

# Participation, Legitimacy and Fiscal Capacity in Weak States: Evidence from Participatory Budgeting\*

Kevin Grieco<sup>†</sup>      Abou Bakarr Kamara<sup>‡</sup>      Niccolò F. Meriggi<sup>§</sup>  
Julian Michel<sup>¶</sup>      Wilson Prichard<sup>||</sup>

January 10, 2025

## Abstract

Building durable fiscal capacity requires that the state obtains compliance with its tax demands, a struggle for weak states that lack enforcement capacity. One potential option for governments in weak states is to enhance their legitimacy and thereby foster voluntary compliance. In this study, we report results from a participatory budgeting policy experiment in Sierra Leone that attempted to increase legitimacy and tax compliance by inviting public participation in local policy decision-making. In phone-based town halls, participants shared policy preferences with neighbors and local politicians and then voted for local public services that were subsequently implemented. We find that the intervention durably increased participants' perceptions of government legitimacy. However, against influential models of tax compliance, we find a robust null effect on tax compliance behavior. In exploratory analyses, we document that partisan affiliation strongly conditions the interventions' effects on tax compliance and attitudes towards paying taxes.

---

\*This project was supported by generous funding from the International Growth Centre and the International Centre for Tax and Development. This study received IRB approval at UCLA (IRB #20-000380) and in Sierra Leone (approved 3/25/2020; amendment approved 5/28/2020). We thank Kate Baldwin, Graeme Blair, Jean-Paul Carvalho, Katherine Casey, Darin Christensen, Cesi Cruz, Evgeniya Dubinina, Adrienne LeBas, Giulia Mascagni, Nara Monkam, Oyebola Okunogbe, Daniel Posner, Soledad Prillaman, Paulina Seelmann, Jon Weigel, and Noam Yuchtman for helpful comments. We would also like to thank participants at the CaliWEPS IV workshop, APSA 2022, the 2022 Global Development Conference, the 2024 World Bank Land Conference, the 2024 IIPF Congress, the 9th Zurich Conference on Public Finance in Developing Countries and seminars at LMU Munich, Oxford and UCLA for their insightful feedback. We are grateful to Dheekshi Arvind, 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 pre-analysis plan associated with this project is registered at: <https://osf.io/dhkfe>.

<sup>†</sup>Department of Political Science, University of California, Los Angeles

<sup>‡</sup>International Growth Center, Sierra Leone

<sup>§</sup>CSAE, Department of Economics, University of Oxford

<sup>¶</sup>Department of Political Science, Ludwig Maximilian University of Munich

<sup>||</sup>Department of Political Science and The Munk School of Global Affairs, University of Toronto

# 1 Introduction

The weakness of many states in sub-Saharan Africa is a key barrier to economic development and political stability (Michalopoulos and Papaioannou 2020; Besley and Persson 2011).<sup>1</sup> Weak tax systems across the continent are, in turn, both effect and cause: state weakness limits the ability of states to raise revenue effectively, while weak revenue collection limits investments in state building (Besley and Persson 2013). How can governments in weak states break out of this pernicious cycle of low state capacity and insufficient revenue collection?

The conventional answer has been to focus on strengthening tax enforcement (e.g., Kleven et al. 2011; Slemrod 2019; Bergeron et al. 2024; Kapon et al. 2024). However, relying on enforcement alone has proven challenging due to limited capacity to pursue non-compliant taxpayers and political resistance, which has made expanding tax enforcement politically unattractive (Christensen and Garfias 2021; Dom et al. 2022).<sup>2</sup>

In this paper, we propose instead that governments can advance efforts to build fiscal capacity—and state capacity more broadly—by increasing their legitimacy. This focus on legitimacy reflects two key channels through which higher legitimacy could enable capacity building. First, citizens are more likely to comply with the demands of legitimate governments (Levi 1988, 1997; Besley 2020; Timmons and Garfias 2015). This is likely to be especially important in weak states, where governments must rely more on quasi-voluntary compliance due to limited enforcement capacity. Second, more legitimate governments are likely to face less political opposition to potentially contentious efforts to build state capacity (Besley and Dray 2024). Resistance to reform efforts to build fiscal capacity can come from taxpayers who oppose higher tax burdens and lack trust that governments will deliver tangible benefits in return (Gottlieb and Hollenbach 2018; Prichard 2015; Christensen and Garfias 2021; Robinson 2023) or from government officials within the administration itself who benefit from the status quo (Prichard et al. 2019). In short, building government legitimacy can enhance fiscal capacity by increasing the political acceptability of efforts to strengthen it or by raising quasi-voluntary tax compliance.

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). Surveying America's young democracy, Tocqueville concluded that when citizens participate in law-making, “law thereby acquires a great authority” (De Tocqueville 2010, pg. 393). Indeed, the link between public participation and tax compliance is central to seminal accounts of the development of fiscal capacity in early

---

<sup>1</sup>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 1960s. 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).

<sup>2</sup>For this reason, an emerging literature explores how non-punitive service delivery interventions affect tax compliance (Kresch et al. 2023; Brockmeyer et al. 2024; Carrillo et al. 2021; Krause 2020).

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.

This paper correspondingly 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). 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 state capacity building in fragile states. The intervention sought to give participants greater voice in, and control over, policy decisions regarding local development projects. Program participants joined WhatsApp chat groups—referred to as *Digital Town Halls* (DTHs)—alongside up to 36 other property owners from their neighborhood. Within these groups, they discussed service preferences, shared these preferences with local politicians, and then voted on the services (valued at approximately US\$1,500) they wanted to see implemented in their neighborhood. The 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 durably 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 all nine legitimacy outcomes at the endline survey, which was conducted soon after services were implemented and seven months after the conclusion of the DTHs. Importantly, while substantial legitimacy gains are observed at midline, citizens' perceptions of the government's administrative competency do not improve until endline, following successful service delivery. Observing how legitimacy evolves at these crucial junctures is a key design innovation of our study. Additionally, we demonstrate that these legitimacy gains are consistent across partisan groups.

Turning to tax compliance behavior, we find large and significant heterogeneous treatment effects conditional on participants' (pretreatment) partisan affiliation. Among copartisans 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. By contrast, we find that the treatment *lowered* compliance for non-copartisans by 4.0 percentage points. In our case these countervailing forces are relatively balanced, leading to no average effect on compliance. This null

effect is robust to alternative estimation specifications or operationalizations of tax compliance. Why does partisan affiliation moderate the intervention’s effect on tax compliance? In standard models, public participation in political affairs increases citizens’ compliance by strengthening their belief that leaders will pursue policies in their interests. We propose an alternative channel in which political participation impacts compliance through *elite opinion leadership*: through participation, citizens learn where political actors stand on specific issues, prompting them to update their expectations about the policy’s benefits (Zaller 1992; Broockman and Butler 2017). When leaders from opposing parties have divergent views on a policy—as was the case in Free-town during the highly politicized tax reform—increased citizen participation leads individuals to adjust their expectations differently, depending on their political allegiance. Consistent with this interpretation, we find that partisanship also moderates the intervention’s effect on the expected benefits of taxation, as measured by participants’ willingness to trade more taxes for improved services. Of course, since partisanship is not randomly assigned, these observed heterogeneous effects could be driven by confounding factors associated with partisan affiliation. We consider and rule out the possibility that the observed conditional treatment effects are driven by differences in demographic characteristics or service preferences between partisan groups.

Ultimately, our study makes four key contributions.

First, we contribute to the literature on participatory institutions and development (Putnam 1993; Acemoglu et al. 2001; North and Weingast 1989) by showing that direct democracy in the form of participatory budgeting increases government legitimacy in weak states. Previous field experimental research on whether participatory institutions in lower-income countries can increase political legitimacy has largely yielded null (Casey et al. 2012; Fearon et al. 2015; Humphreys et al. 2019; Khan et al. 2022) or mixed (Olken 2010) results.<sup>3</sup> While Beath et al. (2017) find that increasing citizen participation in community development projects boosts approval of political leaders in Afghanistan, the effect is confounded by a reduction in elite control.<sup>4</sup> By holding the delivery of selected services constant across treatment and control, our design allows us to isolate effects of town hall participation. Ultimately, our findings suggest a more optimistic view of what participatory fora can achieve: governments can use participatory budgeting to increase participants’ perceptions of government legitimacy.

Second, we demonstrate that partisanship moderates impacts of direct democracy on tax compliance. Prominent lab experiments have identified a “democratic dividend,” where individuals are more likely to comply with rules they had a role in creating (Bó et al. 2010; Sutter et al.

<sup>3</sup>A vast literature on “community-driven development” studies forms of participatory interventions other than participatory budgeting: plebiscites (Olken 2010; Beath et al. 2017), village council elections (at times, bundled with trainings in local democratic practices) (Fearon et al. 2015; Humphreys et al. 2019) and the solicitation of citizen service preferences (Khan et al. 2022). For a review, see Casey (2018).

<sup>4</sup>Treated villages, with higher participation, experience less elite influence, making it unclear whether the increase in approval is driven by participation itself or the shift in project allocation.

2010; Alm et al. 1993).<sup>5</sup> Yet, a recent review of the experimental literature finds mixed evidence for the democratic dividend, casting doubt on simple narratives linking participation to enhanced compliance, and arguing for greater attention to moderating factors that may shape “when and why dividends of democracy emerge” (Markussen and Tyran 2023, pg. 9). Consistent with this call for greater nuance, we demonstrate that partisanship can moderate participation’s impact on compliance. The idea that political participation may lead to backfiring effects among out-partisans is, while intuitive, to our knowledge, absent from the existing literature.<sup>6</sup> More broadly, our results suggest the need to rethink models that view citizens’ tax compliance solely as a function of government performance (e.g., Besley 2020; Levi 1988) or enforcement capacity (Allingham and Sandmo 1972), without considering its partisan composition.<sup>7</sup>

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 the legitimacy of 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 effectively used as a platform for citizen engagement.

Finally, our study highlights that fully understanding the impact of participation on long-term state-building requires examining both citizens’ compliance behavior and broader attitudes toward political authorities. Interventions that boost citizen participation in policymaking could immediately impact state-building by increasing tax compliance, thereby raising revenue and financing subsequent capacity investments. The intervention we study does not impact state capacity through this channel in our context because countervailing treatment effects result in a null average effect on tax compliance. However, there is a second, often overlooked, path from participation to state-building: by increasing citizens’ support for government, participation can give government the *political* capital needed to invest in building long-term fiscal, and state, capacity (Besley and Dray 2024; Christensen and Garfias 2021). Our results suggest a tradeoff for governments seeking to build capacity: actively involving non-copartisans in participatory processes can build the legitimacy needed for medium- and long-term state-building, but it risks behavioral backfires in the short term.

---

<sup>5</sup>Several observational studies also link participation in policy-making to tax compliance (Pommerehne and Weck-Hannemann 1996; Torgler 2005; ?).

<sup>6</sup>Our result that participatory budgeting decreases tax compliance and support for expanded taxation among political opponents is similar to “backfiring” effects 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 tax bill nudges (Castro and Scartascini 2015; De Neve et al. 2021).

<sup>7</sup>Though see Cullen et al. (2021) on political alignment and tax compliance in the United States.

## 2 Interventions: Digital Town Halls and Service Delivery

This research takes place in cooperation with the Freetown City Council (FCC) in the context of a city-wide property tax reform two of the authors 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](#)). It resulted in large overall increases in taxation, with assessed tax liabilities increasing five fold—concentrated among higher value properties—and revenue collection increasing three fold in the first year of the reform ([Prichard et al. 2020](#)).

The mayor publicly announced that Digital Town Halls (DTHs) would be held starting in January 2021. In her messaging, she emphasized that the DTHs would be key for securing citizen participation in decision-making about service delivery, in the context of expanded revenue raising. She also stressed that she intended to institutionalize the DTHs with future DTHs being assigned 20% of property taxes raised per ward ([Freetown City Council 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.

### 2.1 Digital Town Halls

DTHs were WhatsApp group chats where property owners discussed pressing development challenges with other property owners in their ward and then communicated these challenges to their political representatives.<sup>8</sup> The groups then deliberated on how to allocate a budget of 15 million leones (about US\$ 1,500) for 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, reflecting key elements of effective deliberative processes ([Mansbridge 1999; Fishkin 2002](#)):

#### 1. Horizontal Deliberation (January 15 - 19, 2021):

Participants received introductory videos from the Mayor of Freetown and their respective ward councilor.<sup>9</sup> These videos explained the overall process, highlighted the link

<sup>8</sup>We completed a pilot DTH in one ward before scaling the DTHs up to our 30 study wards. In Appendix A.1 we lay out potential advantages and disadvantages of *Digital* Town Halls vis-à-vis in-person Town Halls.

<sup>9</sup>Videos from political representatives were shared with DTH participants by being posted in the WhatsApp group and also indirectly via a Qualtrics link. The research team hired a local team to act as moderators, who were

between property tax payments and service delivery, and invited participants to start discussing development concerns within their group. Group moderators introduced themselves and prompted participants with the following question: *What do you think is the greatest development problem in your ward?* This phase involved purely horizontal deliberation, as participants were informed that political representatives would not be involved or have access to the discussions during 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 them to (i) identify the two greatest development challenges in their ward and (ii) propose projects to address these challenges. Participants were instructed to consider only projects that could be completed within a budget of US\$ 1,500 and to submit their proposals via written message or short voice recording. The DTH facilitators, with the participants' knowledge, aggregated this information and presented memos outlining the concerns and proposed solutions to both the Mayor and the ward councilor. This approach allowed participants to anonymously communicate their preferences to their representatives. Through this process, it became clear that water access was the most pressing concern for many communities.<sup>10</sup>

## **3. Vertical Interaction (February 13 - 16, 2021):**

Participants received separate videos from both their councilor and the Mayor. In these videos, the representatives responded to participants' proposals, justified their preferred services, and explained past and future delivery goals. 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. The Mayor and councilors explained 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.

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

In this phase, participants cast their vote for their preferred project anonymously through

---

supervised by project research assistants. DTH facilitators requested that participants use the chat only between 7 a.m. and 10 p.m. daily to ensure a facilitator would be present at all times. Participants could choose their preferred form of communication (text, voice or video messages) but were asked to contribute in Krio or English.

<sup>10</sup>Other common service preferences included roads, street lights, dustbins, public toilets and the upgrading of drainage systems.

a Qualtrics survey (Appendix Figure B2).<sup>11</sup> After four days of voting, 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. After the announcement of the winning projects, group moderators thanked participants for their contributions and halted participants' ability to post messages in the DTHs.

There was active participation in the Digital Town Halls. We confirmed that 1,457 of the 1809 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.<sup>12</sup> Roughly two-thirds of those who joined voted for their preferred service (68%) and posted at least one message (63%).<sup>13</sup> Across all groups, participants exchanged approximately 2,000 messages. Most messages were text (55.25%) and voice messages (40.2%), with the remainder consisting of images and videos.

Participants 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). Additionally, respondents generally agreed that “participants felt comfortable to make their views known even when their views differed from those of other participants” (3.82/5). While respondents were positive about their DTH experience, they were also realistic about its limitations. Most believed that the DTH budget was insufficient to significantly improve the delivery of the selected service (2.86/5) (see Appendix Table A5).

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.<sup>14</sup> Respondents overwhelmingly reported they believed that the FCC organized the DTH (89%), implemented services (96%) and funded the services (84%). Of the respondents who said the FCC funded the project, 87% thought it was funded through taxes (either from inside or outside the ward), 6% from government transfers, 4% from development partners and 3% from foreign aid (see Appendix Table A6).

---

<sup>11</sup>A “how to” video was posted in each group that provided step-by-step instruction of the voting process. We also gave participants the option to inform moderators of their vote in bilateral conversation.

<sup>12</sup>Only 5% of respondents who joined reported they never accessed the group and another 5% reported they accessed the DTH only once.

<sup>13</sup>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.

<sup>14</sup>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. For a discussion of research ethics, see Appendix G.

## **2.2 Service Delivery & Notification**

Each participating ward received a service project, benefiting both the treated and control units within that ward. Construction began in most wards in October 2021 and was completed in all but one by the end of the year.<sup>15</sup> After the vote, the engineering company determined that fixing neighborhood water pipes was infeasible, despite previous assurances to the contrary. As a result, wards that had selected this service received an alternative water-related project: either a hand pump or a 5,000-liter community water tank.<sup>16</sup> 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). More details on service delivery can be found in Appendix A.4.

To ensure that DTH participants were aware of the successful project implementations, we made notification calls on behalf of the FCC to all treated units. By contacting only the treated units and not the control units, we incorporated these notification calls into our treatment.<sup>17</sup> We successfully reached approximately 70% of treated units to inform them about the implemented services. These calls began in mid-November and were staggered across wards, starting only after service delivery was completed in each ward. The endline survey was similarly staggered, commencing after the notification calls were completed and never earlier than one week after the completion of service delivery.

## **3 Research Design**

### **3.1 Sampling, Randomization and Balance**

To estimate causal effects, we randomized an invitation to join a DTH across a sample of property owners in Freetown. We constructed a sampling frame using FCC administrative records, which provided information on property characteristics and owner contact details. To be eligible for the intervention, property owners needed to own property in one of the 30 study wards, have WhatsApp on their phone and be scheduled to receive a tax bill in the first year of the reform—though properties below the median value were exempt from this tax due to COVID-19-related policy. Out of the 15,977 property owners we contacted, 4,860 were verified to have WhatsApp on one of their phones, making them eligible for the Digital Town Hall intervention. It is important to note that the sample of property owners we contacted was *not* strictly random, as we filtered out some properties to limit geographic spillovers and could not reach owners whose contact information was missing from the FCC records. Appendix B.1

---

<sup>15</sup>In the remaining ward, construction was finished in February 2022.

<sup>16</sup>One ward voted to fix potholes, but due to implementation difficulties, it received 50 meters of truck tracks.

<sup>17</sup>This decision is informed by Khan et al. (2022) who expressed concern that a lack of awareness about service delivery diminished the impact of their preference elicitation and service delivery intervention. We made notification calls to rule out this concern, thereby simplifying the interpretation of our findings.

provides more details about sampling.

To capture property owner level covariates and measure attitudinal outcomes we conducted three rounds of phone-based survey data collection: prior to the DTH (100% response rate), following DTH participation but prior to service delivery (91.3%) and following service delivery (79.4%).<sup>18</sup> Conducting surveys before and after service delivery is a key design innovation of the study, as it allows us to capture the importance of subsequent service delivery in shaping response to participation. 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.

To mitigate potential spillover effects, we drew a restricted sample from the 3,859 eligible property owners who had completed a baseline survey, ensuring that each property was at least 15 meters from the nearest study property. This resulted in a final sample size of 3,618. We then assigned treatment status using a matched-pair design, leveraging baseline survey data to match similar observations into groups of two ([King et al. 2007](#)). We created 1809 pairs and then assigned one observation in each matched pair to treatment and the other to control. Appendix B provides more details on the restricted sampling, matching procedure and treatment assignment.

Table 1 reports balance across baseline attitudinal outcomes, immutable demographic covariates and property characteristics (29 covariates total). We observe imbalance on two variables, no more than is to be expected through chance. As our preregistered specification for survey-based outcomes includes the baseline measure of the dependent variable, we control for these slight imbalances when estimating treatment effects.

---

<sup>18</sup>The response rate at baseline is 100% because only baseline respondents were eligible for the intervention.

Measure	Mean		SD		Difference		Observations	
	C	T1	C	Raw	Std.	p-val	C	T1
<b>Survey Outcomes</b>								
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
<b>Political Party Affiliation</b>								
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
<b>Property Characteristics</b>								
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
<b>Demographics</b>								
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 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:** \*  $p < 0.10$

Table 1: Balance Table

### 3.2 Estimation and Inference

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:

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

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 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  $j$ ;  $\mathbf{X}$  is a set of preregistered property-level characteristics that we include for covariate adjustment only when  $Y$  is property tax compliance behavior.<sup>19</sup>  $\delta$  is a vector of ward fixed effects and  $\epsilon_i$  is the error term.

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

$$DTH_{ij} = \beta_1 D_i + \eta Y_{ijt_1} + \sum_{j=1}^{1809} \mu_j PAIR_{ji} + \zeta_w + \xi \mathbf{X}_i + \nu_i \quad (2)$$

Where  $D_i$  is the randomly assigned treatment indicator, which instruments for  $DTH_i$  in equation 1. Our quantity of interest is  $\alpha_1$  (equation 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.

We estimate treatment effects on various attitudinal outcomes, organizing them into hypothesis families. To adjust for multiple tests, we implement false discovery rate (FDR) corrections as outlined by [Anderson \(2008\)](#) and report the sharpened FDR  $q$ -values alongside conventional  $p$ -values for each indicator (see also [Benjamini et al. 2006](#)). A key feature of sharpened  $q$ -

---

<sup>19</sup>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 1 controls for survey-based outcomes that we expect to predict compliance through the inclusion of block dummies.

values is that they can be *smaller* than unadjusted  $p$ -values when corrections are made within a set of hypotheses where there are many rejections. The intuition is that, since the FDR aims to control the proportion of false discoveries across all hypotheses, knowing that many hypotheses have little risk of false rejection (i.e., very low  $p$ -value) allows for a higher tolerance of false rejections in the remaining hypotheses, and for  $p$ -values to be adjusted downwards.<sup>20</sup>

These corrections are applied within each hypothesis family. This paper combines analyses from two separate preanalysis plans (PAPs), which include several shared indicators. For indicators appearing in only one PAP, we adjust them within the hypothesis family they were originally assigned to. For indicators linked to both PAPs, we assign them to the hypothesis family described in the more recent PAP for adjustment. Appendix B.6 provides further details on how we form hypothesis families to implement these corrections.

## 4 Results

### 4.1 Effects on Legitimacy

We first look at the impact of the treatment on multiple indicators of government legitimacy, which we expect, in turn, to drive changes in tax compliance. Following [Levi et al. \(2009\)](#), we classify those indicators into four categories: Policy Influence, Service Delivery and Responsiveness, Government Administrative Competence and Approval of Political Representatives (Table 2). We discuss results for each in turn.

We examine first citizens' perceptions of their ability to influence policy ([Scharpf 1997; Tyler 2000](#)), focusing on two pre-registered outcomes: 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 ( $\beta = 0.064$  SDUs) and endline ( $\beta = 0.073$  SDUs) and, while the unadjusted  $p$ -values lie at the threshold of conventional levels, the adjusted  $q$ -values suggest these effects are statistically significant.<sup>21</sup>

Second, we study perceptions of service delivery and responsiveness ([Scharpf 1997; Easton 1975; Gilley 2006](#)). We find that the intervention significantly increased treated citizens' perceptions that the local government was responsive to citizens' needs and demands both directly after the DTH (midline:  $\beta = 0.141$  SDUs;  $p$ -value <0.001) and after service implementation

---

<sup>20</sup>As [Anderson \(2008\)](#) notes in his code, "Sharpened FDR q-vals can be LESS than unadjusted p-vals when many hypotheses are rejected, because if you have many true rejections, then you can tolerate several false rejections too."

<sup>21</sup>RI  $p$ -values are 0.096 and 0.113 at midline and endline, respectively.

(endline:  $\beta = 0.116$  SDUs;  $p$ -value = 0.014). 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 ( $\beta = 0.182$  SDUs;  $p$ -value <0.001) and endline ( $\beta = 0.146$  SDUs;  $p$ -value = 0.004). The adjusted  $q$ -values indicate that treatment effects on these outcomes are statistically significant at midline and endline.

Third, we explore perceptions of the ability of governments to administer their constituencies competently (Hutchison and Johnson 2011; Rothstein and Stolle 2008; Magalhães 2014). Our survey data show that, before the intervention, respondents perceived the FCC as fairly incompetent: the average respondent perceived the FCC as *not* transparent (1.36/3) and of middling efficiency (2.86/4) and corruption (3.53/5). Note that for each measure, a higher score indicates better performance. The intervention improves respondents' perceptions of FCC administrative competence across all measures, though notably these improvements come largely at endline, after successful service delivery. While perceptions of transparency show a modest improvement at midline ( $\beta = 0.085$  SDUs;  $p$ -value = 0.109;  $q$ -value = 0.070), this effect increases nearly fourfold by endline ( $\beta = 0.319$  SDUs;  $p$ -value = 0.001). In terms of perceived efficiency in the use of funds for public administration and development, we observe *no effect* at midline ( $\beta = 0.037$  SDUs;  $p$ -value = 0.314;  $q$ -value = 0.181), but a clear positive impact by endline ( $\beta = 0.129$  SDUs;  $p$ -value = 0.007;  $q$ -value = 0.008). Finally, for perceptions of corruption, we find a similar, though more extreme, change: at midline the treatment *increases* participants' perceptions that the FCC is corrupt ( $\beta = -0.141$  SDUs;  $p$ -value <0.001;  $q$ -value = 0.001), but after services are implemented, treated participants *positively* update their views of FCC corruption relative to the control group ( $\beta = 0.087$  SDUs;  $p$ -value = 0.075;  $q$ -value = 0.048). This shift in perception is likely due to citizens initially suspecting that new local development funds would be diverted to patronage and corruption, only to revise their expectations positively once services were actually delivered. The overarching message from these results is clear: for governments to reap the full legitimacy benefits of expanding participation, they must follow through on their service delivery promises. Citizens understand that talk is cheap; they respond to tangible action.

Fourth, we focus on whether participants approved or disapproved of how both the Mayor and their ward councilor have performed over the past twelve months (Levi et al. 2009; Norris 2017). 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 midline and 0.19 SDUs ( $p$ -value <0.001) at endline. That we observe these effects is particularly impressive given that at baseline 44% of the sample gave maximum approval ratings. By contrast, the modal respondent (41%) "disapproved" of their ward 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 <0.001) and endline (0.17 SDUs;  $p$ -value <0.001). Adjusted  $q$ -values indicate that treatment effects at midline and endline are statistically significant for both outcomes.

In summary, Table 2 provides unambiguous evidence that the intervention increases perceptions of government legitimacy. Importantly, we find that the full impact of participation on legitimacy depends crucially on treated individuals seeing evidence of promised service delivery. 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](#)).

Outcome	Baseline		Midline			Endline			
	Mean	Mean	Effect	N	$q$ -val	Mean	Effect	N	$q$ -val
<b>Policy Influence</b>									
Opportunities for voice	2.13 (1.00)	2.33 (0.92)	0.377*** (0.038)	3,288	0.001	2.16 (0.92)	0.251*** (0.046)	2,849	0.001
Ease of participating in political activities	1.75 (1.14)	1.62 (1.02)	0.064 (0.040)	3,298	0.040	1.63 (1.02)	0.073* (0.046)	2,863	0.034
<b>Service Delivery and Responsiveness</b>									
FCC responsiveness to citizens' demands	3.17 (1.19)	3.36 (1.06)	0.141*** (0.038)	3,251	0.001	3.31 (1.16)	0.116** (0.048)	2,830	0.015
Satisfaction with FCC service provision	3.64 (1.17)	3.61 (1.06)	0.182*** (0.040)	3,302	0.001	3.47 (1.21)	0.146*** (0.050)	2,864	0.009
<b>Govt. Administrative Competence</b>									
FCC transparency	1.36 (0.69)	1.42 (0.77)	0.085 (0.052)	3,288	0.070	2.16 (1.34)	0.319*** (0.101)	2,834	0.002
FCC efficiency	2.86 (0.71)	2.86 (0.56)	0.037 (0.038)	3,233	0.181	2.79 (0.70)	0.129*** (0.048)	2,791	0.008
FCC corruption	3.53 (1.00)	3.62 (0.90)	-0.141*** (0.043)	3,177	0.001	3.45 (0.93)	0.087* (0.048)	2,736	0.048
<b>Approval of Political Representatives</b>									
Mayor approval	4.23 (0.89)	4.08 (0.82)	0.149*** (0.042)	3,296	0.001	3.91 (0.94)	0.194*** (0.051)	2,855	0.001
Councilor approval	2.73 (1.22)	2.73 (1.17)	0.193*** (0.040)	3,278	0.001	2.74 (1.22)	0.171*** (0.047)	2,841	0.001

**Table 2** reports the effect of the treatment on political attitudes. 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 conventional  $p$ -values. Columns 5 and 9 report corrected  $q$ -values, which adjust for multiple hypothesis testing, following [Anderson \(2008\)](#). Reported effects are standardized effects.

**Significance:** \*  $p < 0.10$ ; \*\*  $p < 0.05$  \*\*\*  $p < 0.01$

Table 2: Effect on Legitimacy

## 4.2 Average Effects on Tax Compliance

Turning to the impacts on compliance, we first present the average effects (Table 3), followed by an exploration of heterogeneous effects across key sub-groups in the next subsection. We report treatment effects on tax compliance for both 2021 and 2022, though our preregistered primary outcome of interest is 2022. While the Digital Town Hall (DTH) was launched in 2021, service delivery was not completed until after the 2021 tax payment deadline, making 2022 the first tax season following the full treatment of participation and service delivery. We observe compliance behavior for all units.

Panel A of Table 3 reports average treatment effects for the full sample. 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).

Outcome	Mean (1)	Effect (2)	$p$ -value (3)	N (4)
<b>Panel A: Tax Compliance Behavior</b>				
<i>Did the owner pay any taxes?</i>				
2022	0.291 (0.018)	-0.012	0.496	3,618
2021	0.315 (0.019)	-0.007	0.723	3,618
<b>Panel B: Fiscal Exchange Attitudes</b>				
<i>Willingness to pay more taxes for better services</i>				
Midline	4.001 (1.253)	0.066 (0.047)	0.163	3,296
Endline	4.030 (1.293)	-0.075 (0.053)	0.155	2,872

**Table 3** reports treatment effects on tax compliance behavior (Panel A) and attitudes towards fiscal exchange (Panel B). Column 1 reports control group means. Column 2 presents treatment effects estimates. In Panel A these effects are reported in raw percentage points; in Panel B presented effects are standardized effects. Column 3 reports  $p$ -values and Column 4 reports the number of non-missing observations.

**Significance:** \*  $p < 0.10$ ; \*\*  $p < 0.05$  \*\*\*  $p < 0.01$

Table 3: Effect on Tax Compliance

These null effects are robust to different model specifications. Our main, preregistered specification includes ward-fixed effects and a set of property characteristics as control variables. Results are similar when we estimate effects using (i) only the treatment indicator and 2020 (pretreatment) compliance behavior; (ii) only ward-fixed effects and pretreatment compliance; (iii) only property characteristics and pretreatment compliance; and (iv) when we add to our primary specification a dummy indicating the owner has zero tax liability.<sup>22</sup>

Results are also robust to different operationalizations of tax compliance. While Table 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. These robustness results are reported in Appendix Tables C1 and C2, respectively.

These nulls are also precisely estimated and we can rule out all but small effects: estimated standard errors imply that the upper limit of the 95% confidence interval is 2.3 percentage points. Still, we might worry that a (small) true effect exists, but we 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.<sup>23</sup> While the interpretation of outcome is less straightforward—the group mean compliance, pooling across years—this effect is causally identified. The point estimate is close to zero (-1.1pp) and is not statistically significant (*p*-value = 0.45) and the upper limit of the 95% confidence interval is 1.8pp. 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 potential real impacts are almost certainly substantively small.

This null result runs against most existing research, which predicts a consistent link from increased participation and legitimacy to greater tax compliance (Levi 1988, 1997; Besley 2020). It is doubly surprising given that we *do* observe strong and durable positive impacts on government legitimacy. Why do we see positive impacts on legitimacy but not on compliance? One simple and mechanical explanation is that treated property owners *want* to pay more taxes, but face a sharp budget constraint. 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*. However, as presented in Panel B in Figure 3, we do not find evidence that the intervention increases property owners' attitudes towards fiscal exchange. This finding also dispels the possibility that attitudinal effects are driven by experimenter demand (Zizzo 2010), rather than true changes in beliefs. If experimenter demand had shaped results in Table 2, we should have found that treated respondents *say* they would be more willing to pay taxes;

---

<sup>22</sup>Property owners can have no liability in a given year if they paid more than was due the previous year. In 2022, nine property owners had zero tax liability and in 2021, 121 property owners had no liability.

<sup>23</sup>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.

we do not find this.

Another possibility is that the intervention *negatively* impacted other key mediating mechanisms that, although not the primary targets, could plausibly have been affected. We preregistered two additional channels through which the intervention might influence compliance: (i) perceptions of fairness and equity and (ii) enforcement. If the intervention *diminished* participants' views of the tax system's fairness or reduced the perceived likelihood that noncompliers would be punished, this could have counteracted the positive effects on government legitimacy. However, at endline, we find no evidence of lasting treatment effects on either the fairness or enforcement outcomes (Appendix Table C7). While midline results are more varied, that they do not persist to endline makes them unlikely explanations for the null results on compliance.<sup>24</sup>

Third, we explore whether the discrepancy between the observed positive legitimacy effects and null compliance effects results from differences in the samples used. Since tax compliance is measured through administrative records, we estimate treatment effects on compliance using the full sample. In contrast, legitimacy effects are estimated using survey data where respondents can attrit. Appendix Figure C3 reanalyzes compliance treatment effects, excluding property owners who attrited from the survey. The point estimate for compliance *decreases* slightly, ruling out this potential explanation.

A potential concern is that survey respondents may not be the individuals responsible for making property tax payment decisions. However, we guarded against this possibility by implementing a rigorous verification process to confirm that respondents were property owners. In an initial verification survey (prior to the baseline survey, see Section 3.1), we compiled a list of confirmed property owners involved in financial decision-making for the property that used WhatsApp. We then independently verified that the respondent's phone number was linked to an active WhatsApp account. During the baseline survey, we re-verified that the respondent's name matched that of the verified property owner.<sup>25</sup> Additionally, before enrolling participants in DTHs, we re-contacted the selected property owner to verify their identity, ownership status, WhatsApp usage, and WhatsApp phone number.

A final possibility is that true treatment effects exist but are washed out by spillovers. In Appendix Section E, we use a design-based approach to estimate spillover effects, leveraging compliance data from 74,352 properties outside of our study. While we find little evidence of spillovers using our preregistered specifications, there is some suggestive evidence of small spillover effects for properties within 20 meters of a treatment property ( $\beta = 1.8$  percentage points; RI  $p$ -value = 0.13). However, this is unlikely to meaningfully downward bias our estimates, as only 4.3% of control units in our study are within 20 meters of a treated unit (see Appendix Figure B2).

---

<sup>24</sup>These results are discussed in greater detail in Appendix Section C.4.

<sup>25</sup>There was little incentive for someone other than the owner to claim to be the owner because respondents were told that participation was purely voluntary and would not be rewarded.

Given the absence of a simple explanation for the null effect on compliance despite significant increases in legitimacy, we investigate the possibility that our null result may disguise heterogeneous effects across groups.

### 4.3 Partisanship Moderates Participation’s Impact on Tax Compliance

Existing research posits a relatively straightforward link between participation, legitimacy and compliance (Levi 1988; Besley 2020; Beath et al. 2017; Bó et al. 2010; Alm et al. 1993). In this body of work, citizens are more likely to comply with government policy when leaders open channels for public participation because this strengthens belief that leaders will pursue policies that are in citizens’ interests and that citizens will benefit from compliance. From this perspective, a homogeneous citizenry responds to a unitary political actor; increased participation boosts tax compliance by raising citizens’ expectations that government will use revenue to benefit citizens. This could be called the *government legitimacy* channel through which participation increases compliance.

Real-world politics, however, are often contentious. In the context we study, the central government opposed the property tax policy that the Mayor championed and the Mayor publicly clashed with the Ministry of Finance over the Freetown City Council’s (FCC) legal authority to adjust property tax rates without central government approval (Luke 2020). Party politics were at the center of this conflict: while the Mayor’s party, the All People’s Congress (APC), controlled the Freetown City Council (FCC), their primary political opposition, the Sierra Leone People’s Party (SLPP), controlled the central government and therefore the Ministry of Finance.

These partisan dynamics may also shape how citizens form expectations about the benefits of government policy. According to seminal work in political science, citizens often have weak or malleable policy preferences and rely on cues from trusted political elites to form them (Zaller 1992). For instance, supporters of Party A may raise their support for a policy when they learn Party A backs it, while supporters of Party B may lower theirs upon discovering Party B opposes it (Broockman and Butler 2017; Flores et al. 2022; Tappin et al. 2023). It is intuitive that partisanship conditions citizens’ expected benefits of a given policy because partisanship is often associated with perceptions of government performance (Stantcheva 2021). Indeed, we find that APC supporters report higher approval of both APC leaders and the APC-led Freetown City Council (Appendix Table D1).<sup>26</sup> While partisan differences in perceived government performance may be due, in part, to conative biases (e.g., Gaines et al. 2007), these differences may also stem from a rational belief that the government will target benefits at their core supporters (Golden and Min 2013; Cox and McCubbins 1986). That is, partisanship may condition

<sup>26</sup>While we find that partisanship is highly correlated with perceptions of government performance, we do not find evidence that partisanship is associated with normative beliefs about taxation (see Stantcheva 2021, for an excellent examination of partisanship and taxation in the United States). Specifically, partisanship is not associated with the belief that the rich should be taxed more to pay for services for all (Appendix Table D1).

expected benefits of taxation because property owners expect taxes to fund services that will be distributed as pork to core supporters. Although the locations of the public service projects in our study were determined by our research team, working in conjunction with a construction firm, correlational evidence suggests some degree of partisan targeting: APC supporters, on average, are closer to the public service projects implemented in our intervention than SLPP supporters (see Appendix Table D4).<sup>27</sup>

This suggests an alternative way that participation may affect compliance. Through political participation, citizens learn where political actors stand on specific issues, which prompts them to update their expectations about the policy’s benefits. We refer to this as the *elite opinion leadership* channel through which participation can impact compliance. Importantly, this channel implies that when party leaders disagree on their support for a policy, increased citizen participation leads individuals to adjust their expectations about the policy in different ways, depending on their political allegiance.

It seems likely that DTH participants were more exposed to the views of party leaders on the controversial tax policy. Within the DTHs, the Mayor strongly advocated for property taxation, emphasizing its importance for improving local public services. For instance, in the Mayor’s first DTH video, she encouraged participants to pay taxes, assuring them that the FCC would use the revenue to deliver services to Freetown residents.<sup>28</sup> Along with greater exposure to the Mayor’s views, DTH participants were likely also more exposed to competing perspectives from the central government. This is because the treatment led to an immediate increase in participants’ political interest and engagement (Appendix Table C5). For example, in the midline survey DTH participants showed greater interest in politics overall ( $p$ -value <0.001;  $q$ -value = 0.001) and in FCC activities specifically ( $p$ -value <0.001;  $q$ -value = 0.001), suggesting they became more attuned to the politicized debate surrounding property taxation in Freetown.

Given the APC’s support for the tax reform, the elite opinion leadership channel predicts that participation leads APC partisans to raise their expectations about the benefits of taxation more than non-copartisans, thereby also making them more likely to increase their tax compliance. Figure 1 presents evidence in line with these predictions. Plot A shows that participants’ (pre-treatment) partisan affiliation—measured in the baseline survey—conditions how the treatment impacts their tax compliance behavior. Plot A presents predicted marginal effects from a model that interacts treatment with a copartisan indicator variable.<sup>29</sup> The interaction between treat-

---

<sup>27</sup>Project research assistants agree that local politicians could plausibly influence project location at the margins. They met with and took input from local politicians. In a few instances, politicians strongly advocated for projects in specific locations.

<sup>28</sup>Specifically, the Mayor promised that the “FCC will use [tax revenue] to deliver services to the people of Freetown.” In a separate video, the Mayor also reminded participants, “If everyone pays their property rate, you can imagine what type of investment we can make in your ward.”

<sup>29</sup>Other model specifications remain the same as in our main specification. The copartisan variable is equal to 1 for respondents who self-report affinity towards the All People’s Congress; all other respondents are coded as 0. 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

ment and co-partisanship is statistically significant ( $p$ -value = 0.033;  $\beta$  = 0.11). For copartisans 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%. In contrast, treatment effects are *negative* for non-copartisans: treatment lowers compliance by 4.0 percentage points.

Our preferred specification uses the full sample, coding respondents who report feeling personally supportive of and close to the APC as "copartisan," with all others coded as "opposition." Results are robust to alternative partisanship codings. Appendix Figure C2 shows similar results when excluding from the "opposition" group respondents who did not answer the partisanship question. Additionally, Appendix Table C3 reports treatment effects by five partisan sub-groups, demonstrating that results are not driven by a single non-partisan group: sub-group treatment effects are negative for *all* non-partisan groups.

If participation leads to heterogeneous effects on compliance by updating participants' expectations about the benefits of taxation along partisan lines, we should also observe that partisanship moderates the impact of participation on perceptions of those benefits. One way to measure the perceived benefit of taxation is by examining respondents' willingness to exchange taxes for services—the respondent is more willing to make this trade if they believe they will benefit from it. Prior to the DTH, the majority (57.4%) of surveyed respondents reported that they "strongly approved" of expanding taxation for improved services, while a significant minority (14%) opposed this idea.

Panel B (Figure 1) shows predicted marginal effects from an interaction model, where the outcome of interest is the respondent's support for expanding taxation (i.e., their tax policy preference). To increase power for estimating this interaction, the predicted outcome is the respondent's *average* support for expanded taxation across midline and endline surveys. The interaction between treatment and co-partisanship is statistically significant ( $p$ -value = 0.063;  $\beta$  = 0.223 SDUs), and again we see heterogeneous impacts by partisanship: copartisans increase their support for expanding taxation for improved services, while non-copartisans decrease their support for this policy. Appendix Figure C1 shows estimates from an interaction model using midline or endline data separately. Estimated marginal effects display similar patterns.

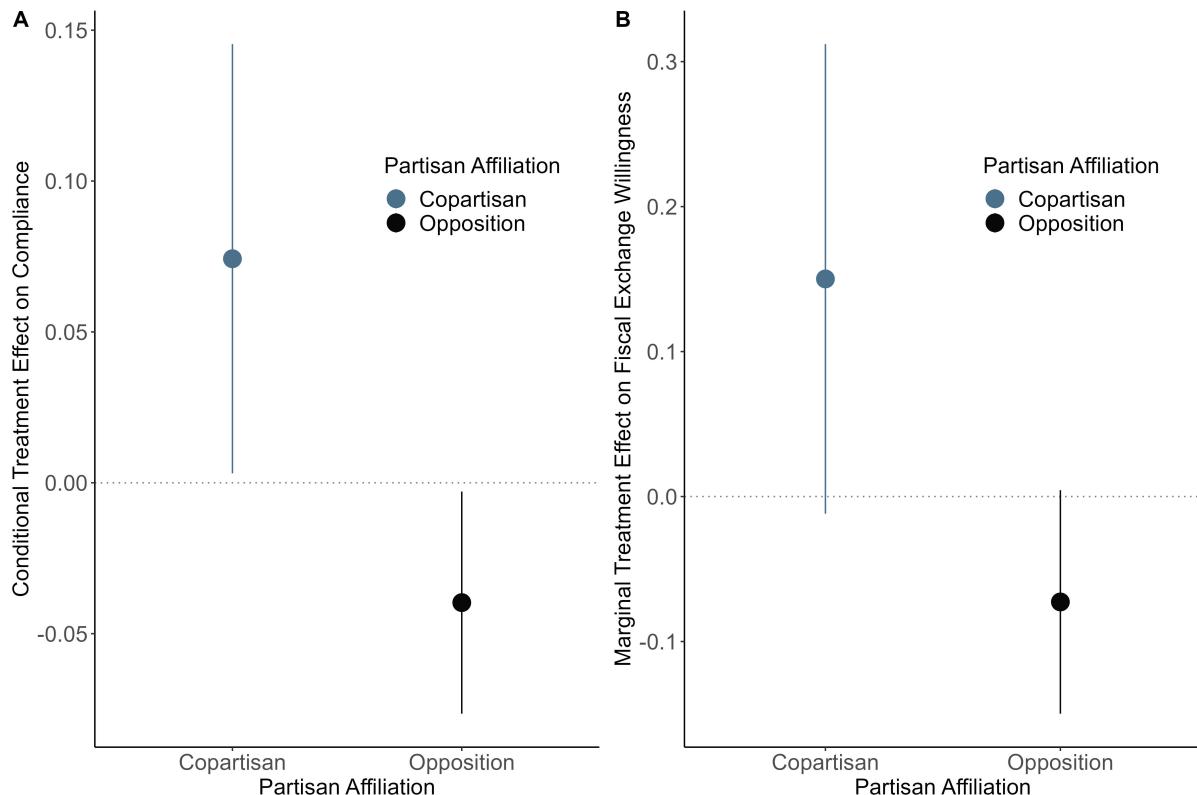
The consistency of patterns in both behavioral (Panel A) and attitudinal (Panel B) outcomes suggests that these subgroup effects are not merely statistical artefacts. However, because partisanship is not randomly assigned, these observed heterogeneous effects could be driven by other confounding factors associated with partisan affiliation. While APC supporters do

---

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."

differ statistically from out-group party members along several demographic dimensions, these differences are substantively small (Appendix Table D2). APC-partisans in our sample are three percentage points less likely to be female and five percentage points less likely to have a degree from an institution of higher education; APC partisans own slightly more properties on average, but with slightly lower taxable value. The small magnitudes of these differences make them unlikely candidates for driving the observed heterogeneity by partisanship. More importantly, Appendix Table D3 (Columns 3-4) shows that these variables do not moderate treatment effects on compliance.

Figure 1: 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 "copartisans." All other respondents are coded as "opposition." Point estimates are presented with 90% confidence intervals.

Another potential confounder is service preferences. For example, if partisan affiliation proxies for service preferences, it could be that treatment effects are larger for copartisans because they get the service they want and lower for non-copartisans because they do not. Appendix Figure D1 shows participants' vote choice in the DTH for local development projects by partisan group. There is little indication that service preferences vary meaningfully by partisanship. Therefore, service preferences are unlikely to be confounding the observed heterogeneous ef-

fects reported in Figure 1.

There are at least two alternative explanation for why partisanship moderates treatment effects on compliance. First, given the politicized nature of the reform, it could be that the DTH increased affective polarization between partisan groups (Iyengar and Westwood 2015).<sup>30</sup> While we find some evidence of polarization at the midline survey, these effects dissipate at endline (Appendix Table C8). We therefore consider it unlikely that the observed heterogeneous effects on compliance are driven by affective polarization. Note, however, that affective polarization and our proposed channel of elite opinion leadership are not mutually exclusive and these dynamics could be operating in tandem.

Second, the partisan targeting of service projects implemented by the DTH may have led to higher compliance from APC supporters but backlash from political out-group members. However, we find this unlikely, as the observed degree of targeting appears substantively small—APC supporters are, on average, only about 50 meters closer to the implemented projects than SLPP supporters (see Appendix Table D4). Furthermore, we find no evidence that treatment effects are greater for APC supporters in town halls where projects exhibited stronger APC targeting (i.e., where APC partisans were, on average, closer to implemented projects than SLPP partisans; see Appendix Figure D2).

In this section, we presented exploratory analyses suggesting that partisanship moderates the effect of participation on compliance. We argue these heterogeneous effects are driven by divergent partisan updating about the benefits of taxation stemming from participants' enhanced understanding of where their party stands on the issue of taxation. We understand this mechanism, which we call elite opinion leadership, as occurring alongside the government legitimacy channel that is commonly posited in the existing literature. Not only do participants positively update about government legitimacy on average (Table 2), Appendix Table C4 shows that positive updating occurs for both copartisans and non-copartisans.<sup>31</sup> That we observe negative treatment effects for non-copartisans on compliance and expected benefits suggests that the elite opinion leadership channel dominates the government legitimacy channel in our context. These negative sub-group effects align with research indicating that, while cues from political leaders may be persuasive for the political ingroup, they can generate backlash from the outgroup (Haas and Khadka 2020; Nicholson 2012).

---

<sup>30</sup>Affective partisan polarization is defined by an emotional attachment to the political in-group and an animosity towards the out-group party (see Iyengar et al. 2019, for a review). From this perspective, participation strengthens APC partisans' affective attachment to their party, thereby increasing their compliance with an APC-backed policy. At the same time, participation heightens opposition partisans' animosity towards the APC, leading to a decrease in their compliance with the APC-backed policy. Previous research has documented these dynamics in town hall settings (Hobolt et al. 2024).

<sup>31</sup>Specifically, we observe positive point estimates for non-copartisans on all legitimacy outcomes at endline. In fact, for the only outcome where treatment effects between copartisans and non-copartisans are statistically distinguishable (Mayor approval), the effects are larger for non-copartisans, likely because copartisans face ceiling effects.

## 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 and advocating for more fiscal capacity. Many governments in poor countries appear mired in similar, pernicious situations of low government legitimacy, low taxpayer compliance, and limited political support for strengthening tax systems. How can governments break out of this vicious cycle?

In this paper, we propose that governments can use direct democracy to overcome legitimacy constraints on state capacity building. We report results from a large-scale, digital participatory budgeting intervention developed to support a weak local government build fiscal capacity in both the short and medium term. In the short term, it aimed to support immediate improvements in tax compliance by increasing public confidence in government. In the medium term, it sought to enhance public perceptions of the legitimacy of the government in order to enable the government to pursue and sustain policy reform efforts (Besley and Dray 2024).

We present two primary findings. First, our results highlight that participatory interventions can improve citizens' attitudes towards government and bolster political legitimacy. These positive effects are consistent across political supporters and opponents of the mayor.

Second, despite those relatively universal and durable impacts on legitimacy, we find that impacts on compliance are heavily moderated by partisanship. We do not find any average effect of the intervention on tax compliance. Instead, we find significant positive impacts among co-partisans of the mayor, but significant negative impacts among non-copartisans. This adds substantial nuance to influential models of tax compliance (Levi 1988, 1997), the literature on the democratic dividend (Bó et al. 2010; Sutter et al. 2010), and studies of participatory budgeting (Pommerehne and Weck-Hannemann 1996; ?), all of which suggest a simpler link from expanded participation to increased tax compliance.

What does this imply for governments considering similar participatory interventions? One might conclude, focusing narrowly on short-term average compliance effects, that this inter-

vention was ineffective. However, we caution against this interpretation as there are several reasons that suggest the *total* effects of similar interventions may be positive. First, we find suggestive evidence that our intervention led to positive spillover effects on people who did not directly participate (Appendix E). Second, the long-term compliance impacts of participatory interventions may differ from their short-term effects. While we observed no immediate impact on compliance, the significant and durable increases in perceived government legitimacy suggest that the long-term effects could be more promising. 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, 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 participatory budgeting with sampling frames that allow for self-selection into eligibility or targeted at those populations we identified as most likely to react positively.

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, or if the Mayor—and messaging around compliance—had been less central to the participatory processes? By capturing the more nuanced impacts of participation on compliance, and broader prospects for building state capacity, our research also points toward this additional set of relatively unexplored questions.

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 budgeting can be used to create more supportive environments for governments which want to carry out ambitious, politically contentious investments in fiscal capacity.

## References

- Acemoglu, D., Johnson, S., and Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review*, 91(5):1369–1401.
- 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. *Lincoln Institute of Land Policy Working Paper*.
- Bates, R. H. and Lien, D.-H. (1985). A note on taxation, development, and representative government. *Politics & Society*, 14(1):53–70.
- Beath, A., Christia, F., and Enikolopov, R. (2017). Direct democracy and resource allocation: Experimental evidence from afghanistan. *Journal of Development Economics*, 124:199–213.
- Benjamini, Y., Krieger, A. M., and Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3):491–507.
- Bergeron, A., Tourek, G., and Weigel, J. L. (2024). The state capacity ceiling on tax rates: Evidence from randomized tax abatements in the drc. *Econometrica*, 92(4):1163–1193.
- Besley, T. (2020). State capacity, reciprocity, and the social contract. *Econometrica*, 88(4):1307–1335.
- Besley, T. and Dray, S. (2024). Trust and state effectiveness: The political economy of compliance. *The Economic Journal*, 134:2225–2251.
- Besley, T. and Persson, T. (2011). *Pillars of prosperity: The political economics of development clusters*. Princeton University Press.
- Besley, T. and Persson, T. (2013). Taxation and development. In *Handbook of Public Economics*, volume 5, pages 51–110. Elsevier.
- Besley, T. and Persson, T. (2014). Why do developing countries tax so little? *Journal of Economic Perspectives*, 28(4):99–120.
- 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.
- Boulianne, S. (2019). Building faith in democracy: Deliberative events, political trust and efficacy. *Political Studies*, 67(1):4–30.
- 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., Garfias, F., and Serrato, J. C. S. (2024). The fiscal contract up close: Experimental evidence from mexico city. 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.
- Callen, M., Weigel, J. L., and Yuchtman, N. (2023). Experiments about institutions. Technical report, National Bureau of Economic Research.
- Carrillo, P. E., Castro, E., and Scartascini, C. (2021). Public good provision and property tax compliance: Evidence from a natural experiment. *Journal of Public Economics*, 198:104422.
- Casey, K. (2018). Radical decentralization: does community-driven development work? *Annual Review of Economics*, 10:139–163.
- Casey, K., Glennerster, R., and Miguel, E. (2012). Reshaping institutions: Evidence on aid impacts using a preanalysis plan. *The Quarterly Journal of Economics*, 127(4):1755–1812.
- 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. 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.

- Cox, G. W. and McCubbins, M. D. (1986). Electoral politics as a redistributive game. *The Journal of Politics*, 48(2):370–389.
- Cullen, J. B., Turner, N., and Washington, E. (2021). Political alignment, attitudes toward government, and tax evasion. *American Economic Journal: Economic Policy*, 13(3):135–166.
- 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.
- De Tocqueville, A. (2010). *Democracy in America*. Liberty Fund.
- Dom, R., Custers, A., Davenport, S., and Prichard, W. (2022). *Innovations in tax compliance: Building trust, navigating politics, and tailoring reform*. World Bank Publications.
- Easton, D. (1975). A re-assessment of the concept of political support. *British Journal of Political Science*, 5(4):435–457.
- Fearon, J. D., Humphreys, M., and Weinstein, J. M. (2015). How does development assistance affect collective action capacity? results from a field experiment in post-conflict liberia. *American Political Science Review*, 109(3):450–469.
- Fishkin, J. S. (2002). Deliberative democracy. In *The Blackwell guide to social and political philosophy*, pages 221–238. Wiley Online Library.
- 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.
- Freetown City Council (2021). Transform freetown: Second year report. Accessed at: <https://fcc.gov.sl/transform-freetown-second-year-report-2020-2021/>.
- Gaines, B. J., Kuklinski, J. H., Quirk, P. J., Peyton, B., and Verkuilen, J. (2007). Same facts, different interpretations: Partisan motivation and opinion on iraq. *The Journal of Politics*, 69(4):957–974.
- 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.
- Gilley, B. (2006). The determinants of state legitimacy: Results for 72 countries. *International*

- Political Science Review*, 27(1):47–71.
- Golden, M. and Min, B. (2013). Distributive politics around the world. *Annual Review of Political Science*, 16(1):73–99.
- Gottlieb, J. and Hollenbach, F. M. (2018). Fiscal capacity as a moderator of the taxation-accountability hypothesis. Technical report, University of Texas Working Paper.
- Grieco, K., Meriggi, N. F., Michel, J., Prichard, W., and Stewart-Wilson, G. (2019). Simplifying property tax administration in africa: Piloting a points-based valuation in freetown, sierra leone. *ICTD Summary Brief No. 19*.
- 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.
- Hobolt, S. B., Lawall, K., and Tilley, J. (2024). The Polarizing Effect of Partisan Echo Chambers. *American Political Science Review*, 118(3):1464–1479.
- Humphreys, M., de la Sierra, R. S., and Van der Windt, P. (2019). Exporting democratic practices: Evidence from a village governance intervention in eastern congo. *Journal of Development Economics*, 140:279–301.
- Hutchison, M. L. and Johnson, K. (2011). Capacity to trust? institutional capacity, conflict, and political trust in africa, 2000–2005. *Journal of Peace Research*, 48(6):737–752.
- Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., and Westwood, S. J. (2019). The Origins and Consequences of Affective Polarization in the United States. *Annual Review of Political Science*, 22(1):129–146.
- Iyengar, S. and Westwood, S. J. (2015). Fear and Loathing Across Party Lines: New Evidence on Group Polarization. *American Journal of Political Science*, 59(3):690–707.
- 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.
- Kapon, S., Del Carpio, L., and Chassang, S. (2024). Using divide-and-conquer to improve tax collection. *The Quarterly Journal of Economics*, 139(4):2475–2523.
- 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 health insurance 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.
- Krause, B. (2020). Balancing purse and peace: tax collection, public goods and protests. *Berkeley, CA: Agricultural and Resource Economics, University of California, Berkeley*.
- Kresch, E. P., Walker, M., Best, M. C., Gerard, F., and Naritomi, J. (2023). Sanitation and property tax compliance: Analyzing the social contract in brazil. *Journal of Development Economics*, 160:102954.
- 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.
- Locke, J. (1690). *Second Treatise of Government*. Awnsham Churchill; Project Gutenberg, 2021.
- Luke, A. (2020). Freetown City Council starved of much needed funds to deliver public services. *The Sierra Leone Telegraph*.
- Magalhães, P. C. (2014). Government effectiveness and support for democracy. *European Journal of Political Research*, 53(1):77–97.
- Mansbridge, J. (1999). Should blacks represent blacks and women represent women? a contingent“ yes”. *The Journal of Politics*, 61(3):628–657.
- Markussen, T. and Tyran, J.-R. (2023). Is there a dividend of democracy? experimental evidence from cooperation games. *CESifo Working Paper*.
- 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.
- 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*, 104(3):566–583.

- Nicholson, S. P. (2012). Polarizing cues. *American Journal of Political Science*, 56(1):52–66.
- Norris, P. (2017). The conceptual framework of political support. In *Handbook on Political Trust*, pages 19–32. Edward Elgar Publishing.
- North, D. C. and Weingast, B. R. (1989). Constitutions and commitment: the evolution of institutions governing public choice in seventeenth-century england. *Journal of Economic History*, pages 803–832.
- 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.
- Olken, B. A. (2010). Direct democracy and local public goods: Evidence from a field experiment in indonesia. *American Political Science Review*, 104(2):243–267.
- 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.
- 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. (2015). *Taxation, responsiveness and accountability in Sub-Saharan Africa: the dynamics of tax bargaining*. Cambridge University Press.
- 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).
- Prichard, W., Kamara, A. B., and Meriggi, N. (2020). Freetown just implemented a new tax system that could quintuple revenue. *African Arguments*.
- Putnam, R. D. (1993). *Making democracy work: Civic traditions in modern Italy*. Princeton University Press.
- Robinson, J. A. (2023). Tax aversion and the social contract in africa. *Journal of African Economies*, 32(Supplement\_1):i33–i56.
- Rothstein, B. and Stolle, D. (2008). The state and social capital: An institutional theory of generalized trust. *Comparative Politics*, 40(4):441–459.
- Scharpf, F. W. (1997). Economic integration, democracy and the welfare state. *Journal of*

- European Public Policy*, 4(1):18–36.
- Sexton, R. (2017). The unintended effects of bottom-up accountability: Evidence from a field experiment in peru. Technical report, Working Paper.
- 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.
- Slemrod, J. (2019). Tax compliance and enforcement. *Journal of Economic Literature*, 57(4):904–954.
- Stantcheva, S. (2021). Understanding tax policy: How do people reason? *The Quarterly Journal of Economics*, 136(4):2309–2369.
- 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.
- 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.
- Timmons, J. F. and Garfias, F. (2015). Revealed corruption, taxation, and fiscal accountability: Evidence from brazil. *World Development*, 70:13–27.
- Torgler, B. (2005). Tax morale and direct democracy. *European Journal of Political Economy*, 21(2):525–531.
- Tyler, T. R. (2000). Social justice: Outcome and procedure. *International Journal of Psychology*, 35(2):117–125.
- Zaller, J. (1992). *The Nature and Origins of Mass Opinion*. Cambridge University.
- Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, 13:75–98.

# Online Appendix

## Table of Contents

---

<b>A Intervention Appendix</b>	<b>34</b>
A.1 Digital Town Halls: Pros and Cons . . . . .	34
A.2 Participation and Experience in DTH . . . . .	36
A.3 Voting and Project Implementation . . . . .	43
A.4 Project Pictures . . . . .	45
A.5 Project Timeline . . . . .	47
<b>B Research Design</b>	<b>48</b>
B.1 Sampling . . . . .	48
B.2 Survey Data Collection . . . . .	49
B.3 Matching for Treatment Assignment . . . . .	50
B.4 Treatment Assignment Map . . . . .	52
B.5 Distance to Closest Study Property . . . . .	53
B.6 Inference . . . . .	55
<b>C Additional Analyses of Treatment Effects</b>	<b>57</b>
C.1 Tax Compliance . . . . .	57
C.2 Legitimacy . . . . .	63
C.3 Political Engagement . . . . .	64
C.4 Fairness and Enforcement . . . . .	66
C.5 Polarization and Cohesion . . . . .	67
<b>D Comparing Partisan Groups</b>	<b>69</b>
D.1 Differences by Partisan Group . . . . .	69
D.2 Partisan Targeting . . . . .	73
<b>E Spillover Analysis</b>	<b>75</b>
<b>F Notification Calls</b>	<b>77</b>
<b>G Research Ethics</b>	<b>79</b>

---

## A Intervention Appendix

### A.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 ([Sexton 2017](#), 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, 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.

## A.2 Participation and Experience in DTH

Table A1 provides more details on participants' participation behavior. Panel A displays participants' self-reported participation behavior: how frequently they access the DTH chat group. Panel B reports behavioral measures of participation.

Panel A: Self-Reported Participation Frequency	%
Daily	53.6
Four to six times per week	8.0
Two or three times per week	22.7
Once per week	5.3
Never	5.0
Panel B: Behavioral Participation	
Voted for service [%]	0.68
Sent any message in DTH [%]	0.63
Median messages sent	2.00
Mean messages sent	3.84

**Table A1** reports DTH participation behavior. Panel A displays the self-reported frequency of accessing the DTH group, while Panel B presents behavioral participation measures, including voting and sending messages. The sample for these statistics consists of property owners who joined a DTH (i.e., participants).

Table A1: DTH Participation

Explanatory Variable	Dependent Variable: Sent at Least One Message			
	(1)	(2)	(3)	(4)
<i>Demographics:</i>				
Female	0.018 (0.026)		0.019 (0.026)	
Age	-0.003*** (0.001)		-0.003*** (0.001)	
Property value (log)	-0.033** (0.015)		-0.030** (0.015)	
Education: Attended some school	0.088** (0.037)		0.073* (0.037)	
<i>Partisan Affiliation:</i> (Baseline group = APC)				
SLPP		-0.019 (0.035)	-0.022 (0.035)	
Third Party (NGC/other)		0.059 (0.072)	0.056 (0.073)	
Independent		-0.004 (0.032)	0.021 (0.033)	
Did not answer		0.069** (0.034)	0.077** (0.034)	
Political Interest [1-4]			0.033*** (0.011)	0.034*** (0.011)
Num. Obs.	1801	1809	1794	1787
R <sup>2</sup>	0.014	0.004	0.005	0.023

Table A2 presents the results from OLS regressions, where the dependent variable is an indicator equal to 1 if the participant sent at least one message in the DTH, and 0 otherwise. Columns 1-3 report results from models that include only demographic characteristics (Column 1), only partisan affiliation (Column 2), and only political interest (Column 3). Column 4 presents results from a model that includes all three sets of variables. The *education* variable has three categories: some education, no education, and missing value. Education data is missing for 115 out of 1809 participants in the DTH sample; missingness (i.e., the “missing” category) is positively correlated with participation.

**Significance:** \*  $p < 0.10$ ; \*\*  $p < 0.05$  \*\*\*  $p < 0.01$

Table A2: Correlates of Participation: Who Messages?

Explanatory Variable	Dependent Variable: Voted for Preferred Service			
	(1)	(2)	(3)	(4)
<i>Demographics:</i>				
Female	-0.021 (0.026)			-0.021 (0.026)
Age		-0.003*** (0.001)		-0.003*** (0.001)
Property value (log)		0.000 (0.015)		0.004 (0.015)
Education: Attended some school		0.104*** (0.038)		0.096** (0.038)
<i>Partisan Affiliation:</i> (Baseline group = APC)				
SLPP		-0.033 (0.035)		-0.034 (0.035)
Third Party (NGC/other)		0.071 (0.070)		0.061 (0.069)
Independent		-0.058* (0.032)		-0.048 (0.033)
Did not answer		0.022 (0.034)		0.026 (0.034)
Political Interest [1-4]			0.026** (0.011)	0.019 (0.011)
Num. Obs.	1801	1809	1794	1787
R <sup>2</sup>	0.013	0.005	0.003	0.021

Table A3 presents the results from OLS regressions, where the dependent variable is an indicator that equals 1 if the participant voted for their preferred service, and 0 otherwise. Columns 1-3 report results from models that include only demographic characteristics (Column 1), only partisan affiliation (Column 2), and only political interest (Column 3). Column 4 presents results from a model that includes all three sets of variables. The *education* variable has three categories: some education, no education, and missing value. Education data is missing for 115 out of 1809 participants in the DTH sample; missingness (i.e., the “missing” category) is positively correlated with voting.

**Significance:** \*  $p < 0.10$ ; \*\*  $p < 0.05$  \*\*\*  $p < 0.01$

Table A3: Correlates of Participation: Who Votes for Services?

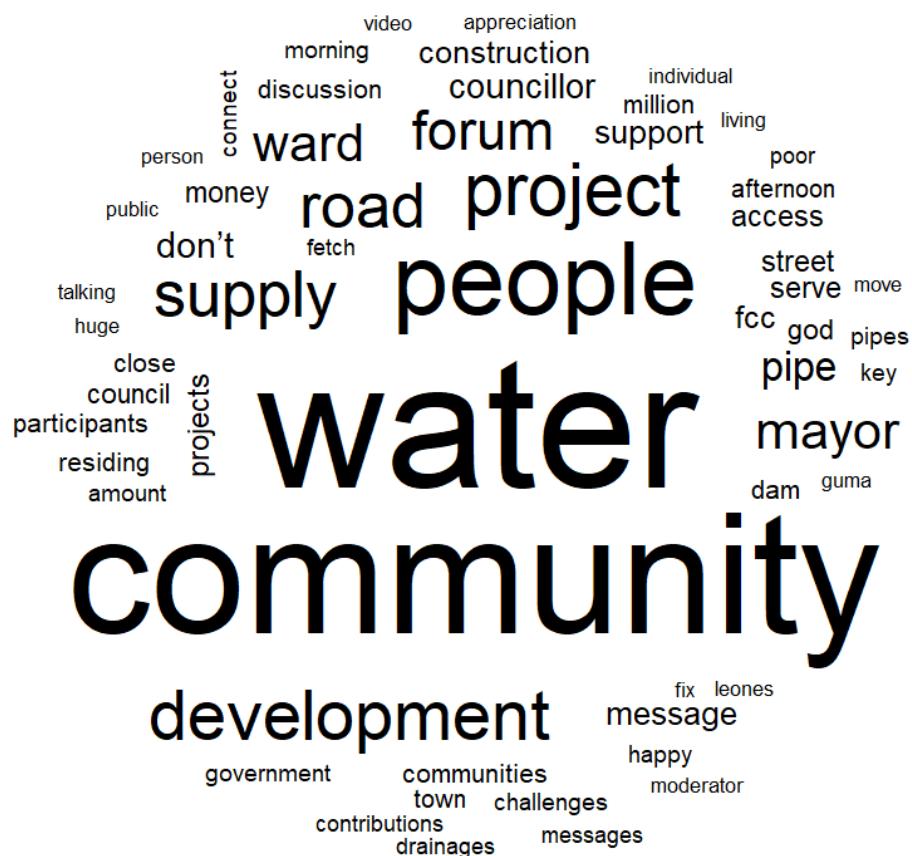
	Dependent Variable: Accessed Chat Group Daily			
	(1)	(2)	(3)	(4)
<i>Demographics:</i>				
Female	-0.038 (0.025)			-0.034 (0.025)
Age	-0.003*** (0.001)			-0.003*** (0.001)
Property value (log)	-0.013 (0.014)			-0.012 (0.015)
Education: Attended some school	0.156*** (0.033)			0.151*** (0.033)
<i>Partisan Affiliation:</i> (Baseline group = APC)				
SLPP		-0.037 (0.034)		-0.039 (0.034)
Third Party (NGC/other)		0.031 (0.072)		0.019 (0.071)
Independent		-0.045 (0.031)		-0.030 (0.032)
Did not answer		-0.015 (0.034)		-0.004 (0.034)
Political Interest [1-4]			0.023** (0.011)	0.017 (0.011)
Num. Obs.	1801	1809	1794	1787
R <sup>2</sup>	0.021	0.002	0.003	0.023

Table A4 presents the results from OLS regressions, where the dependent variable is an indicator equal to 1 if the participant accessed the WhatsApp chat group daily (as self-reported), and 0 otherwise. Columns 1-3 report results from models that include only demographic characteristics (Column 1), only partisan affiliation (Column 2), and only political interest (Column 3). Column 4 presents results from a model that includes all three sets of variables. The *education* variable has three categories: some education, no education, and missing value. Education data is missing for 115 out of 1809 participants in the DTH sample; missingness (i.e., the “missing” category) is positively correlated with daily access.

**Significance:** \*  $p < 0.10$ ; \*\*  $p < 0.05$  \*\*\*  $p < 0.01$

Table A4: Correlates of Participation: Who Checks DTH Daily?

Figure B1: Word Cloud: Content of DTH Messages



*Note:* Figure B1 relies on a randomly drawn sample of 100 DTH messages which were transcribed by the research team. The R package tidytext was used to identify and remove stopwords. Numbers were also removed. Only words are shown that were mentioned at least ten times in this sample of messages. Due to the labor intensiveness of transcribing, we decided to limit the analysis to 100 messages.

We asked respondents seven questions about their experience in the DTH, on a five-point Likert scale (Table A5). These questions were asked in both positive and negative forms, so as to limit social desirability bias in the average response. For example, we asked some respondents if they agreed with the following statement: “The Town Hall allowed me to let my political representatives know about my views.” We asked other respondents if they agreed with the negative version of that statement: “The Town Hall **did not** allow me to let my political representatives know about my views.” Table reports the seven statements and the average agreement with each statement.

We also asked participants which actors they believed were responsible for organizing, implementing, and funding the DTHs (Appendix Table A6). For these questions, respondents were allowed to name multiple actors they thought might be involved.

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 (LE 15 Million) sufficient to meaningfully 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

*Note:* We asked respondents seven questions about their experience in the DTH. 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.

Table A5: DTH Experience

Activity	Perceived Responsible Actor				
	FCC	Govt.	Researchers	Citizens	Other
Organized	89.3	1.9	12.6	0.2	1.4
Implemented	96.1	4.5	1.8	1.7	0.6
Funded	84.2	10.6	2.3	11.5	5.6

*Note:* This table 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 A6: Organization, Implementation, Funding

### A.3 Voting and Project Implementation

Table A7 shows how participants voted. Most participants voted for water-related projects and these were the winning projects in nearly all wards. Most participants got the project they voted for, or at least one fairly similar to what they voted for. Even if a respondent did not get exactly the project they voted for, the selected project addressed a similar issue. For example, hand pumps and fixing water pipes both improve water services.<sup>32</sup> Considered this way, 75% of the voting participants got the project they wanted, or one close to it.

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. Implementation was further delayed in Tengbeh Town where construction was delayed because the implementing construction company wanted additional assurances from the FCC regarding potential liability issues.

	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

*Note:* The top row (“Votes”) describes the number of votes for each project. The middle row (“Won”) shows how those votes translate to the number of projects won for each project type. The bottom row (“Built”) indicates how many projects were delivered by type.

Table A7: Project Votes, Winning Projects, Implemented Projects

---

<sup>32</sup>Or, if a respondent voted for 50 meters of tire tracks, a project that fixes potholes still addresses the issue of road repair.

## Figure B2: Menu of Services

Q1.

Which project would you like to be implemented in your ward?

Each project is worth 15 million leones.

Fixing of potholes

A new water hand pump

2 new solar street lights

Fixing of water pipes

50m of truck tracks



## A.4 Project Pictures

Figure B3: Project Implemented in Ward 418

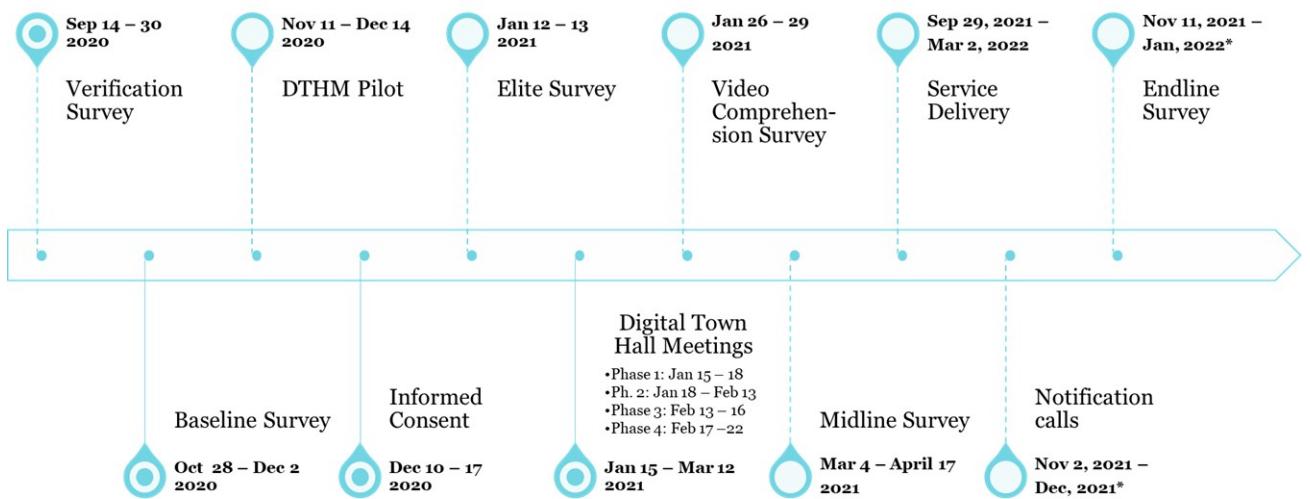


Figure B4: Project Implemented in Ward 442



## A.5 Project Timeline

Figure B5: 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.

## B Research Design

### B.1 Sampling

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. 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).

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. As a response to COVID-19, the FCC waived 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.

Second, we removed some properties from the sample frame to limit geographic spillovers. This was largely motivated by a previous version of our research design, where 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, 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.

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

## B.2 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 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.<sup>33</sup>

**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%).<sup>34</sup>

**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%).

---

<sup>33</sup> Appendix Figure B5 documents the broader data collection and project timeline.

<sup>34</sup> We incentivized midline and endline survey responses by offering packages of mobile data.

### B.3 Matching for Treatment Assignment

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 (“optGreedy”) matching algorithm to find best matches along Mahalanobis distance. We weight certain variables higher than others when matching, as we expect that certain variables are a stronger predictor of our outcomes of interest. We place the greatest weight on unconditional tax morale which we expect to be the strongest predictor of tax compliance, in line with its common use as a 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,<sup>35</sup> and (vi) the number of these five variables that were imputed.

Table B1 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. The last column displays the number of observations that were imputed for matching. Note that in general, the number of imputed responses is low.

---

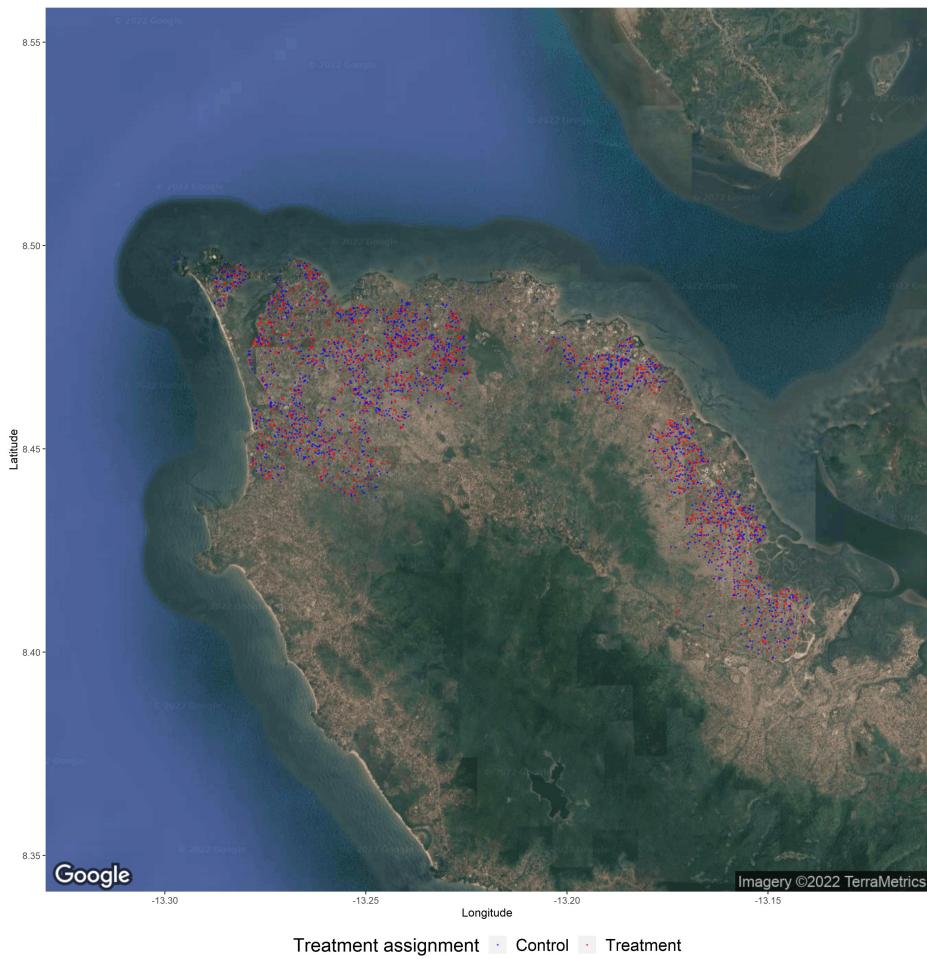
<sup>35</sup>We create a three-level ordinal variable based on two survey items. A first group consists of respondents who have heard of the reform and strongly/somewhat support it; a second group consists of respondents who (a) have heard of the reform and feel neutral towards it and (b) have not heard of the reform; a third group consists of respondents who have heard of the reform and somewhat/strongly oppose it.

<b>Variable Name</b>	<b>Weights</b>	<b>Mean</b>	<b>SD</b>	<b>Min</b>	<b>Max</b>	<b>N Imputed</b>
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

Table B1: Summary Statistics of Matching Variables

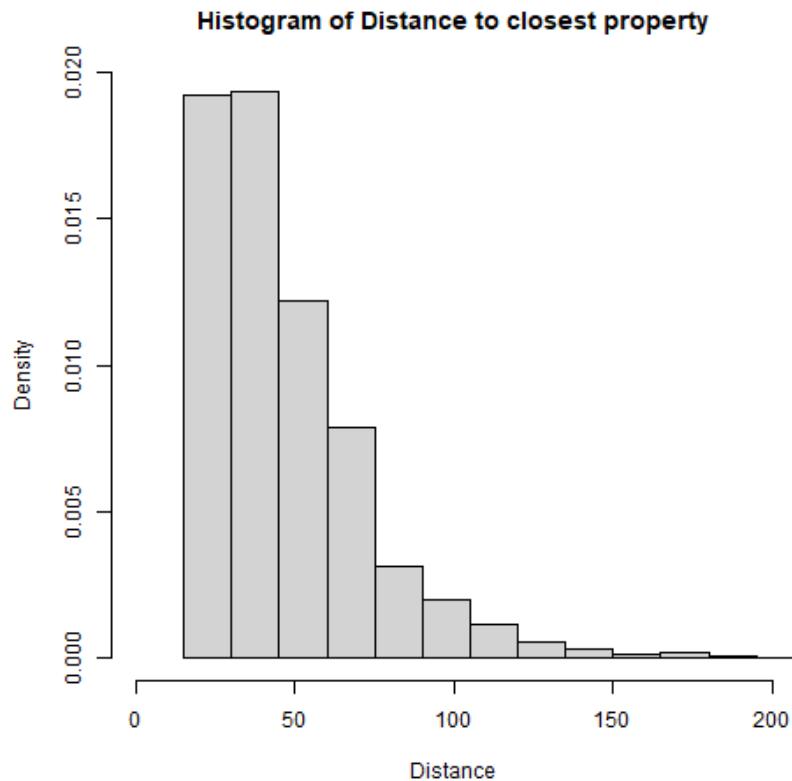
## B.4 Treatment Assignment Map

Figure B1: Digital Town Hall Treatment Assignment in Freetown (Red = Treatment)



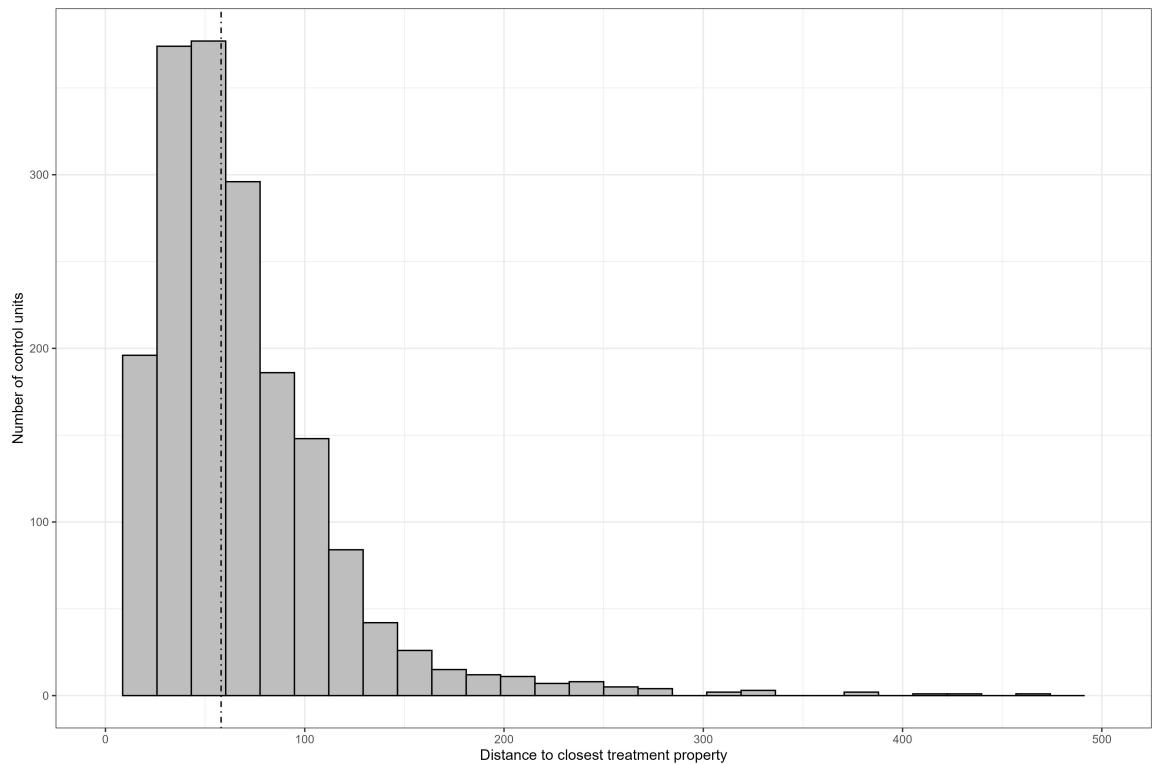
## B.5 Distance to Closest Study Property

Figure B2: Histogram of Minimum Distance (in Meters) Between Study Properties



*Note:* Figure B2 shows the distribution of the distance from each property to the closest property in the sample.

Figure B3: Distribution of Distance Between Control Units and the Closest Treatment Unit



*Note:* Figure B3 plots the distribution of distance between control units and the closest treatment unit. Eight control units have a minimum distance greater than 500 meters and are excluded from Figure B3.

The dotted vertical line shows the median distance (58 meters).

## B.6 Inference

This paper integrates analyses from two separate pre-analysis plans (PAPs). The first PAP investigates the effectiveness of DTHs as tools for enhancing political accountability during COVID-19 (registered at: <https://osf.io/cg738>). The second PAP focuses on tax compliance. Notably, five indicators from the COVID-19 PAP were reshuffled into new hypothesis families in the tax compliance PAP.<sup>36</sup>

To adjust for multiple comparisons, we use the two-step correction method outlined by [Anderson \(2008\)](#), which involves grouping hypotheses into families and then applying corrections within these families. For outcomes that appear in only one PAP (either COVID-19 or tax compliance), we adjust them within the hypothesis family they were originally assigned to. For indicators linked to both PAPs, we assign them to the hypothesis family described in the tax compliance PAP for adjustment. Appendix Table B2 maps the indicators to their respective hypothesis families for adjustment. Column 1 lists all attitudinal outcomes from both PAPs; Column 2 indicates whether the outcome is included in the COVID-19 PAP (C), the tax compliance PAP (T), or both; Column 3 shows the hypothesis family used for adjusting *p*-values, and Column 4 identifies the table number where results for each indicator are presented.

---

<sup>36</sup>We had analyzed effects of the DTH on midline survey outcomes at the time of writing the tax PAP, but we had not yet analyzed any endline data. We note this for transparency reasons, but do not believe that the analysis of midline outcomes impacted our analysis plan for the tax compliance study in any meaningful way.

<b>Outcome</b>	<b>PAP</b>	<b>Family</b>	<b>Table</b>
Fiscal exchange attitudes	T	Fiscal Exchange	2
Satisfaction with FCC service delivery	Both	Fiscal Exchange	1
Opportunities to voice opinions to govt.	Both	Political Efficacy	1
Ease of participating in political activities	Both	Political Efficacy	1
FCC responsiveness to citizens' demands	Both	Political Efficacy	1
Reform improves tax system fairness	T	Fairness of Taxation	17
Number of neighbors who will pay property tax	T	Fairness of Taxation	17
Likelihood detected noncompliers are punished	Both	Enforcement	17
Mayor approval	C	Attitudes Towards Govt.	1
Councilor approval	C	Attitudes Towards Govt.	1
FCC efficiency	C	Attitudes Towards Govt.	1
FCC corruption	C	Attitudes Towards Govt.	1
FCC transparency	C	Attitudes Towards Govt.	1
Satisfaction with the political system	C	Attitudes Towards Govt.	N/A
Support for direct democracy	C	Attitudes Towards Govt.	N/A
Knows ward councilor name	C	Political Knowledge and Efficacy	14
Attempted to contact ward councilor	C	Political Knowledge and Efficacy	14
Attempted to contact MP	C	Political Knowledge and Efficacy	14
Level of interest in politics	C	Political Knowledge and Efficacy	14
Level of interest in FCC activities	C	Political Knowledge and Efficacy	14
Attended political meeting	C	Political Knowledge and Efficacy	14
Level of trust in neighbors	C	Polarization and Cohesion	18
Level of connection of neighbors	C	Polarization and Cohesion	18
Ease of befriending someone in opposition party	C	Polarization and Cohesion	18
Ease of believing opposition party member	C	Polarization and Cohesion	18

Table B2: Inference

## C Additional Analyses of Treatment Effects

### C.1 Tax Compliance

Model	2022			2021		
	Est	SE	p-value	Est	SE	p-value
Baseline Compliance	-0.012	0.018	0.515	-0.010	0.019	0.584
Baseline Compliance + ward FE	-0.010	0.018	0.574	-0.008	0.019	0.679
Baseline Compliance + prop. covs	-0.012	0.018	0.512	-0.008	0.019	0.660
Main spec. + zero liability dummy	-0.011	0.017	0.530	-0.007	0.019	0.707

Table C1 reports treatment effects on tax compliance behavior for 2022 and 2021, using alternative model specifications. Columns 1 and 4 present treatment effects estimates in raw percentage points for 2022 and 2021, respectively. Columns 2 and 5 report standard errors; Column 3 and 6 report *p*-values.

**Significance:** \*  $p < 0.10$ ; \*\*  $p < 0.05$  \*\*\*  $p < 0.01$

Table C1: Effect on Tax Compliance: Alternative Specifications

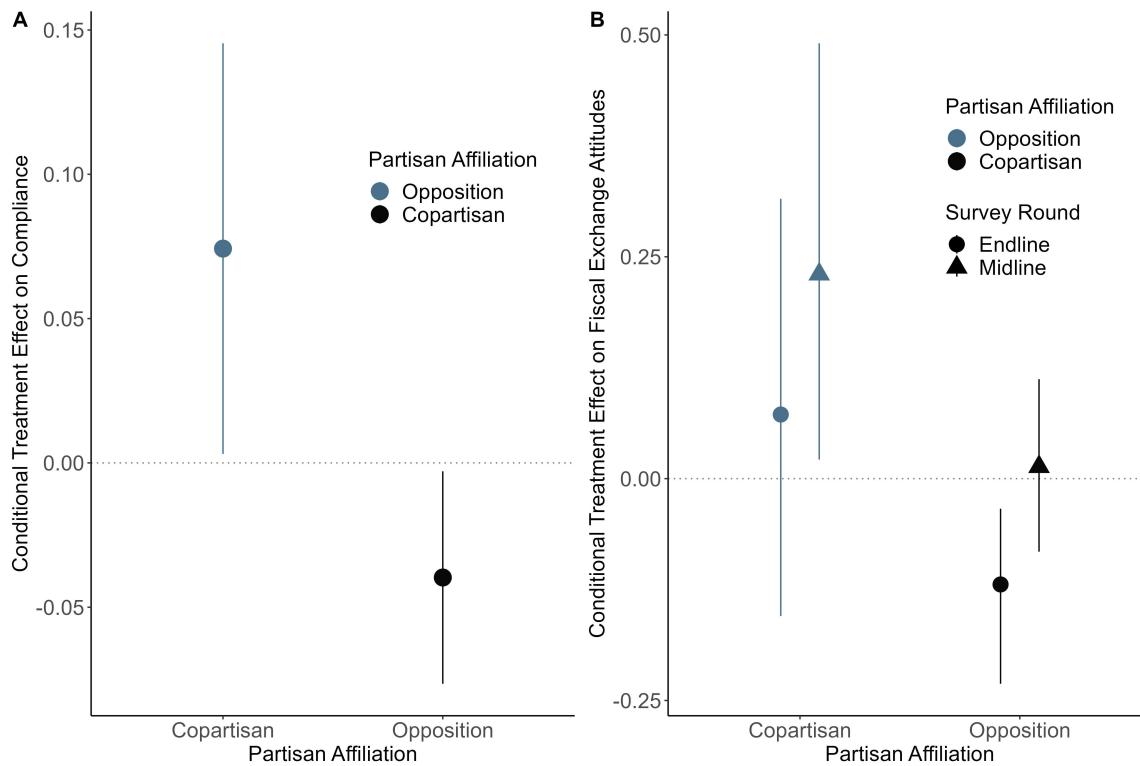
Outcome	2022			2021		
	Est	SE	p-value	Est	SE	p-value
Total paid (USD)	2.944	3.845	0.444	3.922	2.999	0.191
Log total paid (USD)	-0.005	0.011	0.664	-0.001	0.012	0.900

Table C2 reports treatment effects on tax compliance behavior for 2022 and 2021 using alternative operationalizations of the dependent variable. Columns 1 and 4 present treatment effects 2022 and 2021, respectively. Columns 2 and 5 report standard errors; Column 3 and 6 report *p*-values. We convert local currency (SLL) to USD at a rate of 10,000:1, which reflects the exchange rate in January, 2021.

**Significance:** \*  $p < 0.10$ ; \*\*  $p < 0.05$  \*\*\*  $p < 0.01$

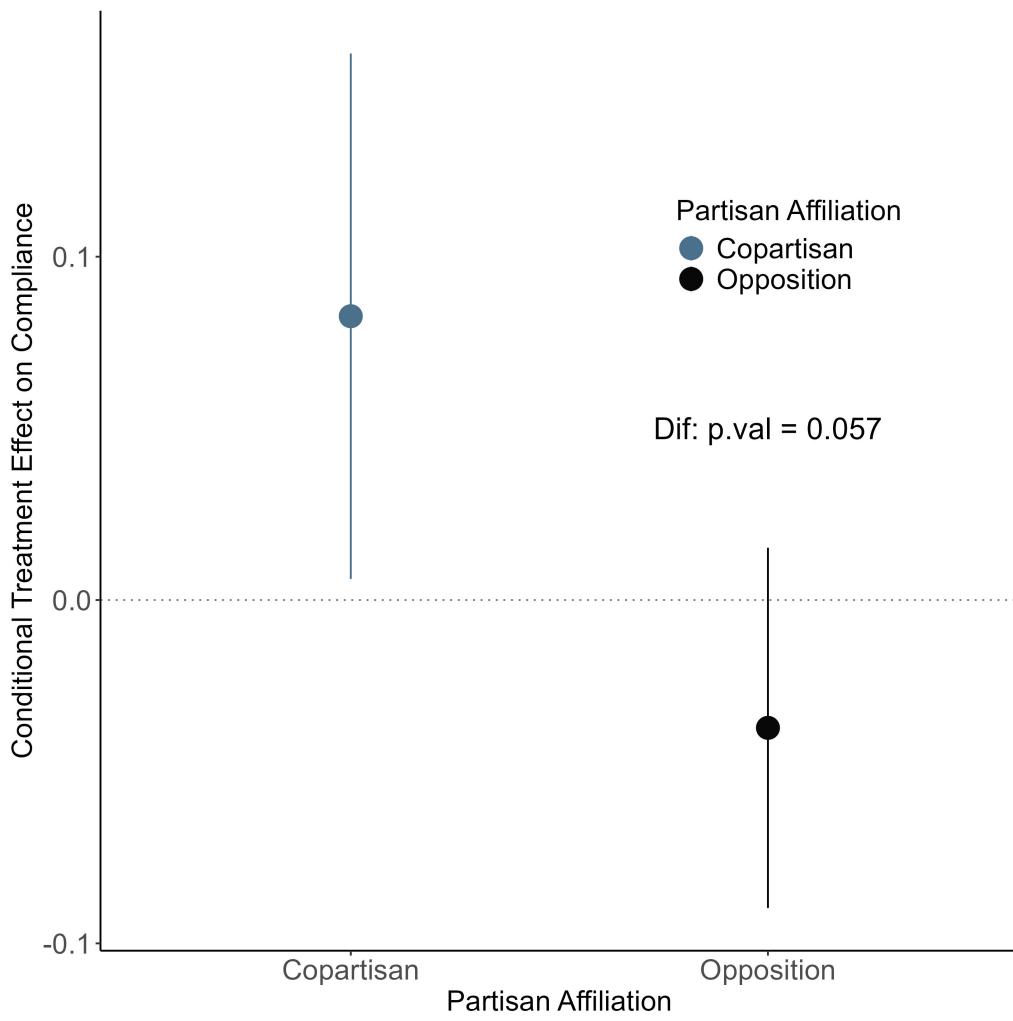
Table C2: Effect on Tax Compliance: Alternative Operationalizations

Figure C1: 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."

Figure C2: Treatment Effects Conditional on Partisan Affiliation (Alternative Coding)



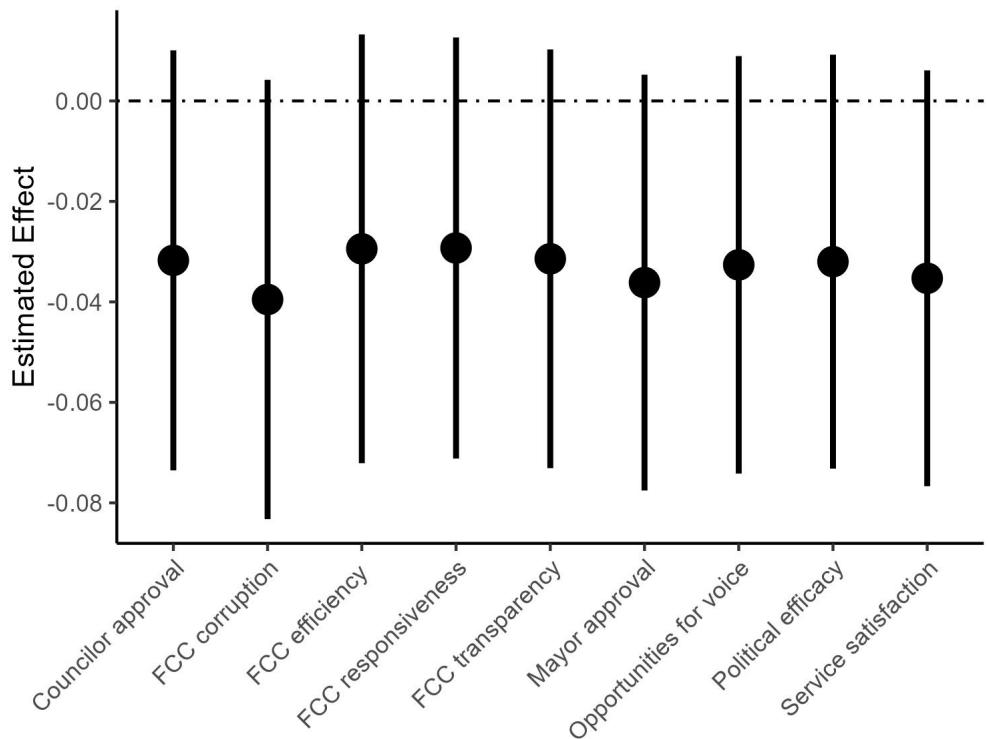
*Note:* Figure C2 presents treatment effects on tax compliance behavior, conditional on partisan affiliation using an alternative coding. Respondents who self-report affinity towards the All People's Congress are coded as "copartisans." Respondents who report affinity to a party other than APC, or who report no affinity towards any party, are coded as "opposition." Respondents who do not answer this question are dropped.

Partisan Group	Sub-group Estimates			Dif APC
	CATE	SE	n	<i>p</i> -value
APC	0.057	0.034	880	0.095
SLPP	-0.035	0.041	720	0.104
Third Party (NGC/Other)	-0.100	0.120	99	0.212
Independent	-0.012	0.033	1115	0.170
No Response	-0.036	0.039	804	0.085

**Table C3** reports treatment effects on compliance by partisan group. Column 1 shows the sub-group treatment effect, Column 2 reports the standard error, and Column 3 displays the number of observations in each sub-group. For all groups other than APC, Column 4 reports the *p*-value for the difference with the APC sub-group treatment effect. For the APC effect, Column 4 reports the *p*-value for the difference from zero.

Table C3: Effect on Tax Compliance: By Party

Figure C3: Reanalysis of Compliance Effect Dropping Attritors



*Note:* Figure C3 reanalyzes treatment effects on compliance, excluding property owners who attrited from the survey. Since the sub-sample of respondents varies by survey question, Figure C3 presents compliance effects separately for each question's responding sub-sample.

## C.2 Legitimacy

Outcome	Midline (CATE)			Endline (CATE)		
	Copart.	Non-Copart.	Diff	Copart.	Non-Copart.	Diff
<b>Policy Influence</b>						
Opportunities to voice opinion to govt	0.398 (0.097)	0.368 (0.048)	0.030 (0.119)	0.186 (0.122)	0.266 (0.058)	-0.081 (0.148)
Ease of participating in political activities	0.186 (0.127)	0.037 (0.056)	0.149 (0.152)	0.217 (0.155)	0.037 (0.063)	0.180 (0.182)
<b>Service Delivery and Responsiveness</b>						
FCC responsiveness to citizens' demands	0.189 (0.112)	0.160 (0.058)	0.029 (0.140)	0.222 (0.157)	0.111 (0.072)	0.111 (0.190)
Satisfaction with FCC service provision	0.239 (0.120)	0.204 (0.059)	0.034 (0.147)	0.285 (0.144)	0.133 (0.074)	0.152 (0.177)
<b>Government Administrative Competence</b>						
FCC corruption	-0.067 (0.114)	-0.165 (0.054)	0.097 (0.138)	0.174 (0.117)	0.060 (0.061)	0.114 (0.144)
FCC efficiency	0.030 (0.066)	0.025 (0.034)	0.004 (0.083)	0.109 (0.088)	0.085 (0.044)	0.024 (0.108)
FCC transparency	0.259 (0.104)	-0.005 (0.045)	0.264 (0.124)	0.298 (0.197)	0.197 (0.087)	0.101 (0.236)
<b>Approval of Political Representatives</b>						
Mayor approval	0.104 (0.097)	0.141 (0.047)	-0.037 (0.119)	0.014 (0.116)	0.220 (0.058)	-0.206 (0.143)
Councilor approval	0.397 (0.135)	0.184 (0.060)	0.213 (0.162)	0.097 (0.147)	0.244 (0.071)	-0.147 (0.177)

**Table C4** reports treatment effects on legitimacy outcomes, conditional on partisan affiliation, at the midline and endline survey. Columns 1 and 2 report treatment effects at midline for copartisans and non-copartisans, respectively. Column 3 reports the difference in treatment effects between subgroups. Treatment effects are reported and standard deviation units and standard errors in parentheses. Columns 4-6 report similar estimates for the endline survey. Respondents who self-report feeling “close to” APC are defined as copartisans; all other respondents are coded as non-copartisans.

**Significance:** \*  $p < 0.10$ ; \*\*  $p < 0.05$  \*\*\*  $p < 0.01$

Table C4: Effects on Legitimacy Outcomes Conditional on Partisan Affiliation

### C.3 Political Engagement

Outcome	Baseline		Midline			Endline			
	Mean	Mean	Effect	N	q-val	Mean	Effect	N	q-val
Knows ward councilor name	0.360 (0.480)	0.383 (0.486)	0.104*** (0.040)	3,618	0.004	0.313 (0.463)	0.044 (0.039)	3,618	0.292
Attempted to contact ward councilor	0.193 (0.395)	0.188 (0.391)	0.214*** (0.044)	3,299	0.001	0.264 (0.440)	0.084 (0.052)	2,865	0.292
Attempted to contact MP	0.112 (0.316)	0.092 (0.289)	0.142*** (0.043)	3,297	0.002	0.159 (0.365)	-0.023 (0.055)	2,865	0.693
Level of interest in politics	1.841 (1.090)	1.818 (1.001)	0.161*** (0.040)	3,299	0.001	2.028 (1.096)	0.091* (0.055)	2,575	0.292
Level of interest in FCC activities	2.952 (1.097)	3.103 (0.896)	0.428*** (0.034)	3,300	0.001	3.157 (0.954)	-0.009 (0.043)	2,871	0.704
Attended political meeting	1.157 (0.501)	1.175 (0.545)	0.051 (0.049)	3,301	0.052	1.150 (0.516)	0.104* (0.058)	2,558	0.292

**Table C5** 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 report 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 last six months.

**Significance:** \*  $p < 0.10$ ; \*\*  $p < 0.05$  \*\*\*  $p < 0.01$

Table C5: Political Engagement

Outcome	Midline			Endline		
	Copart.	Non Copart.	Diff	Copart.	Non Copart.	Diff
Knows ward councilor name	0.023 (0.051)	0.058 (0.024)	-0.035 (0.062)	-0.025 (0.049)	0.036 (0.024)	-0.060 (0.061)
Attempted to contact ward councilor	0.055 (0.049)	0.094 (0.022)	-0.039 (0.058)	0.065 (0.055)	0.022 (0.026)	0.042 (0.066)
Attempted to contact MP	0.033 (0.040)	0.049 (0.017)	-0.016 (0.048)	0.040 (0.046)	-0.023 (0.022)	0.063 (0.056)
Level of interest in politics	0.292 (0.123)	0.139 (0.053)	0.153 (0.145)	0.289 (0.165)	0.036 (0.075)	0.253 (0.199)
Level of interest in FCC activities	0.541 (0.099)	0.446 (0.047)	0.095 (0.120)	0.153 (0.123)	-0.060 (0.059)	0.214 (0.150)
Attended political meeting	0.071 (0.069)	0.011 (0.030)	0.060 (0.081)	0.070 (0.083)	0.046 (0.035)	0.024 (0.098)

**Table C6** reports treatment effects on political engagement, conditional on partisan affiliation, at the midline and endline survey. Columns 1 and 2 report treatment effects at midline for copartisans and non-copartisans, respectively. Column 3 reports the difference in treatment effects between subgroups. Treatment effects are reported in standard deviation units and standard errors are in parentheses. Columns 4-6 report similar estimates for the endline survey. Respondents who self-report feeling “close to” APC are defined as copartisans; all other respondents are coded as non-copartisans.

**Significance:** \*  $p < 0.10$ ; \*\*  $p < 0.05$  \*\*\*  $p < 0.01$

Table C6: Effects on Political Engagement Outcomes Conditional on Partisan Affiliation

## C.4 Fairness and Enforcement

At endline, we find no evidence of persistent treatment effects on either fairness or enforcement mechanism outcomes (Table C7). 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.

Outcome	Baseline		Midline			Endline			
	Mean	Mean	Effect	N	<i>q</i> -val	Mean	Effect	N	<i>q</i> -val
<b>Fairness</b>									
Reform improves tax system fairness	2.113 (0.796)	2.152 (0.691)	0.125** (0.057)	2,252	0.017	2.381 (0.782)	-0.005 (0.049)	2,852	1.000
Number of neighbors who will pay property tax	5.100 (2.381)	5.971 (2.289)	-0.209*** (0.052)	2,878	0.001	5.919 (2.448)	-0.006 (0.060)	2,489	1.000
<b>Enforcement</b>									
Likelihood detected noncompliers are punished	4.060 (1.105)	4.241 (0.983)	-0.316*** (0.044)	3,301	0.001	4.136 (1.042)	0.043 (0.046)	2,857	0.493

**Table C7** reports the effect of the treatment on the alternative mechanisms of fairness and enforcement. Columns 1, 2, and 6 report 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 7 presents treatment effects estimates at the endline survey. Column 4 and 8 report the number of non-missing observations in the midline survey and endline survey, respectively. Reported effects are standardized effects. 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.

**Significance:** \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table C7: Effect on Fairness and Enforcement

## C.5 Polarization and Cohesion

Outcome	Baseline		Midline			Endline			
	Mean	Mean	Effect	N	q-val	Mean	Effect	N	q-val
Level of trust in neighbors	3.23 (0.81)	3.15 (0.69)	0.131*** (0.038)	3,225	0.011	3.17 (0.74)	0.017 (0.045)	2,799	1.000
Level of connection of neighbors	4.56 (0.70)	4.49 (0.68)	0.016 (0.044)	3,228	0.193	4.53 (0.70)	-0.053 (0.049)	2,798	1.000
Ease of befriending out-party members	3.45 (0.94)	4.22 (0.94)	-0.178*** (0.052)	3,214	0.011	4.13 (1.04)	-0.008 (0.058)	2,738	1.000
Ease of believing out-party members	2.98 (1.50)	2.52 (1.31)	-0.068 (0.041)	3,206	0.076	2.62 (1.31)	-0.053 (0.046)	2,672	1.000

**Table C8** reports treatment effects on affective polarization at the midline and endline surveys. 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 report 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. **Significance:** \*  $p < 0.10$ ; \*\*  $p < 0.05$  \*\*\*  $p < 0.01$

Table C8: Affective Polarization

Outcome	Midline			Endline		
	Copart.	Non-Copart.	Diff	Copart.	Non-Copart.	Diff
Level of trust in neighbors	0.193 (0.084)	0.078 (0.038)	0.115 (0.101)	0.100 (0.100)	-0.014 (0.044)	0.114 (0.118)
Level of connection of neighbors	0.107 (0.087)	-0.019 (0.039)	0.125 (0.105)	0.051 (0.082)	-0.067 (0.042)	0.118 (0.099)
Ease of befriending opposition party members	-0.264 (0.133)	-0.136 (0.062)	-0.128 (0.163)	0.068 (0.152)	-0.033 (0.068)	0.100 (0.182)
Ease of believing opposition party members	-0.050 (0.163)	-0.118 (0.077)	0.068 (0.197)	-0.114 (0.172)	-0.069 (0.088)	-0.045 (0.211)

**Table C9** reports treatment effects on political polarization and social cohesion, conditional on partisan affiliation, at the midline and endline surveys. Columns 1 and 2 report treatment effects at midline for copartisans and non-copartisans, respectively. Column 3 reports the difference in treatment effects between subgroups. Treatment effects are reported in standard deviation units and standard errors are in parentheses. Columns 4-6 report similar estimates for the endline survey. Respondents who self-report feeling “close to” APC are defined as copartisans; all other respondents are coded as non-copartisans.

**Significance:** \*  $p < 0.10$ ; \*\*  $p < 0.05$  \*\*\*  $p < 0.01$

Table C9: Effects on Political Polarization and Social Cohesion Conditional on Partisan Affiliation

## D Comparing Partisan Groups

### D.1 Differences by Partisan Group

Measure	Mean		All	Difference		Observations	
	Opp.	APC		Raw	Std.	p-val	Opp.
<b>Approval of APC Political Representatives</b>							
Mayor approval	4.17	4.41	0.89	0.24	0.27***	0.00	2,669
Councilor approval	2.66	2.96	1.22	0.30	0.25***	0.00	2,633
<b>FCC Service Delivery and Responsiveness</b>							
Satisfaction with FCC service provision	3.58	3.83	1.17	0.25	0.21***	0.00	2,710
FCC responsiveness to citizens' demands	3.14	3.27	1.19	0.13	0.11***	0.01	2,575
<b>FCC Administrative Competence</b>							
FCC corruption	3.49	3.65	1.00	0.16	0.16***	0.00	2,176
FCC efficiency	2.82	2.98	0.71	0.16	0.22***	0.00	2,310
FCC transparency	1.37	1.33	0.69	-0.04	-0.06	0.11	2,594
<b>Ideology</b>							
Tax rich more for services	2.08	2.07	0.95	-0.01	-0.01	0.69	2,710

**Table D1** presents approval of APC political leaders and APC-led government institutions by partisan group. Columns 1 and 2 report group means for supporters of the opposition (“opp.”) and APC partisans, respectively. Column 3 reports the standard deviation (pooled). Column 4 reports the raw difference in means between APC supporters and opposition supporters and Column 5 standardizes this difference by the standard deviation. Column 6 reports the *p*-value on this difference. Columns 7 and 8 report non-missing observations for opposition and APC supporters, respectively. All respondents who self-report feeling “close to” APC are defined as APC supporters; all other respondents are coded as opposition. A higher value of the outcome *Tax Rich More for Services* indicates *disagreement*.

**Significance:** \*  $p < 0.10$    \*\*  $p < 0.05$    \*\*\*  $p < 0.01$

Table D1: Approval of Political Representatives by Partisan Group

Measure	Mean		ALL	Difference			Observations	
	Opp.	APC		Raw	Std.	p-val	Opp.	APC
<b>Demographics</b>								
Female	0.31	0.28	0.46	-0.03	-0.07**	0.05	2,738	880
Age	51.84	51.52	12.96	-0.32	-0.02	0.52	2,728	879
Married	0.71	0.74	0.45	0.03	0.07	0.15	2,729	880
Higher education	0.41	0.36	0.49	-0.05	-0.10**	0.03	2,539	840
<b>Property Characteristics</b>								
Total property tax owed (USD, 2021)	96.91	86.95	183.81	-9.96	-0.05	0.11	2,738	880
Number of properties with tax liability (2021)	1.89	1.99	1.47	0.10	0.07*	0.10	2,738	880
Property has water	0.48	0.44	0.50	-0.04	-0.08*	0.07	2,738	880
Property has drainage	0.36	0.35	0.48	-0.01	-0.02	0.35	2,738	880
In informal settlement	0.06	0.06	0.23	0.00	0.00	0.63	2,738	880

**Table D2** presents demographic variables and property characteristics by partisan group. Columns 1 and 2 report group means for supporters of the opposition (“opp.”) and APC partisans, respectively. Column 3 reports the standard deviation (pooled). Column 4 reports the raw difference in means between APC supporters and opposition supporters and Column 5 standardizes this difference by the standard deviation. Column 6 reports the *p*-value on this difference. Columns 7 and 8 report non-missing observations for opposition and APC supporters, respectively. All respondents who self-report feeling “close to” APC are defined as APC supporters; all other respondents are coded as opposition. The variable *Higher education* is an indicator variable equal to 1 if the respondent has a degree from a university, polytechnic school, or teachers college.

**Significance:** \*  $p < 0.10$     \*\*  $p < 0.05$     \*\*\*  $p < 0.01$

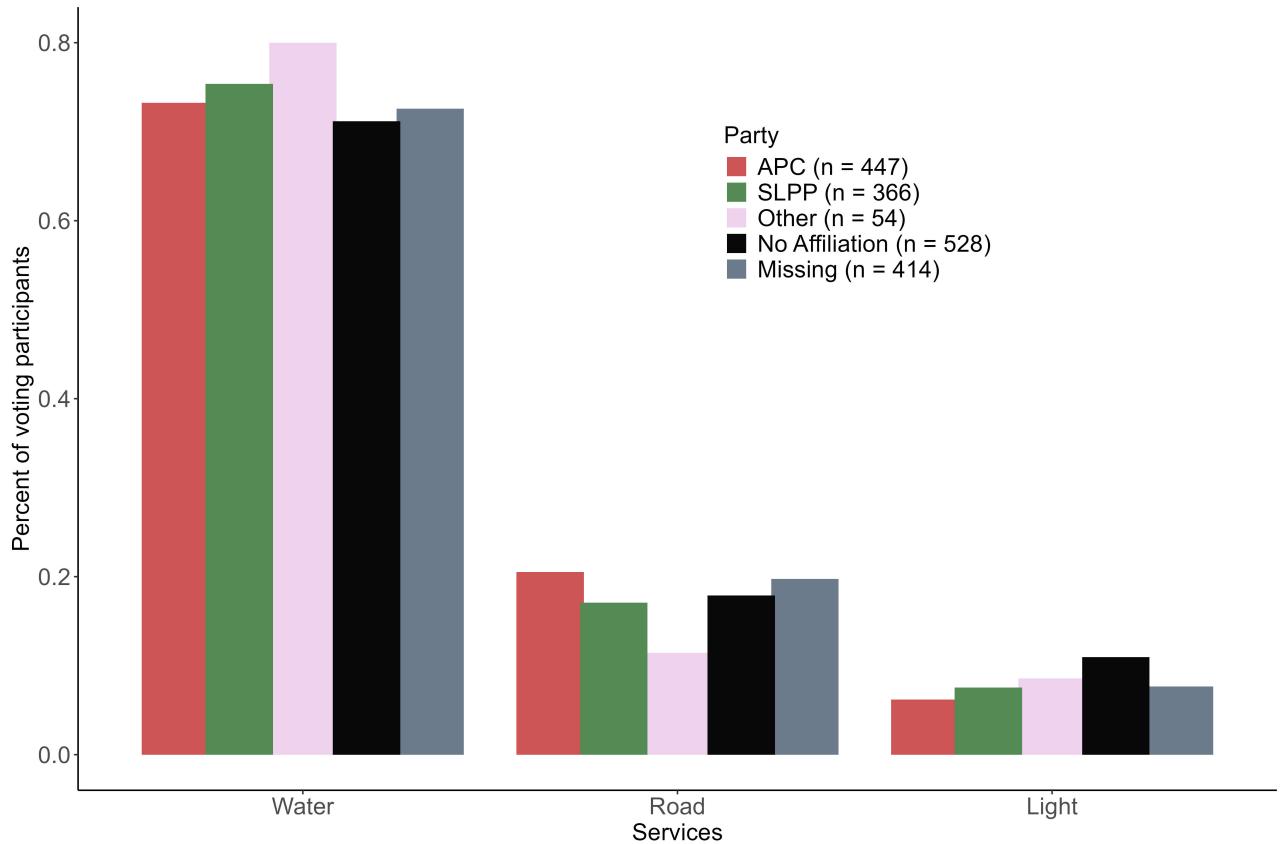
Table D2: Demographics by Partisan Group

Outcome	Conditional Effect		Difference	
	CATE (1)	SE (2)	Dif (3)	p-value (4)
<b>Gender</b>				
Male	-0.001	0.021		
Female	-0.038	0.032	-0.036	0.345
<b>Higher Education</b>				
No	-0.015	0.026		
Yes	0.019	0.034	0.034	0.457
<b>Total Tax Owed</b>				
Below Median	-0.024	0.028		
Above Median	0.000	0.030	0.024	0.603
<b>Water Access</b>				
No	-0.029	0.029		
Yes	0.008	0.033	0.037	0.469
<b>Number of Properties with Tax Liability</b>				
One	-0.048	0.031		
More than one	0.011	0.030	0.059	0.233

**Table D3** reports conditional treatment effects for five demographic variables where baseline levels differ by partisanship (see Appendix Table D2). Column 1 reports conditional average treatment effects for the two groups comprising each variable; Column 2 reports the standard error. Column 3 reports the difference between sub-group conditional effects and Column 4 reports the *p*-value of that difference. Note that to estimate treatment effects conditional on *Number of Properties with Tax Liability* we drop the 121 property owners that had no tax liability for 2021.

Table D3: Effects on Compliance Conditional on Demographic Variables

Figure D1: Votes for Services by Partisan Affiliation



*Note:* Figure D1 presents votes for each service, broken out by partisan affiliation. There is little indication that votes for services differ meaningfully by partisan group. Vote share calculated from voting participants. Whether a participant votes is also similar across partisan groups, and is as follows:

APC = 58%; SLPP = 54%; No affiliation = 52%; Other = 65%; Missing = 60%.

## D.2 Partisan Targeting

Party	Model 1		Model 2		Model 3	
	EST	SE	EST	SE	EST	SE
SLPP	78.96***	23.73	78.82***	23.73	51.60**	20.10
Independent (no reported party)	15.25	20.88	12.40	20.87	0.78	18.26
Third Party Supporter	18.10	47.14	16.67	47.36	2.12	44.31
Did not answer	22.08	22.32	20.53	22.29	14.63	19.15

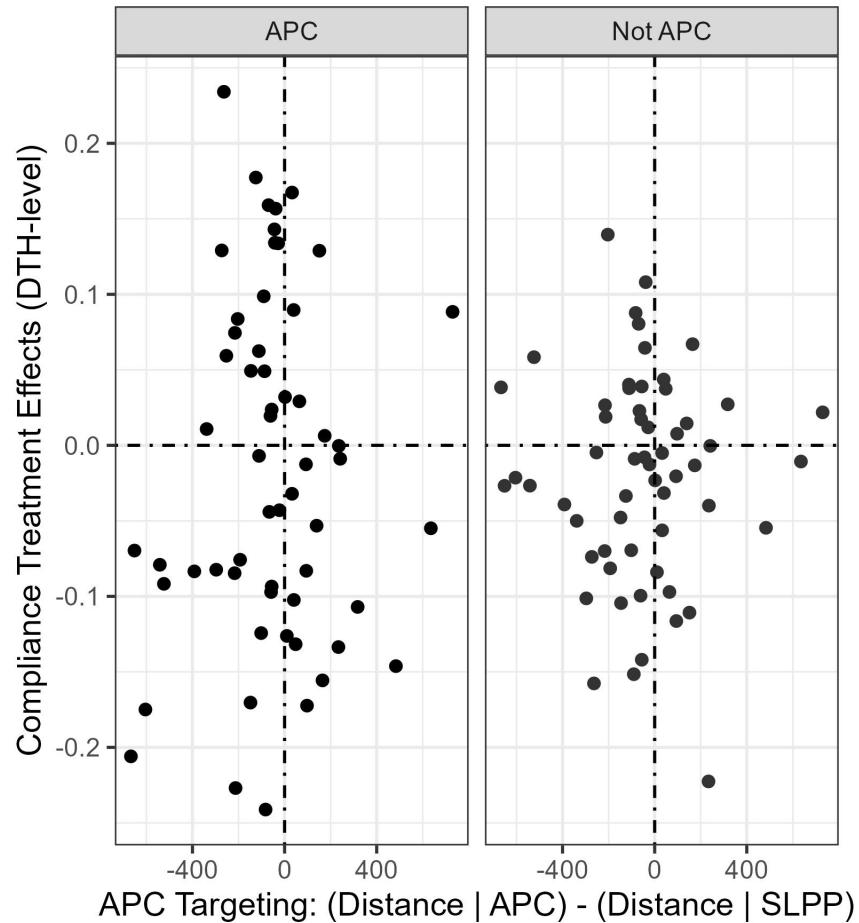
<i>Controls:</i>	
Property Value logged	✓
Ward Fixed Effect	✓

**Table D4** reports the relationship between respondents' partisan affiliation and their distance from services implemented by the DTH intervention. The reference category is APC partisans. Point estimates represent the average distance (in meters) that a respondent from a given partisan group is from the implemented service, relative to an APC supporter. Positive point estimates indicate that respondents are further away. These estimates exclude 25 respondents identified in the administrative data as being over 3.5 kilometers from the implemented service, as this suggests incorrect geo-location. The estimates are not sensitive to the threshold used for dropping respondents.

**Significance:** \*  $p < 0.10$ ; \*\*  $p < 0.05$  \*\*\*  $p < 0.01$

Table D4: Distance to Services

Figure D2: APC Service Targeting and Compliance Effects



*Note:* Figure D2 plots the DTH-level treatment effects against the degree of APC partisan targeting exhibited by the projects implemented in each town hall. APC targeting is defined as the difference between the average distance of APC supporters to the service and the average distance of SLPP supporters to the service. If APC targeting were responsible for the observed heterogeneous treatment effects, we would expect a positive relationship between APC targeting and treatment effects for APC partisans (left panel), and a negative relationship between APC targeting and treatment effects for non-APC supporters (right panel). However, we do not observe these patterns.

## E Spillover Analysis

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.<sup>37</sup> Our approach compares non-study properties geographically proximate to treated study properties to non-study properties proximate to control study properties.<sup>38</sup> We estimate spillovers with the following equation:

$$Y_{i_{2022}} = \beta_1 SPILL_i + \gamma Y_{i_{2020}} + \lambda \mathbf{X}_i + \delta_w + \epsilon_i \quad (3)$$

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,  $\beta_1$  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;  $\delta$  is a vector of ward fixed effects;  $\mathbf{X}$  is the set of property-level characteristics described in Section 3.2, 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 3) 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.<sup>39</sup> To address this, we weight observations by the inverse 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 to a 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.<sup>40</sup>

---

<sup>37</sup>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.

<sup>38</sup>See [Miguel and Kremer \(2004\)](#) for a prominent example of using non-experimental units (i.e., units that are not themselves part of the randomization) to estimate spillovers.

<sup>39</sup>Imagine, for example, potential differences in compliance behavior between densely packed informal settlements and spacious affluent neighborhoods.

<sup>40</sup>Absent a theory-driven procedure for selecting the threshold distance ( $D$ ), we opt for a pragmatic approach.

Table E1 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. The point estimate at this threshold is positive but small, about a third of a percentage point, and is statistically insignificant (RI  $p$ -values in Column 2). As noted in our pre-analysis 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 thresholds change, both above and below our preregistered threshold. The estimated effect is positive at all thresholds, but substantively small and not statistically significant at any threshold. There is some suggestive evidence of a positive spillover for properties within 20 meters of a treated unit ( $\beta = 1.8$  percentage points; RI  $p$ -value = 0.13), but at that threshold estimates are noisy and we cannot reject the null of no spillover effect.

Threshold Distance (meters)	2022 Compliance		N Observations	
	Est	RI $p$ -val	Treatment	Control
20	0.018	0.133	3,212	2,885
30	0.010	0.273	7,259	5,860
40	0.008	0.247	12,103	8,378
<b>64</b> (Preregistered Threshold)	<b>0.003</b>	<b>0.597</b>	<b>24,214</b>	<b>10,585</b>
70	0.004	0.483	26,885	10,441
80	0.005	0.437	31,016	9,782
90	0.008	0.233	34,679	8,814
100	0.008	0.227	37,729	7,906

Table E1 reports spillover effects on the compliance behavior of non-study property owners, at different distance thresholds for defining spillover units (Column 1). Column 2 reports spillover treatment effects where the dependent variable is a dummy indicating if the owner paid any tax in 2022. Treatment effects are reported in raw percentage points. Column 3 reports randomization inference  $p$ -values from 300 simulations. Columns 3 and 4 refer to the number of observations in treatment and control, respectively, at a given distance threshold.

Table E1: Spillover Effects

---

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.

## F Notification Calls

As part of our intervention, we called treated property owners to notify them about the implementation of services selected through the DTH (see Section 2.2). Since this information was provided only to treatment units, our primary analysis does not cleanly isolate the effect of this information alone from the other components of the intervention. To estimate the pure effect of the information about services, we randomized the delivery of this information to property owners outside of our study sample.

**Sample:** We constructed a sample frame of 15,217 non-study property owners who met the following criteria: (i) they owned a property in one of the 30 study wards, (ii) had a phone number on file at the FCC, (iii) had not been contacted as part of the initial verification process that selected property owners for the study, and (iv) had not paid taxes in either 2020 or as of October 23, 2021.<sup>41</sup>

**Randomization:** The 15,217 property owners were randomized into treatment and control groups, where treatment is defined as receiving a call notifying the property owner about implemented services. Randomization was performed by blocking on tax rate decile within each ward. Within each block, units were assigned to the treatment group with a probability of 0.588.

**Treatment Text:** Treated respondents receive a call from a surveyor who identifies themselves as calling on behalf of the FCC. After confirming the respondent's personal information, the surveyor provides the respondent with the following information: *"Recently, in your ward [WARD NAME], [PROJECT DESCRIPTION] has been built by a construction firm on behalf of the Freetown City Council. This is at [PROJECT LOCATION]. This project was funded by resources associated with the FCC's property tax reform."*

Surveyors then ask the property owner if they have heard of this project, and if so, if they had visited it. Then, surveyors conclude the call with the following text: *"We're looking forward to continuing to work with people in your community to better understand the most pressing local development needs. This is one of the steps the FCC is taking to develop the city as part of the FCC's ambitious plan to Transform Freetown. If you have any further questions about the project in your ward, you may contact us at the following phone numbers: XXX or XXX."*

**Estimation:** We estimate ITT treatment effects:

$$Y_{ikt_2} = \beta_1 T_i + \gamma Y_{ijt_1} + \sum_{j=1}^{299} \theta_k Block_k + \delta_w + \lambda \mathbf{X}_i + \epsilon_i \quad (4)$$

Where  $Y_{ikt_2}$  is the post-treatment tax compliance behavior of property owner  $i$  in randomiza-

---

<sup>41</sup>2020 tax compliance was only about 3%.

tion block  $k$ ;  $T_i$  is an indicator variable equal to 1 if owner  $i$  is assigned to treatment and  $\beta_1$  captures the average treatment effect of the Notification Call.  $Y_{ikt_1}$  is the pretreatment compliance measure for owner  $i$ . As service delivery calls were made starting in November 2021, the pretreatment compliance variable is an indicator coded as 1 if a property owner had paid tax in 2020 prior to November, when the campaign started.  $Block_k$  is an indicator variable equal to 1 if owner  $i$  belongs to randomization block  $k$ ;  $\delta$  is a vector of ward fixed effects and  $\epsilon_i$  is the error term.  $\mathbf{X}$  is the same set of property-level characteristics that we use in our main analysis (Section 3.2), included for covariate adjustment.

**Results:** Table F1 presents the results. The estimated treatment effects are small in magnitude, statistically insignificant, and robust to alternative specifications. Information about service provision does not appear to have any effect on compliance behavior.

Model	EST	SE	p-value	N
Preregistered	0.0032	0.0056	0.57	15,202
Baseline compliance only	0.0024	0.0056	0.67	15,202
Ward FE only	0.0023	0.0056	0.68	15,202
Covariates only	0.0032	0.0055	0.56	15,202

Table F1 presents results of the information campaign that called property owners to notify them of service delivery in their wards. The top row presents the preregistered model that contains pre-treatment compliance behavior, ward fixed effects, match-pair dummies, and a set of preregistered controls. As service delivery calls were made starting in November 2021, the pretreatment compliance variable is an indicator coded as 1 if a property owner had paid tax in 2020 prior to November, when the campaign started. The estimated average treatment effect is substantively small and not statistically distinguishable from zero. Rows 2-4 show the robustness of this result to different specifications.

Table F1: Effects on the Service Information Campaign

## G Research Ethics

We find it important to reflect ethically on several dimensions of this project. First, much of this project was carried out during COVID-19. Therefore, we took several steps to minimize in-person contact and the risks associated with that contact. Most fundamentally, we shifted the project’s primary intervention—the town hall—to an online platform, after having originally conceptualized the intervention as a set of in-person town halls. In addition, we conducted data collection through phone interviews, rather than in-person interviews. Phone surveying followed guidelines from the International Growth Centre for conducting research during the pandemic. For example, while our enumeration team met in person to conduct phone interviews, they followed social distancing and sanitation protocols. Finally, all members of our enumeration team received COVID-19 vaccines. We weighed the risks to our enumeration team against the costs associated with calling the project off. Through conversations with research assistants and project supervisors, we believed that much of our enumeration team would be without a paying job during the pandemic if the project was canceled. We reasoned that the costs to enumerators of canceling the project outweighed the risks associated with continuing the project.<sup>42</sup>

Second, only a subset of property owners were eligible to take part in the intervention. We believe that valid equity concerns can be raised about the fact that Freetown citizens who do not own property were not eligible to participate in a participatory budgeting intervention. In addition, eligibility was restricted to property owners (i) with WhatsApp and (ii) with a property above median property value. These latter restrictions were for practical reasons. As the original intervention was originally planned for 2021, and the outcome of interest would be tax compliance in that year, we could only focus on the subset of property owners who received an RDN in 2020.<sup>43</sup> We believe that restricting the intervention to property owners is justified by the scientific goal of the study and because we believe the project has increased the likelihood that all residents of Freetown have a chance to participate in future participatory budgeting programs. Scientifically, we are primarily interested in the relationship between participation in DTH and property tax compliance. Given budget constraints, including citizens not owning property in the intervention would weaken our ability to learn about the effect of the DTH on tax compliance. Moreover, future iterations of the DTHs, to which the Mayor of Freetown has publicly committed, promise to be less restrictive. Freetown residents who were not eligible for this iteration of the DTH are now more likely to be eligible for future participatory budgeting programs, compared to if this DTH project had never taken place. Finally, we do not believe it to be the case that the selected public services only benefit, or even are more likely to benefit,

---

<sup>42</sup>When making our decision to continue with the project during COVID-19, our research team primarily considered the risks and benefits to our enumeration team. However, we can also point out the additional project benefit of delivering key services (totalling over \$45,000) in Freetown.

<sup>43</sup>As described previously, as part of a COVID-19 policy to reduce tax burdens on lower-income households, only property owners in the top half of the assessment distribution received RDNs in 2020.

property owners. For example, community water pumps or street taps benefit everyone in the community, not just property owners. Third, we purposefully did not inform participants that the funds for selected services came from donors. While we were generally ambiguous about the source of the funding, in at least one instance, scripted messages from moderators to participants in DTH referred to DTHs as a way to decide on the allocation of some of “FCC’s budget”. Placards placed at the site of completed projects list the Freetown City Council as the sole implementing partner and the FCC’s logo is the only logo on these placards. We believe this deception to be justified by the scientific benefits of the project. While external donors often play a significant role in bankrolling poor local governments, our goal is to study the fiscal contract between government and citizens. We reasoned that acknowledging the external source of funding would make our results more difficult to interpret. Finally, we note that we are not aware of evidence showing that donor credit claiming for donor-funded projects leads to positive outcomes for citizens; in the absence of such evidence, we follow our instinct that donor credit claiming for our project is not an *ex ante* normatively superior decision.