power have more secure property rights, make more investments in land fertil-

[4]Also related is (Logan, 2013, pg. 33), who finds a positive correlation between respondents' support for traditional leaders and traditional leaders' involvement in land allocation and conflict management. She interprets this relationship as evidence that traditional leaders' "ability to protect community stability is thus a cornerstone of [their] resilience."

ity, and obtain higher yields. But we do not know the percent of farmers whose land has been unfairly reallocated.[5] My approach directly and systematically captures respondents' perceptions and experiences of bias, which allows me to report easily interpretable magnitudes of perceived bias.

3.1 Rule of Law as a Public Good Facilitated by TPIs

As with any political institution, TPIs may be popular because of the public services they provide to constituents (Keele, 2007). Traditional leaders help their communities obtain important services from government (Williams, 2010), provide "a variety of services" (Mengisteab, 2017), and act as "essential safety nets. . . in an environment of poverty and unmet needs" (Koter, 2013).

One key public good provided by TPIs is the rule of law and social stability (Krämer, 2016). The local laws enacted by TPIs may be the product of intensive public participation because these institutions often feature "inclusive decision-making" and "direct forms of democracy," (Baldwin and Holzinger, 2019).[6] If law-making is participatory, the laws on the books are likely to reflect the preferences of large segments of the population. If this is the case, society benefits when individuals follow the law. That is, the rule of law is a public good. TPIs are also involved in applying and enforcing these laws. In a recent Afrobarometer survey (Round 8, conducted 2020), 71% of respondents reported that TPIs have "some" or "a lot" of influence over dispute resolution. Moreover, a recent study on the Ghana-Togo borderland finds that "chiefs are uniformly expected [by citizens] to adjudicate disputes and protect their communities from crime" (Wilfahrt and Letsa, 2023, pg. 440).

As with many public goods, producing the rule of law requires collective action: all citizens must

[5] Baldwin et al. (2022) find an intervention to increase the inclusivity of village-level decision making increases the percent of non-partisans who had disputes resolved to their satisfaction. While they (intuitively) interpret this as evidence that *more* disputes are resolved fairly, we do not know the level of disputes that are resolved fairly. Also note that Acemoglu et al. (2014) do not directly measure TPIs governance performance, but assume that chieftiancies with less political competition (proxied by number of ruling families) are more biased in their governance.

[6] Also see Magaloni et al. (2019) for participatory decision-making in TPIs.

follow the law. But organizing collective action for public goods can be difficult because of a well-known free-rider problem. Individuals must take costly actions to benefit the group. But because group benefits can be enjoyed by individuals who did not contribute to producing them, there is little incentive to make costly contributions. One way that organizations solve this free rider problem is by providing selective incentives, rewarding individuals who contribute to producing the public good and punishing individuals who do not (Olson, 1971).

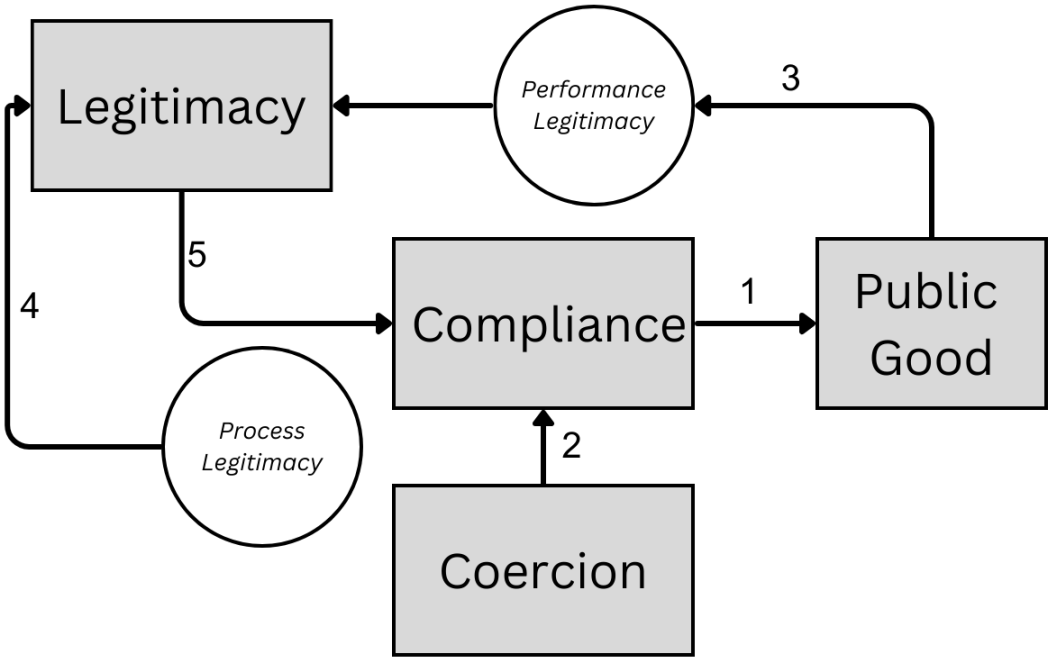
In many parts of the world, the state's coercive capacity is a tool to solve this free rider problem because it raises the cost of noncompliance and because it "persuades [individuals] that others are compelled to pay their share" (Levi, 1988, pg. 54). Where the state is weak, and therefore lacks the capacity to detect and punish noncompliance, TPIs can fill this role. Indeed, there is evidence that TPIs may possess this coercive capacity. According to a recent survey conducted in Malawi, Kenya, and Zambia, 43% of respondents say they make contributions (e.g., money and labor) to community water and sanitation projects (at least in part) because they fear fines from chiefs (Lust et al., 2019). Magaloni et al. (2019) argue that traditional governance systems in Mexico have "an effective system of community service enforced through credible sanctions."

Figure 3.1 visualizes this argument. Compliance with the law is necessary to produce the public good *rule of law* (Arrow 1). To obtain the compliance necessary to produce public goods, TPIs use the stick, punishing noncompliers (Arrow 2). By enforcing the compliance necessary to produce a public good, TPIs generate *performance legitimacy*, as citizens recognize the crucial role played by TPIs in producing the public good (Arrow 3). This legitimacy is a the source of their popularity.[7] In that sense, TPIs' coercive capacity and their legitimacy are complementary: without coercive capacity, TPIs could not organize the public goods that give them legitimacy.[8] TPIs' legitimacy further facilitates citizens' compliance (Arrow 5).

[7]Figure 3.1 also depicts the *process legitimacy* that TPIs from allowing citizens to participate in creating the laws that they enforce (Arrow 4).

[8]As social contract theorists such as Rousseau (1762) point out, the reverse is also likely to be true: coercive capacity would fail to generate meaningful levels of compliance in the absence of legitimacy: "The strongest is never strong enough to be always the master unless he transforms strength into right and obedience into duty."

Figure 3.1: Legitimacy and Coercion as Complementary Drivers of Compliance



3.2 Research Approach

The argument proposed in this chapter developed inductively. In Chapter 2, I found evidence that the effectiveness of traditional leaders in obtaining compliance from citizens with government policy varied across chiefdoms (Figure A.1). That is, chiefs varied in their effectiveness as collaborators. In an attempt to explain this variation, in fall of 2022, I worked with a team of six research assistants to conduct interviews with 300 respondents across 29 villages in four chiefdoms in Kono district (Gbane, Soa, Lei, and Nimikoro).[9] My prior was that this variation may be explained by differences in the way chiefs governed their chiefdoms. Therefore, interviews focused on key elements of chiefdom governance. Before conducting data collection, interviewers were trained during a five-day workshop to follow an interview guide which was structured to cover the following topics:

[9]See Chapter 2 for more details on chiefdom sampling.

- *Local tax:* An existing and widespread poll tax collected by chiefdom authorities. Questions focused on (i) perceived motivations for paying this tax and (ii) monitoring and punishment mechanisms for noncompliance.
- *Local laws:* Interviewer asked respondents to describe common local laws, then focused on the process for creating local laws and respondents' judgment regarding these laws.
- *Perceptions of chief performance:* Interviewers asked respondents to describe things that chiefs did well, things that chiefs could improve and their overall approval of the performance of chiefs. Interviews also asked respondents how they would react if chiefs performed poorly. Interview protocols were designed to ask about specific chiefs individually (i.e., "your section chief"), rather than chiefs generally.
- *Communal labor:* It is common for chiefs call for labor to undertake various activities. Interviewers asked respondents to describe recent projects completed with communal labor and systems of monitoring and punishing noncompliance. Respondents were also asked about their attitudes towards communal labor and who they thought benefited from projects undertaken with communal labor.

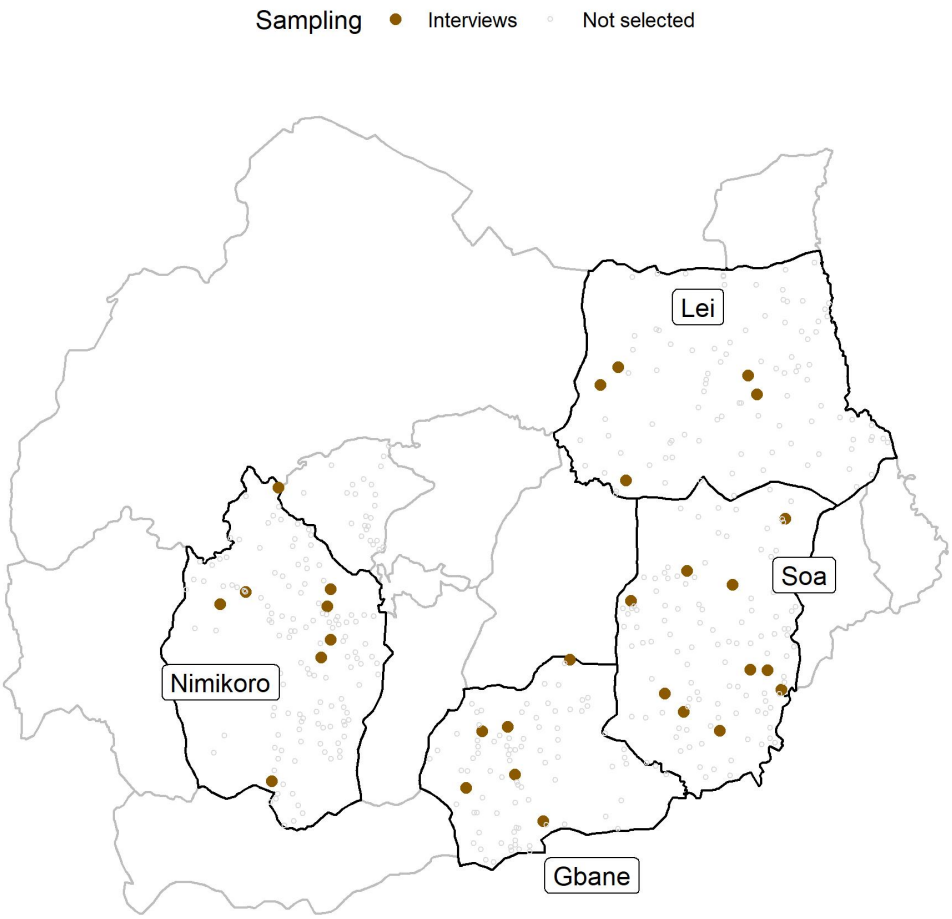
Brown circles in Figure 3.2 represent villages visited for qualitative interviews. To select villages for interviews, I first randomly selected sub-chiefdom administrative units (sections) in each chiefdom. Within each sampled section, I then selected section headquarter town and one other large town for interviews. Interviews were conducted in either Krio or Kono (or a mix of the two), depending on the preference of the respondent. All interviews were recorded and lasted approximately 20 minutes.

Once the interview data was collected, I listened to a random set of (Krio language) interviews, looking for variation in governance that might explain the variation in chiefs' effectiveness as collaborators. I found little evidence of variation in (1) public participation in law-making or (2) TPIs' enforcement capacity; respondents reported overall high levels of participation and enforcement capacity. Therefore, I ruled these out as drivers of chiefs' effectiveness as collaborators. I did find,

however, significant variation in citizens' perceptions of traditional leaders' enforcement of local laws: some respondents complained that chiefs enforced rules unfairly. Focusing on complaints about law enforcement, I went back and I listened to all Krio language interviews, village-by-village. I took notes on each interview, and then summarized notes for all (Krio) interviews fore each village. In general, respondents were pleased with the way that chiefs enforced laws and often cited this as chiefs' most important role. However, there were serious exceptions to this and my notes revealed that in one chiefdom in particular (Lei) there were consistent complaints about law enforcement. I found much less complaints in Soa and Nimikoro and no serious complaints in Gbane. Where respondents took issue with unfair law enforcement, I also documented that they were more likely to criticize the performance of their chiefs. That is, where chiefs were unfair they were also unpopular.

My argument arose out of these documented patterns. To validate these patterns, I developed a coding scheme to capture respondent's (i) perceptions of enforcement mechanisms for local tax and communal labor, (ii) participation in byelaw creation and attitudes towards byelaws, (iii) judgments of leaders' performance, and (iv) descriptions of and attitudes towards projects undertaken with communal labor. A team of three research assistant conducted the coding. Patterns in coding matched the patterns in my notes. In the empirical section, I present summary statistics from the (research assistant) coded interviews alongside my interpretation of these patterns.

Figure 3.2: Sampling Map



3.3 Enacting and Enforcing Laws Outside the State: Empirical Findings

My empirical discussion progresses as follows. Part (i) introduces the making and enforcing of local laws in Sierra Leone. This section demonstrates that citizens' are heavily involved in making laws and that they are generally satisfied with the way chiefs enforce them. Part (ii) explores an omnipresent law in detail: communal labor. This section documents at a granular level TPIs' capacity to monitor participation in communal labor and to punish non-compliers. This section further demonstrates that people anticipate broad social benefits from the projects undertaken with

this labor. Part (iii) uses case studies of two chiefdoms to further explore my argument. This section first highlights variation in the extent to which chiefs enforce laws fairly. In line with my argument, I present evidence that chiefs are less popular where they fail to fairly enforce laws.

3.3 (i) Making and Enforcing Local Laws

In rural Sierra Leone the government struggles to communicate laws and has little capacity to enforce them; government law enforcement officers are absent in rural areas. Instead, TPIs have the legal mandate to establish and enforce local laws (byelaws), which can be enacted at the village level by the village chief or more broadly throughout the chiefdom by chiefdom level authorities. According to interviews, the most common byelaws are those that prohibit theft (mentioned by 65% of respondents), abusive language (51%), and fighting (39%). It was also commonly mentioned that local laws mandate labor contributions to community projects (27%) or regulate livestock (9%) and the harvesting of crops (10%). Table 3.1 provides a more comprehensive list of byelaws mentioned during interviews.

Table 3.1: Common Bylaws (Village, Section, Chiefdom)

Mentioned byelaw	Village (%)	Section (%)	Chiefdom (%)	Any (%)
Any law	**94**	**71**	**68**	**97**
No stealing	41	34	39	65
No abusive language	40	17	19	51
No fighting	28	13	15	39
Participate in communal labor	13	14	14	27
No adultery	15	6	6	19
No rape	3	8	10	15
Law regulating crop harvest	3	5	4	10
Law regulating livestock	0	8	6	9

Table 3.1 reports byelaws most commonly mentioned by respondents during semi-structured interviews. Columns Village, Section, and Chiefdom present the percent of respondents who mention a given law at this administrative level. The column Any is the percent of respondents who mention this law at any administrative level. Percentages are rounded to the nearest integer.

These laws are created with direct and indirect citizen participation. Specifically, I find that (1) byelaws are developed and enacted at meetings, rather than behind closed doors; (2) local representatives are invited to attend meetings for chiefdom byelaws; and (3) participants in these policy-making meetings have space to actively engage. These findings (summarized in Table 3.2) are in line with recent research that documents inclusive decision-making processes in TPIs (Baldwin and Holzinger, 2019).

Table 3.2: Legitimacy: Participation in Local Policy-making

	% Agree	% Disagree	% Unclear
Meetings called to discuss byelaws			
Village	97.8	0.0	2.2
Chiefdom	97.1	0.0	2.9
Representatives invited to byelaw meetings			
Chiefdom	93.6	1.8	4.7
Active participation for meeting attendees			
Village	83.9	8.6	7.5
Chiefdom	78.3	3.7	18.0

Table 3.2 describes respondents' perceptions of participation in local policy making. Interviewers asked respondents about the creation of town, section and chiefdom level byelaws, without explicitly mentioning meetings. For ease of exposition, I combine responses about section and chiefdom level meetings. For the first outcome, respondents coded as "agree" ("disagree") explicitly mention that a meeting was (not) called. For the second outcome respondents are coded as "agree" if they say that representatives would be called to attend either section or chiefdom level meetings. For the third outcome, research assistants coded interviews for evidence of active participation ("agree"), evidence of lack of participation ("disagree"), or no evidence for either ("unclear"). Throughout, respondents are coded as "unclear" for a given outcome when their response is ambiguous or when they don't answer a given prompt. If the interviewer did not ask the question the respondent is removed for that outcome.

Informants highlighted that authorities "made laws in consultation with the people"[10] and that the law making process was open to all community members: "whether you have a [leadership]

[10]Interview: 52

position in the town or not...it is us all that sit and make [the laws]."[11] While chiefs may be the "chief coordinators of the law" they "do not make [laws] alone."[12] Byelaws are developed during open meetings, rather than behind closed doors. To discuss village byelaws authorities "invite the entire community"[13] or "the whole town"[14] to meetings. Similarly, chiefdom authorities call a "general meeting"[15] to formulate chiefdom byelaws. To elicit these responses, interviewers asked informants how byelaws were created without mentioning or making reference to meetings. Nearly all respondents (97%) explicitly mentioned that meetings are called when byelaws are created, either within the village or more broadly within the chiefdom.[16] Moreover, these meetings are themselves often the product of citizens' participation, called to discuss a proposed law put forward by an individual or community groups.[17]

When meetings are called to discuss chiefdom byelaws, all villages are represented by local chiefs and community leaders. Chiefdom authorities invite "town chiefs, section chiefs, youth chairmen, and the mammy queen [i.e, women leaders]" from each village.[18] These invitations are extended "in a form of a letter"[19] or by "sending young men to every community."[20] Informants offered these responses after interviewers prompted: *"when chiefdom byelaws are created, is anyone from your village usually involved?"* Ninety-four percent of respondents said that a representative from their village or section would be invited to attend these meetings and only 1.8% say that they would not.[21]

These policy-making meetings are spaces where attendees are "given a chance to talk"[22] about

[11]Interview: 404
[12]Interview: 61
[13]Interview: 400
[14]Interview: 404
[15]Interview: 401
[16]These statistics, as well as the other statistics in this sub-section, are presented in Appendix Table 3.2.
[17]Interview: 403; 406; 405
[18]Interview: 404; Says another respondent, describing the universal representation from villages in their area: "all the nineteen villages are invited. No one is left out" (Interview: 10).
[19]Interview: 128
[20]Interview: 238
[21]Percentages are conditional on the respondent saying that a meeting would be called.
[22]Interview: 120

"burning issues."[23] Participants can "ask questions and make suggestions"[24] about byelaws that chiefdom authorities or other participants are putting forward. Research assistants coded interviews for evidence of active participation based on respondents' description of policy-making meetings. At the village level, 84% of respondents describe meetings as forums for discussion between village authorities and villagers, compared to 8.6% who report that these meetings are only a space for village authorities to *inform* the village's residents about a byelaw. For chiefdom meetings, 78% of respondents describe these meetings as containing active participation from attendees; only 3.7% of respondents report that these meetings are not open for active participation.

In interviews, respondents readily connect laws to public order and social stability, crucial public goods that improve community welfare. "If laws are not made, the town would not [be in] control. Everybody will just be doing things the way they want."[25] Many informants shared the perspective that, "it is the law that binds us together,"[26] brings "peace and protection,"[27] and "made the town function well."[28]

Moreover, citizens credit traditional leaders with enforcing these laws (Table 3.3, Panel A). To generate more systematic evidence on respondents' perceptions of the role of traditional leaders in law enforcement, research assistants coded responses to the following interview prompts: (1) *"Please tell me something the leaders of this section (chiefdom) are doing well?"* and (2) *"Please tell me something the leaders of this section (chiefdom) could be doing better, that is, could improve?"*. Over a quarter of respondents (26%) cited law enforcement as something that chiefs do well; only 4% of respondent cited law enforcement as something chiefs needed to improve on, or commented that laws were being unfairly or incorrectly enforced or implemented. We might be concerned that respondents avoid making critical remarks about chiefdom authorities, in general. Panel B (Table

[23]Interview: 404
[24]Interview: 32
[25]Interview: 70
[26]Interview: 100
[27]Interview: 112. Another respondent noted laws were put in place so that communities "could have peace" (Interview: 18).
[28]Interview: 37

3.3) suggests that this is not that case: over 60% of respondents were critical of the job traditional leaders were doing to bring development to their community.

In summary, citizens are actively involved in policy-making; view laws as crucial for public order and social stability; and credit TPIs for enforcing these laws. This is in keeping with my argument that TPIs derive their legitimacy from enforcing laws that are welfare-enhancing. In the next section I focus on one particular law, communal labor, and provide further evidence that citizens believe this law is socially beneficial and TPIs are crucial for enforcing this law.

Table 3.3: Perceptions of TPIs' Law Enforcement Performance

	% Agree	% Disagree
Panel A: TPIs enforce laws well		
Section	15.2	1.44
Chiefdom	10.1	2.16
Either	26.0	4.10
Panel B: TPIs are doing well to bring development		
Section	28.7	44.8
Chiefdom	36.5	38.1
Either	50.8	61.9

Table 3.3 provides evidence that citizens approve of the way TPIs enforce laws. Panel A presents citizens' perceptions of the performance of traditional authorities. The prompts were: "Please tell me something the leaders of this section (chiefdom) are doing well?" and "Please tell me something the leaders of this section (chiefdom) could be doing better, that is, could improve?" Responses are coded as "agree" if the informants named law enforcement (top outcome) or development (bottom outcome) as something authorities were doing well and "disagree" if the informants stated that authorities could improve on those dimensions.

3.3 (ii) Communal Labor: Enforcement and Public Good

While the Sierra Leonean government does provide certain public works in rural areas, their capacity to do so is limited. For example, while government may build schools and hospitals, or occasionally undertake large-scale rehabilitation for primary rural roads, they are unlikely to re-

pair school roofs that have been blown in by heavy wind or maintain rural roads that rainy season has left overgrown with brush and cratered with potholes. Instead, many public works in rural areas are organized locally, undertaken with communal labor—mandatory labor contribution that can be demanded by chiefs.

Communal labor is directed towards projects that likely have broad social benefits (Table 3.4), such as clearing vegetation from roadways ("road brushing," 68%) and road maintenance (39%).[29] Indeed, when respondents were asked what their traditional leaders were doing well, many pointed to road maintenance projects undertaken with communal labor. One respondent lauded their section chief for "greatly improving our roads by organizing communal labor"[30] and other respondents praised section leaders for "the maintenance of roads to connect our communities"[31] and maintaining "a good road network."[32] Respondents stated that chiefdom authorities, "make sure that our roads are good for safe movement"[33] and praised how chiefdom authorities "mobilized the youth for road maintenance."[34] While respondents do report instances of chiefs calling communal labor for personal projects—most notably to work on private farms—these occurrences are rare: only 1% of respondents mention that town communal labor has been used for work on a private farm and only 4% of respondents mention that chiefdom communal labor has been used this way.

Moreover, respondents themselves report that communal labor is directed towards projects with broad public benefits.[35] At the village, section, and chiefdom level respondents describe communal labor as being directed towards projects that benefit the public (Table 3.5). At the town level, over 80% of respondents describe communal labor as devoted exclusively towards public project and only 7% say communal labor is sometimes or often devoted to projects that do not benefit

[29]As traditional leaders at each administrative level (i.e., village, section, chiefdom) can demand communal labor, interviewers separately asked respondents about communal labor at each level.

[30]Interview: 75

[31]Interview: 18

[32]Interview: 123

[33]Interview: 21

[34]Interview: 110

[35]The interview prompt asked respondents if communal labor was used "in a fair way that benefits the community or is it used in an abusive way that benefits only a few people?"

the public.[36] The majority of respondents also say that section authorities (75%) and chiefdom authorities (66%) always use communal labor for public benefits. A minority of respondents say that section authorities (11%) and chiefdom authorities (18%) sometimes or often divert communal labor towards private projects.[37]

Table 3.4: Common Communal Labor Projects

Projects	Village (%)	Section (%)	Chiefdom (%)
Any Project	**98**	**82**	**69**
Road brushing	68	27	12
Road maintenance (e.g., fix potholes)	39	46	31
Cleaning (Town / building)	28	14	16
Construction of building	12	19	20
Labor on private/personal farm	1	3	4

Table 3.4 reports projects to which communal labor is devoted, according to respondents in semi-structured interviews. Respondents were asked to name up to three recent projects carried out with communal labor. This table report the percent of respondents that name a given type of project. This table presents a non-exhaustive list. Percentages are rounded to the nearest integer.

Participation in communal labor is mandatory and I find qualitative evidence that TPIs have enforcement mechanisms in place to detect and punish noncompliance.[38] According to respondents, the Youth Leader is commonly responsible for monitoring attendance and participation (mentioned by 67%) and reporting to the authorities: "we have the youth leader, he reports to the chief."[39] Authorities also monitor attendance themselves, either relying on their knowledge of community members to identify who has failed to show up (46%) or keeping an attendance list (33%). Says one respondent, "if you failed to go, your town chief will know because the town chief knows everyone."[40] Taken together, 90% of respondents described at least one monitoring mechanisms at

[36]Communal labor is most frequently called by town chiefs, where 98% of respondents named a recent project undertaken with communal labor.

[37]Note that these percentages do not add up to 100% because some responses were ambiguous.

[38]In Appendix B.1, I present additional evidence of TPIs' enforcement capacity by focusing on the enforcement of a widespread chiefdom tax.

[39]Interview: 32

[40]Interview: 401

either the town or chiefdom level.[41]

Table 3.5: Who Benefits from Communal Labor?

	% Agree	% Disagree	% Unclear
Communal labor directed towards broad public benefits			
Town	82.6	7.0	10.5
Section	74.6	11.3	14.1
Chiefdom	65.5	18.2	16.4

Table 3.5 presents respondents' perceptions regarding the benefits of mandatory communal labor. Interviewers prompted "Is the Communal Labor called for by the Town (Section; Chiefdom) Leaders usually used in a fair way that benefits the community or is it used in an abusive way that benefits only a few people?" Responses are coded as "agree" respondents say community labor is used for broad community benefits, "disagree" if respondents say community labor is sometimes or often used for narrow benefits, and "unclear" is the response is ambiguous or there was no direct response to the question.

Nearly all respondents agreed that individuals will be punished if they are caught missing community labor, with 90% of respondents explicitly describing the punishment as a fine.[42] Only two respondents mentioned non-fine punishments—both involved punishing the offender by impeding their livelihood (e.g., preventing them from farming) until they had undertake community labor. No respondent mentioned that noncompliance would be punished with bias in future dealings with chiefs. Table 3.6 breaks out statistics on monitoring and punishing noncompliance with communal labor demands at the village and chiefdom level.

[41] Interviewers asked respondents, "What happens if someone who was supposed to participate in Community Labor does not? Would the Section (chiefdom) Leaders find out?"

[42] Respondents noted that sickness or unavoidable travel were legitimate excuses for missing communal labor.

Table 3.6: Communal Labor: Monitoring and Punishment Strategies

	Village	Chiefdom	Either
	(%)	(%)	(%)
Communal Labor			
Monitoring: Any	**60**	**83**	**90**
Youth Leader informs authorities	46	44	67
Authorities personally identify	8	42	46
Authorities keep attendance list	9	24	33
Punishment			
Fines	60	83	90

Table 3.6 presents qualitative evidence of TPIs' monitoring and punishment strategies for Communal Labor. Interviewers' prompt: "What happens if someone who was supposed to participate in Community Labor does not? Would the Village (chiefdom) Leaders find out? Would they do anything?" Percentages are rounded to the nearest integer.

3.3 (iii) Unfair Enforcement Weakens Support for Chiefs

In the previous sections I provided evidence that TPIs enforce laws that increase community welfare. But are TPIs popular and legitimate because they do this? While respondents are more likely than not to approve the way TPIs enforce laws (Table 3.3) and administer communal labor (Table 3.5), there is some variation; not all respondents approve of TPIs' law enforcement. If traditional leaders are popular because they effectively enforce popular laws, they should be less popular if they fail to effectively enforce laws or enforce them unfairly.

In this section, I document variation in citizens' perceptions of traditional leaders' enforcement of local laws. In the chiefdom level case studies that follow, I focus on Gbane Chiefdom and Lei Chiefdom because they represent the extremes in terms of perceived fairness of law enforcement. In Lei Chiefdom, I find systematic complaints regarding the enforcement of local laws: there are widespread complaints against chiefdom leaders regarding their handling of ongoing conflict between crop farmers and cattle rearers. Conversely, in Gbane Chiefdom, I do not find complaints

against enacted local laws or their implementation. According to my argument, TPIs' popularity should be lower in Lei Chiefdom than in Gbane Chiefdom. Indeed, I find evidence that this is the case.

Lei Chiefdom

In Lei Chiefdom a local law has been passed that outlines compensation for crop farmers when cattle damage their crops and compensation due to herders if a farmer attacks a cow.[43] In three of the five villages where our team conducted interviews in Lei, multiple respondents highlighted frustration with chiefdom authorities regarding their handing of the conflict between crop farmers and cattle rearers. A respondent in one village takes issue not with the law itself but with its implementation.

> I have a problem with one [law] that has not been implemented fairly. This is occurring during the process of adjudicating on matters where a livestock farmer's animal has eaten a farmer's crop. In matters like that, the crop farmer's complaint is not treated seriously or followed through on according to the byelaw and most times unreasonable [i.e., very low] compensation is made. On the other hand, if a crop farmer kills a cow of a livestock farmer, [that crop farmer] will be beaten, molested, and treated poorly. There is no equity in [chief's] judgment of this byelaw. Cattle rearers are favored against crop farmers.[44]

Says another respondent in the same village agrees, "The laws between the cattle owners and the crops farmers are very fine in writing and when reading them, but its implementation is very bad."[45] A third respondent from the same village takes issue with the perceived difference in standards applied by chiefdom authorities to crop farmers and cattle rearers, "If a cow eats the rice you've planted, they eat the money that you would need to pay the children's school fees. If you

[43]Interview: 96. I use "cow" to refer to the singular of cattle, but I do not intend to convey sex. The cattle in this context are both male and female.

[44]Interview: 100

[45]Interview: 130

complain nothing happens. . .But if you kill one cow. . .[inaudible] that's an issue."[46]

This perception that crop farmers are getting the short end of the stick turns up in other villages. Says a respondent in a different village, "If a cow ruins someone's farm, [the authorities] should summon that person [to court]. At times it can take the chief a month to do so, as they are avoiding the case. But if something happens to a cattle, within 30 minutes or an hour, an arrest is made and someone is detained."[47] Another respondent in the same village has similar frustrations with inaction from chiefdom authorities: "If a cow eats my rice, and I make a report to [the section chief] take action! ...[the authorities] should take action, but they don't."[48] In a third village, there are similar complaints, "As a man of the country, I haven't see anything good yet that [chiefdom authorities] have done. Like when those cows ruin our rice, we cry. The money! But when the cattle herder comes [inaudible] he doesn't have money [for us]. The authorities don't do anything."[49] Even Chiefdom leaders admit that this is a problem. In a section headquarter town in Lei, when asked what chiefdom authorities could improve, the first topic discussed by the Section Chief is the "settling of dispute among farmers and cattle rearers."[50]

Gbane Chiefdom

In Gbane Chiefdom, I fail to find similar systematic complaints about local law enforcement. While many of citizens' complaints in Lei Chiefdom focused on planter herder conflict, there were no such issues that cut across the six villages where we conducted interviews in Gbane Chiefdom. In fact, only in one village did respondents' complaints converge on a topic: the role of chiefdom leaders in resolving a boundary dispute with a neighboring village. Three of the eight interviews I reviewed in this village mention the boundary dispute and place negative judgment on chiefs' role in this dispute. Across 30 interviews I reviewed in the remaining five villages, I fail to document strong criticism of local laws (or their implementation). In two of these villages, the

[46]Interview: 10
[47]Interview: 1
[48]Interview: 31
[49]Interview: 106
[50]Interview: 76

strongest criticism I can find against chiefs is that the paramount chief does not live in the chiefdom headquarter town, but the district headquarter. For reference, nearly all paramount chiefs reside most of the time in the district headquarter town (Koidu).[51] In the remaining three villages, the biggest complaints against chiefs are fairly normal demands for development (e.g., improve water access, improve roads), demands that are also commonplace in other chiefdoms.[52] I cannot find a complaint against a law or implementation of a law.

Perceived Performance

Table 3.7 citizens' perceptions of laws, law enforcement, and traditional leader performance, by chiefdom. Nimikoro and Soa Chiefdoms are included for transparency. Table 3.7 validates my understanding of the patterns of law enforcement in the qualitative data. Research assistants code laws as more effectively and fairly enforced in Gbane compared to Lei. Strikingly, in Gbane, 40% of respondents named law enforcement when asked what traditional leaders do well; in Lei not a single person mentioned this (Column 1). Relatedly, this difference is also reflected in the laws on the books. Respondents in Lei were three times more likely to say that they did not like a law than respondents in Gbane (Column 2).

If chiefs are perceived as legitimate because they effectively enforce laws people like, I should find that TPIs' popularity is higher in Gbane Chiefdom Chiefdom than in Lei Chiefdom. I find evidence that this is the case. In Gbane, about 16% of people thought that their own chiefs were performing worse that chiefs in other chiefdoms. That number is twice as high in Lei where over 30% of people say that chiefs are performing better in other chiefdoms.

[51] Traveling from to Koidu to Gbane's chiefdom headquarter town is a several hour trip on bad roads.

[52] These complaints are stronger in one village, where several informants feel left out of development that they say is occurring in other places in the chiefdom. However, there are no complaints about the way laws are implemented.

Table 3.7: Perceptions of Law Enforcement and Leader Performance

Chiefdom	Approve Enforcement	Dislike Law	Disapprove Chief Performance	N
Gbane	41.2%	11.8%	15.7%	51
Lei	0.0%	36.2%	31.2%	48
Nimikoro	33.3%	9.3%	8.0%	77
Soa	25.8%	22.0%	19.3%	85

Table 3.7 reports citizens' perceptions of laws, law enforcement, and traditional leader performance, by chiefdom. Column 1 reports the percent of respondents that mentioned law enforcement when prompted: *"Please tell me something the leaders of this section (chiefdom) are doing well?"* (as in Table 3.3). Column 2 reports the percent of respondents who reported there is a law in their chiefdom they do not like. Column 3 reports the percent of respondents who state the performance of traditional leaders in their chiefdom is *worse* than the performance of traditional leaders in other chiefdoms. Column 4 shows the number of coded interviews.

3.4 Conclusion

TPIs are not always viewed favorably by scholars (e.g., Mamdani, 1996; Richards, 1996; Ntsebeza, 2005). According to Jesse Ribot and his coauthors "many of the 'indigenous' governance systems, when analyzed as political systems rather than being viewed as cultural forms, would be labelled totalitarian, despotic, oppressive, patriarchal, gender biased or gerontocratic" (Ribot et al., 2008). Yet, it is undeniable that these institutions are popular in many countries in Africa (Logan and Amakoh, 2022). Public support for these institutions may be based on their cultural, spiritual, and symbolic importance (Honig, 2019; Williams, 2010) and their real or imagined continuity with the past (Robinson, 2024). Is that the full story? Or does the public also support these institutions because they govern well?

In this article, I argued that traditional leaders, and the institutions they head, are popular because they help communities maintain the rule of law. Enforcing laws increases community welfare because, generally, laws reflect citizens' preferences and are enforced fairly. Tradition leaders play

legislative, judicial, and executive roles: traditional leaders facilitate citizens' direct and representative participation in law-making; adjudicate disputes; and monitor and punish noncompliance with local laws. I support these arguments with qualitative data from 300 interviews conducted in four chiefdoms in Kono district, Sierra Leone.

The primary contribution of this article is to illuminate chiefs' role in law enforcement and show that TPIs' coercive capacity is a key ingredient of their popularity. In much of rural Africa, the state has no teeth to enforce its laws. And when the state does flex its coercive muscles, it is often in the service of repression (Christensen, 2018). Despite the lack of state presence, the African hinterland is far from anarchy. The findings from this article suggest that social order is achieved because TPIs provide a socially approved centralized system of coercion.

A secondary contribution of this article is to provide a more nuanced understanding of public participation in TPIs. Specifically, while I find evidence of direct citizen participation at the village level, citizens' participation in law-making at the section and chiefdom level is through representatives. Yet, as far as I am aware, there is no research on representation in TPIs and this may be a fertile ground for future research. Does the attendance of village leaders at section meetings impact policy outcomes? If so, structural barriers to meeting attendance (e.g., cost of travel) are likely to have distributional consequences.

Finally, I report levels of governance bias in TPIs. While my findings suggest that most people think that their traditional leaders are performing their roles well most of the time, the qualitative data contain numerous allegations of individual chiefs abusing power or otherwise performing poorly. However, I make no attempt to explain why some chiefs govern better than others. We know little about what explains variation in governance quality across TPIs, though Acemoglu et al. (2014) are a notable exception with their argument about electoral competition. Future research should explore this variation, possibly through more systematic data collection on the coercive powers, institutional checks, and governance performance of TPIs.

Chapter 4

Participation, Legitimacy and Tax Compliance:

Evidence From a Participatory Budgeting Experiment

Fiscal capacity, and state capacity more generally, is essential for economic development and political stability (Besley and Persson, 2011). Building durable fiscal capacity requires that the state obtains compliance with its tax demands. One strategy is through enforcement. There is extensive evidence of the effectiveness of enforcement (e.g., Kleven et al., 2011; Slemrod, 2019), but those efforts are administratively and politically costly, can erode public support if they are viewed as unfair, and are often not undertaken owing to a lack of political support.[1] An alternative strategy is to strengthen the willingness of taxpayers to quasi-voluntarily comply with tax demands by expanding the perceived legitimacy of government (Levi, 1988, 1997). Such efforts may focus on building citizen trust that the tax system is fair, that citizens will receive valuable services in return, or that citizens have a voice in shaping how revenues are used (Luttmer and Singhal, 2014;

[1] Levi (1988) considers the administrative costs of revenue extraction as part of the broader "transaction costs" that rulers face. The political costs of increased tax demands include stronger demand-making from citizens (Weigel, 2020; Martin, 2014) and electoral backlash (Christensen and Garfias, 2021).

Prichard, 2023). In the context of weak states with limited enforcement capacity such efforts to encourage quasi-voluntary compliance are likely to be especially important.[2]

One way that governments may cultivate legitimacy is by inviting public participation in political affairs, which is central to both classic notions of legitimate government (Locke, 1690) and modern democratic theory (Pateman, 1970). Indeed, the link between public participation and tax compliance is central to seminal accounts of the development of fiscal capacity in early modern Europe, which posit that political leaders traded expanded political voice to elites in exchange for consistent sources of revenue (North and Weingast, 1989; Bates and Lien, 1985). In contemporary representative democracies, one method for expanding political voice is to allow citizens to *directly* shape policy outcomes, such as through participatory budgeting. By expanding citizens' political participation, participatory budgeting can build fiscal capacity in two ways. First, it may increase tax compliance, thereby generating additional revenue that can be plowed back into capacity-building. Second, it may increase politicians' willingness to take costly steps to build fiscal capacity. This is because citizens are more likely to oppose the government's attempt to increase fiscal capacity when they believe it will not translate to better services (Gottlieb and Hollenbach, 2018) or will enable extraction by those in government (Olson, 1993; Martin, 2023). Knowing this, office-seeking politicians will be more likely to take steps to build fiscal capacity when citizens perceive the government as legitimate because political costs will be lower.[3] These enabling attitudinal effects may be as, or more, important in the long-term for capacity-building as more direct impacts on compliance (Prichard, 2023).

This paper examines the relationship between political participation, legitimacy, and tax compliance in Freetown, Sierra Leone by reporting results from a participatory budgeting field experiment that we designed and implemented in collaboration with the Freetown City Council (FCC).

[2]Conceptually, we follow Migdal (1988, pg.4) who defines state capacity as the capability of the state to "achieve the kinds of changes in society that their leaders have sought through state planning, policies, and actions" (see also Hanson and Sigman, 2021; Cingolani, 2013).

[3]Indeed, there is evidence that politicians face electoral consequences for investing in fiscal capacity (Christensen and Garfias, 2021).

In doing so, we contribute to an emerging literature on institutional experiments (Callen et al., 2023) by providing the first field experimental study of whether participatory budgeting can facilitate capacity-building in fragile states. The intervention sought to give participants greater voice in, and control over, policy decisions regarding local development projects implemented in their neighborhood. Program participants joined phone-based WhatsApp chat groups—which we call *Digital Town Halls* (DTHs)—with up to 36 other property owners in their neighborhood to discuss service preferences, shared these preferences with politicians, and then voted for the services (of approximately US$1,500) they wanted to be implemented in their neighborhood. Selected services were implemented six months later, and participants were informed of this through a phone call. To identify causal effects, we use a matched-pair design (King et al., 2007) to randomize half of 3,618 property owners into treatment. We observed individual-level tax compliance through administrative records and surveyed the treatment and control groups at three stages: before the process, after services were selected but before they were delivered, and after services were delivered.

We find that participating in the DTHs increases perceptions of government legitimacy. In line with standard conceptualizations of legitimacy (Levi, 1997; Levi et al., 2009), we measure citizens' perceptions of (i) their influence over policy, (ii) government service delivery performance, (iii) government administrative competence, and (iv) politicians' performance in three survey waves. The intervention significantly increases eight of our nine legitimacy outcomes (the p-value on the ninth indicator is 0.11) at the endline survey, which was conducted soon after services were implemented and seven months after the conclusion of the DTHs.

However, despite the intervention's positive and durable effect on perceptions of government legitimacy, we find a robust null effect on property tax compliance. Why does the intervention fail to increase tax compliance despite raising perceptions of legitimacy? Empirically, the answer lies in strikingly heterogeneous, and countervailing, impacts of the intervention on different sub-groups of taxpayers. In exploratory analyses, we find that for individuals supporting expanded taxation to enable improved services at baseline (57.4% of the sample), treatment significantly increases

compliance by 3.31 percentage points, a 9.5% increase over the group's baseline compliance rate. However, these effects are washed out by decreasing compliance among those who initially did not support expanded taxation.[4] We find a similar pattern along partisan lines. For co-partisans of the Mayor, the treatment increases compliance by 7.4 percentage points, which is a substantial 27.9% increase over the group's control compliance rate. In contrast, the treatment lowers compliance by 4.0 percentage points for non-co-partisans.[5]

Interestingly, we do not find corresponding heterogeneous treatment effects on perceptions of government legitimacy. That is, while treatment effects on compliance are larger for co-partisans and tax supporters relative to non-co-partisans and tax opponents, respectively, there is no difference in treatment effects on legitimacy outcomes between these sub-groups. How can we make sense of this divergence between the impact of the treatment on attitudes toward government and the impact on both actual compliance and support for expanded taxation?

The most compelling explanation for these heterogeneous effects on compliance is that treatment leads to heterogeneous, and countervailing, effects on groups' policy preferences about taxation. First, an extensive literature finds that individuals respond to elite cues by shifting their policy position closer to the position of their party (Broockman and Butler, 2017; Tappin et al., 2023).[6] Moreover, while these cues may be persuasive for the political ingroup, they can generate backlash from the outgroup (Nicholson, 2012; Haas and Khadka, 2020). In the DTHs, Mayor Aki-Sawyerr pitched her policy vision of more taxes and better services, a message that is also at the core of her administration's *Transform Freetown* agenda.[7] Co-partisans responded to the treatment by increasing their support for expanding taxation for better services; treated non-co-partisans reduced their support for this policy. Second, an established literature in psychology and political science finds that individuals often have a motivation or goal (Kunda, 1990) when processing new information to

[4]The interaction term of a model that interacts treatment with our five-point measure of (baseline) attitudes towards expanded taxation is statistically significant (p-value < 0.001).

[5]The difference in the treatment effects on compliance between the two groups is significant (p-value = 0.034).

[6]Also see Barber and Pope (2019), Bolsen et al. (2014) and Flores et al. (2022).

[7]As the treatment increased individual's political engagement more broadly, treated individuals were more likely exposed to Mayor Aki-Sawyerr's support for expanded taxation outside of the DTHs.

form opinions about policy (Lodge and Taber, 2013; Taber and Lodge, 2006). Individuals "strive to defend and maintain their extant values, identities, and attitudes" (Slothuus and De Vreese, 2010).[8] We find that treatment crystallized policy preferences about taxation: opponents of expanding taxation became more opposed, and supporters became more supportive. If individuals' preferences about tax policy affect tax compliance behavior, the treatment's polarizing effect on tax policy preferences can explain its polarizing effect on compliance behavior.

Ultimately, our study makes four key contributions. First, we provide the first field experiment on whether direct democracy in the form of participatory budgeting can increase (a) support for authorities seeking to expand the state and (b) tax compliance in an already-existing representative democracy. While a related literature finds that public participation in Town Halls improves development outcomes (Gonçalves, 2014; Fujiwara and Wantchekon, 2013), participants' political efficacy (Boulianne, 2019) and knowledge (Esterling et al., 2011), and approval of participating politicians (López-Moctezuma et al., 2022), we do not know if these participatory platforms engender tax compliance and capacity-building. We find that participation in participatory budgeting Town Halls increased the legitimacy of government and approval of participating politicians, but raised tax compliance only among co-partisans of the Mayor.

Our finding that expanded voice does not, on average, increase tax compliance is in contrast with more optimistic accounts offered by existing observational work (Pommerehne and Weck-Hannemann, 1996; Torgler, 2005; Touchton et al., 2019). Our results also contrast with findings from lab experiments, which find that people are more likely to comply with (lab) taxes when they vote over how funds will be spent (Alm et al., 1993) and under rules they can influence (Bó et al., 2010; Sutter et al., 2010). However, observational studies may suffer from endogeneity issues, lab experiments abstract from political context, and both use as dependent variables survey measures, aggregated revenue data, or game-based measures rather than individual-level compliance

[8] According to Taber and Lodge (2006), "people are often unable to escape the pull of their prior attitudes and beliefs, which guide the processing of new information in predictable and sometimes insidious ways." Also see Stanley et al. (2020). Mullinix (2016) studies the effect of partisan cues and prior beliefs together.

data. Our approach improves on existing work by experimentally manipulating political voice in the real world and directly observing the impact on individuals' compliance behavior. Our finding that increased citizen engagement—and accruing legitimacy gains—does not translate to higher tax compliance has important theoretical implications, as it challenges influential models of quasi-voluntary tax compliance (Levi, 1988; Luttmer and Singhal, 2014) and implications of canonical accounts of the taxation-representation nexus in early European state building (Bates and Lien, 1985; North and Weingast, 1989). However, as we discuss, legitimacy gains can open windows of opportunity for subsequent reform, with likely positive consequences for long-term capacity building.[9]

Second, this paper brings politics into the literature studying the impacts of political participation on compliance, by emphasizing the importance of policy preferences (i.e., support for expanded taxation) and political affiliation in conditioning this relationship. The existing literature argues that citizens should be more supportive of, and compliant with, taxes when they can directly decide how revenue would be used. While this logic is intuitive for certain sub-groups, the prediction is less clear for individuals who are in ideological opposition to taxes or political opposition to the government. On one hand, inviting ideological or political opponents into discussions of how revenues are used may help to overcome initial skepticism or resistance. On the other hand, giving voice to these opponents may deepen existing resistance. In many ways this literature, though focused on political participation, has been oddly apolitical in overlooking different political preferences and partisanship.

We advance this literature by collecting a suite of pre-treatment political and attitudinal variables to tease out heterogeneous effects. In doing so, we are, to our knowledge, the first to document that the effect of expanding political voice varies conditional on fundamental political variables. While our finding that participatory budgeting *decreases* tax compliance and support for expanded taxation among ideological and political opponents is not in line with the literature's predictions,

[9]On the enabling role of citizen trust in government for state capacity formation, see Besley and Dray (2024).

similar "backfiring" effects have been documented for other common policy interventions. These include anti-corruption campaigns (Cheeseman and Peiffer, 2022), interventions to correct political misperceptions (Nyhan and Reifler, 2010), and nudges placed on tax bills (Castro and Scartascini, 2015; De Neve et al., 2021).[10]

Third, we contribute to an emerging literature on e-government and the use of technology in public administration. Whereas earlier research documented the potential of digital technology in facilitating and monitoring tax compliance (Okunogbe and Santoro, 2023; Brockmeyer and Sáenz Somarriba, 2022; Okunogbe and Tourek, 2024), we show that phone-based Digital Town Halls increase support for authorities seeking to expand the state. Our findings also emphasize how WhatsApp, a messenger service that figures prominently in discussions of mis- and disinformation (Badrinathan, 2021; Garimella and Eckles, 2020), can be used as a platform for citizen engagement. Here, we also contribute to the literature on Town Halls: Whereas earlier studies almost exclusively examined offline, in-person Town Halls, this study speaks to the role new digital technology and increasing internet availability can play in engaging citizens in deliberation and decision-making.

Finally, our results have important implications for policy and practice. The persistence of low capacity states is widely seen as a critical development challenge in many contemporary democracies (Prichard et al., 2019). In theory, engaging citizens via participatory budgeting could enable governments in weak states to escape low legitimacy-low compliance traps and reach higher levels of fiscal exchange. The participatory budgeting program that we study does not, however, increase short-term average tax compliance among participants. Yet, the total impact of a similar intervention implemented by a government may be different than the one we evaluate for several reasons. First, the impact of the intervention may spill over to those who did not directly participate, a possibility for which we find suggestive evidence (Appendix C.5). Second, our sub-group analyses

[10] Cheeseman and Peiffer (2022) find that anti-corruption campaigns increase bribe payments for those who initially believed that corruption was widespread. In a population-wide tax messaging experiment in Belgium, De Neve et al. (2021) report that tax morale messages have negative treatment effects for certain sub-groups.

suggests that tax compliance could be improved by targeting Town Hall invitations to those most likely to react positively (e.g., tax supporters). Finally, we note that the intervention's clear positive effect on government legitimacy might allow governments to undertake reform (e.g., increase enforcement measures) that would otherwise be politically infeasible. Overall, governments in democracies with weak states can use participatory budgeting to become more legitimate during politically costly reform efforts to increase state capacity.

4.1 Intervention

This research takes place in cooperation with the Freetown City Council (FCC) in the context of a city-wide property tax reform two of us helped lead. The reform served to broaden the tax base—less than 50% of the approximately 120,000 properties had been registered previously in the property cadastre—and to make the tax burden more equitable through the introduction of a more nuanced, consistent and transparent property valuation scheme (Grieco et al., 2019).

The mayor publicly announced that Digital Town Halls (DTHs) would be held starting in January of 2021. In her messaging, she emphasized that these DTHs would be key for securing citizen participation. She also stressed that she intended to institutionalize the DTHs with future DTHs being assigned 20% of property taxes raised in a given ward (FCC, 2021, pg. 26).

In this study, the DTHs serve as part of a broader intervention that contained three components: (i) DTHs, (ii) service delivery, and (iii) notification calls about delivered services. While only the treatment group was invited to participate in the DTHs, the projects implemented are *public* services and thus available to members of both treatment and control groups. However, only the treatment group received a phone call informing them that the selected service had been delivered. This implies that the estimand in our primary analysis is the effect of participating in a DTH plus having received a notification call, conditional on services being delivered.

4.1 Digital Town Halls

In this study, DTHs take the form of WhatsApp group chats where property owners had the opportunity to engage other property owners in their ward and their political representatives about pressing development challenges facing their community.[11] The overarching goal of the DTH was for the group to deliberate and decide over how a budget of 15 million leones (about USD 1,500) should be spent in their ward. Treated participants were assigned to one of 58 ward-specific chat groups, with group sizes ranging from 17 to 37 (median: 24). The DTHs comprised four distinct phases:

1. **Horizontal Deliberation (January 15-19, 2021):** Participants received introductory videos from the Mayor of Freetown and their respective ward councilor.[12] In this first video, the mayor welcomes participants to the DTH and makes an explicit connection between tax payments and service delivery:

   > Welcome Freetownians to our Digital Town Hall meeting. I am happy to be with you here today to talk with you about why we have this meeting. As we all know—and you all know more than I do—Freetown City Council works to deliver services to the residents of the city. But how can we do it?. . .We the Freetownians need to put our money—through the local tax, through market dues, and of course through our property rate—then we the FCC will take that money to deliver services to the people of Freetown.

   In these videos, political representatives also invited participants to discuss development concerns within the group. In the group chats, group moderators introduced themselves to participants and then asked participants to respond to the following prompt: "What do you

[11] We completed a pilot DTH in one ward before scaling the DTHs up to our 30 study wards. In Appendix C.1 we lay out potential advantages and disadvantages of *Digital* Town Halls vis-à-vis in-person Town Halls.

[12] Videos from political representatives were shared with DTH participants in two ways: videos were posted directly in the WhatsApp group and were available via a Qualtrics link, also posted in chat groups.

think is the greatest development problem in your ward?"[13] Across all groups, participants sent approx. 2,000 messages.[14] This phase was purely horizontal because participants were told that political representatives would not be involved and could not observe or learn about what was discussed in this phase.

2. **Preference Articulation and Aggregation (January 20-February 12, 2021)**

   After five days of horizontal deliberation, DTH participants received a video from the Mayor of Freetown asking participants to (i) identify the two greatest development challenges in their ward and (ii) propose a plan to tackle them.[15] Participants were told that they should think of projects that could be completed for US$1,500 and instructed to submit either a written message or a short voice recording. Participants were told that the team facilitating the DTHs would listen to their messages, aggregate information, and present both the mayor and ward councilor with memos outlining the concerns and proposed solutions of their constituents. In doing so, the DTHs allowed participants to (anonymously) articulate their preferences to their representatives. Through this process, it was clear that water access was the most pressing concern facing many communities. This phase was the first time that participants' statements were shared with political representatives. Some common service preferences include:

   - Water tanks should be provided to communities
   - Water wells should receive a hand pump
   - Water pipes that are broken or inoperable should be fixed
   - Pave roads, with emphasis on access roads for vehicles

[13] The research team hired a local team to act as moderators, supervised and managed by project research assistants. DTH facilitators requested that participants only use the chat between 7am and 10pm daily to ensure that a facilitator can be present at all times. Participants are free to choose the form in which they would like to participate (text/ voice/ video messages), but were asked to contribute in Krio or English.

[14] Most messages were text (55.25%) and voice messages (40.2%). Others included images and videos.

[15] Also in this video, the mayor once again links taxes to service provision.

> I want everyone to remember this: 15 million leones is on the table now, so that we can do something small in your community. But if everyone pays their property rate, you can imagine what type of investment we can make in your ward. So we look forward to you giving your statement [about what project you want]. But most importantly, let us all join hands, let us all pay our property rate...

- Dustbins should be provided, maintained, and collected regularly
- Drainage systems should be upgraded to prevent flooding in rainy season
- Public toilets should be installed at strategic locations
- Street lights should be provided to curtail the crime rates and for road safety
- Local market should be rehabilitated with emphasis on enforcing the zinc roofs

3. **Vertical Interaction (February 13-16, 2021):** Participants received videos from both their councilor and the Mayor. Councilors acknowledged input they received from participants, based on the ward-specific summaries prepared by the research team, and positioned themselves to the demands made.[16] This includes highlighting their preferred services, justifications for their service preferences and explaining past and future delivery goals. In her video, the Mayor also acknowledged and responded to participants' demands.[17] In their videos, the Mayor and councilors told respondents that an engineering firm had been assessing the feasibility of their proposed projects and that five projects had been determined as feasible within the budget:
   - Two new solar street lights
   - Fix some potholes
   - 50m of truck tracks
   - Fix some GUMA water pipes
   - Install a new water hand pump

   Participants were also informed that voting would start in four days. Moderators reminded participants that, "This is your chance to make your case to the other people in this WhatsApp group about what to prioritize before everyone places their final vote."

4. **Decision Making (February 17-22, 2021):**

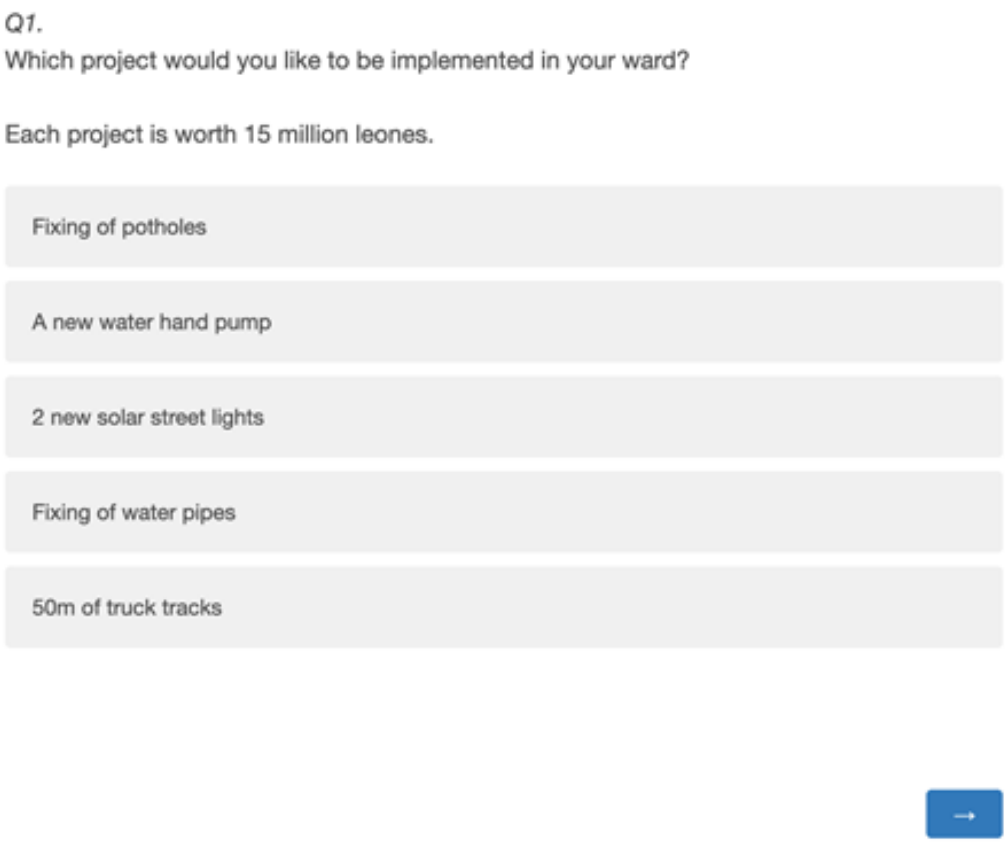
   In this phase, participants cast their vote for the project they wanted to be implemented in

[16]Where a ward held two DTHs, the councilor prepared a video that addressed concerns raised in both DTHs.

[17]We opted for this mediated interaction between citizens and representatives to (i) avoid elite domination of the TH process and (ii) make realistic time demands on representatives.

their ward. This vote could be cast anonymously through a Qualtrics survey (Figure 4.1) and a "how to" video was posted in each group that provided step-by-step instructions of the voting process.[18] Voting was open for four days and was closed on February 22. The mayor announced the winning project for each ward with a ward-specific voice message, which was posted in each DTH alongside a picture of the Mayor in office.[19] After the announcement of the winning projects, group moderators thanked participants for their contributions and then halted participants' ability to post messages in the DTHs. Moderators explained that chats would be used one more time in the future to announce that delivery of the service has concluded.

Figure 4.1: Menu of Services



There was active participation in the Digital Town Halls. We confirmed that 1,457 of the 1809

[18] We also gave participants the option to inform moderators of their vote in bilateral conversation.

[19] In this voice note, the Mayor thanked everyone for their participation and repeated the promise that next year, 20% of the property tax revenues collected in the ward will be made available to the next round of DTHs.

treated property owners joined the DTHs, a compliance rate of 80.5%. The majority of individuals who joined reported that they accessed the DTH daily (54%) and 84.3% reported they accessed it more than once per week.[20] Roughly two-thirds of those who joined voted for their preferred service (68%) and posted at least one message (63%).[21]

Participants also reported that the DTHs were useful and safe spaces for exchanging views with representatives and community members. On average, participating respondents agreed that the DTHs allowed them to "let my political representatives know about my views" (3.94/5) and "better understand views from fellow members of my community" (4.04/5). In addition, respondents on average agreed that "participants felt comfortable to make their views known even when their views differed from those of other participants" (3.82/5). However, respondents were less convinced that the DTH budget was sufficient to improve the delivery of the selected service in a meaningful way (2.86/5) and had middling feelings about the comprehensiveness of the list of services they could vote for (3.33/5) (see Appendix Table C.3).

While the service delivery budget was not drawn from the FCC's regular revenue, this was not communicated to project participants, allowing the Mayor and councilors to claim full credit for the participatory budgeting program and associated service provision.[22] Respondents overwhelmingly reported they believed that the FCC organized the DTH (89%), implemented services (96%), and funded the services (84%).[23] Of the respondents who said the FCC funded the project, 87% thought

[20]Only 5% of respondents who joined reported they never accessed the group and another 5% reported they accessed the DTH only once.

[21]Note that 25 people who did not join the DTH also voted, as we reached out to treated participants bilaterally, and are therefore included in the denominator. The statistics regarding messages include all message formats. The median participant sent out two messages and the mean number of messages sent by participants is just under four. The median number of messages posted per DTH was 70, about evenly split across text and voice messages.

[22]The budget allocated to the DTH did not come from the FCC's regular budget because of (1) the severity of the budget constraint the FCC faced and (2) that property tax revenue would be accrued after the DTHs had taken place. For these reasons, the funds to be decided over were taken from the project's research budget.

[23]See Appendix Table C.2. We asked participants which actors they believed were responsible for organizing, implementing, and funding the DTHs. For these questions, respondents were allowed to name multiple actors they thought might be involved. The next most frequently named actor involved for each activity is as follows: organizing the DTH was "researchers" (12%); for service implementation, central government (4.5%); as for funding, 11% of respondents thought that projects were at least partially funded by private citizens, either inside or outside their ward and 11% thought that the projects were funded by central government.

it was funded through taxes (either from inside or outside the ward), 6% from government transfer, 4% from development partners, and 3% from foreign aid.

4.1 Service Delivery

Each participating ward received a service project—essentially a local public good both treated and control units in that ward could profit from.[24] Construction began in most wards in October 2021 and was completed in all but one ward by the end of the year.[25] In the remaining ward, construction was completed in February 2022.[26]

Despite the multiple rounds of feasibility assessments conducted by the engineering firm to determine the list of feasible projects, at the time of project implementation the engineering firm concluded that some selected projects could not be effectively delivered within some wards. Specifically, we were not able to fix water pipes in wards that selected this service. Of the 19 wards that voted to fix water pipes, we built street taps in 11 of them. In the remaining 8 wards, water street taps were deemed infeasible; in these cases we opted to provide a 5000L Milla Tank instead. This was an item participants could not directly vote for, but one that reflected their preference for improved water provision in affected wards. In addition, one ward voted to "fix pot holes", but implementation proved difficult. Instead, this ward received 50 meters of truck tracks. Appendix Table C.1 summarizes the votes received for each project, the number of selected projects across wards, and the projects that were eventually implemented. Even with these challenges, participants reported they were satisfied with the selected services both after the DTH (4.6/5, midline survey) and after the implementation of the projects (4.2/5, endline survey).

We found it plausible that not all DTH participants would be aware that the projects they selected in the DTH had been successfully implemented.[27] To address this, we made notification calls on

[24] Pictures of implemented projects can be found in Appendix C.2.2.

[25] Implementation was scheduled to start in May 2021, after the midline survey, but was delayed due to negotiations with the delivery firm as well as the complexity of identifying appropriate implementation sites.

[26] In Tengbeh Town, construction was delayed because the implementing construction company wanted additional assurances from the FCC regarding potential liability issues.

[27] See Khan et al. (2022). We also worried that our inability to observe respondents' knowledge of project imple-

behalf of the FCC to all treated units, informing participants of successful implementation. Note that by making these notification calls to treated units but not control, we build the notifications calls into our treatment.[28] We successfully reached approximately 70% of treated units to inform them of the implemented services. These calls started in mid-November and were staggered across wards so that they started once service delivery was completed in that ward. The endline survey similarly was staggered and commenced after notification calls were completed, but never earlier than one week after delivery completion.

4.2 Data Collection

4.2 Sampling

We constructed a sampling frame using FCC administrative records, which contain information about property characteristics and property owner contact details. To be eligible to participate in the Digital Town Hall a property owner must (i) own a property in one of the 30 study wards and (ii) have WhatsApp on their phone.[29] We called 15,977 property owners in the 30 study wards and verified that 4,860 had WhatsApp on one of their phones; these property owners were eligible to be selected for the Digital Town Hall intervention. However, the set of 15,977 property owners we called was *not* a random sample of property owners from the 30 study wards. First, we only attempted to call property owners with above median property values because a COVID-19-related policy in place at the time of these calls waived property tax for below median properties.[30] Second, we removed some properties from the sample frame in an attempt to limit geographic spillovers.[31]

mentation would complicate the interpretation of our findings. For example, to what extent should a null (or perverse) effect be attributed to respondents' (mistaken) belief that selected services had not been implemented?

[28]Thus, treatment effects include the heightened awareness of treated units regarding service implementation.

[29]For property owners that own multiple properties, we coded them as being exclusively eligible for the DTH in the study ward that contains their highest-value property (i.e., highest tax fee).

[30]As a response to COVID-19, the FCC intended to waive property tax for 2020 on properties of below median value. As our intervention was originally scheduled for early 2020, it was necessary to target the DTH intervention at individuals owning properties above the median property value. Politics related to the tax reform caused us to delay the DTH intervention until early 2021. During the calling process, we unintentionally verified 450 individuals owning property below the median value. We included these in our sample.

[31]In a previous version of our research design, we planned to allocate treatment status using a two-stage randomization procedure, to mitigate and estimate geographic spillover (as in Sinclair et al., 2012). Under that research design,

Third, note that we could not contact owners of properties where owner contact information was not listed in FCC records.

4.2 Measurement

We conducted three rounds of phone-based survey data collection:

Baseline: Between October 28 and December 2, 2020, we attempted to survey the 4,860 property owners we had verified as eligible for the study (see Section 4.2.1) and completed baseline surveys with 3,859 individuals (79.4%). Only baseline survey respondents were eligible to receive treatment and were attempted to be surveyed in subsequent rounds.[32]

Midline: After the completion of the DTHs (between March 4 and April 17, 2021) we conducted midline surveys with all study property owners. Importantly, this survey round took place *before* services were implemented. We completed midline surveys with 3,304 study property owners (91.3%).[33]

Endline: After the implementation of the selected services (between November 11, 2021 and January 2022) we conducted endline surveys with all study property owners. We completed endline surveys with 2,872 study property owners (79.4%).

For our measure of tax compliance, we rely on FCC administration data, which allows us to observe individual level tax compliance behavior for the universe of taxable properties in Freetown. Our preregistered measure of tax compliance is a dummy variable equal to 1 if a property owner makes any tax payment in 2022. The control group compliance rate is 29.1% and 31.5% in 2022 and 2021, respectively.

properties were divided into geographic clusters using a grid overlay and properties within five meters of the edge of a grid cell were ineligible for the study. We constructed the call list with this research design in mind, thereby removing properties within five meters of the grid cell edge. While we eventually moved on from this research design, sampling was done with that design in mind.

[32] Appendix Figure C.1 documents the broader data collection and project timeline.

[33] We incentivized midline and endline survey responses by offering packages of mobile data.

4.3 Treatment Assignment and Balance

Before assigning treatment, in an effort to mitigate spillover, we drew a restricted sample from the set of 3,859 eligible property owners with whom we had completed a baseline survey such that each property is at least 15 meters from the closest study property. The restricted sampling left us with a final sample of 3618.[34] We then assigned treatment status using a matched-pair design, leveraging baseline survey data to match similar observations into groups of two.[35] In this matching process we weigh certain variables higher than others, in line with our expectations that they more strongly predict our outcomes of interest. We placed the greatest weight on our measure of unconditional tax morale—considering its common use in the literature as a proxy for tax compliance behavior. Appendix C.3 describes the matching procedure in greater detail. We created 1809 pairs and then assigned one observation in each matched-pair to treatment and the other to control. Appendix Figure C.5 visualizes treatment assignment across Freetown.[36]

Table 4.1 presents balance statistics across a range of (baseline) survey outcomes, demographic covariates, and property characteristics. Observed differences between groups for these variables are no more than we might expect. Given the 29 tests we run, under the null hypothesis of no differences between groups, we would expect 2.9 tests to appear significant at the 90% confidence level. We find only two significant differences: at baseline, the treated group is *less* likely to perceive the FCC as corrupt (*FCC corruption*) and less likely to report they are politically unaffiliated (*No affiliation*). Our preregistered specification for survey-based outcomes includes the baseline measure of the dependent variable. Therefore, when estimating treatment effects on perceptions of corruption we control for the baseline measure in our standard model.

[34] Appendix Figure C.6 shows the distribution of the distance from each property to the closest property in the sample.

[35] We generated matched-pairs using the *blockTools* package in *R*. We use the Optimal Greedy ("optGreedy") matching algorithm to find best matches along mahalanobis distance.

[36] We implemented this randomization in R using the *block_ra* function in the *randomizr* package.

Table 4.1: Balance Table

	Mean		SD	Difference			Observations	
Measure	C	T1	C	Raw	Std.	*p*-val	C	T1
Survey Outcomes								
Opportunities for voice	2.12	2.13	1.00	0.01	0.01	0.75	1,719	1,736
Ease of participating in political activities	1.76	1.74	1.14	-0.02	-0.02	0.62	1,794	1,793
FCC responsiveness to citizens' demands	3.17	3.17	1.18	0.00	0.00	0.91	1,712	1,719
Satisfaction with FCC service provision	3.64	3.64	1.17	0.00	0.00	0.96	1,790	1,796
FCC transparency	1.37	1.35	0.69	-0.02	-0.03	0.34	1,732	1,726
FCC efficiency	2.86	2.87	0.70	0.01	0.01	0.77	1,530	1,577
FCC corruption	3.50	3.57	1.01	0.07	0.07*	0.06	1,481	1,482
Mayor approval	4.23	4.22	0.89	-0.01	-0.01	0.76	1,770	1,774
Councilor Approval	2.73	2.74	1.22	0.01	0.01	0.90	1,751	1,751
Willingness to pay more taxes for better services	4.19	4.18	1.22	-0.01	-0.01	0.78	1,805	1,804
Reform improves tax system fairness	2.12	2.11	0.79	-0.01	-0.01	0.83	1,112	1,129
Number of neighbors who will pay property tax	5.13	5.07	2.41	-0.06	-0.02	0.54	1,138	1,105
Likelihood detected noncompliers are punished	4.06	4.06	1.11	0.00	0.00	0.90	1,788	1,781
Political Party Affiliation								
APC	0.24	0.25	0.43	0.01	0.02	0.59	1,809	1,809
SLPP	0.20	0.20	0.40	0.00	0.00	0.62	1,809	1,809
Other party	0.02	0.03	0.16	0.01	0.06	0.36	1,809	1,809
No affiliation	0.32	0.29	0.47	-0.03	-0.06*	0.03	1,809	1,809
Did not respond	0.22	0.23	0.41	0.01	0.02	0.34	1,809	1,809
Property Characteristics								
Tax compliance 2020	0.07	0.07	0.25	0.00	0.00	0.74	1,809	1,809
Number of properties with tax liability (2021)	1.93	1.89	1.48	-0.04	-0.03	0.37	1,809	1,809
Total property tax owed (USD, 2021)	95.83	93.15	175.59	-2.68	-0.02	0.66	1,809	1,809
Received tax bill (2019 or 2020)	0.80	0.80	0.40	0.00	0.00	0.89	1,791	1,789
Property has water	0.47	0.47	0.50	0.00	0.00	1.00	1,809	1,809
Property has drainage	0.36	0.36	0.48	0.00	0.00	0.81	1,809	1,809
In informal settlement	0.06	0.06	0.23	0.00	0.00	1.00	1,809	1,809
Demographics								
Female	0.31	0.30	0.46	-0.01	-0.02	0.91	1,809	1,809
Age	51.65	51.88	13.00	0.23	0.02	0.60	1,803	1,804
Higher education	0.39	0.40	0.49	0.01	0.02	0.32	1,685	1,694
Married	0.72	0.72	0.45	0.00	0.00	0.68	1,804	1,805

Table 4.1 reports balance across baseline survey outcomes, immutable demographic covariates, and property characteristics. Columns 1-2 report group means; Column 3 reports the control group standard deviation; Columns 4-5 report raw and standardized differences, respectively. Column 6 reports the *p*-value on this difference (not adjusted for multiple comparisons). We convert local currency (SLL) to USD at a rate of 10,000:1, which reflects the exchange rate in January, 2021. A respondent is coded as receiving higher education if they have a university degree, or a degree from a polytechnic school or teacher college. Receiving a tax bill in 2019 and 2020 is self-reported.
Significance: * $p < 0.10$

4.4 Results

The nature of our intervention allows for one-sided noncompliance as property owners must voluntarily join the DTH groups. Of the 1809 property owners assigned to treatment, 1,459 (80.7%) joined WhatsApp groups of the DTH. While Intent-to-Treat (ITT) estimators provide unbiased estimates of being assigned to treatment, the presence of one-sided noncompliance means they will underestimate the effect of *joining* the DTH. Therefore, we estimate the effect of a property owner joining the DTH using an instrumental variable regression framework. Our main equation is:

$$Y_{ijt_2} = \alpha_1 DTH_i + \gamma Y_{ijt_1} + \sum_{j=1}^{1809} \theta_j PAIR_{ji} + \delta_w + \lambda \mathbf{X}_i + \epsilon_i \tag{4.1}$$

Where Y_{ijt_2} is the endline (t_2) outcome of individual *i* in pair *j*; DTH_i is an indicator variable equal to 1 if owner i *joined* the DTH. Y_{ijt_1} is the baseline outcome for owner *i* in pair *j*. When Y is property tax compliance behavior, Y_{t_1} refers to tax compliance behavior in 2020; When Y is a survey outcome, Y_{t_1} refers to the baseline survey outcome. $PAIR_j$ is an indicator variable equal to 1 if owner *i* belongs to pair *j*; **X** is a set of preregistered property-level characteristics that we include for covariate adjustment only when Y is property tax compliance behavior.[37] δ is a vector of ward fixed effects and ϵ_i is the error term.

Using two-stage least squares (2SLS), we jointly estimate:

$$DTH_{ij} = \beta_1 D_i + \eta Y_{ijt_1} + \sum_{j=1}^{1809} \mu_j PAIR_{ji} + \zeta_w + \xi \mathbf{X}_i + \nu_i \tag{4.2}$$

Where D_i is the randomly assigned treatment indicator, which instruments for DTH_i in equation

[37] Preregistered control variables include: (i) log total tax liability, (ii) number of properties with any liability, (iii) access to water, (iv) access to drainage, (v) property in an informal settlement, (vi) property has fencing or gate, (vii) property has garage, (viii) street condition, (ix) street type (x) ease of property access, (xi) window quality, (xii) type of tax bill received. Where covariate data is missing, including baseline values of the outcome, we impute missing data using the baseline mean of that variable. Note that Equation 4.1 controls for survey-based outcomes that we expect to predict compliance and survey outcomes through the inclusion of block dummies.

4.1.[38] Our quantity of interest is α_1 (equation 4.1), which captures the local ATE among the set of individuals who comply with treatment—property owners who joined the DTH. We report estimates with heteroskedasticity-robust standard errors (HC2). As randomization occurs at the level of the observation (property owner), we do not cluster standard errors.

4.4 Legitimacy

We measured four components of standard conceptualizations. First, governments are more legitimate when citizens can influence policy (Levi et al., 2009; Scharpf, 1997). The DTHs were designed to increase citizens' participation in political affairs and their control over political decision-making. We find that the intervention increased participants' perception they could influence policy (Table 4.2). Specifically, we preregistered two indicators and predicted that the treatment should increase respondents' perceptions (1) that they have opportunities to voice their opinions about government matters to government officials and (2) that it is easy to directly engage in political activities. The intervention had large and durable effects on this first indicator, increasing reported *opportunities for voice* by 0.38 standard deviation units (SDUs) at the midline survey and 0.25 SDUs at endline. Given the baseline standard deviation is roughly 1.00, these effect sizes can be interpreted as changes on a 4-point Likert scale. The effect on *ease of participating in political activities* is positive at both midline and endline, with statistical significance at the threshold of conventional levels.[39] Taken together, we interpret these results as strong evidence that the intervention produces durable impacts on participants' perceptions that they can influence policy.

Second, legitimate governments provide citizens with basic services and generally respond to citizens' demands. After citizens voiced their concerns about local development and their preferred services, ward councilors used the DTHs to respond to citizens' concerns. We find that the intervention increased citizens' perceptions that the local government was more *responsive to citizens' needs and demands*. Treated participants reported higher levels of local government responsive-

[38] We estimate these equations using the *iv_robust* package in R.
[39] Conventional p-values are 0.10 and 0.11 at midline and endline, respectively.

ness relative to control respondents, a difference that is statistically significant both directly after the DTH (midline; p-value <0.001) and after service implementation (endline; p-value = 0.016). In addition, the intervention attempted to forge the social and fiscal contract between citizens and politicians by delivering local services that people demanded. We find that the intervention increased citizens' *satisfaction with FCC service provision* at both midline (p-value <0.0001) and endline (p-value = 0.004).

Third, legitimate governments administer their constituencies competently. Our survey data show that, before the intervention, respondents perceived the FCC as fairly incompetent: the average respondent perceives the FCC as *not* transparent (1.36/3), of middling efficiency (2.86/4), and not too corrupt (3.53/5).[40] The intervention, however, improves respondents' perceptions of FCC administrative competence. At baseline, 76% of respondents said it was "difficult" to find out how the FCC spends tax revenue. The intervention improves respondents' perceptions of transparency. While there is a small effect at the threshold of statistical significance at midline (0.085 SDU; p-value = 0.104), that effect balloons by almost a factor of four at endline (0.319 SDU; p-value = 0.0017). At baseline, the modal respondent (62%) reported that the FCC is "somewhat efficient" in the way it uses money for public administration and development. While estimated treatment effects at midline are statistically indistinguishable from zero (p-value = 0.32), at endline we find unambiguous evidence that the intervention increases perceptions of the FCC's efficiency (p-value = 0.0078). For perceptions of corruption, we find a similar, though more extreme, change in treatment effect between midline and endline. At midline, treatment *increases* participants' perceptions that the FCC is corrupt (p-value = 0.001). However, after services have been implemented, treated participants *positively* update their views of FCC corruption, relative to control (p-value = 0.071).

The reversal of the sign on the effect of perceived corruption can be explained with an intuitive logic: in the DTHs, citizens learn of the existence of a development fund for each neighborhood, which is an additional pocket of money prime for corrupt misuse; participants negatively update

[40]For each measure higher scores indicate better performance.

about corruption at the FCC. However, when citizens see these funds being efficiently used to implement development projects, this example of clean governance causes them to update again, such that the FCC is *less* corrupt than they had originally thought. More generally, these three results suggest that the government needs to deliver on its promises of implementing development projects before citizens fully update their attitudes about government administrative competence. Citizens understand that talk is cheap; they respond to action.

Fourth, we measured participants' approval of the performance of local elected officials. Specifically, we asked respondents if they approved or disapproved of how both the Mayor and their ward councilor have performed on the job over the past twelve months. Our data show that the Mayor is popular at baseline: most respondents report they either "strongly approved" (43.4%) or "approved" (44.3%) of the mayor's performance. The intervention increases approval of the Mayor by 0.15 SDUs (p-value $<$0.001) at baseline and 0.19 SDUs (p-value $<$0.001) at endline.[41] Respondents reported much lower approval ratings for their ward councilors at baseline: the modal respondent (41%) "disapproved" of their councilor's performance over the past year. While baseline approval for councilors was low, the intervention increased approval at both midline (0.19 SDUs; p-value $<$0.0001) and endline (0.17 SDUs; p-value $<$0.001). In summary, Table 4.2 provides unambiguous evidence that the intervention increases perceptions of government legitimacy. In the next section, we investigate whether this shift in legitimating beliefs led to a corresponding shift in tax compliance behavior, as would be predicted by the literature (e.g., Levi, 1988).

[41] That we observe these effects is particularly impressive given that (at baseline) 44% of the sample gave maximum approval ratings.

Table 4.2: Effect on Legitimacy Outcomes

	Baseline	Midline			Endline		
Outcome	Mean	Mean	Effect	N	Mean	Effect	N
Political Efficacy							
Opportunities for voice	2.126	2.331	0.377***	3,288	2.161	0.251***	2,849
	(0.995)	(0.920)	(0.038)		(0.922)	(0.046)	
Ease of participating in political activities	1.749	1.623	0.064	3,298	1.631	0.073	2,863
	(1.137)	(1.022)	(0.040)		(1.022)	(0.046)	
Service Delivery and Responsiveness							
FCC responsiveness to citizens' demands	3.172	3.356	0.141***	3,251	3.308	0.116**	2,830
	(1.186)	(1.061)	(0.038)		(1.135)	(0.048)	
Satisfaction with FCC service provision	3.643	3.612	0.182***	3,302	3.471	0.146***	2,864
	(1.168)	(1.059)	(0.040)		(1.213)	(0.050)	
Perceptions of Government Process							
FCC transparency	1.360	1.423	0.085	3,288	2.163	0.319***	2,834
	(0.686)	(0.772)	(0.052)		(1.339)	(0.101)	
FCC efficiency	2.864	2.858	0.037	3,233	2.791	0.129***	2,791
	(0.707)	(0.563)	(0.038)		(0.703)	(0.048)	
FCC corruption	3.532	3.623	-0.141***	3,177	3.454	0.087*	2,736
	(0.997)	(0.897)	(0.043)		(0.928)	(0.048)	
Approval of Political Representatives							
Mayor approval	4.226	4.084	0.149***	3,296	3.907	0.194***	2,855
	(0.888)	(0.815)	(0.042)		(0.937)	(0.051)	
Councilor Approval	2.734	2.730	0.193***	3,278	2.744	0.171***	2,841
	(1.221)	(1.167)	(0.040)		(1.217)	(0.047)	

Table 4.2 reports the effect of the treatment on political attitudes. Columns 1, 2, and 5 reports the control group mean for each indicator at baseline, midline, and endline, respectively, with the standard deviation in parentheses; Column 3 presents treatment effects estimates at the midline survey and Column 6 presents treatment effects estimates at the endline survey. Column 4 and 7 report the number of non-missing observations in the midline survey and endline survey, respectively. Reported effects are standardized effects.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.4 Tax Compliance

Panel A of Table 4.3 reports treatment effects on tax compliance in 2022 (main outcome of interest) and 2021. When interpreting treatment effects recall that, in 2021, the DTH started at the beginning of the tax season and the tax deadline came before any services were delivered. By contrast, the 2022 tax season came after all of the services selected in the DTH were implemented. In that

sense, only 2022 provides a test of the full treatment of participation and service delivery, which is our pre-registered primary outcome of interest. We observe the complete compliance behavior for all units.

Column 1 reports the control group mean compliance rate in 2021 and 2022 and Column 2 reports the effect of the intervention. Focusing first on 2022, the compliance rate in the control group is 29.1%. The estimated treatment effect in 2022 is negative 1.2 percentage points, an effect that is statistically indistinguishable from zero with a *p*-value of 0.5. In 2021, the point estimate on the treatment effect is again negative (-0.78 percentage points) and statistically indistinguishable from zero (*p*-value = 0.72).

Table 4.3: Effect on Tax Compliance

Outcome	Mean	Effect	*p*-value	N
Panel A: Tax Compliance Behavior				
Did the owner pay any taxes?				
2022	0.291	-0.012	0.496	3,618
		(0.018)		
2021	0.315	-0.007	0.723	3,618
		(0.019)		
Panel B: Fiscal Exchange Attitudes				
Willingness to pay more taxes for better services				
Midline	4.001	0.066	0.163	3,296
	(1.253)	(0.047)		
Endline	4.030	-0.075	0.155	2,872
	(1.293)	(0.053)		

Table 4.3 reports treatment effects on tax compliance behavior in 2021 and 2022 (Panel A) and attitudes towards expanding taxation (Panel B). Column 1 (panel A) reports the control group means. Column 2 presents treatment effects estimates, with effects in raw percentage points for tax compliance (Panel A) and standard deviation units for tax attitudes (Panel B). Column 3 reports *p*-values and Column 4 reports the number of non-missing observations.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

These null effects are robust to different model specifications. Our main specification includes

ward-fixed effects and a set of property characteristics; results are similar when we estimate effects using (i) only the treatment indicator and 2020 (pretreatment) compliance behavior; (ii) only ward-fixed effects; (iii) only property characteristics, and (iv) when we add to our primary specification a dummy indicating the owner has zero tax liability. Results are also robust to different operationalizations of tax compliance. While Table 4.3 presents our preregistered dependent variable, which is a dummy equal to one if the owner paid *any* tax, results are robust to using the (i) the total amount paid and (ii) the log total amount paid as the dependent variable.[42]

These nulls are also precisely estimated and any real effect is likely to be small: For 2022, we can say with 95% confidence that the treatment effect is less than 2.3 percentage points. Still, we might worry that a (small) true effect exists but are insufficiently powered to detect it. We can improve statistical power by pooling compliance behavior across 2021 and 2022, thereby leveraging all of our compliance data in a single estimate. In this case, the dependent variable is the mean of compliance dummies in 2021 and 2022.[43] While the interpretation of outcome is less straightforward—the group mean compliance, pooling across years—note that this effect is causally identified. The point estimate for this effect is close to zero (-1.1pp) and is not statistically significant (p-value = 0.45). In summary, we find no evidence that the treatment, on average, impacts compliance behavior. Given the robustness of this finding and the precision of our estimates, any real impacts are likely to be substantively small.

This null result runs against our expectations and the framework of tax compliance that motivated our hypotheses. In many models of tax compliance, government legitimacy is presumed to lead to higher expected service provision from government and therefore greater willingness to pay (Levi, 1988, 1997; Besley, 2020). This trade of taxes for improved services is commonly referred to as "fiscal exchange" (e.g., Timmons, 2005). One potential explanation for the mismatch between effects on attitudes and effects on behavior is an extreme budget constraint on property owners: it

[42] See Appendix Table C.5.

[43] Such that the dependent variable is equal to 0 if they paid in neither year, 0.5 if they paid in one year, and 1 if they paid in both years.

could be that treated property owners *want* to pay more taxes, but they simply have no disposable income to do so. If this were the case, we should see positive impacts on respondents' willingness to pay more taxes for better services, which we refer to as their *attitude towards fiscal exchange*.

We measured property owners' attitudes towards fiscal exchange by asking them if they agreed or disagreed with the following statement: *I would be willing to pay additional taxes in order to receive improved services.*[44] Panel B in Figure 4.3 presents the effect of the treatment on this attitude. We do not find evidence that the intervention increases property owners' attitudes towards fiscal exchange. The estimates at endline, which directly precedes our main compliance outcome, suggest that DTH participants are no more willing to engage in fiscal exchange than control respondents—if anything, participants are *less* disposed towards fiscal exchange (p-value = 0.16). Midline estimates move in the opposite direction, though, again, these effects are not statistically significant (p-value = 0.16).

This finding also dispels another potential explanation for the mismatch between effects on attitudes and the null effect on compliance behavior. We might be concerned that the effect on attitudes is driven by experimenter demand (Zizzo, 2010), rather than true changes in beliefs. If that was the case, we should also find that treated respondents *say* they would be more willing to pay taxes, even when in fact they do not. We found no evidence of this in Table 4.3.

4.4 Political Factors Condition the Effect of Participation on Compliance

Why does the intervention fail to increase tax compliance behavior? Our preferred explanation is that participants interpret the treatment as a partisan policy pitch for expanding taxation, which in turn produces countervailing heterogeneous treatment effects conditional on participants' (1) baseline support for the policy of expanded taxation and (2) partisan affiliation. These heterogeneous effects wash out any average treatment effect.

[44] Ideally, we would measure citizens' attitudes towards fiscal exchange at the current levels of taxes and services. However, in formulating our survey instrument, we found the concept of fiscal exchange easier to communicate in the context of expanding fiscal exchange (i.e., *more* taxes for *more* services). We believe this slight mismatch is not problematic because preferences about fiscal exchange at current levels and sightly expanded levels should be similar.

The DTH intervention is an explicit attempt by politicians to persuade property owners to engage in a "fiscal exchange" by trading taxes for higher quality services. During the DTH the Mayor encouraged participants to pay taxes with the promise that "FCC will take that money to deliver services to the people of Freetown." In a separate video, the Mayor reminded participants that "if everyone pays their property rate, you can imagine what type of investment we can make in your ward."

While the majority (57.4%) of surveyed respondents reported that they "strongly approved" of a policy of expanding taxation for improved services, a significant minority opposed (14%) this idea. An established literature in psychology and political behavior suggests that it should be difficult to persuade these tax skeptics to change their policy preferences. In general, people have a motivation or goal when forming attitudes and beliefs, which are often "directional" in that they lead the receiver towards a particular conclusion (Kunda, 1990; Lodge and Taber, 2013). One typical motivation is that individuals "strive to defend and maintain their extant values, identities, and attitudes" (Slothuus and De Vreese, 2010).[45] Further, previous research indicates that attempts to persuade people that their existing beliefs are incorrect may even backfire (Nyhan and Reifler, 2010).[46] If the intervention failed, or even backfired, in its attempt to convince tax skeptics of a policy of expanded taxation, we should observe null, or even negative, treatment effects for tax skeptics on support for taxation and compliance behavior.

Plot A in Figure C.7 presents predicted marginal effects from a model that interacts treatment with our five-point measure of (baseline) attitudes towards fiscal exchange.[47] The interaction between treatment and attitudes towards fiscal exchange is statistically significant (p-value <0.001; $\beta = 0.052$). Baseline attitudes towards fiscal exchange strongly shape treatment effects on tax compliance in 2022. For respondents that strongly support expanded taxation (five on the five-

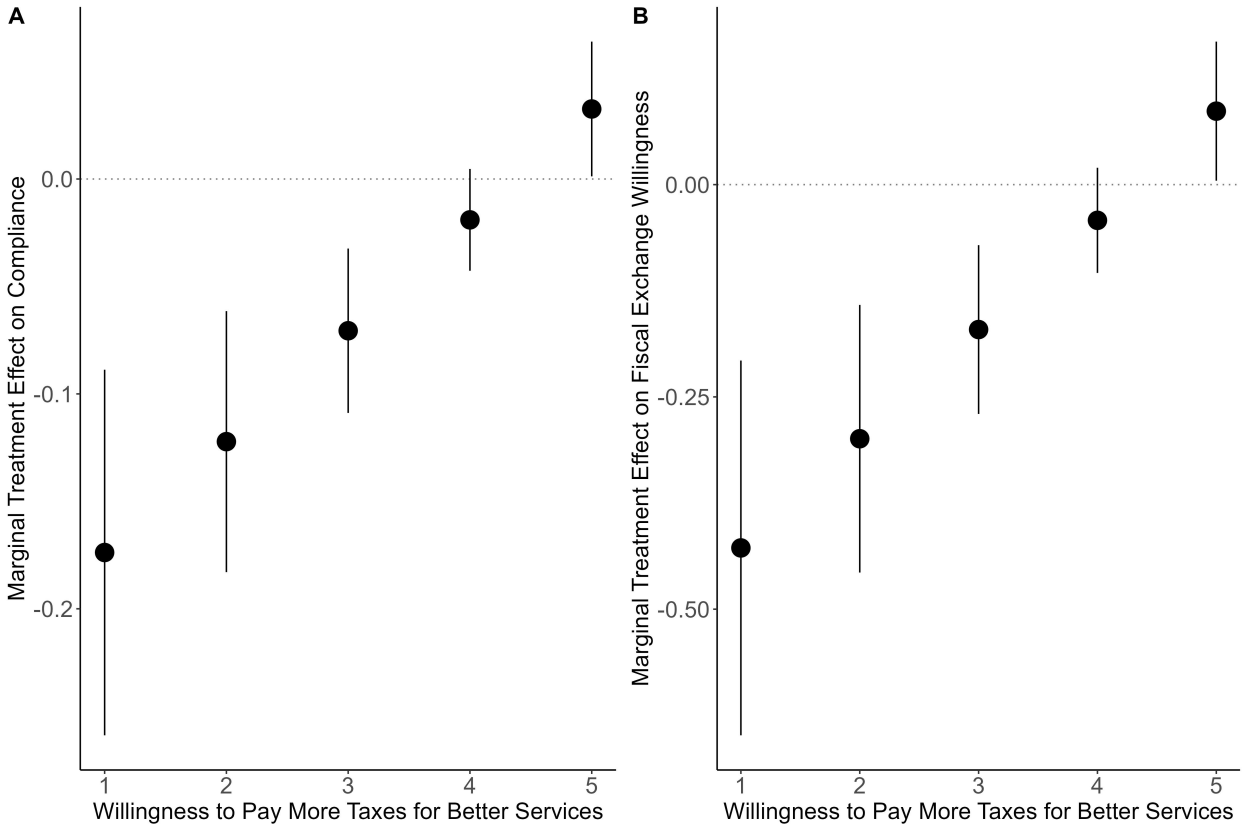
[45] Also see Stanley et al. (2020). Lodge and Taber (2013) lay out several potential mechanisms: (i) a tendency to seek out information that confirms one's prior beliefs ("confirmation bias"); (ii) giving more weight to new information that accords with existing beliefs ("prior attitude effect"); (iii) more heavily scrutinizing information contrary to existing beliefs ("Disconfirmation bias"). See also the discussion in Druckman and McGrath (2019).

[46] See Nyhan (2021) for a review of this backfiring effect.

[47] Other model specifications remain the same as in our main specification.

point scale) the estimated treatment effect is 3.31 percentage points, a 9.5% increase over the group's baseline compliance rate of 34.6%. By contrast, treatment effects are negative for property owners who did not support more fiscal exchange at baseline. For property owners who "somewhat disagreed" or "strongly disagreed" with paying more taxes for improved services, we estimate marginal treatment effects of -12.2 and -17.4 percentage points, respectively.

Figure 4.2: Marginal Treatment Effects by Attitudes Towards Fiscal Exchange



Note: Panel A reports marginal treatment effects on compliance conditional on baseline attitudes towards fiscal exchange. Panel B reports marginal treatment effects on attitudes towards fiscal exchange, conditional on baseline attitudes towards fiscal exchange.

In keeping with this logic, tax skeptics respond to treatment by hardening their opposition to expanded taxation. Plot B (Figure 4.2), presents predicted marginal effects on attitudes towards fiscal exchange from a model that interacts treatment with our five-point measure of (baseline) attitudes towards fiscal exchange. To increase power for estimating this interaction, the predicted outcome

is the respondent's *average* attitude towards fiscal exchange across midline and endline surveys.[48] The interaction term is statistically significant (p-value = 0.0016; $\beta = 0.13$ SDUs) and the pattern of marginal effects on attitudes towards fiscal exchange (Panel B) mirrors the pattern of marginal effects tax compliance behavior (Panel A). Treated property owners who were initially opposed to fiscal exchange *do not* positively update their attitudes towards fiscal exchange. In fact, we find evidence that the intervention deepens the opposition of those initially opposed to fiscal exchange. By contrast, treated respondents who support more fiscal exchange at baseline become *more* supportive relative to the control group.

A second factor that may produce heterogeneous, and countervailing, treatment effects is a participant's partisan affiliation. The property tax reform in Freetown was a highly politicized affair. The Mayor—whose All People's Congress party controlled the Freetown City Council and sat in opposition to the Sierra Leone People's Party that controlled the central government—publicly battled with the Ministry of Finance over the Freetown City Council's legal authority to adjust property tax rates without approval from central government.

It was in this politicized environment that the DTHs took place and the mayor made her direct appeals for fiscal exchange. Indeed, we find that the treatment increased participants' political engagement (Appendix Table C.6), making it more plausible that they understood the Mayor's appeal to support expanded taxation in this politicized environment. An extensive literature finds that individuals respond to elite cues by shifting their policy position closer to the position of their party (Broockman and Butler, 2017; Tappin et al., 2023; Flores et al., 2022). Moreover, while these cues may be persuasive for the political ingroup, they can generate backlash from the outgroup (Nicholson, 2012; Haas and Khadka, 2020). Both of these dynamics suggest that the Mayor's appeal for a policy of expanded taxation should increase tax compliance, and support for taxation, for co-partisans; in contrast, the treatment should decrease tax compliance, and support for expanded taxation, for non-co-partisans.

[48] Note that estimated marginal effects display similar patterns when estimated as an interaction model using midline or endline data separately. See Appendix Figure C.7.

In Figure 4.3, we see evidence that participants' partisan affiliation conditions how the treatment impacts their compliance behavior and support for expanded taxation.[49] Plot A presents predicted marginal effects from a model that interacts treatment with a co-partisan indicator variable.[50] The interaction between treatment and co-partisanship is statistically significant (p-value = 0.033; $\beta = 0.11$). For co-partisans of the Mayor (i.e., APC supporters) the treatment increases compliance by 7.4 percentage points, which is a substantial 30% increase over the group's control compliance rate of 24.4%. This effect is statistically distinguishable from zero (p-value = 0.086). In contrast, treatment effects are *negative* for non-co-partisans of the Mayor: treatment lowers compliance by 4.0 percentage points (p-value = 0.076).[51] Partisanship strongly conditions treatment effects on tax compliance in 2022.

Panel B (Figure 4.3) presents marginal effects on attitudes towards fiscal exchange from a model that interacts treatment with the co-partisanship indicator. As in Figure 4.2, the predicted outcome is the respondent's *average* attitude towards fiscal exchange across midline and endline surveys.[52] The interaction between treatment and co-partisanship is statistically significant (p-value = 0.063; $\beta = 0.223$ SDUs). In line with expectations of heterogeneous impacts of partisan cues, treated property owners who are self-reported co-partisans update their attitudes towards fiscal exchange more positively than do non-co-partisan property owners.

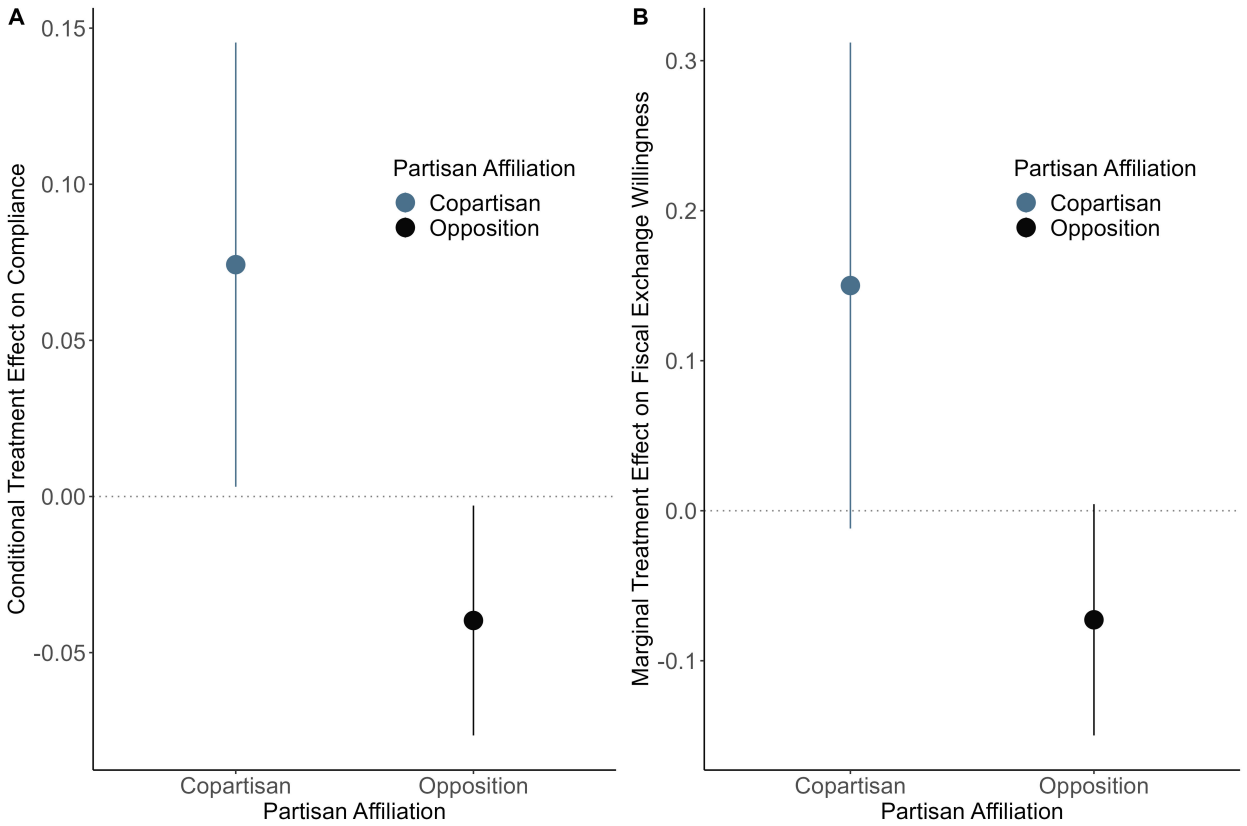
[49] In our baseline survey, we asked respondents which political party (if any) they "personally support and feel close to." Just under half of all respondents reported they had a partisan leaning (47.7%), with 24.3% and 19.9% declaring themselves for the APC (the incumbent party at FCC) and SLPP, respectively. Less than 3% of all respondents declared themselves for a party other than APC or SLPP, with the majority of third-party partisans being affiliated with the NGC. The modal respondent claimed they did not support any party (30.1%) and an additional 22.2% of respondents opted not to answer this question and are labeled as "missing."

[50] This variable respondents is equal to 1 for respondents who self-report affinity towards the All People's Congress; all other respondents are coded as 0.

[51] As noted, the interaction term in the model is statistically significant, which implies the difference in the treatment effects between the two groups is significantly different. In the results shown here, all those respondents who self-report being affiliated with the APC are coded as "co-partisan", while all other respondents are coded as "not co-partisan". Results are unchanged if we drop respondents who don't answer the question.

[52] Appendix Figure C.8 shows estimates from the interaction model using midline or endline data separately. Estimated marginal effects display similar patterns.

Figure 4.3: Treatment Effects Conditional on Partisan Affiliation



Note: Panel A reports marginal treatment effects on tax compliance behavior, conditional on partisan affiliation. Panel B reports marginal treatment effects on attitudes towards fiscal exchange, conditional on partisan affiliation. In both panels, respondents who self-report affinity towards the All People's Congress are coded as "co-partisans." All other respondents are coded as "opposition." Point estimates are presented with 90% confidence intervals.

4.4 Alternative Explanations for Conditional Effects

In Section 4.4.3 we explored why the intervention did not positively impact tax compliance (Table 4.3), despite generating unambiguously positive effects on legitimacy attitudes (Table 4.2) that are commonly thought to drive tax compliance. Our preferred explanation for the null effect on compliance emphasizes that treatment effects are conditional on baseline beliefs; we argue those beliefs shape the way people update their tax attitudes and behaviors in light of the intervention. There are, however, several alternative explanations for why the intervention fails to increase tax compliance. In this section, we explore those alternative explanations.

Does the treatment negatively affect perceived fairness or enforcement?

It is possible that the intervention had a *negative* effect on other key mediating mechanisms that were not the target of our intervention. We preregistered two other mechanism channels through which the intervention might plausibly impact compliance: (i) fairness and equity and (ii) enforcement. If the intervention *reduced* participants' perceptions of the fairness of the tax system or the likelihood noncompliers would be punished, this could have countered the positive attitudinal effects of the intervention reported in Table 4.2.

At endline, we find no evidence of persistent treatment effects on either fairness or enforcement mechanism outcomes (Table 4.4). However, at midline, treatment effects on alternative mechanisms are more varied. We see contradictory results for the fairness and equity mechanism. Before services are delivered treatment respondents believe (i) that the tax system is *more* fair and (ii) that their neighbors are *less* likely to pay, compared to respondents in the control condition. However, after services are delivered, these results both vanish towards zero. With respect to enforcement, at midline we see strong evidence that the treatment group believes they are *less* likely to be punished if they don't pay property tax, relative to control. Again, by the time services have been delivered, this difference in beliefs about enforcement disappears. In summary, while we do see short-term effects on these alternative mechanisms, we see no evidence that these effects persist after services have been delivered, which is the period that directly precedes tax compliance behavior.

Table 4.4: Effect on Alternative Mechanisms

	Baseline	Midline			Endline		
Outcome	Mean	Mean	Effect	N	Mean	Effect	N
Fairness							
Reform improves tax system fairness	2.113	2.152	0.125**	2,252	2.381	-0.005	2,852
	(0.796)	(0.691)	(0.057)		(0.782)	(0.049)	
Number of neighbors who will pay property tax	5.100	5.971	-0.209***	2,878	5.919	-0.006	2,489
	(2.381)	(2.289)	(0.052)		(2.448)	(0.060)	
Enforcement							
Likelihood detected noncompliers are punished	4.060	4.241	-0.316***	3,301	4.136	0.043	2,857
	(1.105)	(0.983)	(0.044)		(1.042)	(0.046)	

Table 4.4 reports the effect of the treatment on the alternative mechanisms of fairness and enforcement. Column 1 reports the control group mean for each indicator, with the standard deviation in parentheses; Column 2 presents treatment effects estimates at the midline survey and Column 4 presents treatment effects estimates at the endline survey. Column 3 and 5 reports the number of non-missing observations in the midline survey and endline survey, respectively. Reported effects are standardized effects.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Baseline beliefs are not related to service preferences

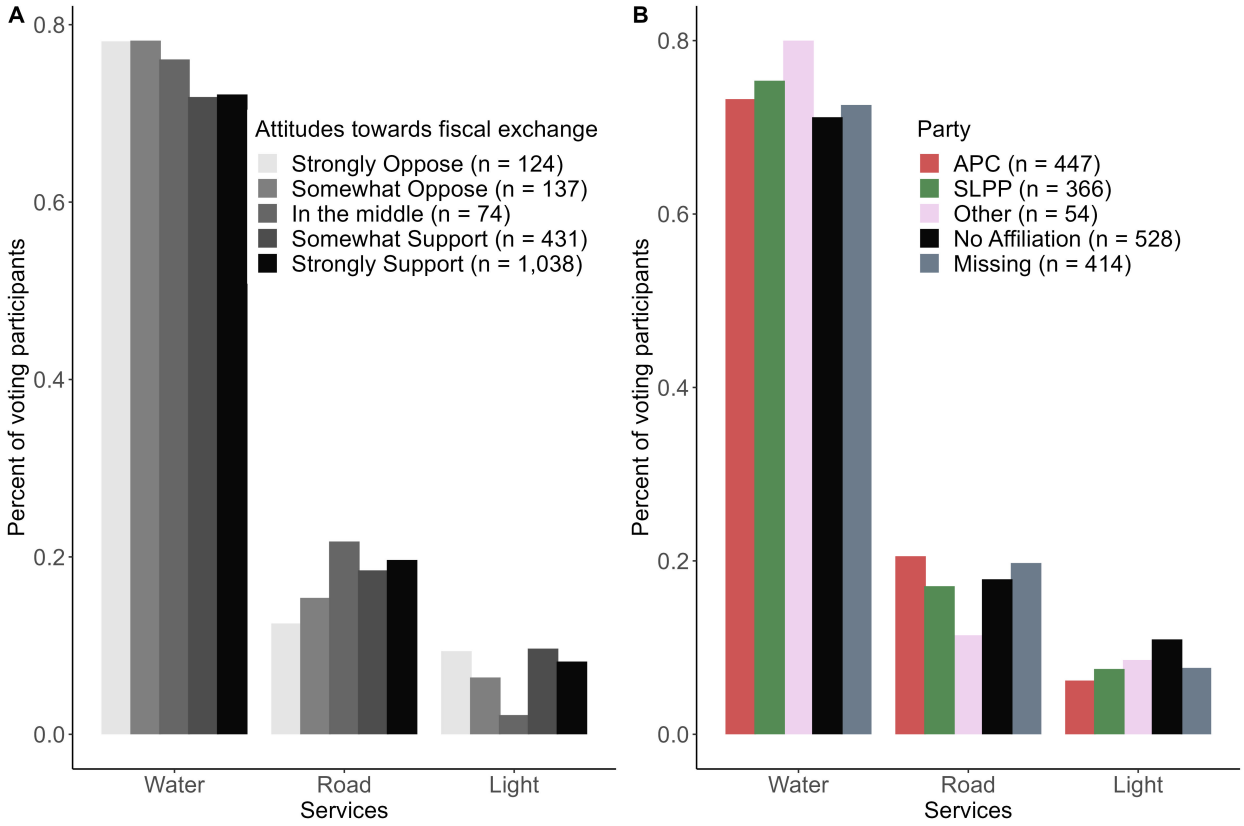
Figures 4.2 and 4.3 show that treatment effects on tax compliance behavior are moderated by participants' baseline attitudes towards fiscal exchange and partisan affiliation, respectively. Our argument is these existing beliefs shape the way participants process the informational signal of the treatment and therefore the treatment's effectiveness. However, these conditional estimates do not causally identify the impact of the moderating variable; factors associated with baseline attitudes towards fiscal exchange or partisan affiliation could be driving the observed conditional effects.

One potential confounder is service preferences. Treatment effects may be conditional on the match between a participant's preferred service and the implemented service. For example, effects may be positive for participants whose preferred service is implemented and negligible (or negative) for participants whose preferred service is not. It could be that fiscal exchange attitudes and partisan affiliation are related to service preferences; when we observe treatment effect heterogeneity conditional on these variables, differences in service preferences are really driving the

difference.

If this were true, we should see that service preferences differ by attitudes towards fiscal exchange and/or partisan affiliation. Figure 4.4 shows the percentage of voting participants who voted for a given project by group.[53] There is little indication that service preferences vary meaningfully across fiscal exchange attitudes (Panel A) or partisan affiliation (Panel B).

Figure 4.4: Votes for Services by Attitudes Towards Fiscal Exchange and Partisan Affiliation



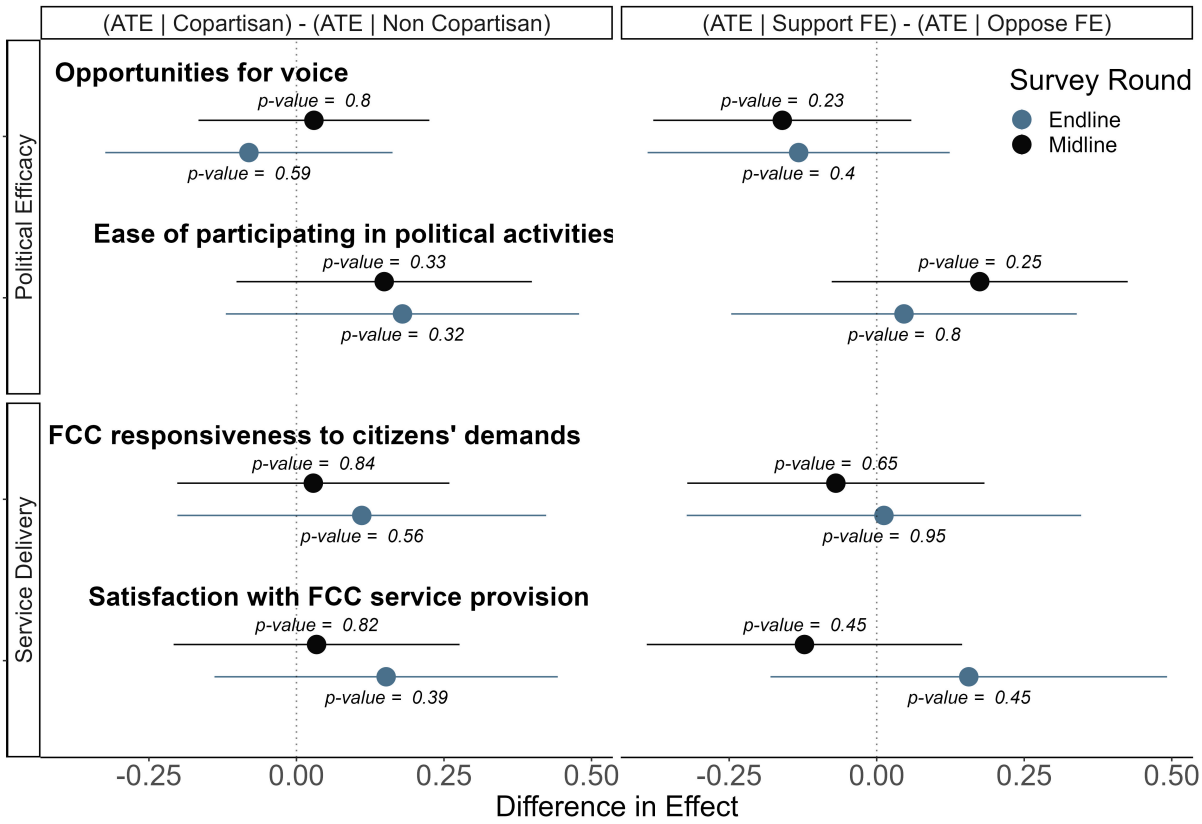
Note: Panels A and B present votes for each service, broken out by attitudes towards fiscal exchange and partisan affiliation, respectively. For both variables, there is little indication that votes for services differ meaningfully by group.

[53] Voting behavior is also similar across groups. For baseline attitudes towards fiscal exchange the percent of participants who voted is as follows: strongly oppose = 52%; somewhat oppose = 57%; in the middle = 62%; somewhat support = 55%; strongly support = 56%. For partisan affiliation, percent voting is: APC = 58%; SLPP = 54%; No affiliation = 52%; Other = 65%; missing = 60%.

Baseline beliefs do not moderate effects on mechanism outcomes

Citizens may be more willing to pay taxes when they believe they will get more of the services they want. We predicted the intervention would increase owners' expected benefits from taxation and we find evidence for this prediction in the form of positive average treatment effects on legitimacy outcomes (Table 4.2). Attitudes towards fiscal exchange and partisan affiliation may moderate the treatment effect on compliance because they moderate key outcomes, which lead to countervailing effects on compliance that swamp any average effect. However, when we break out treatment effects on mechanism outcomes by baseline attitudes towards fiscal exchange and political affiliation we do not find evidence that subgroups have different treatment effects. Relatedly, fiscal exchange attitudes and partisanship are not correlated with participants' reported satisfaction with delivered services (Appendix Table C.9).

Figure 4.5: Differences in Treatment Effects by Sub-Group



Note: The left panel reports the difference between treatment effects conditional on partisan affiliation for key mediating variables. Respondents whose partisan affiliation is missing are coded as *non copartisan*. (This coding decision does not impact results.) Blue and black points are differences for a given outcome at endline and midline, respectively. The right panel reports the difference between treatment effects conditional on baseline attitudes towards fiscal exchange. For these estimates, the group of respondents who "strongly agree" or "agree" that they would be "willing to pay additional taxes in order to receive improved services" are coded as *support*. Respondents who "strongly disagree", "disagree", or are "in the middle" are coded as *oppose*.

4.5 Conclusion

It is well known that poor countries collect less taxes than richer ones (Lee and Gordon, 2005; Besley and Persson, 2014). This disparity is most acute in local government. Property taxes are, almost everywhere in the world, the foundation for effective revenue raising to fund local governments. In lower income countries in particular, the performance of property taxes has lagged dramatically behind their potential (Bahl and Vazquez, 2008). Whereas many high-performing

wealthier countries collect 2 to 3% of GDP in recurrent property taxes, most lower-income countries appear to collect less than 0.2% of GDP from those same taxes. That makes property taxes the most under-performing major tax type across lower-income countries (Brockmeyer et al., 2021).

This under-performance not only undermines revenues but also the broader development of strong local social contracts. With little revenue, local governments are unable to be responsive to the needs and priorities of local citizens; citizens view unresponsive governments as illegitimate and have little interest in paying greater taxes. Many governments in poor countries appear mired in similar, pernicious low legitimacy-low compliance equilibria.

This paper reports results from a large-scale, digital participatory budgeting intervention that was developed to break local government out of this low legitimacy-low compliance trap. We present three primary findings. First, our results highlight that participatory interventions can improve citizens' attitudes towards government and bolster political legitimacy. Second, contrary to influential models of compliance (Levi, 1988, 1997) and implications of canonical accounts of the development of fiscal capacity (North and Weingast, 1989; Bates and Lien, 1985), we show that this increase in citizens' perceptions of government legitimacy does not lead to a corresponding increase in tax compliance. Third, in exploratory analyses, we show that treatment effects on tax compliance behavior and support for expanding taxation are conditional on preexisting policy preferences and political affiliation and that the countervailing impacts of these heterogeneous effects bury an average treatment effect.

What does this imply for governments considering similar participatory interventions? One might conclude, focusing narrowly on average compliance effects, that this intervention was ineffective. However, we caution against this interpretation for several reasons. First, while we do not find average compliance effects of participating in the intervention, its *total* effects may be positive because the intervention's impact spills over to people who did not directly participate but learn about it from those who did. Indeed, in Appendix C.5, we provide suggestive evidence for positive spillover effects. Second, the medium or long-term compliance impacts of participatory

interventions may differ from their short-run impacts. While we report null short-term effects on compliance, given the significant increases in the perceived legitimacy of government one might imagine that long-term impacts are encouraging, as behavior may respond to attitudinal change with a lag. Third, there is the question of participant selection. To rigorously estimate population average treatment effects, we randomly sampled property owners into our intervention. By contrast, in the real world participants often self-select into participatory programs. Given the large treatment effect heterogeneity that we document, self-selection may produce much different average treatment effects. Future research should explore compliance effects of Town Halls with (a) sampling frames allowing for self-selection into eligibility or targeted at those populations we identified as most likely to react positively, (b) bipartisan elite participation, and (c) unmediated interaction between participants and representatives. Our results also call for more research into how sub-populations may require different policy interventions: Could negative subgroup effects for ideological and political opponents have been avoided if the participatory intervention had been coupled with enforcement-based strategies? Finally, governments considering implementing similar interventions care deeply about outcomes other than compliance, such as how they are perceived by voters. We find large, durable treatment effects on perceptions of government legitimacy. Thus, participatory interventions can be used to create more supportive environments for governments who want to carry out ambitious, politically contentious investments in fiscal capacity.

Appendix A

Appendix to Chapter 2

A.1 Context

Table A.1: Influence of and Attitudes Towards TPIs (Afrobarometer data)

	Sierra Leone (%)	Other Countries (%)	Sierra Leone (rank of 31)
Panel A: TPIs are politically relevant			
Local governance ("some" or "a lot")	91.5	62.5	1
Solving dispute ("some" or "a lot")	94.9	70.9	1
Land allocation ("some" or "a lot")	78.1	55.3	5
Influence votes ("some" or "a lot")	16.5	41.3	31
Panel B: Attitudes towards traditional leaders			
Trust ("somewhat" or "a lot")	77.1	66.8	8
Job approval ("approve" or "strongly approve")	75.2	69.2	13
Involved in corruption ("none" or "some")	67.7	72.4	22
Traditional Leaders serve community interests	67.3	58.9	12

Table A.1 reports the perceived influence of and attitudes towards Traditional Leaders in 31 African countries (source: Round 8 of the Afrobarometer survey). This table reports only responses from rural residents, to match the context of this study. Therefore, Cape Verde, Tunisia, and Mauritius are excluded because reported questions were not asked. Survey questions Rows 1-4: How much influence do traditional leaders currently have in each of the following areas: (1) Governing your local community? (2) Solving local disputes? (3) Allocating land? (4) Influencing how people in their communities vote? Survey question for Row 5: "How much do you trust each of the following, or haven't you heard enough about them to say: traditional leaders?" Survey question for Row 6: "Do you approve or disapprove of the way the following people have performed their jobs over the past twelve months, or haven't you heard enough about them to say: Your traditional leader?" Survey question for Row 7: "How many of the following people do you think are involved in corruption, or haven't you heard enough about them to say: traditional leaders?" Survey question for Row 8: "Which of these statements is closest to your own opinion? traditional leaders: (i) mostly look out for what is best for the people in their communities; (ii) mostly serve the interests of politicians and government officials; (iii) mostly look out for their own personal interests.

A.2 Using Beans to Measure Perceived Probabilities

For many questions in our survey, we asked respondents to gauge their expectations or perceptions on a 10-point scale. To make this scale more concrete to survey respondents, all enumerators were given 10 beans and a plastic plate, which served as a visual aid regarding the 10-point scale.

Before entering the main modules of the survey, enumerators guided respondents through several sample questions to familiarize respondents with this scale. The response patterns to these practice questions were encouraging and suggested that respondents understood and were comfortable using the 10-point scale. Average responses were low to unlikely events ("chance that the president visits this community tomorrow") and high for highly likely events ("chance that you will drink water this month"). In addition, the response patterns were in keeping with basic laws of probability—respondents overwhelmingly reported that they had an equal or greater likelihood of visiting the district headquarters town in the next 30 days than in the next seven days: Just 4% of respondents report they are more likely to travel to the district capital over the next seven days than over the next thirty days. Table A.2 in reports responses to four key sample question for respondents in the control group.

Table A.2: Responses to Practice Questions (control group)

Question	**Average beans**
Likelihood of drinking water this month	8.55
Likelihood the president will visit this community tomorrow	1.82
Likelihood of traveling to district capital this week	5.33
Likelihood of traveling to district capital this month	7.44

This exercise also provides insight into how respondents interpreted the levels of my measurement scale. While our enumeration team coached respondents that each bean represented 10 percentage points of probability ("each bean is one chance out of 10"), it seems more likely that respondents

understood each bean as an increase (or decrease) in relative likelihood, rather than representing exactly 10 percentage points. This means that between-respondent differences in measured outcomes may represent differences in the way respondents map perceived probabilities to the 10-point scale—in addition, of course, to representing real differences in beliefs.[1] Therefore, responses to these practice questions might predict responses to other survey questions, a relationship that can be leveraged to reduce noise when estimating treatment effects. In include several of these measures in the pre-specified covariate adjustment.

[1] For example, a respondent who believes it to be very unlikely that the President will visit their community tomorrow may represent this believe with zero or one beans. Note that enumerators were trained to emphasize repeatedly that respondents could put as many or as little beans as they like and were allowed to put all ten beans or no beans at all into the plastic plate.

A.3 Outcomes Description and Summary Statistics

Table A.3: Description of Outcome Variables

Indicator Name	Variables Description
Self-reported propensity to pay tax	A survey question that directly asks respondents how likely they are to pay their full tax liability if a tax collector comes to their house today. Respondents are asked to express this likelihood using the beans.
Coins given to KDC's development fund	Number of coins (out of five) donated during the donation game, a dictator game in which the giver is the respondent, and the receiver is the property tax revenue fund. The value of each coin is about US$0.05.
Perceived neighbors' propensity to pay tax	Proportion of other property owners in the respondent's village that the respondent thinks will pay their property tax. Expressed using the beans.
Tax morale (secondary outcome)	The respondent is asked to imagine a situation in which they would not be fined or penalized for not paying their property tax. The respondent is then asked if they think it is (morally) right to pay their tax. Expressed using the beans on a 10-point scale.
Own village will benefit from tax	Perceived likelihood that respondent's village will benefit from the property tax. Expressed using the beans.
Other villages will benefit from tax	Out of 10 towns in the respondent's chiefdom, how many does the respondent think will benefit from the property tax? Expressed using the beans.
Proportion of revenue towards development	Proportion of the revenue collected from the tax that will be used for development. Expressed using the beans.
Ease of discovering how tax revenue was spent	Perceived ease of finding out how property tax revenue has been spent. Expressed using the beans.
Respondent fined by chiefs if non-compliant	Perceived likelihood that chiefs will fine respondent if they fail to pay the property tax. Expressed using the beans.
Villages where chiefs fine noncompliers	Out of 10 towns in respondent's chiefdom where some people did not pay property tax, in how many will the chief fine property owners who did not pay? Expressed using the beans.
Chief will favor compliant in land allocation	Perceived likelihood that chiefs are more willing to allocate land (for farming, construction, etc.) to people who pay their property tax compared to people who do not pay. Expressed using the beans.
Chief will favor compliant in dispute resolution	In a dispute between two people, perceived likelihood that chiefs would favor a person who has paid their property tax over a person who has not paid. Expressed using the beans.

Table A.4: Summary Statistics (control group)

	Mean	SD	Min	Q25	Q50	Q75	Max	N	Missing
Compliance Outcomes									
Self-reported propensity to pay tax	6.73	3	0	5	6	10	10	428	0
Coins given to KDC's development fund	1.66	1.44	0	1	1	2	5	428	0
Perceived neighbors' propensity to pay tax	5.97	2.32	0	5	6	7	10	405	23
Tax Morale (Secondary outcome)	7.36	2.82	0	5	8	10	10	428	0
Legitimacy Outcomes									
Own village will benefit from tax	7.11	2.85	0	5	7	10	10	421	7
Other villages will benefit from tax	6.71	2.84	0	5	7	10	10	404	24
Proportion of revenue towards development	6.18	2.61	0	5	6	8	10	417	11
Ease of discovering how tax revenue was spent	4.46	3.09	0	2	4	7	10	426	2
Coercion Outcomes									
Respondent fined by chiefs if non-compliant	7.51	2.66	0	5	8	10	10	425	3
Proportion of villages where chiefs fine noncompliers	6.39	3.47	0	4	6	10	10	394	34
Chief will favor compliant in land allocation	7.63	2.46	0	6	8	10	10	425	3
Chief will favor compliant in dispute resolution	6.53	2.79	0	5	6	10	10	426	2

A.4 Survey Sampling

In this section I provide additional details for how I sampled chiefdoms, villages, and respondents for surveying. Kono district contains 14 chiefdoms and roughly 1,300 villages. I used geographic cluster sampling to select 123 villages for the study from a set of 434 eligible villages in five chiefdoms.

Sampling Chiefdoms

The research design relies on the development of chiefdom-specific tax awareness videos. Therefore, it was only possible to conduct the study in chiefdoms where we were able to create a tax awareness video with the paramount chief. I reached out to paramount chiefs or senior chiefdom authorities in all 14 chiefdoms and successfully shot videos with senior chiefs in 10 of those 14 chiefdoms. For the four chiefdoms where I did not shoot a video, I was unable to schedule a recording session in the three-day period for which I had hired a professional filmmaker.[2] In three chiefdoms, we shot videos not with the paramount chief, but with his representative: I excluded these chiefdoms.[3] I excluded one chiefdom (Gorama Kono) because I judged that the language used by the paramount chief in the video differed too much from the agreed script. Finally, I decided to exclude Toli Chiefdom for practical budgetary reasons. Toli contains less than 2% of the villages in Kono and is the most sparsely populated and least accessible chiefdom in the district. I determined that enumeration costs in Toli would be too high to warrant inclusion in the study. This left five chiefdoms included: Soa, Lei, Gbane, Nimikoro, and Kamara.

[2]The paramount chief of Sandor was traveling; the paramount chief of Gbane Kandor did not come to Koidu (the district headquarters); and the paramount chiefs of Nimiyama and Tankoro were unable to meet due to scheduling conflicts.

[3]In Mafindor Chiefdom, we filmed the video with the acting regent chief, as the paramount chief recently passed away and a new one has not been elected. In Fiama Chiefdom, we filmed the video with the chiefdom speaker, as the paramount chief is the Kono paramount chief representative in parliament. In Gbense Chiefdom, we filmed the video with the chiefdom speaker at the request of the paramount chief.

Sampling of Eligible Villages

Here I provide more details on my cluster sampling strategy. I first grouped the 434 eligible villages into 155 geographical clusters, dropping three isolated villages.[4] I then sampled clusters, and within sampled clusters, I sampled villages.

One goal of my sampling process was to generate a final sample that had sufficient variation in two village level characteristics: (i) the distance to the chiefdom headquarters town, where the paramount chief resides, and (ii) the size of the village.

I thus coded each of the 155 clusters along these two dimensions. Within each chiefdom, I coded each cluster into one of six strata that combined three levels on the distance dimension and two levels on the village size dimension. On the distance dimension, villages could be near, middle, or far from the chiefdom's headquarters town. On the village size dimension, clusters were coded as either containing a large village or not, with "large" defined as at or above the 75th percentile in terms of population. To increase variation along the distance dimension, I dropped clusters coded as a middle distance from the chiefdom's headquarters town. This leaves me with clusters in four strata from which to draw my sample:

1. Clusters near the chiefdom's headquarters town that contain a large village.
2. Clusters near the chiefdom's headquarters town that do not contain a large village.
3. Clusters far the from chiefdom's headquarters town that contain a large village.
4. Clusters far the chiefdom's headquarters town that do not contain a large village.[5]

I then wrote a sampling procedure that aimed to balance my final number of observations across each of the four strata. The specifics of the sampling procedure are as follows:

[4]Clustering was done within each chiefdom, so that villages were not clustered across chiefdom boundaries. After initial clustering, 25 villages were in clusters of their own. I placed these villages in the closest cluster. In three instances, these one-cluster villages were more than three kilometers from the closest village in their new cluster; I dropped these three villages.

[5]In one chiefdom (Kamara), there were no eligible clusters in the stratum representing large village and near the chiefdom's headquarters town. Therefore, I have 19 total strata from which to draw clusters.

- First, I drew two clusters in each strata. (There are two strata that contain only one cluster of villages—in these I drew one cluster).[6]
- Second, I selected two village in each stratum. In strata that contain large villages, I selected one large and one small village.
- Third, I checked whether the number of potential observations in each stratum was at least 100. As a proxy for the number of potential observations in each village, I used the number of structures recorded in the 2015 census.
- Fourth, in strata where the target number of potential observations was not been met, I drew an additional village from the set of sampled clusters.[7]
- It remains possible that the maximum number of potential observations in a given strata did not reach 100. In this case, I drew an additional cluster from the appropriate cluster stratum.

Selecting Respondents

Once in a village, the enumeration team used a random walk strategy to select respondents for the survey. The protocol for this strategy was as follows:

- The enumeration team arrived in the village in the morning and went directly to the house of the village chief (or another village authority if the village chief was not present that day). A letter had been dropped off to village authorities within the previous three days specifying the date of the enumeration team's arrival. Enumeration teams ranged between two and six people, depending on the size of the village.
- From the house of the village chief, the enumerators agreed to walk in separate directions. After agreeing which directions they would each travel, enumerators used their tablets to select a distance, which told enumerators whether to interview a respondent at the first, second, third, or fourth house in their chosen direction. If the enumerator found no one home at the relevant house, the enumerator proceeded to the next house in that direction.

[6] Of course, I selected no clusters in the one stratum that contains no clusters.

[7] For example, if after step 3 the not large villages sampled in a given chiefdom contained fewer than 100 structures, I drew another not large village from the set of sampled clusters in that chiefdom.

- The enumerators asked to speak to the person "most responsible and influential" for making decisions related to the property. If that person was home, the enumerator began the informed consent process to start the interview. If that person was not home, the enumerator asked if he or she would return later that day. If so, the enumerator scheduled a time to return to interview that person. If not, the enumerator asked if there is "someone else who is involved in decision-making related to this property." If so, the enumerator asked to interview that person. If not, the enumerator attempted to schedule an appointment for later. If that was not possible, the enumerator moved on to a different property.
- After completing an interview, the enumerator used the tablet to select the direction and distance of the next house. Previously interviewed houses (marked by a sticker) were not included in the count.
- If an enumerator walked past the last structure of the village in a given direction, he or she turned around and finished the count, walking back in the direction they came.
- If an enumeration team completed interviews with all available respondents before the end of the day, they proceeded to their next scheduled village. Otherwise, the enumerators left for their next scheduled village in the morning.[8]

[8]Note that in several large villages, enumeration teams were scheduled to conduct interviews for more than one day.

A.5 Attention / Manipulation Checks

Recalling Number and Identity of Speakers in Video

The tax awareness videos contain information that I expect to modify respondents' beliefs in theoretically important ways. First, I check whether respondents can correctly recall the number and identity of the speakers in the video they watched.[9] 94% of respondents correctly state the number of speakers and 93% correctly identify the speakers.[10] Enumerators asked these questions directly after a respondent watched the video.

Attention Checks

I checked whether respondents can recall theoretically important messages delivered in the video they watched. At the end of the survey, we asked respondents a set of six yes/no questions, regarding whether statements were included in the video. We asked respondents whether the following statements were discussed in the video they watched:

1. A property tax that will be collected on houses. ["Tax"]
2. The Chiefdom Council is working with Kono District Council on this property tax. ["Collaboration"]
3. After taxes are collected chiefdom authorities will call a meeting to discuss how to spend the money collected. ["Spend"]
4. After taxes are collected chiefdom authorities will call a meeting to discuss how to punish non-compliers. ["Punish"]
5. Tax collectors will be paid 10% of the money they collect. ["Salary"]
6. All tax collectors have an identification card with their name and picture. ["ID Card"]

Table A.5 reports property owner responses by treatment condition. Column 2 ("n") refers to the

[9] Respondents who see the control video see only one speaker, District Council Chairman Solomon Bundo. Respondents who see treatment videos see two speakers—Chairman Bundo and the paramount chief of their chiefdom.

[10] If there were two speakers, this meant correctly naming both.

number of observations in each treatment group.[11] The value in each of the remaining six columns is the percent of respondents that affirmed a given message was given in the video. First, let's consider a set of three questions that all respondents should answer in a similar way, regardless of treatment condition. Of course, the central messaging of the video is around a house/property tax. Column 3 ("Tax") tells us that across treatment and control groups 95 to 98 percent of respondents correctly state that the video contained messaging about a house tax.

Respondents were also asked about two statements that did not appear in any video:

- Tax collectors will be paid 10% of the money they collect (Column 7, "Salary")
- All tax collectors have an identification card with their name and picture (Column 8, "ID card")

Respondents did well at identifying statements that were not in the videos. Across treatment and control 85% of respondents correctly state that compensation for tax collectors is not discussed and 78% correctly state that tax collector ID cards are not mentioned in the video. As expected, there does not appear to be meaningful differences between treatment arms.

Table A.5: Attention Check

Treatment Arm	**Tax**	**Collaboration**	**Spend**	**Punish**	**Salary**	**ID Card**	**n**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
C: Tax information	0.95	0.68	0.44	0.42	0.14	0.20	428
T1: TPI collaboration	0.98	0.86	0.53	0.48	0.14	0.22	454
T2: Legitimacy	0.97	0.89	0.78	0.60	0.18	0.24	436
T3: Coercion	0.98	0.90	0.75	0.81	0.15	0.24	433

Next, let's consider responses to three questions that we do expect to change with the respondent's

[11] Note that a one respondent is dropped from T2 group, who responded "I don't know" to these comprehension check questions.

treatment condition. First, recall that the treatment videos attempt to manipulate respondents' perceptions about the collaboration between the state and TPIs. As reference to this collaboration only appears in treatment versions (and not in control), we should see respondents in T1, T2, and T3 more likely to agree that the collaboration between chiefs and state was discussed in the video (compared to control video). Indeed, that is what we see. Respondents in T1, T2, and T3 are respectively 18 percentage points, 21 percentage points, and 22 percentage points more likely to state that collaboration between KDC and chiefdom authorities was mentioned in the video.

The goal of the Legitimacy treatment (T2) is to prime respondents to legitimacy mechanisms in TPIs and the T2 video the paramount chief says he will call a meeting to discuss with his sub-chiefs on how the collected revenue will be spent. Respondents who watched T2 videos are 25 percentage points more likely to affirm that their videos referred to these meetings, compared to respondents who watched the T1 video.[12] In the Coercion treatment (T3) the paramount chief said that he would call a meeting with chiefdom authorities to discuss how to punish non-compliers. Respondents who viewed the T3 video are 33 percentage points more likely to state that their video contained this message, compared to respondents who watched the T1 video.[13]

The response patterns from the comprehension check exercise are encouraging. Overall, respondents are good at identifying messaging content that was or was not in their video and responses vary in predicted ways with the video version that respondents watched. However, for questions that involve TPIs, the rate of "false positives"—respondents affirming that a message was delivered in their video when in fact it was not—is high. For example, 68% of respondents in the control video (column "Collaboration") affirm the video discussed collaboration between the local government and the chiefdom council when this was in fact not the case. What should we make of this rather high "false positive" rate?

I argue that the six recall questions should be considered a hard test and that the high rate of false

[12]This increase jumps to 34 percentage points when comparing T2 to the pure control.
[13]This increase jumps to 39 percentage points when comparing T3 to the pure control.

positives is indicative of the difficulty of the test, rather than a lack of respondent comprehension. First, the recall questions are designed as leading questions ("was X in the video?"), which likely generates the confirmation bias that I am here calling a "false positive". This seems to be only part of the explanation, as this confirmation bias should be consistent across all questions, but we see higher rates of false positives for recall questions about TPIs. Second, recall questions were asked at the end of the survey, whereas the video was shown at the beginning of the survey. The motivation for putting these recall questions at the end of the survey is to avoid priming respondents before measuring outcomes. For example, asking respondents if the video mentioned collaboration between government officials and TPIs might prime control respondents to think about chiefs when they otherwise would not have. The tradeoff is that by placing comprehension questions at the end, survey questions that come prior to the comprehension check questions can also prime respondents. More concretely, respondents are first asked a host of survey questions about TPIs, then at the end of the survey respondents are asked if the video they watched contained messaging about TPIs. It is possible that respondents infer that the subjects they were asked about in the survey (ie., TPIs) are likely to have been addressed in the video. Third, placing the recall questions at the end of the survey creates a time lapse between the video and the recall questions, which may lower recall.

If the high false positive rate is driven mainly by the placement of the recall questions at the end of the survey (rather than general lack of comprehension), we should see much lower false positives if the recall questions were asked directly after the video. Prior to undertaking my primary data collection, I conducted a pilot survey where we did ask recall questions directly after the video. Table A.6 shows results from that pilot. False positive rates in the control group plummet. Only 30% of respondents who watch the control video incorrectly state their video discussed collaboration, down from the 68% we saw in our true study. Rates of false positives drop across each of the other four comprehension check measures.

Table A.6: Attention Checks (pilot)

Treatment Arm	**Tax**	**Collaboration**	**Spend**	**Punish**	**Salary**	**ID Card**	**n**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
C: Tax information	0.89	0.30	0.20	0.07	0.09	0.04	46
T1: TPI collaboration	0.93	0.72	0.10	0.14	0.00	0.00	29
T2: Legitimacy	0.92	0.76	0.70	0.24	0.00	0.03	37
T3: Coercion	0.94	0.81	0.50	0.50	0.06	0.11	36

A.6 Intervention Development

The tax awareness video was developed in a series of meetings that were attended by members of the Kono Revenue Mobilization Team between 2018 and 2021. Early meetings focused on the development of tax collection protocols and program infrastructure. After 2019, the focus of the meetings shifted toward strategies to improve tax compliance. An awareness-raising campaign ("sensitization plan") was one oft-discussed strategy for increasing tax compliance. At a July 2020 meeting, the KDC requested that I develop a proposal for that campaign.

At a November 2020 (Zoom) meeting, I presented a proposal for a video-based property tax awareness campaign, involving both KDC officials and traditional leaders. I had two motivations for including traditional leaders in the tax awareness video. First, there was consensus amongst the Revenue Mobilization Team that (i) property owners were unaware that their traditional leaders were collaborating with KDC on the property tax, and (ii) effectively communicating traditional leader involvement to property owners would increase compliance. Second, several interviewed tax collectors reported that they believed property owners would be more willing to pay property tax if they knew their traditional leaders were collaborating with KDC on the property tax.[14] The proposal was met with general approval, and I was directed to continue developing plans for a tax awareness video that contained local government officials and traditional leaders.

At a January 2021 meeting in Kono, the Revenue Mobilization Team agreed on the basic contours of a tax awareness video, including the three key messages that traditional leaders should emphasize (collaboration with KDC, accountability, and enforcement). I then met with senior paramount chiefs to workshop the video script.

[14] In late 2019 and early 2020, I worked with a research assistant to interview tax collectors to understand what they thought would increase compliance. Several stated that they were more successful collecting taxes when property owners knew that their traditional leaders supported and were involved in the property tax. For example, a tax collector from Tankoro chiefdom noted, "well, if the paramount chief calls a meeting. When the chiefs are more strongly backing this thing [the property tax], that's going to make people pay. You know that the community people fear/respect their authorities" (my translation).

Chiefdom-specific tax awareness videos were recorded and edited in March 2021. Kono district is named for its predominant ethno-linguistic group, and I expected Kono speakers to dominate our sample. That said, Kono district also contains a significant non-Kono speaking population. Therefore, we filmed the tax awareness videos in both Kono and Krio (an English-based creole that is the country's lingua franca). As Kono is not a written language, the video script was written in Krio. When filming, we first walked chiefs through the Krio script and then filmed the Krio version. Before filming the Kono version, chiefs listened to a prepared Kono recording and practiced the script with a Kono-speaking senior research assistant. We recorded the videos segment by segment; when chiefs deviated meaningfully from the script, we reshot the segment.

We shot tax awareness videos with traditional leaders in 10 chiefdoms, and I conducted this study in five of those chiefdoms. In four of these chiefdoms, we filmed both Kono and Krio versions. In one (Nimikoro), we only filmed a Kono version, so property owners who did not speak Kono were excluded.[15]

[15] In Nimikoro, as in other chiefdoms, we first discussed the script in Krio. However, the paramount chief requested that we film the Kono version before the Krio version. After finishing the filming of the Kono version, the chief left to attend another appointment; we were unable to meet again to film the Krio version while the filmmaker was in Kono district.

A.7 Additional Coercion Results

In Table A.7, I attempt to tease out the specific tools of TPIs' coercive power by examining the impact of the Coercion treatment (T3) on six intermediate experimental outcomes. A first measure captures the perceived likelihood that respondents will be fined if they don't comply; a second measure captures the percent of villages in the chiefdom where noncompliant owners will be fined by chiefs; the third and fourth indicators measure perceptions that noncompliers will be fined by the *local government*. The fifth and sixth indicators measure governance bias, capturing respondents' perceptions that chiefs will favor compliant property owners in decisions of land allocation and dispute resolution, respectively. The comparison group is the pure control (C; Tax information).[16]

Point estimates are positive, though not individually statistically significant, for the indicators that capture perceptions that noncompliant property owners will be fined (outcomes 1-4); point estimates are negative (though not statistically significant) for the indicators that capture beliefs that noncompliant property owners will face future bias in decisions regarding land or disputes (outcomes 5-6).[17] Note that Table A.7 includes two indicators that were not preregistered, both measuring the probability that noncompliers will be fined by the local government. I have included these outcomes to provide additional evidence that the Coercion treatment primes fine-based punishment for noncompliance.[18] While the results in Table A.7 are murky, there is more evidence that the Coercion treatment (T3) is priming beliefs about fine-based punishment, rather than governance bias-based punishment.

[16]Using the Collaboration treatment (T1) as the comparison group is problematic because questions about punishment are likely to prime respondents to TPIs' coercive capacity, undercutting the treatment effect of T3. For example, asking respondents about the probability that chiefs will fine noncompliance after the respondent has watched a video containing the paramount chief, is likely to prime respondents to TPIs' propensity for issuing fines.

[17]I preregistered an index as the main outcome for hypothesis testing these intermediate coercion outcomes. The point estimate on the index is near zero and not statistically significant. However, I drop the index from Table A.7 because my exposition and analysis of intermediate coercion outcomes focuses on variation in effects between sub-indicators, rather than their average effect.

[18]These two outcomes are the only two punishment related outcomes that I measured but did not preregister for analysis.

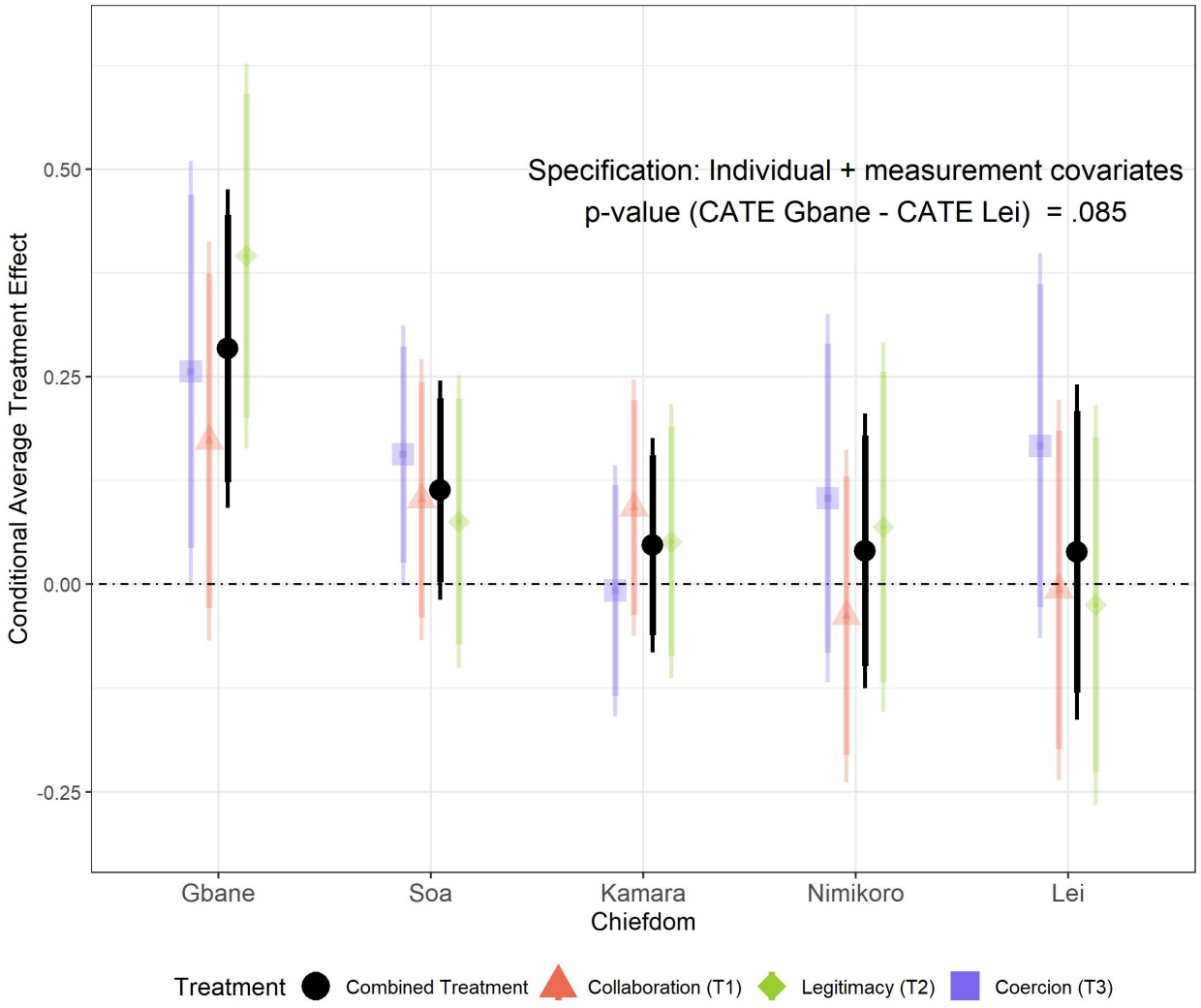
Table A.7: Effects of Coercion Treatment (T3) on Intermediate Outcomes

Secondary Outcome	**Mean** (1)	**T3-C** (2)	**N** (3)
Punishment through Fines			
Respondent fined by chiefs if non-compliant	7.506	0.079	1,748
	(2.660)	(0.062)	
Proportion of villages where chiefs fine noncompliers	6.393	0.026	1,621
	(3.474)	(0.057)	
+Respondent fined by local govt if non-compliant	7.154	0.086	1,737
	(2.711)	(0.060)	
+Proportion of villages where local govt fines noncompliers	6.300	0.019	1,585
	(3.451)	(0.057)	
Punishment through Biased Governance			
Chief will favor compliant in land allocation	7.628	-0.031	1,744
	(2.460)	(0.054)	
Chief will favor compliant in dispute resolution	6.526	-0.034	1,742
	(2.793)	(0.056)	

Table A.7 reports the effect of the Coercion treatment (T3) on the secondary coercion outcomes. Column 1 reports the control group mean for each indicator, with the standard deviation in parentheses. Column 2 present standardized treatment effects for T3, relative to control. Models are estimated using OLS with preregistered specifications. Column 3 reports the number of non-missing observations.
[+] Outcome not preregistered.

A.8 Variation in Treatment Effect By Chiefdom

Figure A.1: Combined Treatment Effects by Chiefdom



A.9 Kono Videos Translation / Transcriptions

This appendix provides translations for Kono versions of the tax awareness videos. For each of the five chiefdoms, we created three treatment video segments. The control video (with only the District Council Chairman) is the same across all chiefdoms. There are two translations for each video segment. One translation was done by my lead research assistant. A second translation was done by the founder of a local research and capacity building NGO.

Control: Tax information

Translation 1: Good morning, good evening, good night. It's your son Solomon Sahr Bundo, Chairman on top the Kono District Council. As you all know, going further – in the morning hours people are calling for water well, roads and other things, they spoke of going further– and as you know going further does not happen for free, it involves money. And this money, we as Kono people, we can come together and gather our money, that will make us to go further.

Everybody that constructs a house, any kind of house, it can be a stick house, it can be zinc house or story building, you should give money for that house. This will make the country to go further.

For this reason, we sit and discuss what you should give for your house for a year– not a month, but a year. The money that I am calling now is for a year. This is what we give for our house for a year. This money you give is not for a month, but a year:

- Thatch house, stick house. The money you give, they call it in Leones twenty-thousand [calls amount in Krio]. As I say, the money you give for stick house and thatch house is 20,000 Leones [This time calls amount in Kono].
- A stick house with a zinc roof is 30,000 leones [calls amount in Krio].
- If you construct a dirty block house, without giving it cement, you only put your block, but you did not plaster it and cover it with a zinc, you pay 40,000 leones.

- The dirty block you plaster with cement, you plaster both in and outside of the house– that is 50,000 Leones.
- If you construct a house, you plaster both in and outside, with a toilet inside– there are many of these now in our villages– you pay 80,000 Leones.
- Now if you want to construct a house and you want to do it with cement block, but you don't put a toilet inside the house, you pay 120,000 Leones.
- If you construct it now, a concrete house with a toilet inside – they call it "self-contain concrete house" [Krio]", we call concrete, you construct it with cement, you have the toilet inside the house, you pay 150,000 Leones.
- If you construct your house as a one story building, whatever happens it's a story building – no matter, they can't build a story building with mud, they only construct cement block with a story building– you pay 300,000 Leones.
- If you have more than one story in that particular house, it can be one, two, or three or even you touch the sky, we have put this in one category and you pay 400,000 Leones.

Please, I am apologizing. Let's gather our small money so we can construct our roads, we can maintenance our water wells and build our schools. That will make Kono to go further. It's me, as I started that is how I am going to end it, it is your son Solomon Sahr Bondu, Chairman of Kono District Council. Thank you very much for listening.

Translation 2: Greetings my people! Good morning, good afternoon and good evening. This is your son Solomon Sahr Gbondo who is heading the Kono District Council. As all of you may know, people are calling for development projects like boreholes, roads rehabilitation and other many more things. But it is worthy of note that, development projects come with a price, it does not happen out of nothing, funds are needed to finance it, these projects. With regards to that, we the people of Kono should come together, work in unity to raise these funds to achieve our development goals.

I want people to know that those that own houses, whether it is constructed with sticks and mud-bricks, covered with local roofs or zinc, cement houses and concrete story buildings, must pay property taxes for that particular structure. After series of engagement, we have agreed that house owners should pay the following amount annually or yearly, please note that, it is not a monthly but yearly and should be done once a year.

- Stick house, with local roof. That is, palm trees leaves. You pay 20,000 Leones.
- Unpaved mud-bricks house with zinc roof. You pay 30,000 Leones.
- Mud-bricks house paved with cement in and out and covered with zinc roof. You pay 40,000 Leones.
- Mud-bricks house paved with cement and has toilet facilities inside. You pay 50,000 Leones.
- House built with cement bricks without toilet facilities inside. You pay 80,000 Leones.
- House built with cement bricks with toilet facilities inside. This is also known as "self-contain house". You pay 120,000 Leones.
- One story building. You pay 300,000 Leones.
- More than one story building. It can have multiple floors. You pay 400,000 Leones.

Please, let us pay our taxes in order for us to able to carry out development projects in the district like roads rehabilitation, digging of boreholes, building of schools, and other things. If we commit ourselves to such, we will be able to move on with our development projects for the good of the districts. I will conclude in the similar way as I introduced myself at the beginning of the video: I am your son and Chairman of Kono District Council, Solomon Sahr Gbondo. Thank you all!!

Treatment 1: TPI collaboration

Nimikoro Chiefdom

Translation 1: My people, good afternoon. Nimikoro good afternoon, Kono people good afternoon. This is your chief, Paramount Chief Aiah Denti Formansah Bono III, Nimikoro, Kono. My people, I want to tell you that Nimikoro Chiefdom and Kono District Council are working together to collect tax for our properties, our houses.

Translation 2: Greeting's relatives, Nimikoro and people of Kono! This is your Paramount Chief – doubling as head of the Poro Secret Society in his Chiefdom – Aiah Denton Bona the III of Nimikoro Chiefdom. My people, let me take this opportunity to inform you that Nimikoro Chiefdom Committee is working in collaboration with the Kono District Council to ensure we pay taxes for our houses. It is called "Property Tax" in the White Man's English language.

Gbane Chiefdom

Translation 1: Gbane good afternoon. This is your Chief Aiah Bindi Faefankongor II. Gbane Chiefdom and Kono District Council are working together to collect tax for our properties, especially houses, for us to pay for them.

Translation 2: My Gbane people, I greet you all. This is your Paramount Chief Aiah Bindi Faefankongor the II. Gbane Chiefdom and Kono District Council are working in unity or collaboration to collect property taxes – more importantly taxation for houses– which we should pay.

Lei Chiefdom

Translation 1: My people, good afternoon. This is your Chief Sahr Cheety Mani, Lei Chiefdom. My people - Lei chiefdom is working with Kono District Council to collect property tax for everyone to pay for their house.

Translation 2: Greetings my people! This is your Paramount Chief Sahr Cheety Mani of Lei Chiefdom. My people, the Paramount Chief of Lei Chiefdom is working in collaboration with Kono District Council to ensure people pay taxes for their houses.

Soa Chiefdom

Translation 1: It's me, Paramount Chief Emmanuel Tamba Torcheor Foryoh IV, Soa Chiefdom. Soa Chiefdom Council and the Kono District Council have sat together so that they can collect house rate from us that have built houses. It is called property tax, and it is to be collected to develop our Chiefdom.

Translation 2: I am Paramount Chief Tamba Emmanuel Torcheor Foryoh the IV of Soa Chiefdom. The Soa Chiefdom Council and the Kono District Council held a meeting and have agreed to collect revenue through payment taxes for our houses which we house owners should pay. It is called "Property Tax". We should collect property tax revenues in order for us to be able to fund our development projects in Soa Chiefdom.

Kamara Chiefdom

Translation 1: My people good afternoon. This is your chief, Chief Ngekia, of Kamara Chiefdom. Kamara Chiefdom and Kono District Council are working together so that our taxes will be collected together, and our house rates also together.

Translation 2: My people, greetings. This is your Paramount Chief, Chief Ngekia of Kamara Chiefdom. The Kono District Council and Kamara Chiefdom have agreed to collect taxes, among these taxes are house tax payment.

Treatment 2: Legitimacy

Nimikoro Chiefdom

Translation 1: The reason why I will call Nimikoro Council– we call it in English "Nimikoro Council Committee" – this Nimikoro Council, we'll call everyone and present the money and ask what will we do with the money, so that a single person will not take the money and put it in his own pocket and it will not benefit Nimikoro. If this money is gathered and you didn't see any good thing that the money brings and it didn't bring any benefit in Nimikoro Chiefdom you will not be happy and you will get angry at me.

Translation 2: The reason for this notice is to notify you that after the taxes from property owners have been collected, we will summon everyone to a meeting and present the money for all of us to see the pool of money that has been collected. Then we will inquire of the people what should be done with the revenue collected with regard to undertaking chiefdom development projects. We are doing such to discourage anyone who may have plan to siphon or misappropriate the funds collected for his or her personal gains at the expense of Nimikoro Chiefdom's interest. After the tax revenue collection exercise, if the people understands that nothing significant is done to facilitate development projects from the money collected, it will spur dissenting view in the minds of the people and they will be annoyed with me.

Gbane Chiefdom

Translation 1: After the collection, I will call a meeting with my chiefs so that we will discuss on the use of the money. I believe that if the right work is not done with the money you will not be happy for Gbane.

Translation 2: After we would have finished collecting the tax payment, I will summon a meeting, in this meeting, I shall request the presence of other subordinate chiefs in the chiefdom for us discuss and map out ways of how the money collected (tax funds paid) is going to be utilized. I

am of the belief or conviction that if we do not utilize the funds collected in the best way for the development of the chiefdom, you the chiefdom people will be annoyed.

Lei Chiefdom

Translation 1: When the money is collected, I will call my chiefs for us to sit together and know what we will do with the money for us to develop this chiefdom. I know that this money, if it does not benefit Lei, you will not be happy.

Translation 2: When we are done collecting the taxes. I will summon a meeting that will bring together my subordinate chiefs for us meet, discuss and bring up resolutions on how we are going to use the money collected to foster development projects in the Chiefdom. I am aware that people will not be happy if the chiefdom do not experience a huge benefit out of the collected money.

Soa Chiefdom

Translation 1: The money that would be collected from Soa, here, I will call my chiefs and the Chiefdom Council will sit together and we'll arrange how we will work with the money. I know that the money collected, if we didn't work with it correctly, no one will be happy here in Soa.

Translation 2: When the tax funds would have been collected, I will invite my subordinate Chiefs and some members of the Chiefdom Council to a meeting in a bid for us to discuss and map out resolutions on how the funds will be utilized. I am mindful of the fact that people will not be pleased, if the tax revenues collected are not properly used to facilitate development projects in the chiefdom.

Kamara Chiefdom

Translation 1: When the taxes are collected, I will call a meeting for everyone to come for us to know the money collected, and what work will we do with it. I know if this tax did not bring any benefit to us here in Kamara, we will not be happy.

Translation 2: After the taxes would have been collected, I will summon a meeting where all the people will be invited to understand and decide on what we will do with the tax funds collected. I am aware that the people of Kamara Chiefdom won't be happy if the money collected does not bring benefits to the chiefdom.

Treatment 3: Coercion

Nimikoro Chiefdom

Translation 1: Why, when we finish collecting the money we will all come and sit together– we will fine those who did not agree to pay their own taxes – and discuss what we will do them. We will not sorry for anyone... when we call a meeting, you that didn't agree, we and the other chiefs, starting from me the Paramount Chief down to all the other chiefs in our villages, we will not be happy with anyone who did not pay the tax. If you did not pay we will charge you and we will not feel sorry for you. We will fine you and take you before court.

Translation 2: Furthermore, another meeting will be summoned, where all of us will meet to discuss and take actions against those who may have refused to pay the property tax. We will not be merciful to anyone who is guilty of tax evasion. During that meeting, I and the other subordinate chiefs in all the towns even down to the least hamlet in this chiefdom, will stand tough in ensuring we bring actions against tax evaders if even it is going to an extent of prosecuting them in the court of law.

Gbane Chiefdom

Translation 1: After the collection I will call a meeting with other chiefs for us to discuss and know how to deal with those that didn't pay for their houses. In addition to that, myself and the other chiefs will not be happy with anyone who did not pay for his house.

Translation 2: Also, after the collection of these taxes, I will hold another meeting with the chiefs

to engage or brainstorm on what to do with those that have refused to pay taxes for their houses. Let me emphasize that I and the rest of the chiefs will not be merciful on anyone who have refuse to pay the tax.

Lei Chiefdom

Translation 1: When this money is collected, we will sit with my chiefs for us to decide, those that did not pay for their houses, what we will do with them. Me and the other chiefs we will sit and we will not take kindly to anyone who did not pay for their houses.

Translation 2: Furthermore, after the conclusion of the tax collection exercise, I will summon a meeting again with the chiefs to engage on what actions we should bring up against those that have evaded the payment of taxes for their houses. We will not be merciful or lenient with anyone that do not honor the payment of tax for his or her house.

Soa Chiefdom

Translation 1: In addition to that, when the tax is collected, I will call my chiefs, we will sit together and find out to know, who actually denied to pay the house rate. And these house rates, those who denied to pay, we will find a solution how to deal with them so that tomorrow other people will not deny to pay.

Translation 2: In addition, after tax revenues will have been collected, I will summon my subordinate Chiefs to another meeting again. In this meeting, we will discuss issues pertinent to house owners who may have refused to pay their property tax. Furthermore, we will develop strategies that will discourage the act of property tax evasion to deter people not to evade tax in the future. We will not be tolerant to those who refused to pay their property tax. Such persons will be categorized or listed as individuals who does not like the development of Soa Chiefdom.

Kamara Chiefdom

Translation 1: When the tax is collected, I will call a meeting again for us all to come and sit and know those who didn't pay taxes for their houses, what we will do with them. My self and the other chiefs will take kindly to anyone who did not agree to pay tax for his house.

Translation 2: When the house tax payment exercise concludes, I will summon another meeting where all of the people will be invited to meet and agree on actions that we will take against those people in the chiefdom that have evaded or defaulted in paying their taxes. I and the rest of the chiefs will not be merciful or compassionate on anyone who flout paying their taxes.

Appendix B

Appendix to Chapter 3

B.1 Enforcing Local Tax

Here I present additional qualitative evidence of TPIs' enforcement capacity from an existing, widespread poll tax (called the local tax), which is collected by chiefdom authorities.[1] I find that TPIs have monitoring mechanisms in place to detect noncompliance, and that fines are the typical punishment for noncompliance.

Traditional leaders commonly used roadblocks (mentioned by 32% of respondents) to monitor compliance with the local tax, erected either inside the village or at key junctions on the road network.[2] Another common monitoring strategy, noted by one village chief, is for authorities to "go house-to-house to check for tax payers" (24%).[3] Informants also reported that authorities keep records of who has paid (24%). While village chiefs can monitor compliance directly by, for example, making lists of compliant community members,[4] chiefdom authorities can monitor *villages*

[1] Chiefdom authorities are entitled to keep most of the local tax revenue; a small percent is transferred to the local government.

[2] One respondent explained that chiefdom authorities "erect check points in collaboration with the chiefdom police, especially when the compliance rate is low" (Interview: 405). A Village Chief noted, "we erect checkpoints on the roads" to monitor compliance (Interview: 404).

[3] Interview: 404

[4] Interview: 406

by tracking the number tax receipts and associated revenue turned in by a given village.[5] In total, 68% of respondents described at least one strategy that authorities used to monitor compliance with the local tax, at either the village or chiefdom level.[6]

The majority of respondents (55%) report that individuals found to have not paid their local tax will be issued a fine by authorities: "either you buy the tax, or you pay a fine."[7] Other respondents note that non-compliers can be taken to higher authorities (45%), a situation also likely to end with a fine.[8] The most commonly mentioned non-fine form of punishment is for village authorities to prevent noncompliant community members from accessing their farms (9%), thus cutting off a major source of income. Taken together, 78% of respondents believe non-compliant individuals will face some consequences at the hand of either village or chiefdom level authorities.[9] Appendix Table B.1 breaks out these statistics at the village and chiefdom level.

[5]Chiefdom authorities distribute receipt books to village chiefs based on the village's population. Village chiefs then return completed receipt books along with the associated tax revenue. Therefore, chiefdom authorities can easily identify low compliance villages as the villages to which they have given receipt books that have not been completed and returned. To give this monitoring teeth, Chiefdom authorities may also levy fines on the village chief of low compliance villages or force these village chiefs to "buy" additional tax receipts, which the chief must then "sell" to his people (Interviews: 4; 27; 53; 207).

[6]Interviews prompted respondents with the following question: "Did village (Section/Chiefdom) leaders do anything to check if people had paid Local Tax this year (2022)? Or do they not do anything like that?". Fifty-two percent of respondents described village level monitoring mechanisms and 49% of respondents described monitoring mechanisms outside the village.

[7]Interview: 402

[8]After being taken to authorities respondents may be held for several hours ("they will detain you for one hour or two hours"(Interview: 404)

[9]Interviewers asked respondents "When village (section/chiefdom) leaders found out that someone had not paid, did they anything about it, or did they not do anything?"

Table B.1: Local Tax: Monitoring and Punishment Strategies

	Village	Chiefdom	Either
	(%)	(%)	(%)
Local Tax			
Monitoring: Any	**52**	**49**	**68**
Roadblocks / Checkpoint	11	28	32
Door-to-door checks	23	5	24
Authorities keep records	20	6	24
Punishment: Any	**69**	**55**	**78**
Fines	36	40	55
Taken to higher authorities	34	25	45
Banned from farming	9	2	9

Table B.1 presents qualitative evidence of TPIs' monitoring and punishment strategies for Local Tax. Monitoring interview prompt: "Did village (chiefdom) leaders do anything to check if people had paid Local Tax this year (2022)? Or do they not do anything like that?" Punishment interview prompt: "When village (chiefdom) leaders found out that someone had not paid, did they anything about it, or did they not do anything?" Percentages are rounded to the nearest integer.

Appendix C

Appendix to Chapter 4

C.1 Digital Town Halls: Pros and Cons

To begin with, participation can be less costly: If access to WhatsApp already exists, participants only need to invest a modest amount of time and mobile data to enter the DTH. Whereas offline THs enable participation only for a short and fixed time period, DTHs can be accessed for weeks and whenever it is convenient for participants. This flexibility reduces the oft significant opportunity costs of participation (Casey, 2018). Intuitively, transportation costs—traditionally a barrier to participation especially in rural settings (**?**, p.35)—are not incurred. Remarkable improvements in internet activity in developing countries—31 % of Sierra Leoneans in 2018 own a phone with internet access (Afrobarometer 2018)—have led to an explosion in social media usage (21.5% of Sierra Leonean report obtaining news through Facebook or Twitter at least "a few times a week" (Afrobarometer 2018). As our study population is property owners in the capital city, we expect these numbers to be even higher in our setting. In our model of mediated interaction through WhatsApp, participation is less costly for political representatives too: All that is required of them is to read a summary of participant contributions and to respond in a limited number of video and voice messages.

Second, perhaps counter-intuitively, we argue that DTHs hold more deliberative promise: In the Habermasian ideal type of deliberative democracy, participants engage in potentially endless communicative action (an exchange of reasoned arguments) as equals until the best argument prevails (Habermas, 1975). In offline THs, attendants regularly find themselves unable to make their views known in front of representatives as time constraints only allow for a limited number of contributions. Statements, especially from members of marginalized groups, are often interrupted by other participants (Parthasarathy et al., 2019). In contrast, DTHs allow all participants to make their views known without running the risk of interference by others. Importantly, DTHs alleviate the constraint of limited attention spans on successful argumentative reasoning: While it is easy to forget what a participant argued a few minutes ago in an offline TH, participants in WhatsApp can just scroll back. Whereas immediate reactions are required offline to ensure that the conversation stays on topic, DTHs enable participants to first reflect on their statement—in theory for multiple days—before posting it. Therefore, the longer time frame in a DTH should increase the argumentative quality of contributions and facilitate perspective taking (as the need for immediate reactions in offline DTHs precludes taking the time to reflect on where someone else's argument is coming from). Finally, we can avoid face-to-face interactions which in group settings under time constraints lend themselves to emotionalized exchanges (more cues are visible—e.g., body language and facial expressions—which make it harder to focus on the merits of the argument alone). Third, DTHs can alleviate one dimension of the well-known gap in political participation by targeting the relatively young who usually are less likely to participate in conventional forms of political engagement. Yet, it is to be expected that DTHs—just like their offline analogue—display additional participation biases (higher ability and willingness to participate among those able to afford smart phones and internet usage, the more educated and literate, those with higher political efficacy (on self-selection in offline TH participation, see Boulianne, 2019; Neblo et al., 2010).

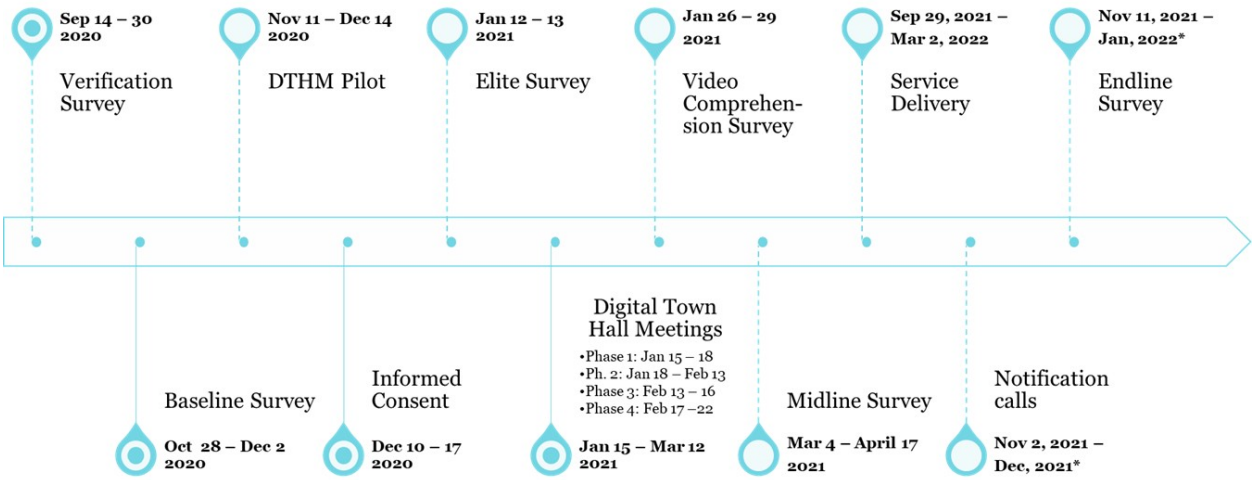
However, there are also potential relative disadvantages to the DTH format: The relative anonymity decreases the (reputational) cost of disruptive behavior as participants can choose how much identifying information they provide through their WhatsApp profile. Furthermore, moderating chats

can be costly, constrained by the functionalities provided by WhatsApp (messages can only be deleted by who wrote them) and, if done poorly, runs the risk of altering the conversation. The absence of face-to-face interactions can lead to questioning that one is actually talking to ones' representatives and fellow community members. Fortunately, this is less of a concern here as political representatives have prominently associated themselves with the DTH intervention in public. One may argue that voice- and text-based communication is less rich when other cues cannot be observed (e.g., the eyes as an indicator of the sincerity of the speaker). The mediated interaction between participants and representatives relies on trust in the intermediary that is aggregating the information. Perhaps most crucially, while DTHs reduce participation costs for many, those lacking internet/ WhatsApp access cannot participate. Finally, the brevity of text messages may not be conducive to the articulate elaboration of arguments (Jaidka et al., 2019). However, there are no length limitations in WhatsApp and participants have the option to record voice and video messages as well. Through our endline survey and by capturing all DTH conversations, we can measure many of the aforementioned potential disadvantages how prevalent they were.

C.2 Intervention Appendix

C.2 Project Timeline

Figure C.1: Project Timeline



Note: Notification calls and endline surveys in one ward, Tengbeh Town, were delayed by two months due to contractual issues with the construction firm.

C.2 Project Pictures

Figure C.2: Project implemented in Ward 418

Figure C.3: Project Implemented in Ward 442

Figure C.4: Project Implemented in Ward 444

C.2 Participation and Experience in DTH

Table C.1: Project Votes, Winning Projects, Implemented Projects

	Projects for vote					Replacement Projects	
	Water		Road repair		Solar	Water	
	Pipes	Pump	Tracks	Potholes	Street lights	Tank	Tap
Votes	429	313	138	51	83		
Won	19	9	2	1	0		
Built	0	9	3	0	0	8	11

Table C.1 reports projects the types of projects voted for, won, and built.

Table C.2: Organization, Implementation, Funding

	Perceived Responsible Actor				
Activity	FCC	Govt	Researchers	Citizens	Other
Organized	0.893	0.019	0.126	0.002	0.014
Implement	0.961	0.045	0.018	0.017	0.006
Funded	0.842	0.106	0.023	0.115	0.056

Table C.2 reports participants' perceptions of which actor(s) organized, implemented, and funded the DTHs. Participants were allowed to name multiple actors. Data from midline survey.

Table C.3: Participants' DTH Experiences

Question	Agree [0-5]
DTH gave space to voice views to political representatives	3.94
DTH facilitated better understanding of community members views	4.04
Budget (LE15 Million) sufficient to meaningful improve selected service	2.86
Participants comfortable making views known	3.82
Menu of services reflected services community wanted improved, given budget	3.33
Selected service will be delivered in the near future	3.58
Vote was fair and gave every participant the same influence	3.83

Table C.3 reports participants' experiences in the DTHs. We asked respondents seven questions. We asked questions in both positive and negative forms, so as to limit confirmation bias in the average response. Questions in the table are presented in the positive form.

C.3 Randomization and Treatment Assignment Appendix

C.3 Matching

We match property owners using the following covariates:

- Unconditional tax morale
- Service conditional tax morale
- Perceived probability of punishment for non-compliance
- Satisfaction with FCC service provision
- Tax reform awareness and support
- RDN received in 2019 or 2020
- Opportunities to voice opinion about FCC governance
- Willingness to believe member of opposing party
- Mayor approval
- FCC councilor approval
- Gender
- FCC responsiveness
- Age
- Property value
- Education

We generated matched-pairs using the *blockTools* package in *R*. We use the Optimal Greedy ("opt-Greedy") matching algorithm to find best matches along mahalanobis distance. In this matching process we weight certain variables higher than others, in line with our expectations that certain variables are a stronger predictor of our outcomes of interest. We place the greatest weight on our measure of unconditional tax morale—we expect this to be the strongest predictor to tax compliance, in line with the common use of this variable as proxy for tax compliance behavior. We place equal weight on another set of six measures from our baseline survey. Three of these measures are important factors in the literature on tax compliance: (i) service conditional tax morale, (ii) perceived likelihood of punishment for non-compliance, and (iii) satisfaction with FCC service

provision. We also place equal weight on the (iv) gender of the property owner, (v) their awareness and support of the property tax reform,[1] and (vi) the number of these five variables that were imputed.[2]

Table C.4 presents descriptive statistics and match weights for our matching variables. If a respondent refused to answer a question or said they "did not know" we imputed the value as the unconditional mean of the variable.[3] The last column displays the number of observations that were imputed for matching. Note that in general, the number of imputed responses is low.

Table C.4: Summary Statistics of Matching Variables

Variable name	**Weights**	**Mean**	**SD**	**Min**	**Max**	**n imputed**
Unconditional tax morale	1.10	3.77	1.55	1.00	5.00	25
Service conditional tax morale	1.00	1.96	0.96	1.00	3.00	11
Perceived probability of punishment	1.00	4.06	1.11	1.00	5.00	52
Satisfaction with FCC service provision	1.00	3.64	1.17	1.00	5.00	35
Gender (female = 1)	1.00	0.31	0.46	0.00	1.00	0
Reform awareness / support	1.00	2.38	0.67	1.00	3.00	19
RDN delivered 2019 or 2020	0.90	0.83	0.38	0.00	1.00	0
Opportunities for voice	0.10	2.13	0.99	1.00	4.00	174
Mayor approval	0.10	4.23	0.89	1.00	5.00	79
Councilor approval	0.10	2.73	1.22	1.00	5.00	122
FCC responsiveness	0.10	3.17	1.19	1.00	5.00	199
Believe opposition member	0.10	3.00	1.55	0.00	5.00	132
Age	0.09	51.77	12.93	20.00	100.00	11
Property tax value (USD)	0.09	60.25	87.45	2.88	1281.85	0
Education [0-2]	0.09	1.31	0.62	0.00	2.00	259

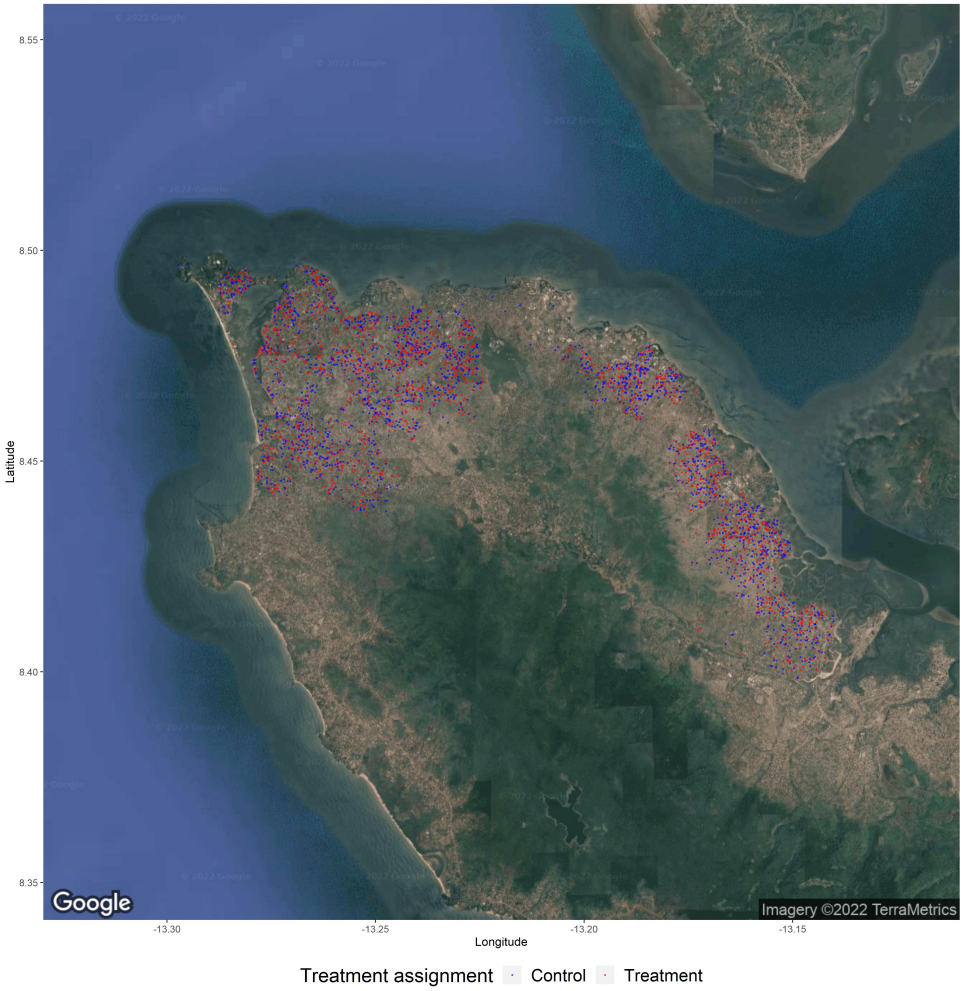
[1]We create a three level ordinal variable based on two survey items. A first group consists of respondents who have heard of the reform and strongly/somewhat support it; a second group consists of respondents who (a) have heard of the reform and feel neutral towards it and (b) have not heard of the reform; a third group consists of respondents who have heard of the reform and somewhat/strongly oppose it.

[2]This avoids matching observations with missing values on these key variables to observations that have non-missing values close to the mean.

[3]Following suggestions of: https://egap.org/resource/10-things-to-know-about-missing-data/

C.3 Treatment Assignment Map

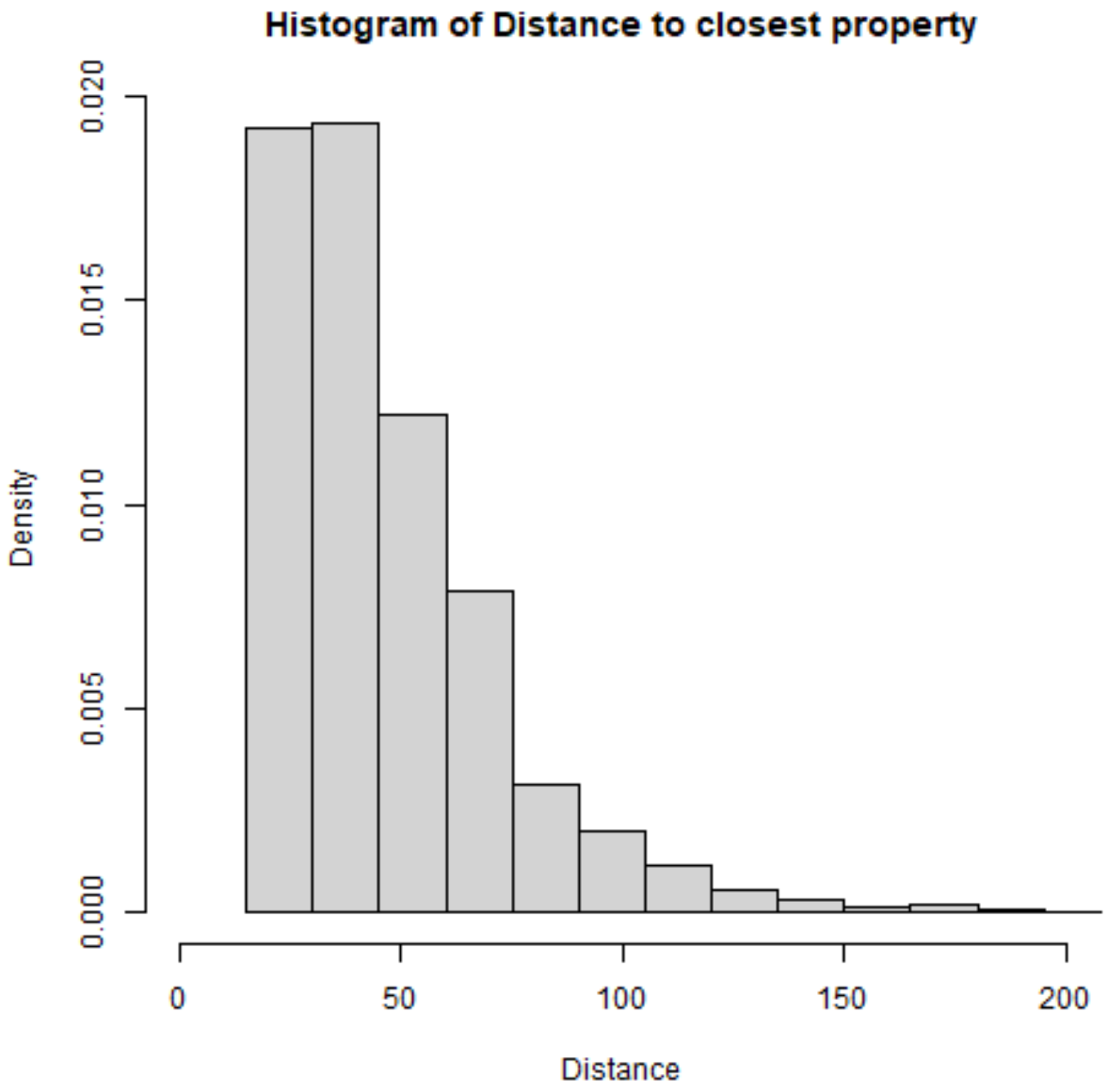
Figure C.5: Treatment Assignment Map



Note: This map plots treatment assignment across Freetown. Red points are treatment properties and blue points are control.

C.3 Distance to Closest Study Property

Figure C.6: Distance to Closest Study Property



Note: Histogram shows the minimum distance (in meters) between study properties. We used a restricted sampling technique to ensure the minimum distance was 15 meters.

C.4 Additional Tables and Figures

C.4 Compliance Treatment Effects

Table C.5: Alternative Operationalizations of Tax Compliance

	Total Paid		**Log Total Paid**		
Year	Est	p-value	Est	p-value	N
2021	39.22	0.19	-0.015	0.90	3618
	(29.99)		(0.12)		
2022	29.44	0.44	-0.050	0.66	3618
	(38.45)		(0.12)		

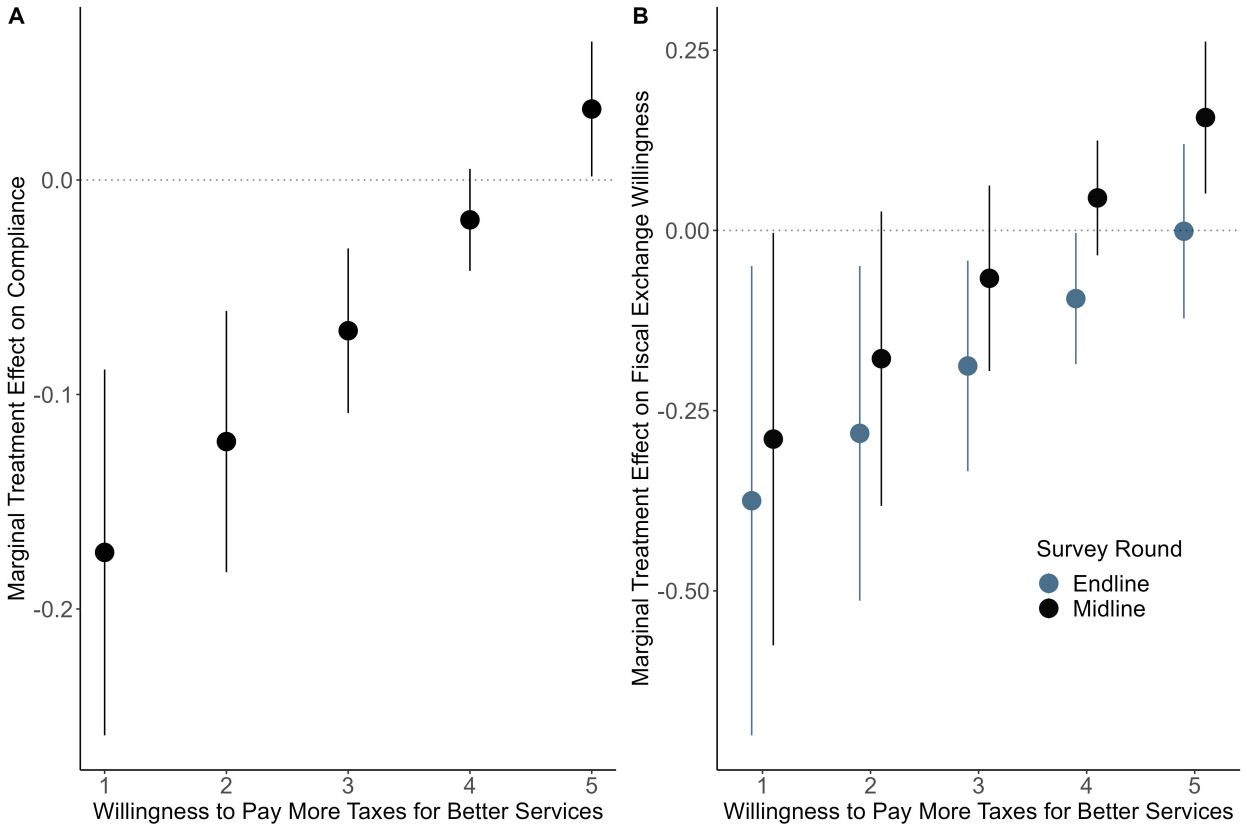
Table C.5. Columns 1 and 2 report the estimates and conventional p-values for total paid. Columns 3 and 4 report the estimates and conventional p-values for log total paid. Columns 1 and 3 report standard errors in parentheses. Treatment effects are reported in raw percentage points.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C.4 Conditional Treatment Effect

Plot B (Figure C.7), presents predicted marginal effects on attitudes towards fiscal exchange from a models that interacts treatment with our five-point measure of (baseline) attitudes towards fiscal exchange, using both midline and endline data. The interaction term is statistically significant (p-value = 0.034; $\beta = 0.11$ SDUs) in the model using midline data and at the threshold of conventional levels of statistical significance when using endline data (p-value = 0.12; $\beta = 0.093$ SDUs). In line with the expectations of motivated reasoning, treated property owners who were initially opposed to fiscal exchange *do not* positively update their attitudes towards fiscal exchange. In fact, in both midline and endline surveys, we find evidence that the intervention hardens the opposition of those initially opposed to fiscal exchange. By contrast, treated respondents who support fiscal become *more* supportive in the mid than their peers in the control group. While this effect disappears at endline, marginal effects of those supportive of fiscal exchange are larger than for those who are not supportive. Note that detecting positive treatment effects for the highly supportive group

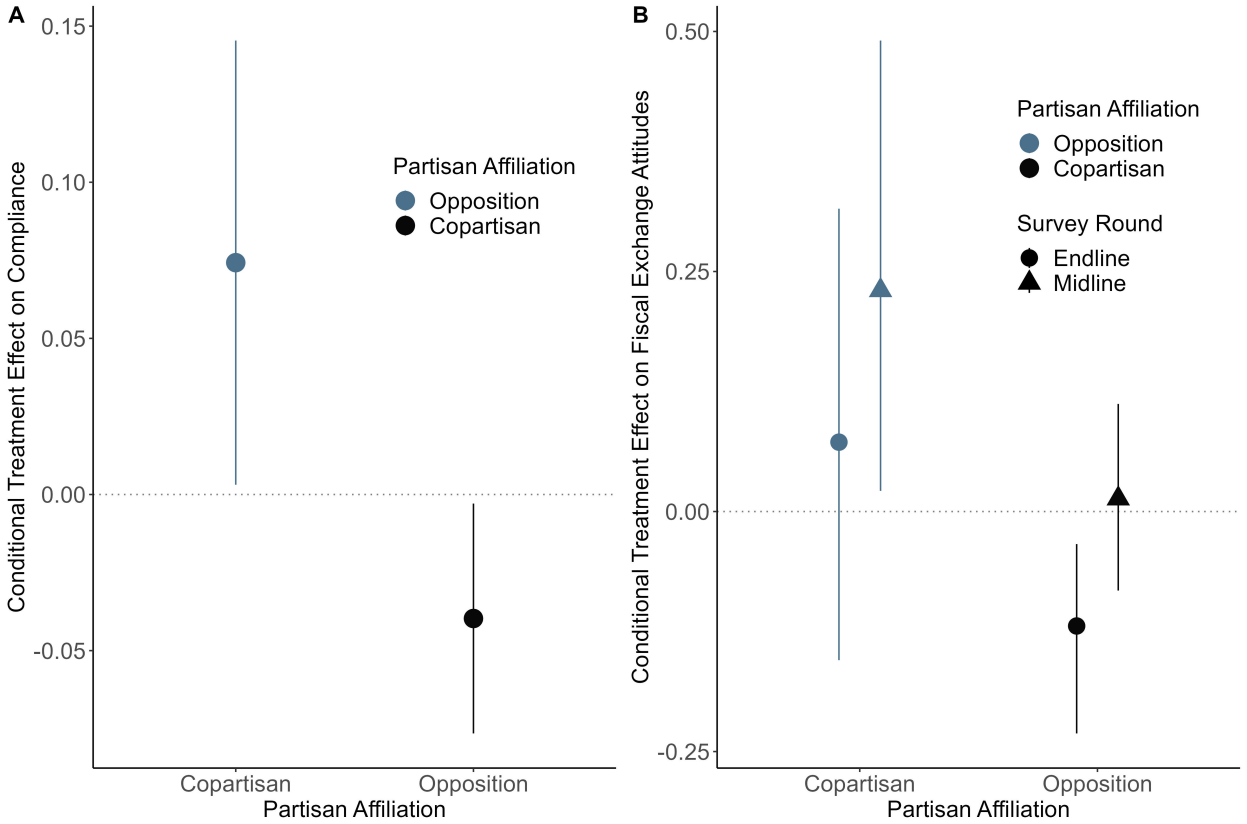
is difficult due to ceiling effects: for respondents who are fully supportive at baseline, positive treatment effects can only be observed by treated respondents being more likely to *remain* fully supportive.

Figure C.7: Marginal Treatment Effects by Attitudes Towards Fiscal Exchange



Note: Panel A reports marginal treatment effects on compliance conditional on baseline attitudes towards fiscal exchange. Panel B reports marginal treatment effects on attitudes towards fiscal exchange, conditional on baseline attitudes towards fiscal exchange.

Figure C.8: Treatment Effects Conditional on Partisan Affiliation



Note: Panel A reports marginal treatment effects on tax compliance behavior, conditional on partisan affiliation. Panel B reports marginal treatment effects on attitudes towards fiscal exchange, conditional on baseline attitudes towards fiscal exchange. In both panels, respondents who self-report affinity towards the All People's Congress are coded as "copartisans." All other respondents are coded as "opposition."

C.4 Political Engagement

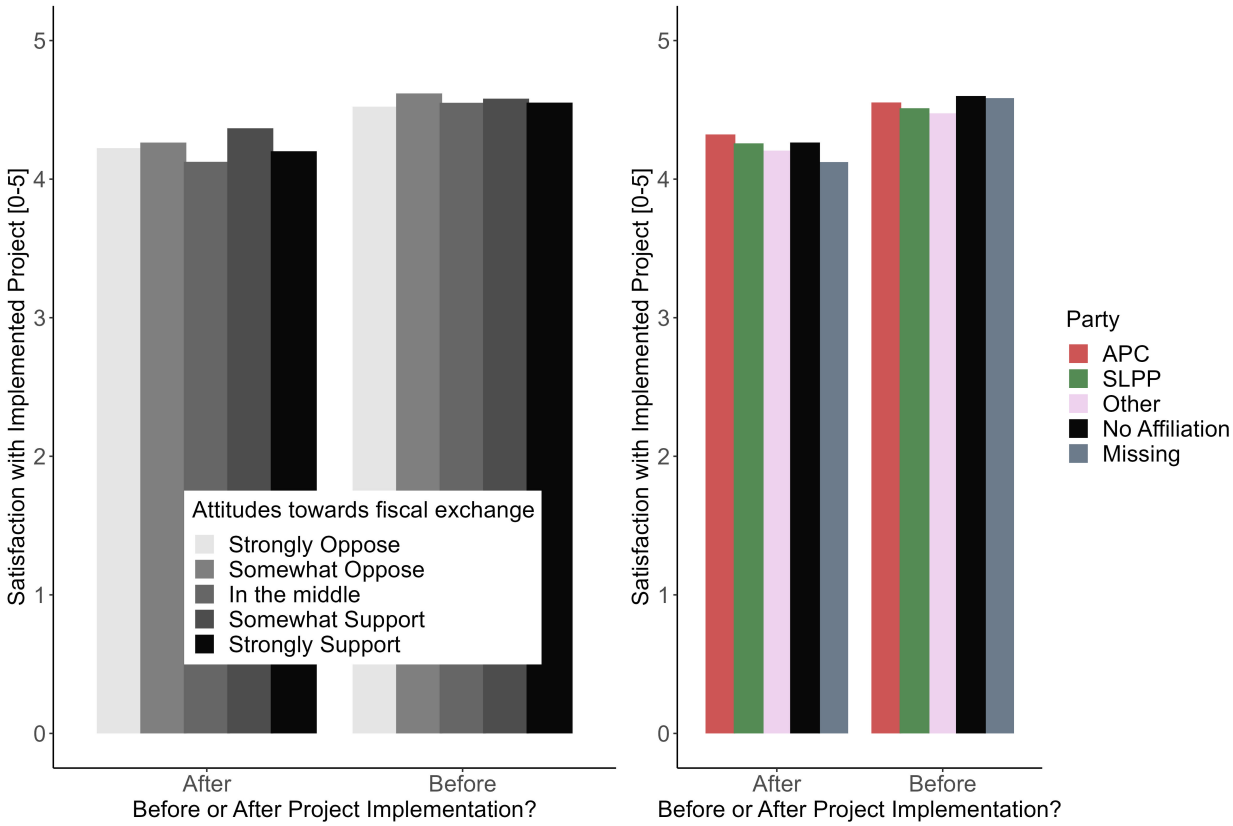
Table C.6: Political Engagement

	Baseline	Midline				Endline			
Outcome	Mean	Mean	Effect	N	q-val	Mean	Effect	N	q-val
Knows ward councillor name	0.360	0.383	0.104**	3,618	0.007	0.313	0.044	3,618	0.283
	(0.480)	(0.486)	(0.040)			(0.463)	(0.039)		
Attempted to contact ward councillor	0.193	0.188	0.214***	3,299	0.001	0.264	0.084	2,865	0.283
	(0.395)	(0.391)	(0.044)			(0.440)	(0.052)		
Attempted to contact MP	0.112	0.092	0.142***	3,297	0.001	0.159	-0.023	2,865	0.582
	(0.316)	(0.289)	(0.043)			(0.365)	(0.055)		
Level of interest in politics	1.841	1.818	0.161***	3,299	0.001	2.028	0.091	2,575	0.283
	(1.090)	(1.001)	(0.040)			(1.096)	(0.055)		
Level of interest in FCC activities	2.952	3.103	0.428***	3,300	0.001	3.157	-0.009	2,871	0.582
	(1.097)	(0.896)	(0.034)			(0.954)	(0.043)		
Attended political meeting	1.157	1.175	0.051	3,301	0.048	1.150	0.104*	2,558	0.283
	(0.501)	(0.545)	(0.049)			(0.516)	(0.058)		

Table C.6 reports the effect of the treatment on political engagement measures. Columns 1, 2, and 6 report the control group mean for each indicator for the baseline, midline, and endline surveys, respectively, with the standard deviation in parentheses. Column 3 presents treatment effects estimates at the midline survey and Column 7 presents treatment effects estimates at the endline survey. Columns 4 and 8 reports the number of non-missing observations in the midline survey and endline survey, respectively. Stars refer to randomization inference p-values. Columns 5 and 9 report corrected q-values, which adjust for multiple hypothesis testing, following Anderson (2008). Reported effects are standardized effects. Attempts to contact MP or Councilor, or attendance at political meeting, are for the last six months.
Significance: $^{*}p < 0.10$; $^{**}p < 0.05$ $^{***}p < 0.01$

C.4 Service Satisfaction by Group

Figure C.9: Service Satisfaction by Group



C.5 Spillover

We use a design-based strategy to estimate spillovers that occur due to geographic proximity between properties. For this analysis, we focus on tax compliance spillovers from treated properties to 74,352 properties outside of our study.[4] Our approach compares non-study properties geographically proximate to treated study properties to non-study properties proximate to control study properties.[5] We estimate spillovers with the following equation:

$$Y_{i_{2022}} = \beta_1 SPILL_i + \gamma Y_{i_{2020}} + \lambda \mathbf{X}_i + \delta_w + \epsilon_i \tag{C.1}$$

Where $Y_{i_{2022}}$ is the binary tax compliance outcome of non-study property owner *i* in 2022; $SPILL_i$ is a dummy variable equal to 1 if there is at least one treated study property *close* to non-study property owner *i*. Therefore, β_1 captures the spillover effect on tax compliance of being close to a treated property owner. $Y_{i_{2020}}$ is the tax compliance behavior of property owner *i* in 2020; δ is a vector of ward fixed effects; $\mathbf{X}$ is the set of property-level characteristics described in Section 4.4, included as covariate adjustment.

As the density of buildings varies across the city, the probability of being assigned to "spillover treatment" (i.e. the probability that $SPILL_i$ is equal to one in equation C.1) varies across properties. That is, non-study properties in denser areas are more likely to be assigned to spillover treatment because they are more likely to be close to more study units. In this context, unweighted regressions can be biased because building density (and therefore treatment assignment) may also be correlated with compliance behavior.[6] To address this, we weight observations by the inverse

[4] While we observe compliance outcomes for 95,769 properties that are not eligible for the intervention, some individuals own multiple properties. Intuitively, the effects of the DTH should only spillover to affect the compliance behavior of a proximate non-study property when the property owner is living there. As we lack data on the residence of property owners who own more than one property, we assume that these multiple property owners are living in their highest value property. Therefore, our spillover analysis is restricted to the set of 74,352 non-study properties that are the highest value property registered to a given property owner.

[5] See Miguel and Kremer (2004) for an example of a (prominent) study that uses non-experimental units (i.e., units that are not themselves part of the randomization) to estimate spillovers.

[6] Imagine, for example, potential differences in compliance behavior between densely packed informal settlements

probability of being assigned their spillover treatment condition, where assignment probability is calculated by re-simulating treatment assignment of study properties (Blattman et al., 2021; Gerber and Green, 2012; Chen et al., 2010). Note that this implies non-study properties that are not close study property are weighted zero (i.e., not used to calculate spillover effects). Estimating spillovers crucially depends on choosing a distance threshold to define non-study properties as "close" to study properties. We pre-specified this distance as 64 meters, believing that it would maximize the precision of our estimates, without downward biasing them.[7]

Panel A in Table C.7 shows spillover effects on compliance behavior at this preregistered threshold distance. Column 1 shows results for compliance behavior in 2022, our preregistered primary dependent variable for the spillover analysis. While the point estimate at this distance threshold is positive, about a third of a percentage point, it is statistically insignificant (the RI p-value is in brackets under the estimate). As noted in our preanalysis plan, the selection of this preferred distance threshold is somewhat arbitrary; the additional results in Column 1 show the estimated spillover effect when the distance threshold is defined below (Panel B) or above (Panel C) the preregistered threshold. The estimated effect is positive at all thresholds, and approaches statistical significance at some, but the results are at best suggestive. While we cannot reject the null of no spillover effect at our preregistered threshold, one interpretation of the results in Column 1 is that there is a positive spillover effect, but our estimation is noisy.

and spacious affluent neighborhoods.

[7] Absent a theory-driven procedure for selecting the threshold distance (D), we opt for a pragmatic approach. While the overall number of non-study properties used in the spillover estimation increases with higher values of D, the number of spillover control units is maximized when D equals 64 meters. Values of D greater than 64 have increasing units in the spillover treatment condition, but decreasing units of spillover control units. Given that the motivation for selecting higher values of D is to increase precision, selecting a value of D greater than 64 meters requires that the loss of precision brought on by the decline of units in the control arm is outweighed by increase in precision due to additional units entering into the treatment arm. When D is equal to 64 meters the treatment spillover arm has 24,177 units, compared to 10,637 units in spillover control; therefore, we privilege maintaining control units over gaining treatment units.

Table C.7: Spillover Effects

Threshold Distance	Effect on Compliance		N observations	
	2022	'22 or '23	Treatment	Control
Preregistered Threshold Distance				
64	0.34	1.20*	24,214	10,585
	[0.60]	[0.08]		
Below Preregistered Threshold				
60	0.66	1.48**	22,366	10,514
	[0.25]	[0.04]		
50	0.28	1.30*	17,222	9,964
	[0.70]	[0.07]		
40	0.79	1.34**	12,103	8,378
	[0.25]	[0.04]		
35	1.28*	1.58*	9,653	7,264
	[0.08]	[0.05]		
30	1.00	1.39	7,259	5,860
	[0.27]	[0.12]		
Above Preregistered Threshold				
70	0.42	1.01	26,885	10,441
	[0.48]	[0.16]		
80	0.49	0.99	31,016	9,782
	[0.44]	[0.19]		
90	0.75	1.16	34,679	8,814
	[0.23]	[0.22]		
100	0.80	1.21	37,729	7,906
	[0.23]	[0.22]		

Table C.7 reports spillover effects on the compliance behavior of non-study property owners, at different distance thresholds for defining spillover units (Column 1). Column 2-3 reports spillover treatment effects on compliance behavior. In Column 2, the dependent variable is a dummy indicating if the owner paid any tax in 2022. In Column 3, the dependent variable is a dummy indicating if the owner paid any tax in 2022 or 2023. Treatment effects are reported in raw percentage points. Randomization inference *p*-values are below each estimate in brackets. Stars refer to randomization inference *p*-values. Columns 4 and 5 refer to the number of observations in treatment and control, respectively, at a given distance threshold.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

One way to reduce noise is to pool tax compliance behavior across 2022 and 2023, such that the

dependent variable becomes paying taxes in either 2022 or 2023 (Column 2).[8] Treatment effects for these pooled estimates at the preregistered threshold are substantively large (1.2 percentage points) and statistically significant (p-value = .08). Moreover, these results appear robust to the definition of the threshold distance. When the threshold is defined below the preregistered threshold (Panel B), point estimates are larger and estimates are statistically significant. When the threshold is defined above the preregistered threshold (Panel C), point estimates are of a similar magnitude and RI p-values hover at the threshold of statistical significance.

[8]This strategy will usefully increase precision if the spillover effect persists into 2023. Conversely, if the spillover effect only existed in 2022, pooling with 2023 would make it harder to observe effects.

Bibliography

Acemoglu, D., Reed, T., and Robinson, J. A. (2014). Chiefs: Economic Development and Elite Control of Civil Society in Sierra Leone. *Journal of Political Economy*, 122(2):319–368.

Alatas, V., Banerjee, A., Hanna, R., Olken, B. A., Purnamasari, R., and Wai-Poi, M. (2019). Does Elite Capture Matter? Local Elites and Targeted Welfare Programs in Indonesia. 109:334–339.

Allingham, M. G. and Sandmo, A. (1972). Income Tax Evasion: A Theoretical Analysis. *Journal of Public Economics*, 1(3-4):323–338.

Alm, J., Jackson, B. R., and McKee, M. (1993). Fiscal Exchange, Collective Decision Institutions, and Tax Compliance. *Journal of Economic Behavior & Organization*, 22(3):285–303.

Anderson, M. L. (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American statistical Association*, 103(484):1481–1495.

Badrinathan, S. (2021). Educative interventions to combat misinformation: Evidence from a field experiment in india. *American Political Science Review*, 115(4):1325–1341.

Bahl, R. W. and Vazquez, J. M. (2008). The property tax in developing countries: Current practice and prospects. In *In Toward a Vision of Land in 2015: International Perspectives*, pages 23–46. Lincoln Institute of Land Policy.

Balán, P., Bergeron, A., Tourek, G., and Weigel, J. L. (2022). Local Elites as State Capacity: How City Chiefs Use Local Information to Increase Tax Compliance in the Democratic Republic of the Congo. *American Economic Review*, 112(3):762–97.

Baldwin, K. (2016). *The Paradox of Traditional Chiefs in Democratic Africa*. Cambridge University Press.

Baldwin, K. (2019). Elected MPs, Traditional Chiefs, and Local Public Goods: Evidence on the Role of Leaders in Co-production from Rural Zambia. *Comparative Political Studies*, 52(12):1925–1956.

Baldwin, K. and Holzinger, K. (2019). Traditional Political Institutions and Democracy: Reassessing Their Compatibility and Accountability. *Comparative Political Studies*, 52(12):1747–1774.

Baldwin, K., Kao, K., and Lust, E. (2023). Is Authority Fungible? Legitimacy, Domain Congruence, and the Limits of Power in Africa. *American Journal of Political Science*.

Baldwin, K., Muyengwa, S., Mvukiyehe, E., et al. (2022). Accountability and Inclusion in Customary Institutions: Evidence from a Village-Level Experiment in Zimbabwe. *Quarterly Journal of Political Science*, 17(2):283–315.

Baldwin, K. and Raffler, P. (2019). Traditional Leaders, Service Delivery, and Electoral Accountability. *Decentralized Governance and Accountability*, pages 61–90.

Barber, M. and Pope, J. C. (2019). Does Party Trump Ideology? Disentangling Party and Ideology in America. *American Political Science Review*, 113(1):38–54.

Bargain, O. and Aminjonov, U. (2020). Trust and Compliance to Public Health Policies in Times of COVID-19. *Journal of Public Economics*, 192:104316.

Basurto, M. P., Dupas, P., and Robinson, J. (2020). Decentralization and efficiency of subsidy targeting: Evidence from chiefs in rural malawi. *Journal of public economics*, 185:104047.

Bates, R. H. and Lien, D. D.-H. (1985). A note on taxation, development, and representative government. *Politics & Society*, 14(1):53–70.

Beetham, D. (2013). *The Legitimation of Power*. Bloomsbury Publishing.

Besley, T. (2020). State Capacity, Reciprocity, and the Social Contract. *Econometrica*, 88(4):1307–1335.

Besley, T., Burgess, R., Khan, A., and Xu, G. (2022). Bureaucracy and Development. *Annual Review of Economics*, 14:397–424.

Besley, T. and Dray, S. (2024). Trust and State Effectiveness: The Political Economy of Compliance. *The Economic Journal*.

Besley, T. and Persson, T. (2008). Wars and State Capacity. *Journal of the European Economic Association*, 6(2-3):522–530.

Besley, T. and Persson, T. (2011). *Pillars of Prosperity*. Princeton University Press.

Besley, T. and Persson, T. (2014). Why Do Developing Countries Tax So Little? *Journal of Economic Perspectives*, 28(4):99–120.

Bhavnani, R. R. and Lee, A. (2018). Local Embeddedness and Bureaucratic Performance: Evidence From India. *The Journal of Politics*, 80(1):71–87.

Blattman, C., Green, D. P., Ortega, D., and Tobón, S. (2021). Place-based Interventions at Scale: The Direct and Spillover Effects of Policing and City Services on Crime. *Journal of the European Economic Association*, 19(4):2022–2051.

Bó, P. D., Foster, A., and Putterman, L. (2010). Institutions and Behavior: Experimental Evidence on the Effects of Democracy. *American Economic Review*, 100(5):2205–2229.

Bolsen, T., Druckman, J. N., and Cook, F. L. (2014). The Influence of Partisan Motivated Reasoning on Public Opinion. *Political Behavior*, 36:235–262.

Boulianne, S. (2019). Building Faith in Democracy: Deliberative Events, Political Trust and Efficacy. *Political Studies*, 67(1):4–30.

Brierley, S., Lowande, K., Potter, R. A., and Toral, G. (2023). Bureaucratic Politics: Blind Spots and Opportunities in Political Science. *Annual Review of Political Science*, 26:271–290.

Brierley, S. and Ofosu, G. K. (2023). Chiefs' Endorsements and Voter Behavior. *Comparative Political Studies*, pages 1–32.

Brockmeyer, A., Estefan, A., Arras, K. R., and Serrato, J. C. S. (2021). Taxing Property in Developing Countries: Theory and Evidence from Mexico. Technical report, National Bureau of Economic Research.

Brockmeyer, A. and Sáenz Somarriba, M. (2022). Electronic Payment Technology and Tax Compliance: Evidence from Uruguay's Financial Inclusion Reform. *CEPR Discussion Paper No. DP17097*.

Broockman, D. E. and Butler, D. M. (2017). The Causal Effects of Elite Position-taking on Voter Attitudes: Field Experiments with Elite Communication. *American Journal of Political Science*, 61(1):208–221.

Buur, L. and Kyed, H. M. (2007). *State Recognition and Democratization in Sub-Saharan Africa: A New Dawn for Traditional Authorities?* Springer.

Callen, M., Weigel, J. L., and Yuchtman, N. (2023). Experiments About Institutions. *Annual Review of Economics*, 16.

Carlson, E. and Seim, B. (2020). Honor Among Chiefs: An Experiment on Monitoring and Diversion Among Traditional Leaders in Malawi. *The Journal of Development Studies*, 56(8):1541–1557.

Casey, K. (2018). Radical Decentralization: Does Community-driven Development Work? *Annual Review of Economics*, 10:139–163.

Casey, K., Glennerster, R., Miguel, E., and Voors, M. (2018). Skill Versus Voice in Local Development. *The Review of Economics and Statistics*, pages 1–46.

Castro, L. and Scartascini, C. (2015). Tax Compliance and Enforcement in the Pampas: Evidence From a Field Experiment. *Journal of Economic Behavior & Organization*, 116:65–82.

Cheeseman, N. and Peiffer, C. (2022). The Curse of Good Intentions: Why Anticorruption Messaging Can Encourage Bribery. *American Political Science Review*, 116(3):1081–1095.

Chen, J., Humphreys, M., and Modi, V. (2010). Technology Diffusion and Social Networks: Evidence From a Field Experiment in Uganda. *Manuscript, Columbia University*.

Christensen, D. (2018). The Geography of Repression in Africa. *Journal of Conflict Resolution*, 62(7):1517–1543.

Christensen, D. and Garfias, F. (2021). The Politics of Property Taxation: Fiscal Infrastructure and Electoral Incentives in Brazil. *The Journal of Politics*, 83(4):1399–1416.

Cingolani, L. (2013). The State of State Capacity: a review of concepts, evidence and measures.

Crowder, M. (1964). Indirect Rule—French and British style. *Africa*, 34(3):197–205.

Crowder, M. (1968). *West Africa Under Colonial Rule*. Hutchinson.

De Kadt, D. and Larreguy, H. A. (2018). Agents of the Regime? Traditional Leaders and Electoral Politics in South Africa. *The Journal of Politics*, 80(2):382–399.

De Neve, J.-E., Imbert, C., Spinnewijn, J., Tsankova, T., and Luts, M. (2021). How to improve tax compliance? evidence from population-wide experiments in belgium. *Journal of Political Economy*, 129(5):1425–1463.

Díaz-Cayeros, A., Magaloni, B., and Ruiz-Euler, A. (2014). Traditional governance, citizen engagement, and local public goods: Evidence from mexico. *World Development*, 53:80–93.

Doner, R. F., Ritchie, B. K., and Slater, D. (2005). Systemic Vulnerability and the Origins of Developmental States: Northeast and Southeast Asia in Comparative Perspective. *International Organization*, 59(2):327–361.

Druckman, J. N. and McGrath, M. C. (2019). The Evidence For Motivated Reasoning in Climate Change Preference Formation. *Nature Climate Change*, 9(2):111–119.

Englebert, P. (2002). Born-again Buganda or the Limits of Traditional Resurgence in Africa. *The Journal of Modern African Studies*, 40(3):345–368.

Englebert, P. (2005). Back To the Future? Resurgent Indigenous Structures and the Reconfiguration of Power in Africa. *Tradition and Politics: Indigenous Political Structures in Africa*, pages 33–60.

Esterling, K. M., Neblo, M. A., and Lazer, D. M. (2011). Means, Motive, and Opportunity in Becoming Informed About Politics: A Deliberative Gield Experiment With Members of Congress and Their Constituents. *Public Opinion Quarterly*, 75(3):483–503.

Evans, P. B. (1989). Predatory, Developmental, and Other Apparatuses: A Comparative Political Economy Perspective on the Third World State. In *Sociological Forum*, volume 4, pages 561–587. Springer.

Fanthorpe, R. (2004). Chiefdom Governance Reform Programme Public Workshops: An Analysis of the Facilitators' Reports. *Unpublished report, University of Sussex, Anthropology Department*.

Fanthorpe, R., Lavali, A., and Sesay, M. G. (2011). Decentralization in Sierra Leone: Impact, Constraints and Prospects. *DFID Sierra Leone*.

FCC (2021). Transform Freetown: Second Year Report. Accessed at: https://fcc.gov.sl/transform-freetown-second-year-report-2020-2021/.

Fehr, E. and Gächter, S. (2002). Altruistic Punishment in Humans. *Nature*, 415(6868):137–140.

Finan, F., Olken, B. A., and Pande, R. (2017). The Personnel Economics of the Developing State. *Handbook of Economic Field Experiments*, 2:467–514.

Fjeldstad, O.-H., Chambas, G., and Brun, J.-F. (2014). Local Government Taxation in Sub-Saharan Africa: A Review and an Agenda for Research. *CMI Working Paper*.

Flores, A., Cole, J. C., Dickert, S., Eom, K., Jiga-Boy, G. M., Kogut, T., Loria, R., Mayorga, M., Pedersen, E. J., Pereira, B., et al. (2022). Politicians Polarize and Experts Depolarize Public Support for COVID-19 Management Policies Across Countries. *Proceedings of the National Academy of Sciences*, 119(3):e2117543119.

Fujiwara, T. and Wantchekon, L. (2013). Can Informed Public Deliberation Overcome Clientelism? Experimental Evidence from Benin. *American Economic Journal: Applied Economics*, 5(4):241–55.

Fukuyama, F. (2013). What Is Governance? *Governance*, 26(3):347–368.

Garimella, K. and Eckles, D. (2020). Images and Misinformation in Political Groups: Evidence from WhatsApp in India. *Harvard Kennedy School Misinformation Review*.

Gerber, A. S. and Green, D. P. (2012). *Field Experiments: Design, Analysis, and Interpretation*. WW Norton.

Goldstein, M. and Udry, C. (2008). The Profits of Power: Land Rights and Agricultural Investment in Ghana. *Journal of Political Economy*, 116(6):981–1022.

Gonçalves, S. (2014). The Effects of Participatory Budgeting On Municipal Expenditures and Infant Mortality in Brazil. *World Development*, 53:94–110.

Gottlieb, J. and Hollenbach, F. M. (2018). Fiscal capacity as a moderator of the taxation-accountability hypothesis. Technical report, University of Texas Working Paper.

Gottlieb, J., LeBas, A., and Magat, J. (2021). Can Social Intermediaries Build the State? Evidence from an Experiment in Lagos, Nigeria.

Grieco, K. (2024). Can Rural Property Tax Generate Revenue? A Simple Accounting Exercise in Sierra Leone. Technical report.

Grieco, K., Meriggi, N. F., Michel, J., Prichard, W., Stewart-Wilson, G., et al. (2019). Simplifying Property Tax Administration in Africa: Piloting a Points Based Valuation in Freetown, Sierra Leone.

Grieco, K., Yusuf, Y., and Meriggi, N. (2020). Rapid Country Study: Sierra Leone. *COVID-19 Series*.

Haas, N. and Khadka, P. B. (2020). If They Endorse It, I Can't Trust It: How Outgroup Leader Endorsements Undercut Public Support for Civil War Peace Settlements. *American Journal of Political Science*, 64(4):982–1000.

Habermas, J. (1975). *Legitimation Crisis*, volume 519. Beacon Press.

Hanson, J. K. and Sigman, R. (2021). Leviathan's Latent Dimensions: Measuring State Capacity for Comparative Political Research. *The Journal of Politics*, 83(4):1495–1510.

Henn, S. J. (2023). Complements or Substitutes? How Institutional Arrangements Bind Traditional Authorities and the State in Africa. *American Political Science Review*, 117(3):871–890.

Herbst, J. (1990). War and the State in Africa. *International Security*, 14(4):117–139.

Herbst, J. (2014). *States and Power in Africa: Comparative Lessons in Authority and Control*, volume 149. Princeton University Press.

Hetherington, M. J. and Rudolph, T. J. (2008). Priming, Performance, and the Dynamics of Political Trust. *The Journal of Politics*, 70(2):498–512.

Holzinger, K., Kern, F. G., and Kromrey, D. (2016). The Dualism of Contemporary Traditional Governance and the State: Institutional Setups and Political Consequences. *Political Research Quarterly*, 69(3):469–481.

Holzinger, K., Kern, F. G., and Kromrey, D. (2020). Explaining the Constitutional Integration and Resurgence of Traditional Political Institutions in Sub-Saharan Africa. *Political Studies*, 68(4):973–995.

Honig, L. (2019). Traditional Leaders and Development in Africa. In *Oxford research encyclopedia of politics*.

Huntington, S. P. (1968). *Political Order In Changing Societies*. Yale University Press.

Jackson, P. (2007). Reshuffling an Old Deck of Cards? The Politics of Local Government Reform in Sierra Leone. *African Affairs*, 106(422):95–111.

Jaidka, K., Zhou, A., and Lelkes, Y. (2019). Brevity is the Soul of Twitter: The Constraint Affordance and Political Discussion. *Journal of Communication*, 69(4):345–372.

Keele, L. (2007). Social Capital and the Dynamics of Trust in Government. *American Journal of Political Science*, 51(2):241–254.

Khan, A. Q., Khwaja, A. I., Olken, B. A., and Shaukat, M. (2022). Rebuilding the Social Compact: Urban Service Delivery and Property Taxes in Pakistan. IGC.

King, G., Gakidou, E., Ravishankar, N., Moore, R. T., Lakin, J., Vargas, M., Téllez-Rojo, M. M., Hernández Ávila, J. E., Ávila, M. H., and Llamas, H. H. (2007). A 'Politically Robust' Experimental Design for Public Policy Evaluation, with Application to the Mexican Universal HealthInsurance Program. *Journal of Policy Analysis and Management*, 26(3):479–506.

Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S., and Saez, E. (2011). Unwilling or Unable To Cheat? Evidence From a Tax Audit Experiment in Denmark. *Econometrica*, 79(3):651–692.

Kling, J. R., Liebman, J. B., and Katz, L. F. (2007). Experimental Analysis of Neighborhood Effects. *Econometrica*, 75(1):83–119.

Koter, D. (2013). King Makers: Local Leaders and Ethnic Politics in Africa. *World Politics*, 65(2):187–232.

Krämer, M. (2016). Neither Despotic Nor Civil: The Legitimacy of Chieftaincy in its Relationship with the ANC and the State in KwaZulu-Natal (South Africa). *The Journal of Modern African Studies*, 54(1):117–143.

Kunda, Z. (1990). The Case for Motivated Reasoning. *Psychological Bulletin*, 108(3):480.

Lee, Y. and Gordon, R. H. (2005). Tax structure and economic growth. *Journal of Public Economics*, 89(5-6):1027–1043.

Levi, M. (1988). *Of Rule and Revenue*. University of California Press.

Levi, M. (1997). *Consent, Dissent, and Patriotism*. Cambridge University Press.

Levi, M., Sacks, A., and Tyler, T. (2009). Conceptualizing Legitimacy, Measuring Legitimating Beliefs. *American Behavioral Scientist*, 53(3):354–375.

Levi, M. and Stoker, L. (2000). Political Trust and Trustworthiness. *Annual Review of Political Science*, 3(1):475–507.

Lin, W. (2013). Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman's Critique. *The Annals of Applied Statistics*, 7(1):295–318.

Locke, J. (1690). *Second Treatise of Government*. Awnsham Churchill. Project Gutenberg, 2021.

Lodge, M. and Taber, C. S. (2013). *The Rationalizing Voter*. Cambridge University Press.

Logan, C. (2013). The Roots of Resilience: Exploring Popular Support for African Traditional Authorities. *African Affairs*, 112(448):353–376.

Logan, C. and Amakoh, K. (2022). African Citizens' Message to Traditional Leaders: Stay in Development, Stay Out of Politics. *Afrobarometer Dispatch No. 248*.

López-Moctezuma, G., Wantchekon, L., Rubenson, D., Fujiwara, T., and Pe Lero, C. (2022). Policy Deliberation and Voter Persuasion: Experimental Evidence from an Election in the Philippines. *American Journal of Political Science*, 66(1):59–74.

Lust, E., Kao, K., Landry, P., Harris, A., Dulani, B., Metheney, E., Nickel, S., Carlitz, R., Gakii Gatua, J., Jöst, P., Mechkova, V., Mujenja, M. F., Tengatenga, J., Grimes, M., Ahsan Jansson, C., Alfonso, W., Nyasente, D., Ben Brahim, N., Jordan, J., Bauhr, M., Boräng, F., Ferree, K., Hartmann, F., and Lueders, H. (2019). The Local Governance and Performance Index 2019.

Luttmer, E. F. and Singhal, M. (2014). Tax Morale. *Journal of Economic Perspectives*, 28(4):149–168.

Magaloni, B., Díaz-Cayeros, A., and Ruiz Euler, A. (2019). Public Good Provision and Traditional Governance in Indigenous Communities in Oaxaca, Mexico. *Comparative Political Studies*, 52(12):1841–1880.

Mamdani, M. (1996). *Citizen and Subject: Contemporary Africa and the Legacy of Late colonialism*. Princeton University Press.

Manning, R. E. (2009). The Landscape of Local Authority in Sierra Leone: How "Traditional" and "Modern" Justice and Governance Systems Interact. *Justice and Development Working Paper Series 1.1*.

Martin, L. (2014). Taxation, Loss Aversion, and Accountability: Theory and Experimental Evidence for Taxation's Effect on Citizen Behavior. *Unpublished paper, Yale University, New Haven, CT*.

Martin, L. E. (2023). *Strategic Taxation: Fiscal Capacity and Accountability in African States*. Oxford University Press.

Maru, V. (2006). Between Law and Society: Paralegals and the Provision of Justice Services in Sierra Leone and Worldwide. *Yale J. Int'l L.*, 31:427.

Mengisteab, K. (2017). Institutional Fragmentation in Africa and its Implications. In *Traditional Institutions in Contemporary African Governance*, pages 1–13. Routledge.

Michalopoulos, S. and Papaioannou, E. (2020). Historical Legacies and African Development. *Journal of Economic Literature*, 58(1):53–128.

Migdal, J. S. (1988). *Strong Societies and Weak States: State-Society Relations and State Capabilities in the Third World*. Princeton University Press.

Miguel, E. and Kremer, M. (2004). Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities. *Econometrica*, 72(1):159–217.

Mokuwa, E., Voors, M., Bulte, E., and Richards, P. (2011). Peasant Grievance and Insurgency in Sierra Leone: Judicial Serfdom as a Driver of Conflict. *African Affairs*, 110(440):339–366.

Mullinix, K. J. (2016). Partisanship and Preference Formation: Competing Motivations, Elite polarization, and Issue importance. *Political Behavior*, 38:383–411.

Nathan, N. L. (2019). Electoral Consequences of Colonial Invention: Brokers, Chiefs, and Distribution in Northern Ghana. *World Politics*, 71(3):417–456.

Neblo, M. A., Esterling, K. M., Kennedy, R. P., Lazer, D. M., and Sokhey, A. E. (2010). Who Wants to Deliberate–and Why? *American Political Science Review*, pages 566–583.

Nicholson, S. P. (2012). Polarizing Cues. *American Journal of Political Science*, 56(1):52–66.

North, D. C. and Weingast, B. R. (1989). Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in Seventeenth-Century England. *The Journal of Economic History*, 49(4):803–832.

Ntsebeza, L. (2005). *Democracy Compromised: Chiefs and the Politics of the Land in South Africa*, volume 5. Brill.

Nyhan, B. (2021). Why the Backfire Effect Does Not Explain the Durability of Political Misperceptions. *Proceedings of the National Academy of Sciences*, 118(15):e1912440117.

Nyhan, B. and Reifler, J. (2010). When Corrections Fail: The Persistence of Political Misperceptions. *Political Behavior*, 32(2):303–330.

Okunogbe, O. and Santoro, F. (2023). The Promise and Limitations of Information Technology for Tax Mobilization. *The World Bank Research Observer*, 38(2):295–324.

Okunogbe, O. and Tourek, G. (2024). How Can Lower-Income Countries Collect More Taxes? The Role of Technology, Tax Agents, and Politics. *Journal of Economic Perspectives*, 38(1):81–106.

Olson, M. ([1965] 1971). *The Logic of Collective Action: Public Goods and the Theory of Groups*, volume 124. Harvard University Press.

Olson, M. (1993). Dictatorship, Democracy, and Development. *American Political Science Review*, 87(3):567–576.

Ostrom, E., Walker, J., and Gardner, R. (1992). Covenants with and without a sword: Self-governance is possible. *American political science Review*, 86(2):404–417.

Parthasarathy, R., Rao, V., and Palaniswamy, N. (2019). Deliberative Democracy in an Unequal World: A Text-as-data Study of South India's Village Assemblies. *American Political Science Review*, 113(3):623–640.

Pateman, C. (1970). *Participation and Democratic Theory*. Cambridge University Press.

Pepinsky, T. B., Pierskalla, J. H., and Sacks, A. (2017). Bureaucracy and Service Delivery. *Annual Review of Political Science*, 20:249–268.

Pommerehne, W. W. and Weck-Hannemann, H. (1996). Tax Rates, Tax Administration and Income Tax Evasion in Switzerland. *Public Choice*, 88(1):161–170.

Prichard, W. (2023). Unpacking Tax Morale: Distinguishing Between Conditional and Unconditional Views of Tax Compliance. Technical report.

Prichard, W., Custers, A. L., Dom, R., Davenport, S. R., and Roscitt, M. A. (2019). Innovations in Tax Compliance: Conceptual Framework. *World Bank Policy Research Working Paper*, (9032).

Reed, T. and Robinson, J. (2013). The Chiefdoms of Sierra Leone. *Manuscript, Harvard Univ. http://people. fas. harvard. edu/treed/history. pdf*.

Ribot, J. C., Chhatre, A., and Lankina, T. (2008). Introduction: Institutional Choice and Recognition in the Formation and Consolidation of Local Democracy. *Conservation and Society*, 6(1):1–11.

Richards, P. (1996). Fighting for the Rain Forest: War, Youth and Resources in Sierra Leone.

Richards, P. (2005). To Fight or to Farm? Agrarian Dimensions of the Mano River Conflicts (Liberia and Sierra Leone). *African Affairs*, 104(417):571–590.

Richards, P. (2016). *Ebola: How a People's Science Helped End an Epidemic*. Bloomsbury Publishing.

Robinson, A. L. (2024). Inventing Traditional Authority: Lhomwe Chiefs in Malawi. *The Journal of Modern African Studies*, pages 1–27.

Rousseau, J.-J. (1993[1762]). The Social Contract and Discourses.

Sawyer, E. (2008). Remove or Reform? A case for (Restructuring) Chiefdom Governance in Post-conflict Sierra Leone. *African Affairs*, 107(428):387–403.

Scharpf, F. W. (1997). Economic Integration, Democracy and the Welfare State. *Journal of European Public Policy*, 4(1):18–36.

Sinclair, B., McConnell, M., and Green, D. P. (2012). Detecting Spillover Effects: Design and Analysis of Multilevel Experiments. *American Journal of Political Science*, 56(4):1055–1069.

Skalník, P. (1996). Authority Versus Power: Democracy in Africa, Must Include Original African Institutions. *The Journal of Legal Pluralism and Unofficial Law*, 28(37-38):109–121.

Slemrod, J. (2019). Tax Compliance and Enforcement. *Journal of Economic Literature*, 57(4):904–54.

Slothuus, R. and De Vreese, C. H. (2010). Political Parties, Motivated Reasoning, and Issue Framing Effects. *The Journal of Politics*, 72(3):630–645.

Stanley, M. L., Henne, P., Yang, B. W., and De Brigard, F. (2020). Resistance to Position Change, Motivated Reasoning, and Polarization. *Political Behavior*, 42:891–913.

Sutter, M., Haigner, S., and Kocher, M. G. (2010). Choosing the Carrot or the Stick? Endogenous Institutional Choice in Social Dilemma Situations. *The Review of Economic Studies*, 77(4):1540–1566.

Taber, C. S. and Lodge, M. (2006). Motivated Skepticism in the Evaluation of Political Beliefs. *American Journal of Political Science*, 50(3):755–769.

Tappin, B. M., Berinsky, A. J., and Rand, D. G. (2023). Partisans' Receptivity to Persuasive Messaging is Undiminished by Countervailing Party Leader Cues. *Nature Human Behaviour*, 7(4):568–582.

The Economist (2020). A Tax on All Your Houses. https://www.economist.com/middle-east-and-africa/2020/06/18/a-mayor-is-reforming-sierra-leones-rotten-property-tax.

Tilly, C. (1990). *Coercion, Capital, and European States, AD 990-1992*. Basil Blackwell.

Timmons, J. F. (2005). The Fiscal Contract: States, Taxes, and Public services. *World Politics*, 57(4):530–567.

Torgler, B. (2005). Tax Morale and Direct Democracy. *European Journal of Political Economy*, 21(2):525–531.

Touchton, M. R., Wampler, B., and Peixoto, T. C. (2019). Of Governance and Revenue: Participatory Institutions and Tax Compliance in Brazil. *World Bank Policy Research Working Paper*, (8797).

Tyler, T. R. (2006). *Why People Obey the Law*. Princeton University Press.

Ubink, J. (2007). Traditional Authority Revisited: Popular Perceptions of Chiefs and Chieftaincy in Peri-urban Kumasi, Ghana. *The Journal of Legal Pluralism and Unofficial Law*, 39(55):123–161.

Van der Windt, P., Humphreys, M., Medina, L., Timmons, J. F., and Voors, M. (2019). Citizen Attitudes Toward Traditional and State Authorities: Substitutes or Complements? *Comparative Political Studies*, 52(12):1810–1840.

Voors, M., Turley, T., Bulte, E., Kontoleon, A., and List, J. A. (2018). Chief for a Day: Elite Capture and Management Performance in a Field Experiment in Sierra Leone. *Management Science*, 64(12):5855–5876.

Weigel, J. L. (2020). The Participation Dividend of Taxation: How Citizens in Congo Engage More with the State When it Tries to Tax Them. *The Quarterly Journal of Economics*, 135(4):1849–1903.

Wilfahrt, M. and Letsa, N. W. (2023). Judge, Landlord, Broker, Watchman: Assessing Variation in Chiefly Duties and Authority in the Ghana–Togo Borderlands. *The Journal of Modern African Studies*, 61(3):439–462.

Wilkinson, A. and Fairhead, J. (2017). Comparison of Social Resistance to Ebola Response in Sierra Leone and Guinea Suggests Explanations Lie in Political Configurations Not Culture. *Critical Public Health*, 27(1):14–27.

Williams, J. M. (2010). *Chieftaincy, The State, and Democracy: Colitical Legitimacy in Post-apartheid South Africa*. Indiana University Press.

Zizzo, D. J. (2010). Experimenter Demand Effects in Economic Experiments. *Experimental Economics*, 13:75–98.