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

Kevin Grieco[†]

Abou Bakarr Kamara[‡]

Niccolò Meriggi[§]

Julian Michel[¶]

Wilson Prichard[‖]

September 11, 2024

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 phone-based 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 behavior and attitudes towards paying taxes.

*This project was supported by generous funding from the International Growth Center and the International Center 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, Katherine Casey, Darin Christensen, Cesi Cruz, Evgeniya Dubinina, Adrienne LeBas, Giulia Mascagni, Nara Monkam, Oyebola Okunogbe, Daniel Posner, Soledad Prillaman, and Jon Weigel for helpful comments. We would also like to thank participants at the CaliWEPSs IV workshop, APSA 2022, the 2022 Global Development Conference, the 2024 World Bank Land Conference, and the 2024 IIPF Congress for their insightful feedback. We are grateful to Jacques Courbe, Emile Eleveld, Yojin Higashibaba, Xenia Rak, Michael Rozelle, and Ella Tyler for excellent research assistance. Thanks also to our wonderful team of supervisors and enumerators. Of course, this project would not have been possible without the contributions and cooperation of the Freetown City Council. The pre-analysis plan associated with this project is registered at: https://osf.io/dhkfe.

[†]Postdoctoral Fellow, International Centre for Tax and Development

[‡]Senior Country Economist, International Growth Center, Sierra Leone

[§]Postdoctoral Research Fellow, CSAE, Department of Economics, University of Oxford

[¶]Ph.D. Candidate in Political Science, UCLA

[‖]Associate Professor, 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).[1] 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 2011). A critical question, then, is: how can governments in weak states break out of the reinforcing trap of weak 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). However, reliance on enforcement alone has proven challenging owing to limited capacity to pursue non-compliant taxpayers and political resistance that has made expanded tax enforcement politically unattractive (Christensen and Garfias 2021).

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; Dom et al. 2022; 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) or from government officials within the administration itself who benefit from the status quo (Prichard et al. 2019). These channels suggest that 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 modern Europe, which posit that political leaders traded expanded political voice to elites in

[1] 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).

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 eight of our nine legitimacy outcomes (the p-value on the ninth indicator is 0.11) at the endline survey, which was conducted soon after services were implemented and seven months after the conclusion of the DTHs. Critically, 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, by collecting baseline data at the individual level, we demonstrate that these legitimacy gains are consistent across key subgroups, including political opponents.

Turning to tax compliance behavior, we find large and significant heterogeneous treatment effects conditional on participants' 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. This increase also reflects a similar rise in their support for a policy of expanded taxation. By contrast, we find that the treatment *lowered* compliance for non-copartisans by 4.0 percentage points and reduced support for a policy of expanded taxation. In our case these countervailing forces are relatively balanced, leading to no average effect on compliance. Although these findings emerge from exploratory analyses, they are intuitive given the context of the heavily politicized tax reform during which the intervention occurred. We argue that, in this charged environment, the DTH exposed participants to a flurry elite messaging about taxation and that these elite cues shaped participants' attitudes towards taxation and their subsequent tax compliance behavior. We consider, and rule out, the possibility that these conditional treatment effect estimates are confounded by factors such as preferences for specific services or satisfaction with their implementation.

Ultimately, our study makes five 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 results (Casey et al. 2012; Fearon et al. 2015; Humphreys et al. 2019; Khan et al. 2022). 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.[2] 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 offer a novel and more comprehensive perspective of the impact of direct democracy on tax compliance, challenging broader assumptions about both the impact of political participation on behavior and the politics of taxation. 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. 2010; Alm et al. 1993). However, these studies ignore the partisan rancor of real-world politics. Moreover, 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. This finding also contrast with several observational studies that link participation in policy-making to tax compliance (Pommerehne and Weck-Hannemann 1996; Torgler 2005; Touchton et al. 2019). The idea that political participation

[2]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.

may lead to backfiring effects among out-partisans is, while intuitive, to our knowledge, absent from the existing literature.[3] 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.[4]

Third, 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 it does not immediately raise revenue; countervailing treatment effects result in a null average effect. 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).

Fourth, 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 results have important implications for governments introducing participatory budgeting to increase tax compliance and build broader legitimacy and political support for reform. Our results show that while participatory interventions can broadly enhance government legitimacy, they may act as a double-edged sword for short-term compliance, increasing it for some groups but reducing it for others. These results suggest a tradeoff for governments: actively involving non-copartisans in participatory processes can build the legitimacy needed for medium- and long-term state-building, but it risks potential behavioral backfires in the short term. With that said, it is important to note that our research design likely maximized the risk of such backfiring: by randomly selecting participants, and actively encouraging participation, we likely over-represented opponents compared to processes where individuals opt-in. Moreover, the Mayor's prominent role in implementing the intervention likely polarized it more than

[3] 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).

[4] Though see Cullen et al. (2021) on political alignment and tax compliance in the United States.

if it had been introduced by apolitical bureaucrats.

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 us helped lead. The reform served to broaden the tax base—less than 50% of the approximately 120,000 properties had been registered previously in the property cadastre—and to make the tax burden more equitable through the introduction of a more nuanced, consistent and transparent property valuation scheme (Grieco et al. 2019). 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 these 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 in a given ward (FCC 2021, pg. 26).

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

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.[5] 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 respec-

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

tive ward councilor.[6] These videos explained the overall process, highlighted the link between property tax payments and service delivery, and invited participants to start discussing development concerns within their group. Group moderators introduced themselves and then prompted participants with the following question: *What do you think is the greatest development problem in your ward?* Across all groups, participants exchanged approximately 2,000 messages.[7] 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 a project 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.[8]

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.[9] 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

[6] Videos from political representatives were shared with DTH participants in two ways: videos were posted directly in the WhatsApp group and were available via a Qualtrics link. The research team hired a local team to act as moderators, who were supervised and managed 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.

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

[8] Other common service preferences included roads, street lights, dustbins, public toilets, and the upgrading of drainage systems.

[9] 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.

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

4. **Decision Making (February 17-22, 2021):**
In this phase, participants cast their vote for the project they wanted to be implemented in their ward. This vote could be cast anonymously through a Qualtrics survey (Appendix Figure 3) and a "how to" video was posted in each group that provided step-by-step instruction of the voting process.[10] Voting was open for four days and was closed on February 22. The mayor announced the winning project for each ward with a ward-specific voice message, which was posted in each DTH alongside a picture of the Mayor in office. After the announcement of the winning projects, group moderators thanked participants for their contributions and then halted participants' ability to post messages in the DTHs. Moderators explained that chats would be used one more time in the future to announce that delivery of the service has concluded.

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.[11] Roughly two-thirds of those who joined voted for their preferred service (68%) and posted at least one message (63%).[12]

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).[13]

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

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

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

[12]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.

[13]See Appendix Table 4.

credit for the participatory budgeting program and associated service provision.[14] 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 transfer, 4% from development partners, and 3% from foreign aid.[15]

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. In the remaining ward, construction was finished in February 2022. 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.[16] 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 the Appendix B.3.

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.[17] 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

To estimate causal effects, we randomize an invitation to join the DTH across a sample of property owners in Freetown.

We constructed a sampling frame using FCC administrative records, which provided informa-

[14]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.

[15]See Appendix Table 3.

[16]One ward voted to fix potholes, but due to implementation difficulties, it instead received 50 meters of truck tracks.

[17]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 more limited participatory planning intervention. We made notification calls to rule out this concern, thereby simplifying the interpretation of our findings.

tion 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 C.1 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%).[18] 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 C provides more details on the restricted sampling, matching procedure, and treatment assignment.

Appendix Table 9 reports balance across baseline attitudinal outcome, 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.

[18]Note that the response rate at baseline is 100% because only those who answered the baseline were eligible for the intervention.

4 Estimation and Results

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

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

Where Y_{ijt_2} is the endline (t_2) outcome of individual *i* in pair *j*; DTH_i is an indicator variable equal to 1 if owner i *joined* the DTH. Y_{ijt_1} is the baseline outcome for owner *i* in pair *j*. When Y is property tax compliance behavior, Y_{t_1} refers to tax compliance behavior in 2020; When Y is a survey outcome, Y_{t_1} refers to the baseline survey outcome. $PAIR_j$ is an indicator variable equal to 1 if owner *i* belongs to pair *j*; **X** is a set of preregistered property-level characteristics that we include for covariate adjustment only when Y is property tax compliance behavior.[19] δ is a vector of ward fixed effects and ϵ_i is the error term.

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

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

Where D_i is the randomly assigned treatment indicator, which instruments for DTH_i in equation 1.[20] Our quantity of interest is α_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.

[19] 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 and survey outcomes through the inclusion of block dummies.

[20] We estimate these equations using the *iv_robust* package in R.

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 standard conceptualizations, we classify those indicators into four categories: Policy Influence, Service Delivery and Responsiveness, Government Administrative Competence and Approval of Political Representatives (Table 1). We discuss results for each in turn.

We look first at citizens perceptions of their ability to influence policy (Levi et al. 2009; Scharpf 1997), 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 and endline, with statistical significance at the threshold of conventional levels.[21]

Second, we look at perceptions of service delivery and responsiveness. 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; p-value <0.001) and after service implementation (endline; p-value = 0.016). In addition, the intervention attempted to forge the social and fiscal contract between citizens and politicians by delivering local services that people demanded. We find that the intervention increased citizens' satisfaction with FCC service provision at both midline (p-value <0.0001) and endline (p-value = 0.004).

Third, we look at perceptions of the ability of governments to administer their constituencies competently. 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).[22] 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 (0.085 SDU; p-value = 0.104), this effect increases nearly fourfold by endline (0.319 SDU; p-value = 0.0017). In terms of perceived efficiency in the use of funds for public administration and development, we observe *no effect* at midline (p-value = 0.32), but a clear positive impact by endline (p-value = 0.0078). 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 (p-value=

[21] Conventional p-values are 0.10 and 0.11 at midline and endline, respectively.

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

0.001), but after services are implemented, treated participants *positively* update their views of FCC corruption relative to the control group (p-value = 0.071). 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 measured whether participants approved or disapproved of how both the Mayor and their ward councilor have performed on the job over the past twelve months. Our data show that the Mayor is popular at baseline: most respondents report they either "strongly approved" (43.4%) or "approved" (44.3%) of the mayor's performance. The intervention increases approval of the Mayor by 0.15 SDUs (p-value <0.001) at midline and 0.19 SDUs (p-value <0.001) at endline.[23] Respondents reported much lower approval ratings for their ward councilors at baseline: the modal respondent (41%) "disapproved" of their councilor's performance over the past year. While baseline approval for councilors was low, the intervention increased approval at both midline (0.19 SDUs; p-value <0.0001) and endline (0.17 SDUs; p-value <0.001).

In summary, Table 1 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).

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

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

Table 1 reports the effect of the treatment on legitimacy outcomes. Column 1 reports the control group mean for each indicator at baseline, with the standard deviation in parentheses; Column 2 reports the control group mean at midline and Column 3 presents treatment effects estimates at the midline survey. Column 4 reports the number of non-missing observations in the midline survey. Columns 5-7 present similar statistics for the endline survey: Column 5 reports the control group mean at endline, Column 6 presents treatment effects estimates for the endline survey, and Column 7 reports the number of non-missing observations in the endline survey. Reported effects are standardized effects.
$^{*}p < 0.10$; $^{**}p < 0.05$ $^{***}p < 0.01$

Table 1: Effect on Legitimacy

4.2 Average Effects on Tax Compliance

Turning to the impacts on compliance, we first present the average effects (Table 2), 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 2 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)
Panel A: Tax Compliance Behavior				
Did the owner pay any taxes?				
2022	0.291	-0.012 (0.018)	0.496	3,618
2021	0.315	-0.007 (0.019)	0.723	3,618
Panel B: Fiscal Exchange Attitudes				
Willingness to pay more taxes for better services				
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 2 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. $^{*}p < 0.10$; $^{**}p < 0.05$ $^{***}p < 0.01$

Table 2: 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.[24]

[24]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.

Results are also robust to different operationalizations of tax compliance. While Table 2 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 10 and 11, 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.[25] 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 2, 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, we should have found that treated respondents *say* they would be more willing to pay taxes; 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.

[25] 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.

However, at endline, we find no evidence of lasting treatment effects on either the fairness or enforcement outcomes (Appendix Table 17). While midline results are more varied, that they do not persist to endline makes them unlikely explanations for the null results on compliance.[26]

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 has posited 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, political leaders are viewed as unitary actors, and citizens are more likely to comply when they believe these leaders will pursue policies aligned with their preferences.

However, politics is often contentious, with competing politicians pursuing and promoting different policies. Even when citizens know what they want, they are often unsure which policy will achieve it, and use elite cues to form opinions and policy preferences (Zaller 1992). It is not difficult to imagine that inviting public participation on a contentious political issue can politicize behavior and attitudes on that issue by exposing participants to elite messaging on that topic. Indeed, there is good evidence that individuals respond to elite cues by shifting their policy position closer to the position of their political party (Broockman and Butler 2017; Flores et al. 2022; Tappin et al. 2023); and there is growing evidence that, while these cues may be persuasive for the political ingroup, they can generate backlash from the outgroup (Haas and Khadka 2020; Nicholson 2012).

Certainly, the property tax reform in Freetown was a highly politicized affair. The Mayor—whose All People's Congress party controlled the Freetown City Council and sat in opposition to the Sierra Leone People's Party that controlled the central government—publicly battled with the Ministry of Finance over the Freetown City Council's legal authority to adjust property tax rates without approval from central government (Luke 2020). In this politicized environment, the Mayor used the DTHs to promote her policy vision to transform Freetown through taxation and service provision. For example, in one DTH video the Mayor encouraged participants to pay taxes with the promise that "FCC will use [tax revenue] to deliver services to the people of Freetown."[27] Consistent with the idea that the DTH was a *political* environment, we find that the treatment led to an immediate increase in participants' political interest and engagement (Appendix Table 14).

We correspondingly explore the possibility of heterogeneous effects across different parti-

[26]These results are discussed in great detail in Appendix Section D.4.

[27]In a separate video, the Mayor reminded participants that "if everyone pays their property rate, you can imagine what type of investment we can make in your ward."

san groups. Figure 1 presents evidence that participants' partisan affiliation conditions how the treatment impacts their compliance behavior and tax policy preferences (i.e., support for expanded taxation). Plot A presents predicted marginal effects from a model that interacts treatment with a co-partisan indicator variable.[28] The interaction between treatment and co-partisanship is statistically significant (p-value = 0.033; $\beta = 0.11$). For co-partisans of the Mayor (i.e., APC supporters) the treatment increases compliance by 7.4 percentage points, which is a substantial 30% increase over the group's control compliance rate of 24.4%. In contrast, treatment effects are *negative* for non-co-partisans of the Mayor: treatment lowers compliance by 4.0 percentage points. [29]

Panel B (Figure 1) presents evidence that treatment similarly polarizes tax policy preferences along partisan dimensions. Panel B predicted marginal effects from an interaction model, but now with the respondent's tax policy preferences as the outcome of interest. To increase power for estimating this interaction, the predicted outcome is the respondent's *average* support for expanded taxation across midline and endline surveys.[30] Again, 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.

The consistency of patterns in both attitudinal and behavioral 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. One potential confounder is service preferences. If partisan affiliation proxies 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 9 shows participants' vote choice in the DTH for local development projects by partisan group. There is little indication that service preferences vary meaningfully by partisan affiliation. Therefore, service preferences are unlikely to be confounding the

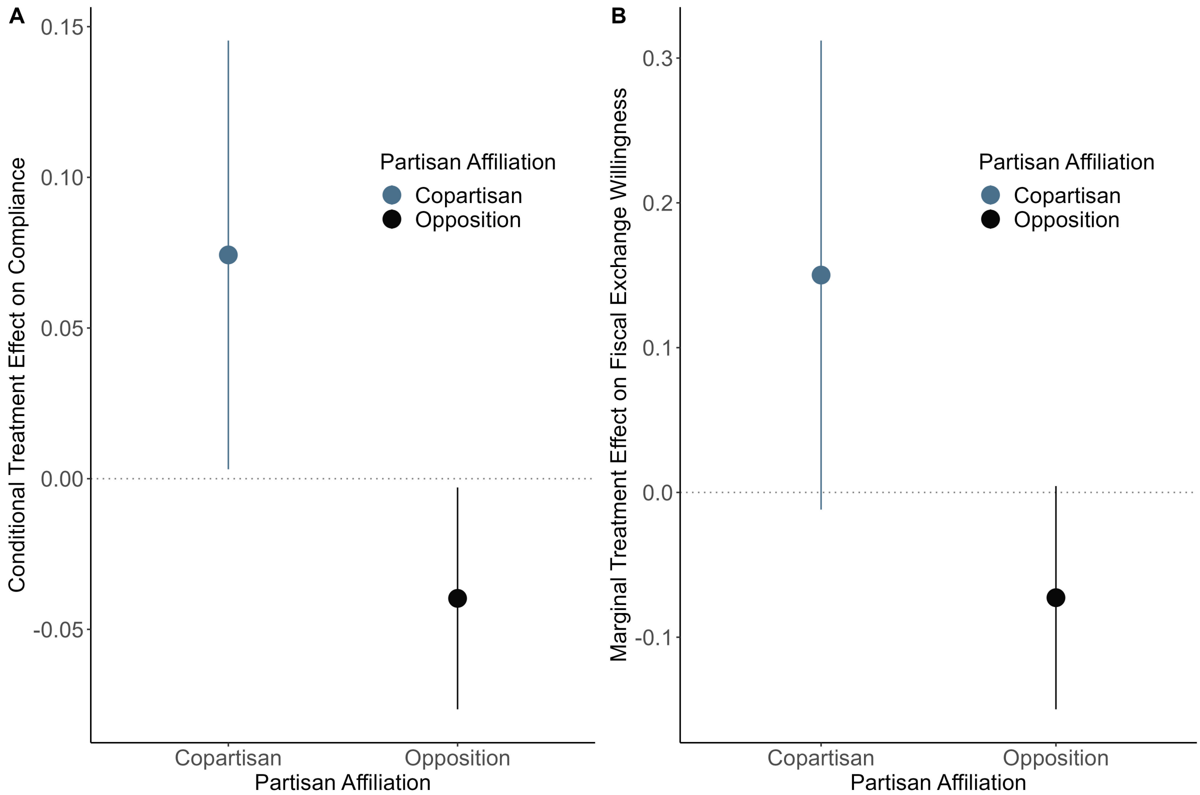
[28] Other model specifications remain the same as in our main specification. The copartisan variable is equal to 1 for respondents who self-report affinity towards the All People's Congress; all other respondents are coded as 0. In our baseline survey, we asked respondents which political party (if 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 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."

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

[30] Appendix Figure 7 shows estimates from the interaction model using midline or endline data separately. Estimated marginal effects display similar patterns.

observed heterogeneous effects by partisanship reported in Figures 1.[31]

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 "co-partisans." All other respondents are coded as "opposition." Point estimates are presented with 90% confidence intervals.

While we have evidence that partisan affiliation moderates treatment effects for compliance, it does not moderate treatment effects for legitimacy outcomes (Appendix Table 12). If legitimacy outcomes were the mediating mechanism between participation and compliance, we would expect to observe positive effects on legitimacy outcomes for copartisans and negative effects for non-copartisans. However, we do not observe negative point estimates for non-copartisans on any legitimacy outcome at endline. In fact, for the only outcome where treatment effects between copartisans and non-copartisans are statistically distinguishable (Mayor approval), the effects are actually *larger* for non-copartisans, likely because copartisans face ceiling effects. This, again, suggests that legitimacy outcomes are not a straightforward link to compliance behavior.

This presents a puzzle: why does the treatment *increase* legitimacy outcomes for non-copartisans

[31] Relatedly, partisanship is not correlated with participants' reported satisfaction with *delivered* services (Appendix Figure 10), ruling out differential satisfaction as a confounder.

while simultaneously reducing compliance behavior and support for a policy of expanded taxation? One possibility is that participants' partisan affiliation influences their responses to politicized position issues, such as taxation, but not to valence issues reflected in legitimacy outcomes that remained above the political fray. Taxation was a *position* issue that was heavily debated at the time of the intervention, with political opponents arguing that property tax reform was ill-timed during COVID-19 and unnecessary for the city's development. In contrast, legitimacy outcomes address *valence* issues, such as voice and government responsiveness, which participants agreed were important and believed the DTHs effectively addressed. Political opponents did not contest the importance of voice or responsiveness as goals, nor did they dispute that the DTHs had achieved these goals. As a result, these outcomes remained non-politicized.

4.3.1 DTHs Crystallize Existing Attitudes Towards Taxation

We argue that partisanship moderates the effect of the intervention on tax compliance because it shifts participants' attitudes and behavior to align with their partisan group. Here we document another pattern that may have emerged from the DTHs politicizing taxation. Figure 2 demonstrates that participants' political views about taxation increasingly shaped their behavior: those who support taxation become more likely to pay, while those who oppose it become less likely to do so (Panel A). Moreover, DTHs *crystallized* these existing attitudes towards taxation (Panel B).

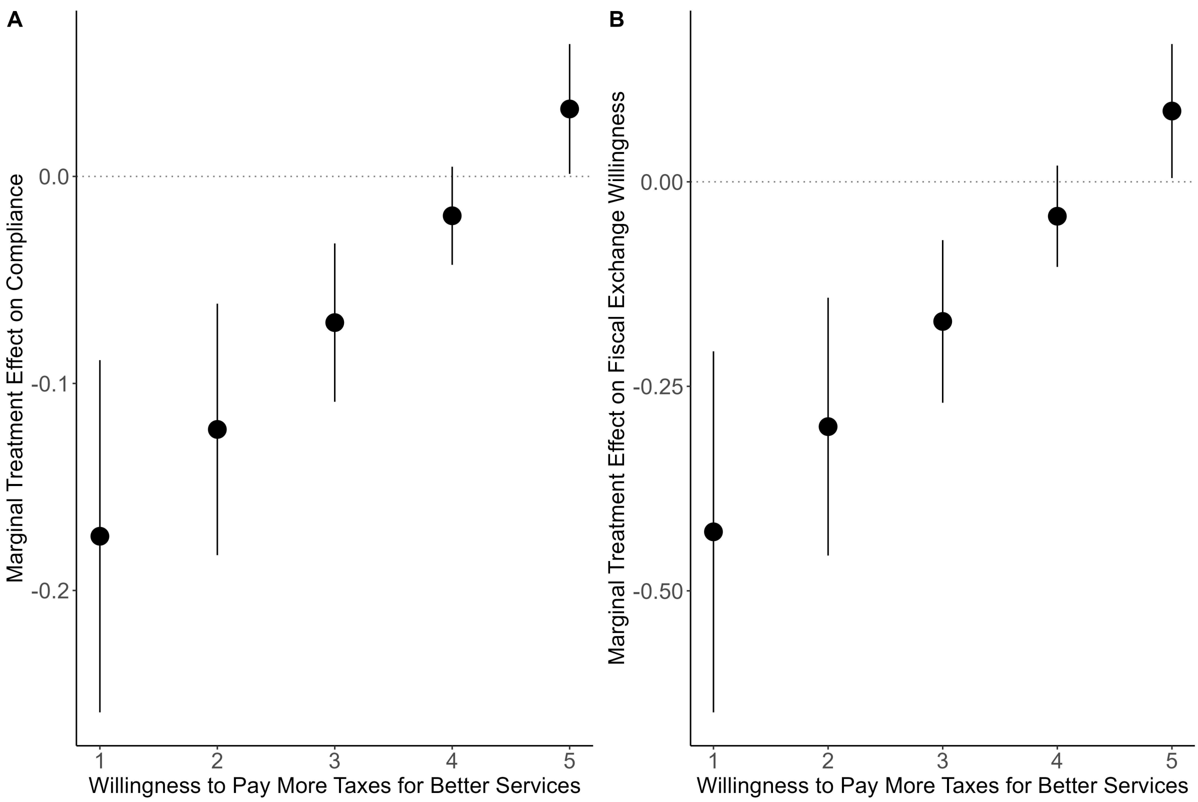
Participants held policy positions about taxation prior to the DTH: while the majority (57.4%) of surveyed respondents reported that they "strongly approved" of a policy of expanding taxation for improved services, a significant minority opposed (14%) this idea. How did these baseline beliefs condition tax compliance behavior? Plot A presents predicted marginal effects from a model that interacts treatment with our five-point measure of (baseline) attitudes towards expanded taxation. The interaction term is statistically significant (p-value <0.001; $\beta = 0.052$). Baseline support for expanded taxation strongly shapes treatment effects on tax compliance in 2022. For respondents that strongly support expanded taxation (five on the five-point scale) the estimated treatment effect is 3.31 percentage points, a 9.5% increase over the group's baseline compliance rate of 34.6%. By contrast, treatment effects are negative for property owners who did not support more fiscal exchange at baseline. For property owners who "somewhat disagreed" or "strongly disagreed" with paying more taxes for improved services, we estimate marginal treatment effects of -12.2 and -17.4 percentage points, respectively.

Moving to the crystallization of existing attitudes, Panel B presents predicted marginal effects on support for expanded taxation from a model that interacts treatment with our five-point measure of (baseline) support for expanded taxation.[32] The interaction term is statistically

[32]To increase power for estimating this interaction, the predicted outcome is the respondent's *average* support for expanded taxation across midline and endline surveys. Note that estimated marginal effects display similar patterns when estimated as an interaction model using midline or endline data separately. See Appendix Figure 8.

significant (p-value = 0.0016; $\beta = 0.13$ SDUs).

Figure 2: Marginal Treatment Effects by Attitudes Towards Fiscal Exchange



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

This section presented results from exploratory analyses. The patterns we uncovered in these analyses suggest that partisanship moderates the effect of the intervention on tax compliance. We argue it does so because treatment intensifies participants' exposure to the politicized issue of property taxation in Freetown. In additional analyses we show that compliance behavior becomes increasingly a function of existing political beliefs—that is, tax compliance behavior becomes more *politicized*. During this process those beliefs about taxation harden.

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, as well as among those who support or oppose the expansion of taxation.

Second, despite those relatively universal and durable impacts on legitimacy, we find that impacts on compliance are heavily moderated by partisanship and pre-existing policy preferences. We do not find any average effect of the intervention on tax compliance. Instead, we find significant positive impacts among co-partisans and those supportive of expanded fiscal exchange, but significant negative impacts among those who are not from the Mayor's party or oppose expanded taxation. This adds substantial nuance to influential models of compliance (Levi 1988, 1997), canonical accounts of the development of fiscal capacity (North and Weingast 1989; Bates and Lien 1985), the literature on the democratic dividend (Bó et al. 2010; Sutter et al. 2010), and studies of participatory budgeting (Pommerehne and Weck-Hannemann 1996; Touchton et al. 2019), 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 intervention 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 interventions can be used to create more supportive environments for governments who 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.

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.

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.

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*, page ueae030.

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

Besley, T. and Persson, T. (2014). Why do developing countries tax so little? *Journal of Economic Perspectives*, 28(4):99–120.

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. and Sáenz Somarriba, M. (2022). Electronic payment technology and tax compliance: Evidence from uruguay's financial inclusion reform.

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.

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.

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.

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

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.

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.

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.

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.

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.

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.

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.

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.

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*.

Markussen, T. and Tyran, J.-R. (2023). Is there a dividend of democracy? experimental evidence from cooperation games.

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*, pages 566–583.

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

North, D. C. and Weingast, B. R. (1989). Constitutions and commitment: the evolution of institutions governing public choice in seventeenth-century england. *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.

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.

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.

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.

Touchton, M. R., Wampler, B., and Peixoto, T. C. (2019). Of governance and revenue: Participatory institutions and tax compliance in brazil. *World Bank Policy Research Working Paper*, (8797).

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

A 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, it is to be expected that DTHs—just like their offline analogue—

display additional participation biases (higher ability and willingness to participate among those able to afford smart phones and internet usage, the more educated and literate, those with higher political efficacy (on self-selection in offline TH participation, see Boulianne 2019; Neblo et al. 2010).

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

B Intervention Appendix

B.1 Participation and Experience in DTH

We asked participants which actors they believed were responsible for organizing, implementing, and funding the DTHs (Table 3. For these questions, respondents were allowed to name multiple actors they thought might be involved.

	Perceived Responsible Actor				
Activity	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 3: Organization, Implementation, Funding

We also asked respondents seven questions about their experience in the DTH, on a five point Likert scale (Table 4). These questions were asked in both positive and negative form, 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.

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

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 4: DTH Experience

Tables 5 and 6 provide more details on participants' participation behavior.

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

Note: Self-reported frequency of accessing the DTH group. Amongst respondents who we confirmed as joining the group.

Table 5: DTH Participation

Item	Percent / Value
Voted for service [%]	0.68
Sent any message in DTH [%]	0.63
Median messages sent	2.00
Mean messages sent	3.84

Table 6: Voting and Messaging Behavior

B.2 Voting and Project Implementation

Figure 3 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.[33] Considered this way, 75% of people 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, 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 7: Project Votes, Winning Projects, Implemented Projects

[33] Or if a respondent voted for 50meters of tire tracks, a project that fixes pot holes still address the issue of road repair.

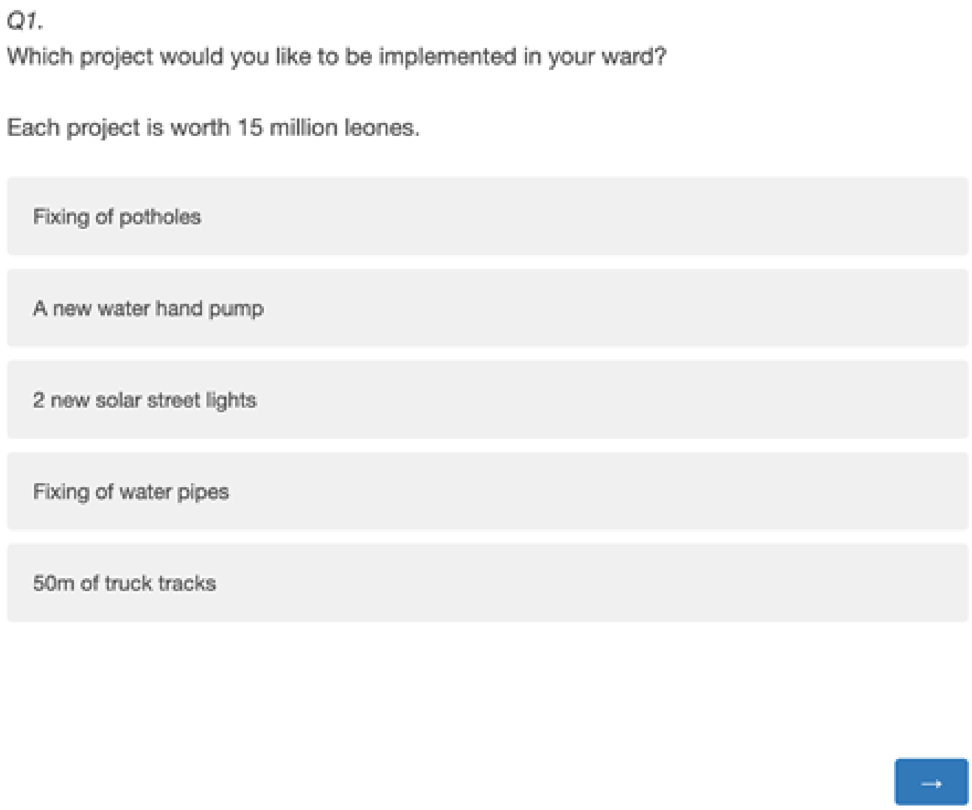


Figure 3: Menu of services

B.3 Project Pictures

Figure 4: Project implemented in Ward 418

C Research Design

C.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 in an attempt 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.

C.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.[34]

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%).[35]

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%).

We provide financial incentives—packages of mobile data—for midline and endline survey takers to minimize attrition.

C.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. In this matching process we weight certain variables higher than others, in line with our expectations that certain variables are a stronger predictor of our outcomes of interest. We place the greatest weight on our measure of unconditional tax morale—we expect this to be the strongest predic-

[34]Appendix Figure **??** documents the broader data collection and project timeline.

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

tor to tax compliance, in line with the common use of this variable as proxy for tax compliance behavior. We place equal weight on another set of six measures from our baseline survey. Three of these measures are important factors in the literature on tax compliance: (i) service conditional tax morale, (ii) perceived likelihood of punishment for non-compliance, and (iii) satisfaction with FCC service provision. We also place equal weight on the (iv) gender of the property owner, (v) their awareness and support of the property tax reform,[36] and (vi) the number of these five variables that were imputed.[37]

Table 8 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.

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

Table 8: Summary Statistics of Matching Variables

[36]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.

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

C.4 Treatment Assignment Map

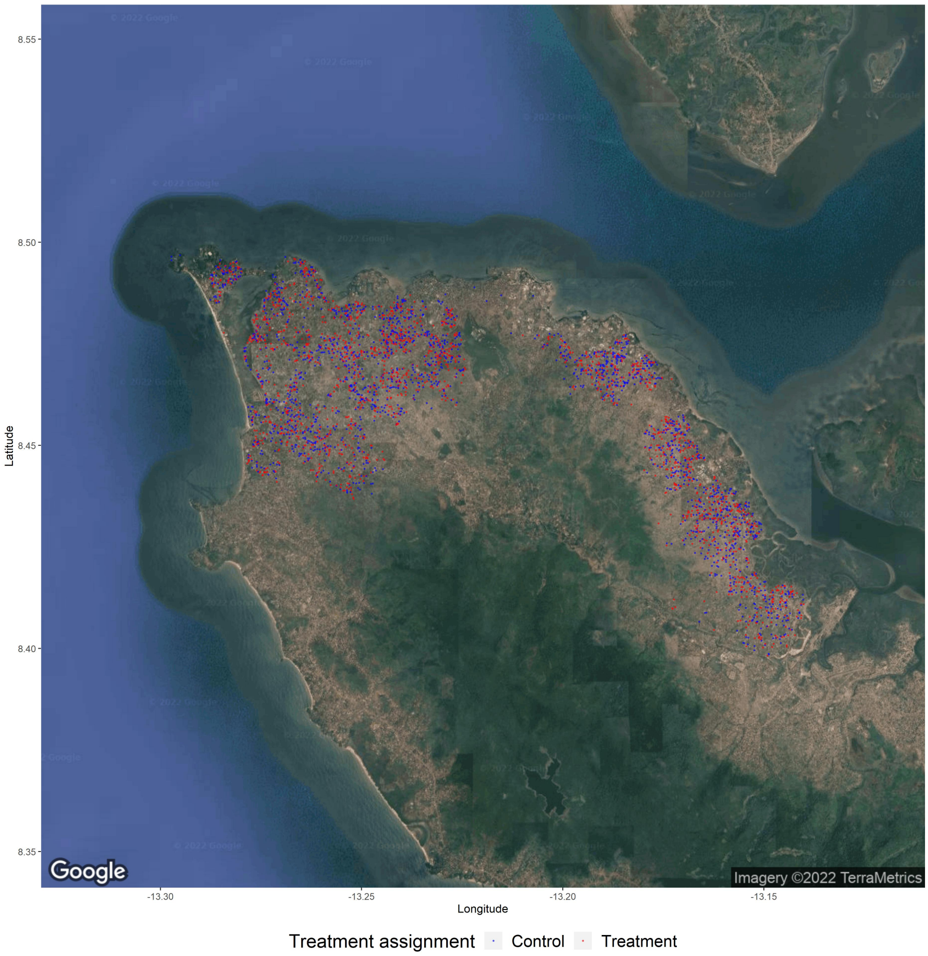


Figure 5: Digital Town Hall treatment assignment in Freetown (red = treatment)

C.5 Distance to Closest Study Property

Figure 6 shows the distribution of the distance from each property to the closest property in the sample.

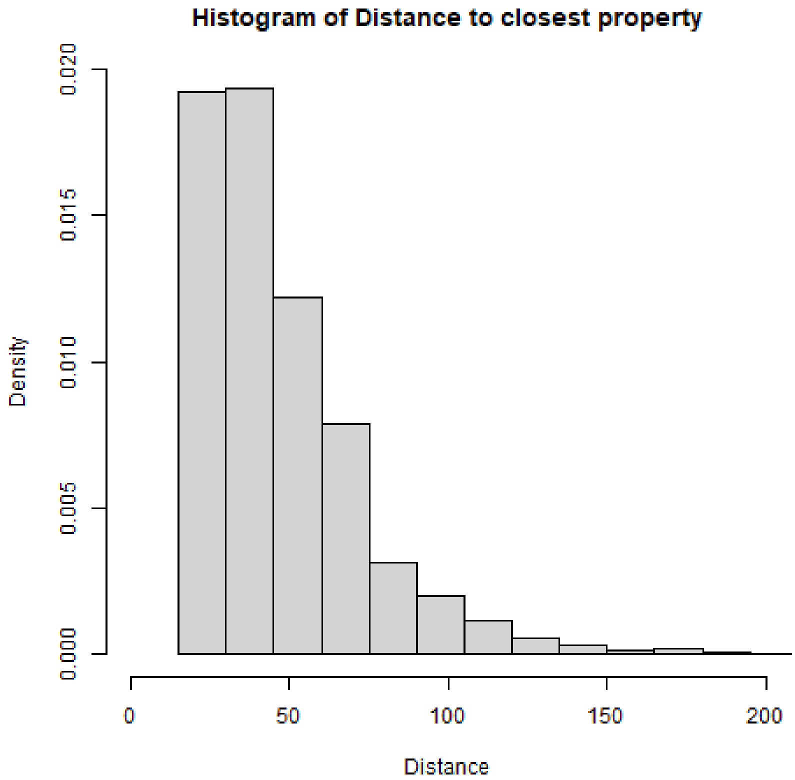


Figure 6: Histogram of minimum distance (in meters) between study properties

C.6 Balance

Table 9 presents balance statistics across a range of (baseline) survey outcomes, demographic covariates, and property characteristics. Observed differences between groups for these variables are no more than we might expect. Given the 29 tests we run, under the null hypothesis of no differences between groups, we would expect 2.9 tests to appear significant at the 90% confidence level. We find only two significant differences: at baseline, the treated group is *less* likely to perceive the FCC as corrupt (*FCC corruption*) and less likely to report they are politically unaffiliated (*No affiliation*). Our preregistered specification for survey-based outcomes includes the baseline measure of the dependent variable. Therefore, when estimating treatment effects on perceptions of corruption we control for the baseline measure in our standard model.

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

Table 9 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 9: Balance Table

D Additional Analyses

D.1 Tax Compliance

	2022			2021		
Model	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 10 reports treatment effects on tax compliance behavior for 2022 and 2021, using alternative model specifications. Columns 1 and 4 presents treatment effects estimates in raw percentage points for 2022 and 2021, respectively. Columns 2 and 5 reports standard errors; Column 3 and 6 reports p-values.
$^{*}p < 0.10$; $^{**}p < 0.05$ $^{***}p < 0.01$

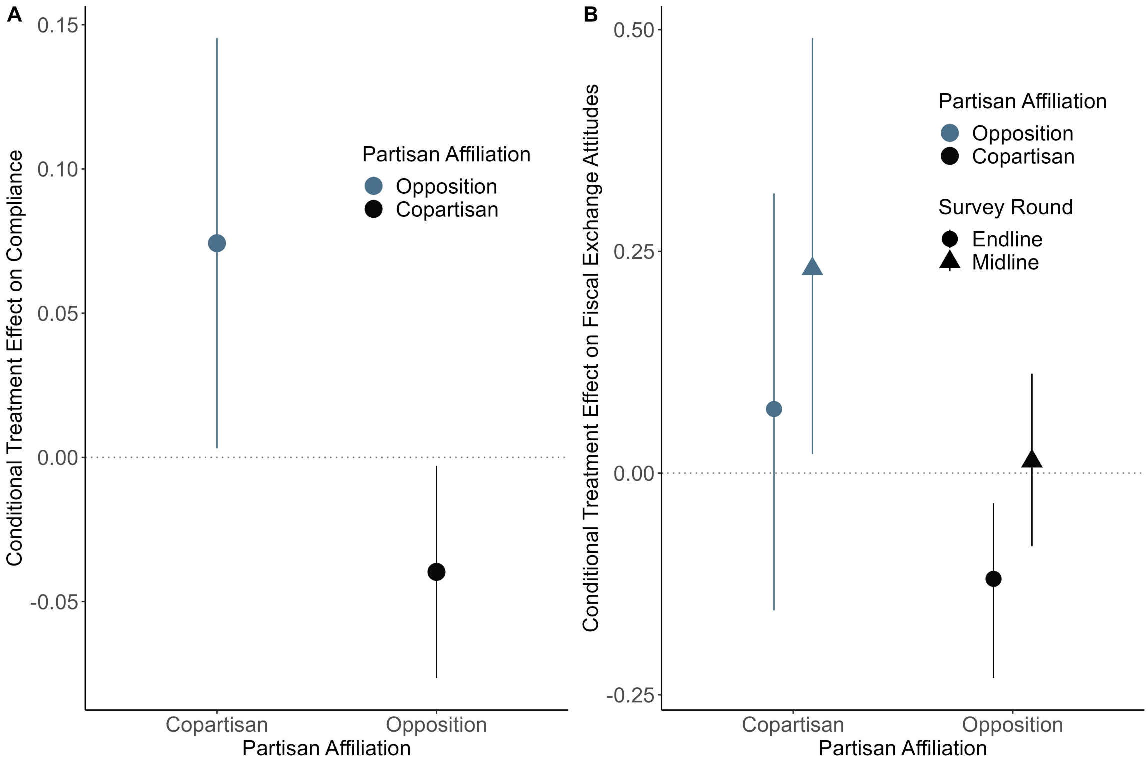
Table 10: Effect on Tax Compliance: Alternative Specifications

	2022			2021		
Outcome	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 11 reports treatment effects on tax compliance behavior for 2022 and 2021 using alternative operationalizations of the dependent variable.Columns 1 and 4 presents treatment effects 2022 and 2021, respectively. Columns 2 and 5 reports standard errors; Column 3 and 6 reports p-values. We convert local currency (SLL) to USD at a rate of 10,000:1, which reflects the exchange rate in January, 2021.
$^{*}p < 0.10$; $^{**}p < 0.05$ $^{***}p < 0.01$

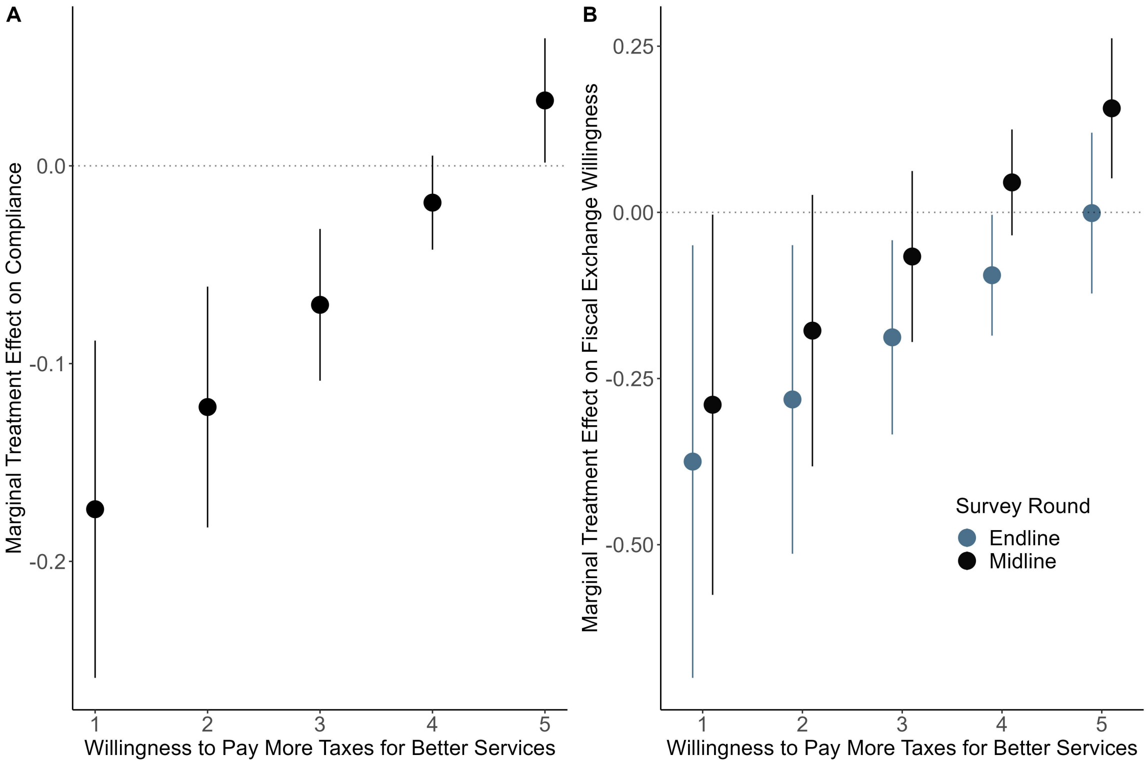
Table 11: Effect on Tax Compliance: Alternative Operationalizations

Figure 7: 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 8: Marginal Treatment Effects by Attitudes Towards Fiscal Exchange



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

D.2 Legitimacy

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

Table 12 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.
$^{*}p < 0.10$; $^{**}p < 0.05$ $^{***}p < 0.01$

Table 12: Effects on Legitimacy Outcomes Conditional on Partisan Affiliation

Outcome	Midline (CATE)			Endline (CATE)		
	Support	Oppose	Diff	Support	Oppose	Diff
Policy Influence						
Opportunities to voice opinion to govt	0.378	0.406	-0.028	0.252	0.259	-0.006
	(0.044)	(0.128)	(0.143)	(0.053)	(0.144)	(0.163)
Ease of participating in political activities	0.083	-0.035	0.118	0.073	0.026	0.047
	(0.046)	(0.128)	(0.144)	(0.053)	(0.147)	(0.167)
Service Delivery and Responsiveness						
FCC responsiveness to citizens' demands	0.152	0.101	0.051	0.125	0.077	0.047
	(0.043)	(0.124)	(0.139)	(0.055)	(0.159)	(0.179)
Satisfaction with FCC service provision	0.186	0.170	0.015	0.178	-0.024	0.202
	(0.046)	(0.135)	(0.151)	(0.056)	(0.167)	(0.184)
Approval of Political Representatives						
Mayor approval	0.167	0.039	0.128	0.235	-0.041	0.276
	(0.047)	(0.146)	(0.161)	(0.060)	(0.159)	(0.181)
Councilor approval	0.172	0.366	-0.194	0.186	0.120	0.066
	(0.045)	(0.134)	(0.150)	(0.053)	(0.154)	(0.171)
Government Administrative Competence						
FCC corruption	-0.133	-0.181	0.047	0.108	-0.028	0.136
	(0.049)	(0.154)	(0.170)	(0.054)	(0.161)	(0.178)
FCC efficiency	0.046	-0.009	0.055	0.133	0.125	0.008
	(0.043)	(0.123)	(0.137)	(0.055)	(0.159)	(0.178)
FCC transparency	0.064	0.181	-0.117	0.405	-0.215	0.619*
	(0.059)	(0.169)	(0.189)	(0.117)	(0.328)	(0.370)

Table 13 reports treatment effects on legitimacy outcomes, conditional on baseline tax preferences, at the midline and endline survey. Columns 1 and 2 report treatment effects at midline for supporters of taxation and opposers of taxation, 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 somewhat (6.9%) or strongly (7.1%) disagree with paying more taxes for improved services are defined as tax opponents; respondents who somewhat (25%) or strongly (57.2%) agree with paying more taxes for better services are defined as supporters. Respondents whose support was "in the middle" (3.8%) are grouped with supporters.

$^{*}p < 0.10$; $^{**}p < 0.05$ $^{***}p < 0.01$

Table 13: Effects on Legitimacy Outcomes Conditional on Tax Attitudes

D.3 Political Engagement

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

Table 14 reports the effect of the treatment on political engagement measures. Columns 1, 2, and 5 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 6 presents treatment effects estimates at the endline survey. Columns 4 and 7 reports the number of non-missing observations in the midline survey and endline survey, respectively. Stars refer to randomization inference p-values. Reported effects are standardized effects. Attempts to contact MP or Councillor, or attendance at political meeting, are for last six months. $^{*}p < 0.10$; $^{**}p < 0.05$ $^{***}p < 0.01$

Table 14: Political Engagement

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

Table 15 reports treatment effects on political engagement, conditional on partisan affiliation, at the midline at 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. $^{*}p < 0.10$; $^{**}p < 0.05$ $^{***}p < 0.01$

Table 15: Effects on Political Engagement Outcomes Conditional on Partisan Affiliation

Outcome	Midline			Endline		
	Support	Oppose	Diff	Support	Oppose	Diff
Knows ward councillor name	0.121	0.029	0.092	0.057	-0.020	0.076
	(0.046)	(0.137)	(0.153)	(0.044)	(0.130)	(0.145)
Attempted to contact ward councillor	0.221	0.199	0.021	0.082	0.118	-0.037
	(0.051)	(0.134)	(0.151)	(0.059)	(0.173)	(0.192)
Attempted to contact MP	0.128	0.258	-0.130	-0.085	0.297	-0.381*
	(0.050)	(0.137)	(0.155)	(0.062)	(0.171)	(0.192)
Level of interest in politics	0.193	-0.006	0.198	0.160	-0.313	0.473*
	(0.046)	(0.132)	(0.147)	(0.064)	(0.162)	(0.185)
Level of interest in FCC activities	0.445	0.327	0.117	-0.055	0.232	-0.287*
	(0.039)	(0.120)	(0.134)	(0.049)	(0.138)	(0.155)
Attended political meeting	0.043	0.089	-0.047	0.112	0.036	0.076
	(0.057)	(0.172)	(0.192)	(0.067)	(0.196)	(0.220)

Table 16 reports treatment effects on political engagement, conditional on baseline tax preferences, at the midline and endline survey. Columns 1 and 2 report treatment effects at midline for supporters of taxation and opposers of taxation, 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 somewhat (6.9%) or strongly (7.1%) disagree with paying more taxes for improved services are defined as tax opponents; respondents who somewhat (25%) or strongly (57.2%) agree with paying more taxes for better services are defined as supporters. Respondents whose support was "in the middle" (3.8%) are grouped with supporters. $^{*}p < 0.10$; $^{**}p < 0.05$ $^{***}p < 0.01$

Table 16: Effects on Political Engagement Outcomes Conditional on Tax Attitudes

D.4 Fairness and Enforcement

At endline, we find no evidence of persistent treatment effects on either fairness or enforcement mechanism outcomes (Table 17). 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.

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

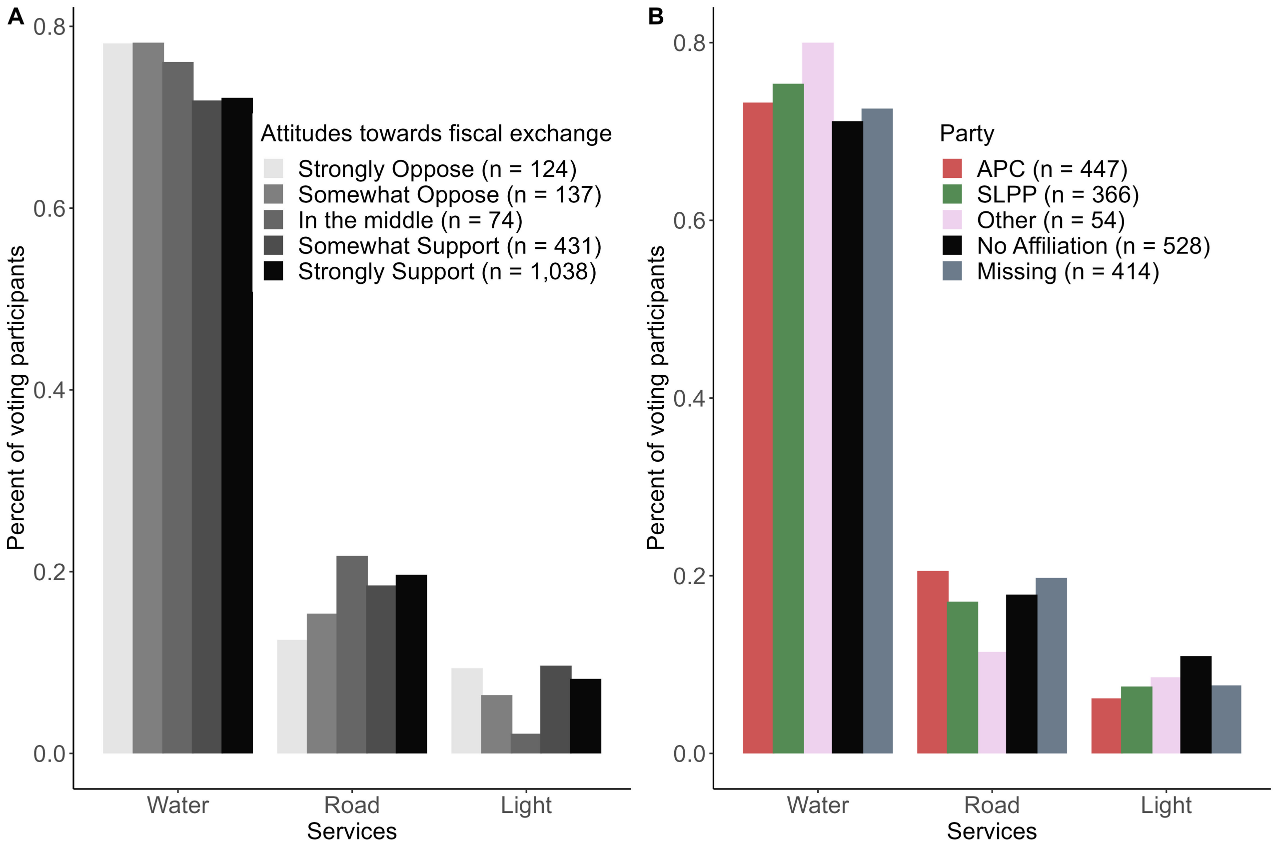
Table 17 reports the effect of the treatment on the alternative mechanisms of fairness and enforcement. Columns 1, 2, and 5 reports the control group mean for each indicator at baseline, midline, and endline, respectively (with the standard deviation in parentheses). Column 3 presents treatment effects estimates at the midline survey and Column 6 presents treatment effects estimates at the endline survey. Column 4 and 7 reports the number of non-missing observations in the midline survey and endline survey, respectively. Reported effects are standardized effects.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 17: Effect on Fairness and Enforcement

D.5 Service Preferences by Group

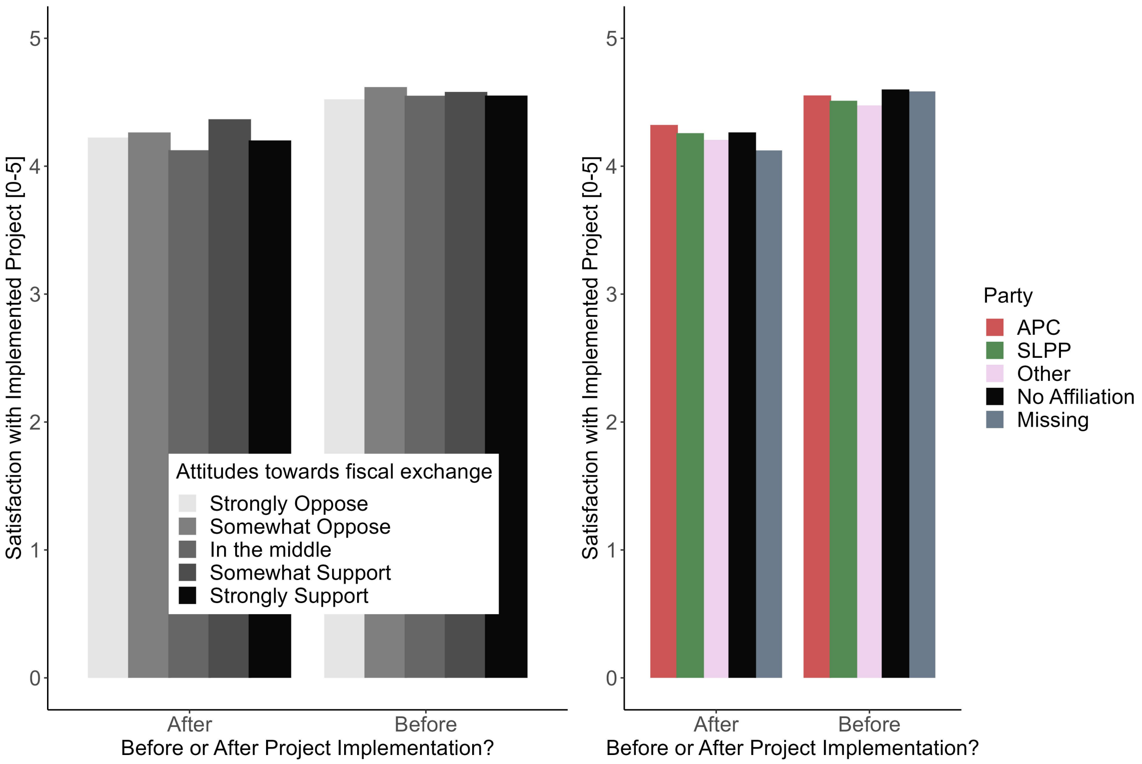
Vote share calculated from voting participants. Voting behavior is also similar across groups. For baseline attitudes towards fiscal exchange the percent of participants who voted is as follows: strongly oppose = 52%; somewhat oppose = 57%; in the middle = 62%; somewhat support = 55%; strongly support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

Figure 9: Votes for Services by Tax Attitudes (Panel A) and Partisan Affiliation (Panel B)



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

Figure 10: Service Satisfaction by Group



D.6 Polarization and Cohesion

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

Table 18 reports treatment effects on political polarization and social cohesion, conditional on partisan affiliation, at the midline at 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.
$^{*}p < 0.10$; $^{**}p < 0.05$ $^{***}p < 0.01$

Table 18: Effects on Political Polarization and Social Cohesion Conditional on Partisan Affiliation

Outcome	Midline			Endline		
	Support	Oppose	Diff	Support	Oppose	Diff
Level of trust in neighbors	0.176	-0.142	0.317**	0.024	-0.023	0.047
	(0.044)	(0.128)	(0.143)	(0.054)	(0.137)	(0.158)
Level of connection of neighbors	0.057	-0.228	0.285*	-0.096	0.204	-0.300*
	(0.050)	(0.144)	(0.161)	(0.057)	(0.148)	(0.170)
Ease of befriending opposition party members	-0.197	-0.086	-0.111	-0.015	0.043	-0.058
	(0.059)	(0.180)	(0.200)	(0.067)	(0.187)	(0.210)
Ease of believing opposition party members	-0.086	0.058	-0.144	-0.085	0.130	-0.215
	(0.047)	(0.137)	(0.153)	(0.053)	(0.154)	(0.172)

Table 19 reports treatment effects on political polarization and social cohesion, conditional on baseline tax preferences, at the midline at endline survey. Columns 1 and 2 report treatment effects at midline for supporters of taxation and opposers of taxation, 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 somewhat (6.9%) or strongly (7.1%) disagree with paying more taxes for improved services are defined as tax opponents; respondents who somewhat (25%) or strongly (57.2%) agree with paying more taxes for better services are defined as supporters. Respondents whose support was "in the middle" (3.8%) are grouped with supporters. $^{*}p < 0.10$; $^{**}p < 0.05$ $^{***}p < 0.01$

Table 19: Effects on Political Polarization and Social Cohesion Conditional on Tax Attitudes

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.[38] Our approach compares non-study properties geographically proximate to treated study properties to non-study properties proximate to control study properties.[39] We estimate spillovers with the following equation:

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

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, β_1 captures the spillover effect on tax compliance of being close to a treated property owner. $Y_{i_{2020}}$ is the tax compliance behavior of property owner *i* in 2020; δ is a vector of ward fixed effects; **X** is the set of property-level characteristics described in Section 4, included as covariate adjustment.

As the density of buildings varies across the city, the probability of being assigned to "spillover treatment" (i.e. the probability that $SPILL_i$ is equal to one in equation 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.[40] 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 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.[41]

[38] 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.

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

[40] Imagine, for example, potential differences in compliance behavior between densely packed informal settlements and spacious affluent neighborhoods.

[41] Absent a theory-driven procedure for selecting the threshold distance (*D*), we opt for a pragmatic approach.

The top panel A in Table 20 shows spillover effects on compliance behavior at this preregistered threshold distance. Column 1 shows results for compliance behavior in 2022, our preregistered primary dependent variable for the spillover analysis. While the point estimate at this distance threshold is positive, about a third of a percentage point, it is statistically insignificant (the RI *p*-value is in brackets under the estimate). As noted in our preanalysis plan, the selection of this preferred distance threshold is somewhat arbitrary; the additional results in Column 1 show the estimated spillover effect when the distance threshold is defined below (Panel B) or above (Panel C) the preregistered threshold. The estimated effect is positive at all thresholds, and approaches statistical significance at some, but the results are at best suggestive. While we cannot reject the null of no spillover effect at our preregistered threshold, one interpretation of the results in Column 1 is that there is a positive spillover effect, but our estimation is noisy.

One way to reduce noise is to pool tax compliance behavior across 2022 and 2023, such that the dependent variable becomes paying taxes in either 2022 or 2023 (Column 2).[42] Treatment effects for these pooled estimates at the preregistered threshold are substantively large (1.2 percentage points) and statistically significant (*p*-value = .08). Moreover, these results appear robust to the definition of the threshold distance. When the threshold is defined below the preregistered threshold (Panel B), point estimates are larger and estimates are statistically significant. When the threshold is defined above the preregistered threshold (Panel C), point estimates are of a similar magnitude and RI *p*-values hover at the threshold of statistical significance.

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.

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

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

Table 20 reports spillover effects on the compliance behavior of non-study property owners, at different distance thresholds for defining spillover units (Column 1). Column 2-3 reports spillover treatment effects on compliance behavior. In Column 2, the dependent variable is a dummy indicating if the owner paid any tax in 2022. In Column 3, the dependent variable is a dummy indicating if the owner paid any tax in 2022 or 2023. Treatment effects are reported in raw percentage points. Randomization inference p-values are below each estimate in brackets. Stars refer to randomization inference p-values. Columns 4 and 5 refer to the number of observations in treatment and control, respectively, at a given distance threshold. $^{*}p < 0.10$; $^{**}p < 0.05$ $^{***}p < 0.01$

Table 20: Spillover Effects