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

Kevin Grieco†
Abou Bakarr Kamara‡
Niccolò Meriggi§
Julian Michel†
Wilson Prichard¶

April 25, 2024

PRELIMINARY DRAFT. PLEASE DO NOT CIRCULATE.

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 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 preexisting attitudes towards paying taxes and partisan affiliation strongly condition 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 Graeme Blair, Katherine Casey, Darin Christensen, Cesi Cruz, 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 Cali-WEPSs IV workshop, APSA 2022, and the Global Development Conference 2022 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

†Ph.D. Candidate in Political Science, UCLA
‡Senior Country Economist, International Growth Center, Sierra Leone
§Research Fellow, Centre for the Study of African Economies, University of Oxford
¶Associate Professor, University of Toronto

1 Introduction

Fiscal capacity, and state capacity more generally, is essential for economic development and political stability (Besley and Persson 2011). Building durable fiscal capacity requires that the state obtains compliance with its tax demands. One strategy is through enforcement. There is extensive evidence of the effectiveness of enforcement (e.g., Kleven et al. 2011; Slemrod 2019), but those efforts are administratively and politically costly, can erode public support if they are viewed as unfair, and are often not undertaken owing to a lack of political support.[1] An alternative strategy is to strengthen the willingness of taxpayers to quasi-voluntarily comply with tax demands by expanding the perceived legitimacy of government (Levi 1988, 1997). Such efforts may focus on building citizen trust that the tax system is fair, that citizens will receive valuable services in return, or that citizens have a voice in shaping how revenues are used (Luttmer and Singhal 2014; Prichard 2023). In the context of weak states with limited enforcement capacity such efforts to encourage quasi-voluntary compliance are likely to be especially important.[2]

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

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

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

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

This paper examines the relationship between political participation, legitimacy, and tax compliance in Freetown, Sierra Leone by reporting results from a participatory budgeting field experiment that we designed and implemented in collaboration with the Freetown City Council (FCC). In doing so, we contribute to an emerging literature on institutional experiments (Callen et al. 2023) by providing the first field experimental study of whether participatory budgeting can facilitate capacity-building in fragile states. The intervention sought to give participants greater voice in, and control over, policy decisions regarding local development projects implemented in their neighborhood. Program participants joined phone-based WhatsApp chat groups—which we call *Digital Town Halls* (DTHs)—with up to 40 other property owners in their neighborhood to discuss service preferences, shared these preferences with politicians, and then voted for the services (of approximately US$1,500) they wanted to be implemented in their neighborhood. Selected services were implemented six months later, and participants were informed of this through a phone call. To identify causal effects, we use a matched-pair design (King et al. 2007) to randomize half of 3,618 property owners into treatment. We observed individual level tax compliance through administrative records and we surveyed the treatment and control groups at three stages: before the process, after services were selected but before they were delivered, and after services were delivered.

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

However, despite the intervention's positive and unambiguous effect on perceptions of government legitimacy, we find a robust null effect on tax compliance behavior. Why does the intervention fail to increase tax compliance despite raising perceptions of legitimacy? Empirically, the answer lies in strikingly heterogeneous, and countervailing, impacts of the intervention on different sub-groups of taxpayers. In exploratory analyses, we find that for property owners who supported (at baseline) expanded taxation to enable improved services (57.4% of the sample), treatment significantly increases compliance by 3.31 percentage points, a 9.5% increase over the group's baseline compliance rate. However, these effects are washed out by decreasing compliance among those who initially did not support expanded taxation.[4] We find a very similar pattern along partisan lines. For co-partisans of the Mayor, the treatment significantly increases compliance by 7.4 percentage points, which is a substantial 27.9% increase over the group's control compliance rate. In contrast, the treatment lowers compliance by 4.0

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

percentage points for respondents who are not copartisans of the Mayor.[5]

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

The most compelling explanation for these heterogeneous effects on compliance is that treatment leads to heterogeneous, and countervailing, effects on groups' policy preferences about taxation. First, an extensive literature finds that individuals respond to elite cues by shifting their policy position closer to the position of their party (Broockman and Butler 2017; Tappin et al. 2023).[6] Moreover, while these cues may be persuasive for the political ingroup, they can generate backlash from the outgroup (Nicholson 2012; Haas and Khadka 2020). In the DTHs, Mayor Aki-Sawyerr pitched her policy vision of more taxes and better services, a message that is also at the core of her administration's *Transform Freetown* agenda.[7] Copartisans responded to the treatment by increasing their support for expanding taxation for better services; treated non-copartisans reduced for support for this policy. Second, an established literature in psychology and political science finds that people often have a motivation or goal (Kunda 1990) when processing new information to form opinions about policy (Lodge and Taber 2013; Taber and Lodge 2006). Individuals often "strive to defend and maintain their extant values, identities, and attitudes" (Slothuus and De Vreese 2010).[8] We find that treatment crystallized policy preferences about taxation: opponents of expanding taxation became more opposed, and supporters become more supportive. If individuals' preferences about tax policy affect tax compliance behavior, the treatment's polarizing effect on tax policy preferences can explain its polarizing effect on compliance behavior.

Ultimately, our study makes several overlapping research contributions. First, we provide the first field experiment on whether direct democracy in the form of participatory budgeting can increase (a) support for authorities seeking to expand the state and (b) tax compliance in an already-existing representative democracy. While a related literature finds that public participation in Town Halls improves development outcomes (Gonçalves 2014; Fujiwara and Wantchekon 2013), participants' political efficacy (Boulianne 2019) and knowledge (Esterling et al. 2011), and approval of participating politicians (López-Moctezuma et al. 2022), we do not

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

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

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

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

know if these participatory platforms engender tax compliance and capacity-building. We find that participation in participatory budgeting Town Halls increased the legitimacy of government and approval of participating politicians, but raised tax compliance only among copartisans of the mayor.

Our finding that expanded voice does not, on average, increase tax compliance is in contrast with the more optimistic accounts offered by existing observational work (Pommerehne and Weck-Hannemann 1996; Torgler 2005; Touchton et al. 2019). Our results also contrast with findings from lab experiments, which find that people are more likely to comply with (lab) taxes when they vote over how funds will be spent (Alm et al. 1993) and with rules when they are decided on in a democratic fashion (Bó et al. 2010; Sutter et al. 2010). However, observational studies may suffer from endogeneity issues, lab experiments abstract from political context and existing observational and lab experimental research use as dependent variables survey measures, aggregated revenue data, or game-based measures rather than individual-level compliance data. Our approach improves on existing work by experimentally manipulating political voice in the real world and directly observing the impact on individuals' compliance behavior. Our finding that increased citizen engagement—and accruing legitimacy gains—does not translate to higher tax compliance has important theoretical implications, as it challenge influential models of quasi-voluntary tax compliance (Levi 1988; Luttmer and Singhal 2014) and canonical accounts of the taxation-representation nexus in early European state building (Bates and Lien 1985; North and Weingast 1989). However, as we discuss, legitimacy gains can open windows of opportunity for subsequent reform, with likely positive consequences for long-term capacity building.

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

We advance the literature by collecting a suite of pre-treatment political and attitudinal variables to tease out these heterogeneous effects. In doing so, we are the first (to our knowledge) to document that the effect of expanding political voice varies conditional on fundamental political variables. While our finding that participatory budgeting *decreases* tax compliance and support for expanded taxation among ideological and political opponents is not in line with the literature's predictions, similar "backfiring" effects have been documented for other common policy interventions, either on average or for large sub-groups. Such instances include anti-corruption campaigns (Cheeseman and Peiffer 2022), interventions to correct political misperceptions (Nyhan and Reifler 2010), and nudges placed on tax bills (De Neve et al. 2021).[9]

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

Finally, these results have important implications for policy and practice. The participatory budgeting program that we study does not, on average, increase tax compliance. However, as we discuss in the conclusion, that does not necessarily imply that policy-makers should eschew this participatory policy tool. The total impacts of a similar intervention implemented by a government may be different than the one we evaluate for several reasons. First, the impact of the intervention may spillover to those who did not directly participate, a possibility for which we find suggestive evidence (Appendix E). Second, medium and long-run effects may be different than the short run effects we study. Third, in the real world, participants are likely to opt-in to a participatory intervention, rather than be randomized into it, as was the case in our study; this may impact average treatment effects. Finally, we note that, looking beyond compliance, the intervention's clear positive effect on government legitimacy might allow governments to undertake reform (e.g., to build fiscal capacity) that would otherwise be politically infeasible.

2 Interventions: Digital Town Halls and Service Delivery

This research takes place in cooperation with the Freetown City Council (FCC) in a 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

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

of a more nuanced, consistent and transparent property valuation scheme.

The mayor publicly announced that digital town halls (DTHs) would be held starting in January of 2021. In her messaging, she emphasized that these DTHs would be key for securing citizen participation. She also stressed that she intended to institutionalize the DTHs and that future DTH would be assigned 20% of the property tax revenue raised in a given ward (see the Freetown City Council's second year Transform Freetown report, pg. 26)

In this study, the digital town halls serve as part of a broader intervention that contained three components: (i) digital town halls, (ii) service delivery, (iii) notification calls about delivered services. While only the the treatment group was invited to participate in the DTHs, the project 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 digital town hall plus having received a notification call, conditional on services being delivered.

2.1 Digital Town Halls

In this study, DTHs take the form of WhatsApp group chats where property owners had the opportunity to engage other property owners in their ward and their political representatives about pressing development challenges facing their community.[10] The overarching goal of the DTH was for the group to deliberate and decide over how a budget of 15 million leones (about USD 1,500) should be spent in their ward. Treated participants were assigned to one of 58 chat groups with other property owners in their own ward. The number of participants in each chat ranged from 17 to 37 (the median chat group size was 24). The DTHs comprised four distinct phases:

1. **Horizontal Deliberation (January 15-19, 2021):** Participants received introductory videos from the Mayor of Freetown and their respective ward councilor.[11] In this first video, the mayor welcomes participants to the DTH and then makes an explicit connections between the payment of taxes (including property tax) and the delivery of services:

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

[10] 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.

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

money to deliver services to the people of Freetown.

In these videos, political representatives also invited participants to discuss development concerns within the group. In the group chats, group moderators introduced themselves to participants (both in the group chat and in one-to-one conversations with participants) and then asked participants to respond to the following prompt: "What do you think is the greatest development problem in your ward?"[12] Across all groups, participants sent 2,000 messages.[13] This phase was purely horizontal because participants were told that political representatives would not be involved and could not observe or learn about what was discussed in this phase.

2. **Preference Articulation and Aggregation (January 20-February 12, 2021)**
After five days of horizontal deliberation, DTH participants received a video from the Mayor of Freetown requesting that participants to (i) identify the two greatest development challenges in their ward and (ii) propose a plan to to tackle that challenge.[14] Participants were told that they should think of projects that could be completed for under US$1,500 and instructed to submit either a written message or a short voice recording. Participants were told that the team facilitating the DTHs would listen to their messages, aggregate information, and present both the mayor and ward councilor with memos outlining the concerns and proposed solutions of their constituents. In doing so, the DTHs allowed participants to (anonymously) articulate their preferences to their representatives. Through this process, it was clear that water and water access were the most pressing concerns facing many communities. This phase is the first time that participants' statements were be shared with political representatives. Some common examples demands communicated to political representatives include:

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

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

[13]The majority of messages sent were text (55.25%) and an additional 40.2% were voice messages. Other types of messages include images and videos.

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

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

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

3. **Vertical Interaction (February 13-16, 2021):** Participants received videos from both their councilor and the Mayor. Councilors acknowledged input they received from participants, based on the ward-specific summaries prepared by the research team, and position themselves to the demands made.[15] This includes highlighting their preferred services, justifications for their service preferences and explaining past and future delivery goals. In her video, the Mayor also acknowledged and responded to participants' demands, though in more generic language, as she prepared one video that was shared with each DTH.[16] In their videos, the Mayor and councilors told respondents that an engineering firm that been assessing the feasibility of their proposed projects and that five projects had been determined as feasible within the budget:

    - Two new solar street lights
    - Fix some potholes
    - 50m of truck tracks
    - Fix some GUMA water pipes
    - Install a new water hand pump

    Participants were also informed that voting would start in four days. Moderators reminded participants that, "This is your chance to make your case to the other people in this WhatsApp group about what to prioritise 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 (Figure 1) and a "how to" video was posted in each group that provided step-by-step instructions of the voting process.[17] Voting was open for four days and was closed on February 22. The mayor announced the winning project for each ward with an ward-specific voice message, which was posted in each DTH alongside a picture of the Mayor in office.[18]

---

[15] Where there were two DTHs in a ward, the councilor prepared a video that addressed concerns raised in both DTHs.

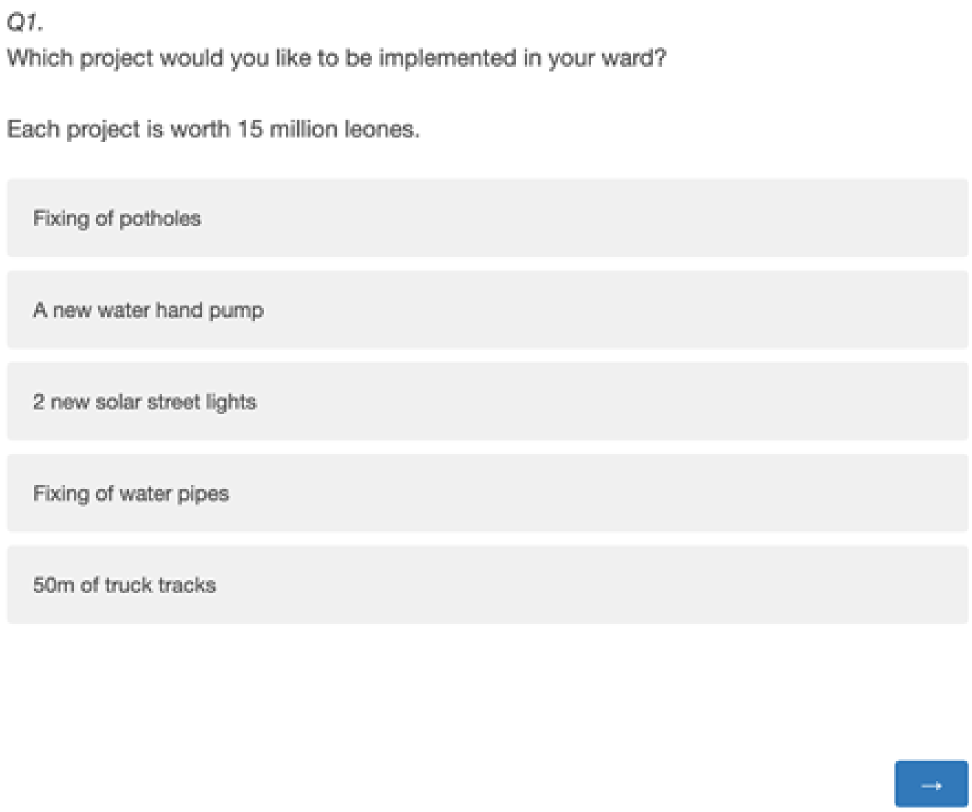
[16] We opted for this mediated interaction between citizens and representatives to (i) avoid elite-domination of the TH process and (ii) make realistic time-demands on representatives.

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

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

After the announcement of the winning projects, group moderators thanked participants for their contributions and then halted participants' ability to post messages in the the DTHs. Moderators explained that chats would be used one more time in the future to announce that delivery of the service has concluded.

Figure 1: Menu of services



There was active participation in the DTHs. We confirmed that 1,457 of the 1809 treated property owners joined the digital town hall, a compliance rate of 80.5%. The majority of property owner who joined the DTH reported that they accessed the groups daily (54%) and and 84.3% reported they accessed the DTH more than once per week.[19] Roughly two-thirds of those who joined the DTH voted for their preferred service to be implemented (68%) and posted at least one message in the DTH (63%).[20]

Participants also report 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)

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

[20] 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 statistic regarding messages includes all forms of messages, such as text and audio. 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.

and "better understand views from fellow members of my community" (4.04/5). In addition, respondents on average agreed that "participants felt comfortable to make their views known even when their views differed from those of other participants" (3.82/5). However, respondents were less satisfied that the DTH budget was sufficient to improve the delivery of the selected service in a meaningful way (2.86/5) and had middling feelings about the comprehensiveness of the list of services they could vote for (3.33/5) (see table 9 in the Appendix).

While the service delivery budget was no drawn from the FCC's regular revenue, this was not communicated to project participants, allowing the Mayor and councilors to claim full credit for the participatory budgeting program and associated service provision.[21] Respondents overwhelmingly reported they believed that FCC organized the DTH (89%), implemented services (96%), and funded the services (84%).[22] 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.

2.2 Service Delivery

Each participating ward received a service project—essentially a local public good both treated and control units in that ward could profit from.[23] Construction began in most wards in October 2021 and was completed in all but one ward by the end of the year.[24] In the remaining ward, construction was completed in February 2022.[25]

Despite the multiple rounds of feasibility assessments conducted by the engineering firm to determine the list of feasible projects, at the time of project implementation the engineering firm concluded that some selected projects could not be effectively delivered within some wards. Specifically, we were not able to fix water pipes in wards that selected this service. Of the 19 wards that voted to fix water pipes, we built street taps in 11 of them. In the remaining 8 wards, water street taps were deemed infeasible; in these cases we opted to provide a 5000L Milla Tank instead. This was an item participants did not have the option to directly vote for in the DTHs,

[21]The budget allocated to the DTH did not come from FCC's regular budget because of (1) the severity of the budget constraint the FCC faces and (2) that the expected increase in property tax revenue will be accrued after the DTHs have taken place. For these reasons, the funds to be decided over are taken from the project's research budget.

[22]See table 8 in the appendix. We asked participants which actors they believed were responsible for organizing, implementing, and funding the DTHs. For these questions, respondents were allowed to name multiple actors they thought might be involved. The next most frequently named actor involved for each activity is as follows: organizing the DTH was "researchers" (12%); for service implementation, central government (4.5%); as for funding, 11% of respondents thought that projects were at least partially funded by private citizens, either inside or outsider their ward and 11% thought that the projects were funded by central government.

[23]Pictures of implemented projects can be found in Appendix B.2.

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

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

but one that reflected the participants' preference for improved water provision in affected wards. In addition, one ward voted to "fix pot holes", but implementation proved difficult. Instead, this ward received 50 meters of truck tracks. Table 1 summarizes the votes received for each project, the number of projects that won the vote across wards, and the projects that were eventually implemented. Even with these challenges, participants reported they were satisfied with the selected services both after the DTH (4.6/5, midline survey) and after the implementation of the projects (4.2/5. endline survey).

	Projects for vote					Replacement Projects	
	Water		Road repair		Solar	Water	
	Fix pipes	Hand pump	Truck tracks	Potholes	Street lights	Tank	Street tap
Votes	429	313	138	51	83	NA	NA
Won	19	9	2	1	0	NA	NA
Built	0	9	3	0	0	8	11

Note: The top row ("Votes") describes the number of votes of each project. The middle row ("Won") shows how those votes translate to number of projects won for each project type. The project "fixing of water pipes" was found to be too difficult to implement—wards that voted for this project received a different water related project: either a 5000 liter tank in the community ("Milla Tank") or a stree tap. One ward voted to "fix pot holes", but implementation proved difficult. Instead, this ward received 50 meters of truck tracks.

Table 1: Project votes, winning projects, implemented projects

We found it plausible that not all DTH participants would be aware that the projects they selected in the DTH had been successfully implemented.[26] To address this, we made notification calls on behalf of the FCC to all treated units, informing participants the project chosen the the DTH had been successfully implemented. Note that by making these notification call to treated units but not control, we build the notifications calls into into our treatment.[27] We successfully reached approximately 70% of treated units to inform them of the implemented services. These calls started in mid-November and were staggered across wards so that they started once service delivery was completed in that ward. The endline survey similarly was staggered and commenced after notification calls were completed, but never earlier than one week after delivery completion.

[26]Further, we worried that our inability to observe respondents' knowledge of project implementation would complicate the interpretation of our findings. For example, to what extent should a null (or perverse) effect be attributed to respondents' (mistaken) belief that services selected in the DTH had not been implemented?

[27]Therefore, the treatment effect includes the heightened awareness of treated units regarding the implementation of the service.

3 Data Collection

3.1 Sampling

We constructed a sampling frame using FCC administrative records, which contains information about property characteristics and property owner contact details. To be eligible to participate in the Digital Town Hall a property owner must (i) own a property in one of the 30 study wards and (ii) have WhatsApp on their phone.[28] 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 owner were eligible to be selected into 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.[29] Second, we removed some properties from the sample frame in an attempt to limit geographic spillover.[30] Third, note that we could not contact owners of properties where owner contact information was not listed in FCC records.

3.2 Measurement

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

> **Baseline:** Between October 28 and December 2, 2020, we attempted to survey the 4,860 property owners that we had verified as eligible for the study (see Section 3.1) and completed baseline surveys with 3,859 eligible property owners (79.4%). Property owners with whom we completed the baseline survey were eligible to receive treatment and we

[28]For property owners that own multiple properties, we coded them as being eligible for the DTH in the ward that contains their highest value property (i.e., highest tax fee). This prevents the same property owner from being assigned to the DTH in multiple wards or being assigned to both treatment and control conditions. We made multiple property owners eligible for the DTH in the study wards where their highest value property was located as we reasoned that they were more likely to be resident of these properties and more likely to be involved in the administering of these properties (and therefore more likely to be directly involved in the decision to pay property tax). Note that there are only a handful of DTH participants who own multiple properties and are in the DTH of their second highest value property. In these instances, the ward in which they have a higher value property is not a study ward.

[29]As a response to COVID-19, the FCC intended to waive property tax for 2020 on properties of below median value. As our intervention was originally scheduled for early 2020, it was necessary to target the DTH intervention at property owners who owned 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 property owners who own a property below the median value. We included these property owners in our sample.

[30]In a previous version of our research design, we planned to allocate treatment status using a two-stage randomization procedure, to mitigate and estimate geographic spillover (as in Sinclair et al. 2012). Under that research design, 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. We we eventually moved on from this research design, sampling was done with that design in mind.

attempted to survey these property owners in subsequent survey rounds.[31]

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%).[32]

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

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

4 Treatment Assignment and Balance

Before assigning treatment, in an effort to mitigate spillover, we drew a restricted sample from the set of 3,859 eligible property owners with whom we had completed a baseline survey such that each property is at least 15 meters from the closest study property. The restricted sampling left us with a final sample of 3618.[33] We then assigned treatment status using a matched-pair design, leveraging baseline survey data to match similar observations into groups of two.[34] 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 placed the greatest weight on our measure of unconditional tax morale—we expected this to be the strongest predictor to tax compliance, in line with the common use in the literature of this variable as proxy for tax compliance behavior. Appendix C describes the matching procedure in greater detail. We created 1809 pairs and then assign one observation in each matched-pair to treatment and the other to control.[35]

[31] Appendix Figure 6 documents the broader data collection and project timeline.

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

[33] Appendix Figure 9 displays the distribution of the distance from each property to the closest property in the sample.

[34] 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.

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

	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 2 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 2: Balance Table

Table 2 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.

5 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 town hall using an instrumental variable regression framework. Our main equation is:

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

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 digital town hall. 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.[36] δ is a vector of ward fixed effects and ϵ_i is the error term.

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

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

Where D_i is the randomly assigned treatment indicator, which instruments for DTH_i in equation 1.[37] Our quantity of interest is α_1 (equation 1), which captures the local ATE among the set of people who comply with treatment, which in this case refers to property owners who joined

[36]Preregistered control variables include: We include the following control variables: (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 type received. Where covariate data is missing, including baseline values of the outcome of interest, 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.

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

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.

5.1 Effects on Legitimacy

We measured four components of standard conceptualizations. First, one factor driving legitimacy is "the extent to which citizens can influence policy" (Levi et al. 2009). The DTHs were designed to build the social contract by increasing citizens' participation in political affairs and increase their control over political decision-making. We find that the intervention increased participants' perception they could influence policy (Table 3). Specifically, we preregistered two indicators and predicted that the treatment should increase respondents' perceptions (1) that they have opportunities to voice their opinions about government matters to government officials and (2) that it is easy to directly engage in political activities. The intervention had large and durable effects on this first indicator, increasing reported *opportunities for voice* by 0.38 standard deviation units (SDUs) at the midline survey and 0.25 SDUs at endline. Given the baseline standard deviation is roughly 1.00, these effect sizes can be interpreted as changes on a 4-point Likert scale. The effect on *ease of participating in political activities* is positive at both midline and endline, with statistical significance at the threshold of conventional levels.[38] Taken together, we interpret these results as strong evidence that the intervention produces durable impacts on participants' perceptions that they can influence policy.

Second, legitimate governments provide citizens with basic services and generally respond to citizens' demands. After citizens voiced their concerns about local development and their preferred services, Ward Councilors used the digital town halls to respond to citizens' concerns. We find that the intervention increased citizens' perceptions that the local government was more *responsive to citizens' needs and demands*. Treated participants reported higher levels of local government responsiveness relative to control respondents, a difference that is statistically significant both directly after the digital town hall (midline; p-value <0.001) and after the selected service is implemented (endline; p-value = 0.016). In addition, the intervention attempted to forge the social and fiscal contract between citizens and politicians by delivering the local services that people demanded. We see find that the intervention increased citizens' *satisfaction with FCC service provision* at both midline (p-value <0.0001)—when services are promised, but as yet not delivered—and endline (p-value = 0.004), after services were implemented.

Third, legitimate governments administer their constituencies competently. Our survey data show that, prior to the intervention, respondents perceive the FCC as fairly incompetent: the average respondent perceives the FCC as *not* transparent (1.36/3), of middling efficiency (2.86/4), and not too corrupt (3.53/5).[39] The intervention, however, improves respondents perceptions

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

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

of FCC administrative competence. At baseline, 76% of respondents said it was "difficult" to find out how the FCC spends tax revenue. The intervention improves respondents' perceptions transparency. While there is a small effect at the threshold of statistical significance at midline (0.085 SDU; p-value = 0.104), that effect balloons by almost a factor of four at endline (0.319 SDU; p-value = 0.0017). At baseline the modal respondent (62%) reported that the FCC is "somewhat efficient" in the way it uses money for public administration and development. While estimated treatment effects at midline are statistically indistinguishable from zero (p-value = 0.32), at endline we find we find unambiguous evidence that the intervention increases perceptions of the FCC's efficiency (p-value = 0.0078). For perceptions of corruption, we find a similar, though more extreme, change in treatment effect between midline and endline. At midline, treatment *increases* participants' perceptions that the FCC is corrupt (p-value = 0.001). However, after services have been implemented, treated participants *positively* update their views of FCC corruption, relative to control (p-value = 0.071).

The reversal of the sign on the effect of perceived corruption can be explained with an intuitive logic: in the DTHs, citizens learn of the existence of a development fund for each neighborhood, which is an additional pocket of money prime for corrupt exploitation; participants negatively update about corruption at the FCC. However, when citizens see these funds being efficiently used to implement development projects, this example of clean governance causes them to update again, such that the FCC is *less* corrupt than they had originally thought. More generally, these three results suggest that government needs to deliver on its promises of implementing development projects before citizens fully update their attitudes about government administrative competence. Nothing could be more rational: citizens understand that talk is cheap; they respond to action.

Fourth, we measured participants' approval of the performance of local elected officials. Specifically, we asked respondents if they approved or disapproved of how both the Mayor and their ward councillor has performed on the job over the past twelve months. Our data show that the Mayor is popular at baseline: most respondents report they either "strongly approved" (43.4%) or "approved" (44.3%) of the mayor's performance. The intervention increases approval of the Mayor by 0.15 SDUs (p-value <0.001) at baseline and 0.19 SDUs (p-value <0.001) at endline.[40] Respondents reported much lower approval ratings for their ward councilors at baseline: the modal respondent (41%) "disaproved" of their councilors performance over the past year. While baseline approval for councilors was low, the intervention increased participants' approval at both midline (0.19 SDUs; p-value <0.0001) and endline (0.17 SDUs; p-value <0.001).

[40] 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 3 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 3: Effect on Legitimacy

In summary, Table 3 provides unambiguous evidence that the intervention increases attitudes commonly associated with legitimacy. In the next section, we investigate whether this shift in legitimating beliefs led to a corresponding shift in tax compliance behavior, as would be predicted by the literature (e.g., Levi 1988).

5.2 Effect on Tax Compliance

Panel A of Table 4 reports treatment effects on tax compliance in 2022 (main outcome of interest) and 2021. When interpreting treatment effects recall that, in 2021, the DTH started at the beginning of the tax season and the tax deadline came before any services were delivered.

By contrast, the 2022 tax season comes after all of the services selected in the DTH are implemented. In that sense, only 2022 provides a test of the full treatment of participation + service delivery and is our pre-registered primary outcome of interest.

Column 1 reports the control group mean compliance rate in 2021 and 2022 and Column 2 report 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). We observe compliance behavior for all units; there are no missing values.

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	0.496	3,618
		(0.018)		
2021	0.315	-0.007	0.723	3,618
		(0.019)		
Panel B: Fiscal Exchange Attitudes				
Willingness to pay more taxes for better services				
Midline	4.001	0.066	0.163	3,296
	(1.253)	(0.047)		
Endline	4.030	-0.075	0.155	2,872
	(1.293)	(0.053)		

Table 4 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 4: Effect on Tax Compliance

These null effects are robust to different model specifications. Our main specification includes ward-fixed effects and a set of property characteristics; results are similar when we estimate effects using (i) only the treatment indicator and 2020 (pretreatment) compliance behavior; (ii) only ward fixed effects; (iii) only property characteristics, and (iv) when we add to our primary specification a dummy indicating the owner has zero tax liability. Results are also robust to different operationalizations of tax compliance. While Table 4 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.[41]

These nulls are also preciously estimated and any real effect is likely to be small: For 2022, we can say with 95% confidence that the treatment effect is less than 2.3 percentage points. Still, we might worry that a (small) true effect exists, but we find a null because we are underpowered. 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.[42] While the interpretation of outcome is a bit awkward—the group mean compliance, pooling across years—note that this effect is causally identified. The point estimate for this effect is close to zero (-1.1pp) and is not statistically significant (p-value = 0.45).[43] In summary, we find no evidence that the treatment, on average, impacts compliance behavior. Given the robustness of this finding, and the precision of our estimates, any real impacts are likely to be substantively small.

This null result runs against our expectations and the framework of tax compliance that motivated our hypotheses. In many models of tax compliance, government legitimacy is presumed to lead to higher expected service provision from government and therefore greater willingness to pay (Levi 1988, 1997; Besley 2020). This trade of taxes for improved services is commonly referred to as "fiscal excahnge" (e.g., Timmons 2005). One potential explanation for the mismatch between effects on attitudes and effects on behavior is an extreme budget constraint on property owners: it could be that treated property owners *want* to pay more taxes, but they simply have no disposable income with whcih to do so. If this were the case, we should see positive impacts on respondents' willingness to pay more taxes for better services, which we refer to as their *attitude towards fiscal exchange*.

We measured property owners' attitudes towards fiscal exchange by asking them if they agreed or disagreed with the following statement: *I would be willing to pay additional taxes in order to receive improved services.*[44] Panel B in Figure 4 presents the effect of the treatment on this attitude. We do not find evidence that the intervention increases property owners' attitudes towards fiscal exchange. The estimates at endline, which directly precedes our main compliance outcome, suggests that DTH participants are no more willing to engage in fiscal exchange than control respondents—if anything, participants are *less* disposed towards fiscal exchange (p-value = 0.16). Midline estimates move in the opposite direction, though, again, these effects are not statistically significant (p-value = 0.16).

[41] Results not presented in this draft.

[42] 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.

[43] Results not shown.

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

This finding also dispels another potential explanation for the mismatch between effects on attitudes and null effect on compliance behavior. We might be concerned that the effect on attitudes is driven by experimenter demand effects (Zizzo 2010), rather than true changes in beliefs. If that was the case, we should also find that treated respondents *say* they would be more willing to pay taxes, even when in fact they do not. We found no evidence of this in Table 4.

5.3 Why Does Participatory Budgeting Fail to Increase Compliance? How Political Factors Condition the Effect of Participation on Compliance

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

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

While the majority (57.4%) of surveyed respondents reported that they "strongly approved" of a policy of expanding taxation for improved services, a significant minority opposed (14%) this idea. An established literature in psychology and political behavior suggests that it will not have been easy to persuade these tax skeptics to change their policy preferences. In general, people have a motivation or goal when forming attitudes and beliefs, which are often "directional" in that they lead the receiver towards a particular conclusion (Kunda 1990; Lodge and Taber 2013). One typical motivation is that individuals "strive to defend and maintain their extant values, identities, and attitudes" (Slothuus and De Vreese 2010).[45] Further, previous research indicates that attempts to persuade people that their existing beliefs are incorrect may even backfire (Nyhan and Reifler 2010).[46] If the intervention failed, or even backfired, in its attempt to convince tax skeptics of a policy of expanded taxation, we should observe null, or even negative, treatment effects for tax skeptics on support for taxation and compliance behavior.

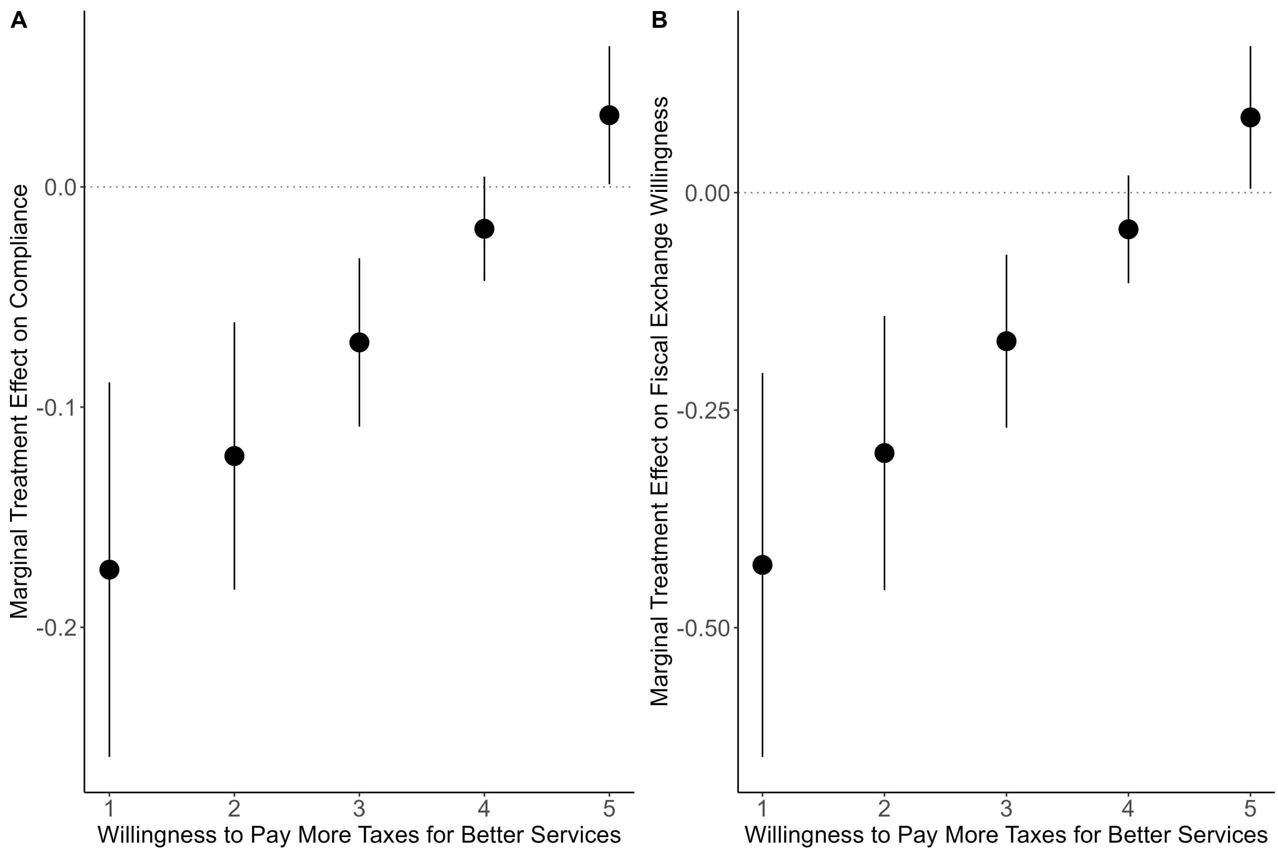
Plot A in Figure 10 presents predicted marginal effects from a model that interacts treatment

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

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

with our five-point measure of (baseline) attitudes towards fiscal exchange.[47] The interaction between treatment and attitudes towards fiscal exchange is statistically significant (p-value <0.001; $\beta = 0.052$). Baseline attitudes towards fiscal exchange strongly shape treatment effects on tax compliance in 2022. For respondents that strongly support expanded taxation (five on the five-point scale) the estimated treatment effect is 3.31 percentage points, a a 9.5% increase over the group's baseline compliance rate of 34.6%. By contrast, treatment effects are negative for property owners who do not approve fiscal exchange at baseline. For property owners who "somewhat disagreed" or "strongly disagreed" with paying more taxes for improved services, we estimate marginal treatment effects of -12.2 and -17.4 percentage points, respectively.

Figure 2: Marginal Treatment Effects by Attitudes Towards Fiscal Exchange



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

In keeping with this logic, tax skeptics respond to treatment by hardening their opposition to expanded taxation. Plot B (Figure 2), presents predicted marginal effects on attitudes towards fiscal exchange from a models that interacts treatment with our five-point measure of (baseline) attitudes towards fiscal exchange. To increase power for estimating this interaction, the pre-

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

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

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

It was in this politicized environment that the DTHs took place and the mayor made her direct appeals for fiscal exchange. Indeed, we find that the treatment increased participants' political engagement (Appendix Table 11), making it more plausible that they understood the Mayor's appeal to support expanded taxation in this politicized environment. An extensive literature finds that individuals respond to elite cues by shifting their policy position closer to the position of their party (Broockman and Butler 2017; Tappin et al. 2023; Flores et al. 2022). Moreover, while these cues may be persuasive for the political ingroup, they can generate backlash from the outgroup (Nicholson 2012; Haas and Khadka 2020). Both of these dynamics suggest that the Mayor's appeal for a policy of expanded taxation should increase tax compliance, and support for taxation, for copartisans; in contrast, the treatment should decrease tax compliance, and support for expanded taxation, for non-copartisans.

In Figure 3, we see evidence that participants' partisan affiliation conditions how the treatment impact their compliance behavior and their support for expanded taxation.[49] Plot A presents predicted marginal effects from a model that interacts treatment with a copartisan indicator variable.[50] The interaction between treatment and copartisanship 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

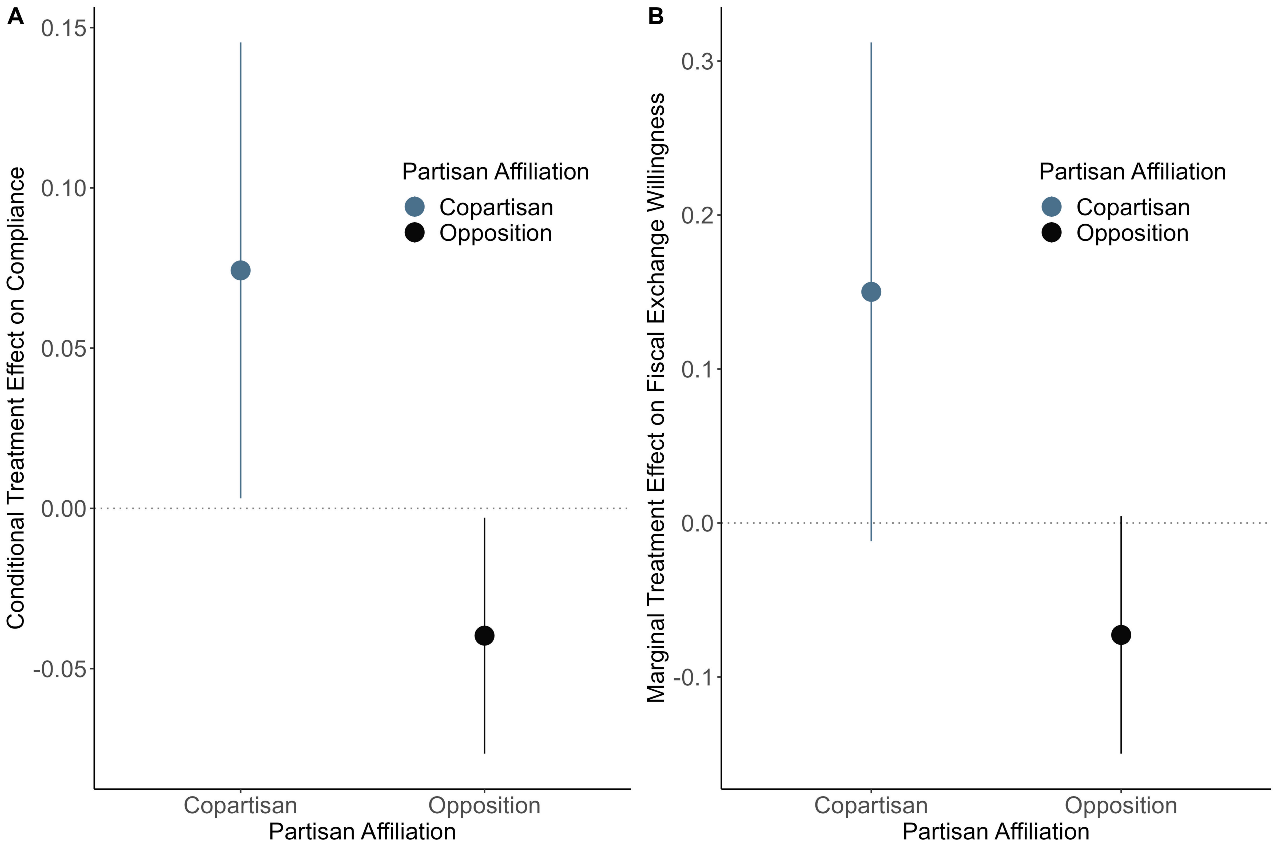
[48] Note that estimated marginal effects display similar patterns when estimated the interaction model using midline or endline data separately. See Appendix Figure 10

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

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

group's control compliance rate of 24.4%. This effect is statistically distinguishable from zero (p-value = 0.086). In contrast, treatment effects are *negative* for non co-partisans of the Mayor (i.e., respondents who do not declare themselves APC supporters): treatment lowers compliance in this group by 4.0 percentage points, an effect that is statistically distinguishable from zero (p-value = 0.076).[51] Partisanship strongly conditions treatment effects on tax compliance in 2022.

Figure 3: Treatment Effects Conditional on Partisan Affiliation



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

Panel B (Figure 3) presents marginal effects on attitudes towards fiscal exchange from a model that interacts treatment with the copartisanship indicator. As in Figure 2, the predicted outcome is the respondent's *average* attitude towards fiscal exchange across midline and endline surveys.[52] The interaction between treatment and co-partisanship is statistically significant (p-

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

[52] Appendix Figure 11 shows estimates from the interaction model using midline or endline data separately.

value = 0.063; $\beta = 0.223$ SDUs). In line with the expectations of the heterogeneous impact of partisan cues, treated property owners who are self-reported copartisans update their attitudes towards fiscal exchange more positively than do non copartisan property owners.

5.4 Why Does Participatory Budgeting Fail to Increase Compliance? Alternative Explanations

In Section 5.3 we explored why the intervention did not positively impact tax compliance (Table 4), despite generating unambiguously positive effects on legitimacy attitudes (Table 3) that are commonly thought to drive tax compliance. Our preferred explanation for the null effect on compliance emphasizes that treatment effects are conditional on baseline beliefs; we argue those beliefs shape the way people update their tax attitudes and behaviors in light of the intervention. There are, however, several alternative explanations for why the intervention fails to increase tax compliance. In this section we explore those alternative explanations.

5.4.1 Does the treatment negatively affect perceived fairness or enforcement?

It is possible that the intervention had a *negative* effect on other key mediating mechanisms that were not the target of our intervention. We preregistered two other mechanism channels through which the intervention might plausibly impact compliance: (i) fairness and equity and (ii) enforcement. If the intervention *reduced* participants' perceptions of the fairness of the tax system or the likelihood noncompliers would be punished, this could have washed out the positive attitudinal effects of the intervention reported in Table 3.

At endline, we find no evidence of persistent treatment effects on either fairness or enforcement mechanism outcomes (Table 5). 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 periods that directly precedes tax compliance behavior.

Estimated marginal effects display similar patterns.

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

Table 5: Effect on Alternative Mechanisms

5.4.2 Baseline beliefs are related to service preferences

Figures 2 and 3 show that treatment effects on tax compliance behavior are moderated by participants' baseline attitudes towards fiscal exchange and partisan affiliation, respectively. Our argument is these existing beliefs shape the way participants' process the informational signal of the treatment and therefore the treatment's effectiveness. However, these conditional estimates do not causally identify the impact of the moderating variable; factors associated with baseline attitudes towards fiscal exchange or partisan affiliation could be driving the observed conditional effects.

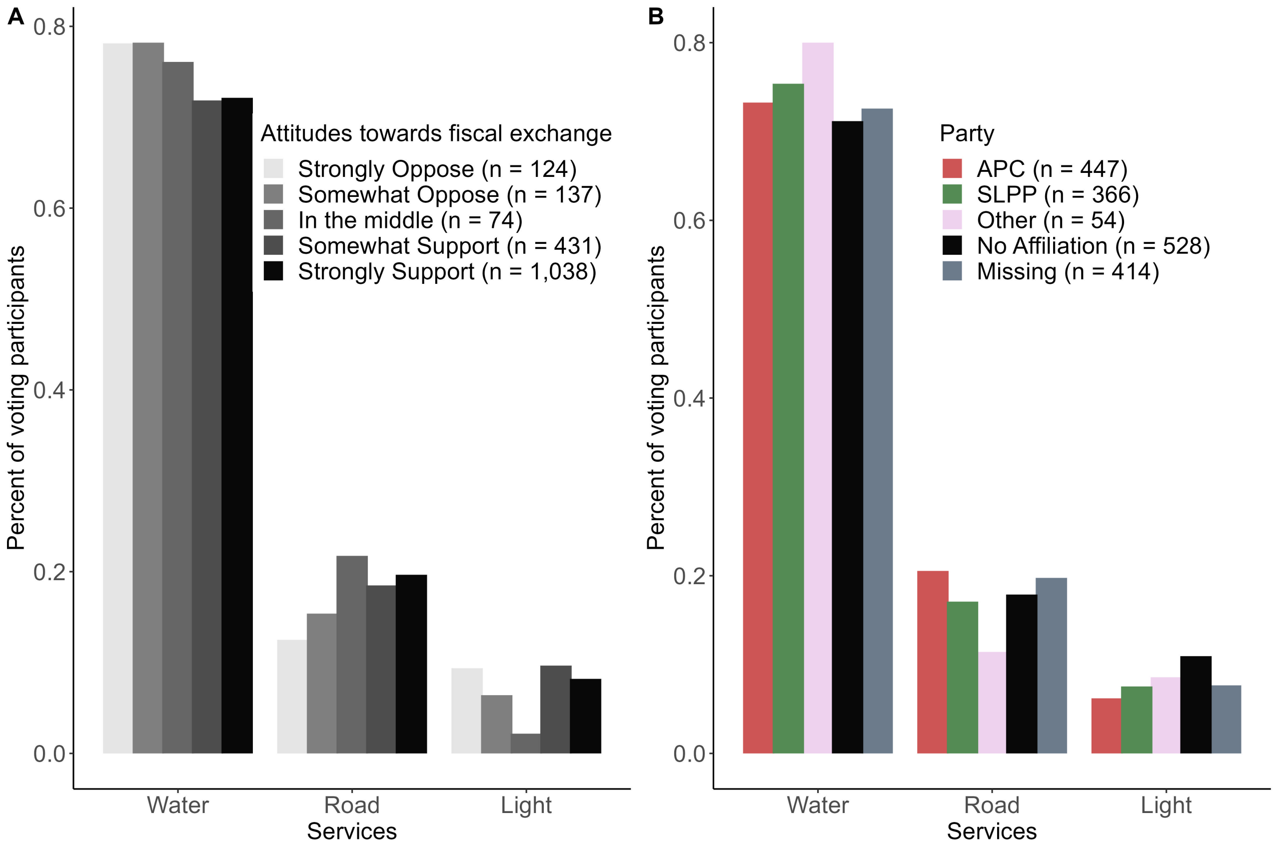
One potential confounder is service preferences. Treatment effects may be conditional on the match between a participants' preferred service and the implemented service. For example, effects may be positive for participants whose preferred service is implemented and negligible (or negative) for participants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

If this was what was going on, we should see that service preferences differ by attitudes towards fiscal exchange and/or partisan affiliation. Figure 4 shows the percent of voting participants who voted for a given project by group.[53] There is little indication that service preference vary

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

meaningfully across fiscal exchange attitudes (Panel A) or partisan affiliation (Panel B).

Figure 4: Votes for Services by Baseline Attitudes Towards Fiscal Exchange (Panel A) and Partisan Affiliation (Panel B)

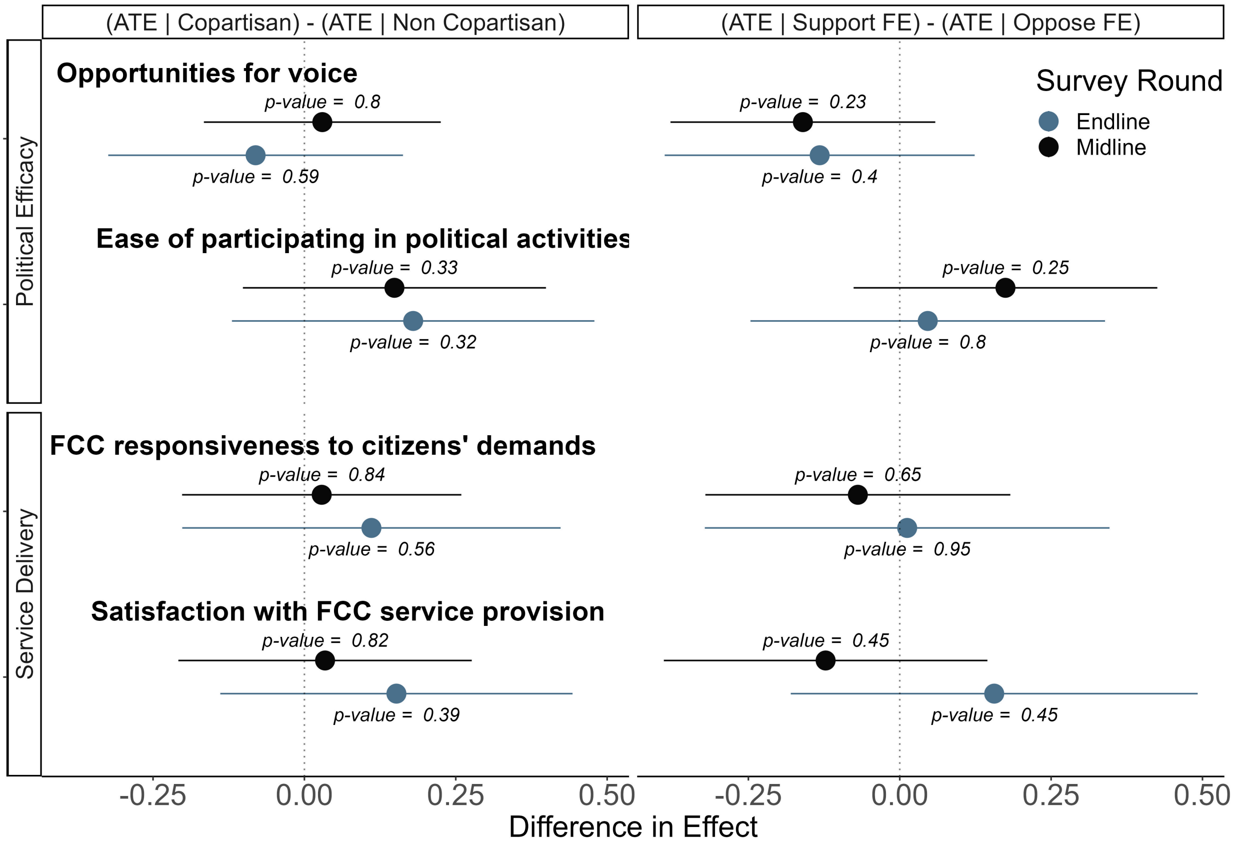


Note: Panels A and B present votes for each service, broken out by attitudes towards fiscal exchange and partisan affiliation, respectively. For both variables, there is little indication that votes for services differ meaningful by group.

5.4.3 Baseline beliefs moderate effects on mechanism outcomes

In our framework citizens are more willing to pay taxes when they believe they will get more of the services they want. We predicted the intervention would increase owners' expected benefits from taxation and we find evidence for this prediction in the form of positive average treatment effects on legitimacy outcomes (Table 3). It could be that attitudes towards fiscal exchange and partisan affiliation moderate the intervention's effect on these mechanism variables and potentially swamp the average effect on compliance. However, when we break out treatment effects on mechanism outcomes by baseline attitudes towards fiscal exchange and political affiliation we do not find evidence that subgroups have different treatment effects. Relatedly, fiscal exchange attitudes and partisanship are not correlated with participants' reported satisfaction with delivered services (Appendix Table 12).

Figure 5: Differences in Treatment Effects by Group



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

6 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 particularly 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. 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.

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

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

What does this imply for governments considering similar participatory interventions? One might conclude, focusing narrowly on average compliance effects, that these interventions have little value. However, we caution against jumping too quickly to that conclusion. There are several reasons why the total treatment effects of a participatory budgeting program may be different, and more encouraging, when implemented by governments.

First, while we do not find any average effect on compliance of participating in the intervention, it is possible that the *total* effects of the intervention are positive because the intervention's impact spills over to people who did not directly participate in the program but learn about it from those who did. Indeed, in Appendix E we provide some suggestive evidence for positive spillover effects. Second, the medium or long-term compliance impacts of participatory interventions pay differ from their short run impacts. While we report null short-term effects on compliance, given the significant increases in the perceived legitimacy of government one might imagine that long-term impacts are more encouraging. Third, there is the question of participant selection. In an effort to rigorously estimate population average treatment effects, we randomly sampled property owners into our intervention. By contrast, participants often self-select into real world participatory programs. Given the large treatment effect heterogeneity that we document, self-selection may produce much different average treatment effects. Future research should explore who opts-in to participatory interventions; estimating treatment effects on that sub-group may be more policy relevant than the quantity we estimate in this study.

Finally, governments considering implementing similar interventions may care about outcomes other than compliance. We find large, durable treatment effects on a suit of legitimacy indicators. In doing so, participatory interventions may create more politically supportive environments for governments who want to carry major reform, such as making investments in fiscal capacity. Future research could investigate if similar interventions generate political capital for politicians to carry out other reform.

References

Alm, J., Jackson, B. R., and McKee, M. (1993). Fiscal exchange, collective decision institutions, and tax compliance. *Journal of Economic Behavior & Organization*, 22(3):285–303.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association*, 103(484):1481–1495.

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

Barber, M. and Pope, J. C. (2019). Does party trump ideology? disentangling party and ideology in america. *American Political Science Review*, 113(1):38–54.

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

Besley, T. (2020). State capacity, reciprocity, and the social contract. *Econometrica*, 88(4):1307–1335.

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.

Bolsen, T., Druckman, J. N., and Cook, F. L. (2014). The influence of partisan motivated reasoning on public opinion. *Political Behavior*, 36:235–262.

Boulianne, S. (2019). Building faith in democracy: Deliberative events, political trust and efficacy. *Political Studies*, 67(1):4–30.

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.

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.

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.

Druckman, J. N. and McGrath, M. C. (2019). The evidence for motivated reasoning in climate change preference formation. *Nature Climate Change*, 9(2):111–119.

Esterling, K. M., Neblo, M. A., and Lazer, D. M. (2011). Means, motive, and opportunity in becoming informed about politics: A deliberative field experiment with members of congress and their constituents. *Public Opinion Quarterly*, 75(3):483–503.

Flores, A., Cole, J. C., Dickert, S., Eom, K., Jiga-Boy, G. M., Kogut, T., Loria, R., Mayorga, M., Pedersen, E. J., Pereira, B., et al. (2022). Politicians polarize and experts depolarize public support for covid-19 management policies across countries. *Proceedings of the National Academy of Sciences*, 119(3):e2117543119.

Fujiwara, T. and Wantchekon, L. (2013). Can informed public deliberation overcome clientelism? experimental evidence from benin. *American Economic Journal: Applied Economics*, 5(4):241–55.

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.

Gonçalves, S. (2014). The effects of participatory budgeting on municipal expenditures and infant mortality in brazil. *World Development*, 53:94–110.

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

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.

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.

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.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological bulletin*, 108(3):480.

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

Levi, M. (1988). *Of rule and revenue*. University of California Press.

Levi, M. (1997). *Consent, dissent, and patriotism*. Cambridge University Press.

Levi, M., Sacks, A., and Tyler, T. (2009). Conceptualizing legitimacy, measuring legitimating beliefs. *American behavioral scientist*, 53(3):354–375.

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

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

López-Moctezuma, G., Wantchekon, L., Rubenson, D., Fujiwara, T., and Pe Lero, C. (2022). Policy deliberation and voter persuasion: Experimental evidence from an election in the philippines. *American Journal of Political Science*, 66(1):59–74.

Luttmer, E. F. and Singhal, M. (2014). Tax morale. *Journal of economic perspectives*, 28(4):149–168.

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

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.

Mullinix, K. J. (2016). Partisanship and preference formation: Competing motivations, elite polarization, and issue importance. *Political Behavior*, 38:383–411.

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. (2021). Why the backfire effect does not explain the durability of political misperceptions. *Proceedings of the National Academy of Sciences*, 118(15):e1912440117.

Nyhan, B. and Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2):303–330.

Okunogbe, O. and Santoro, F. (2023). The promise and limitations of information technology for tax mobilization. *The World Bank Research Observer*, 38(2):295–324.

Okunogbe, O. and Tourek, G. (2024). How can lower-income countries collect more taxes? the role of technology, tax agents, and politics. *Journal of Economic Perspectives*, 38(1):81–106.

Olson, M. (1993). Dictatorship, democracy, and development. *American political science review*, 87(3):567–576.

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. (2023). Unpacking tax morale: Distinguishing between conditional and unconditional views of tax compliance. Technical report.

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.

Slothuus, R. and De Vreese, C. H. (2010). Political parties, motivated reasoning, and issue framing effects. *The Journal of politics*, 72(3):630–645.

Stanley, M. L., Henne, P., Yang, B. W., and De Brigard, F. (2020). Resistance to position change, motivated reasoning, and polarization. *Political Behavior*, 42:891–913.

Sutter, M., Haigner, S., and Kocher, M. G. (2010). Choosing the carrot or the stick? endogenous institutional choice in social dilemma situations. *The Review of Economic Studies*, 77(4):1540–1566.

Taber, C. S. and Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American journal of political science*, 50(3):755–769.

Tappin, B. M., Berinsky, A. J., and Rand, D. G. (2023). Partisans' receptivity to persuasive messaging is undiminished by countervailing party leader cues. *Nature Human Behaviour*, 7(4):568–582.

Timmons, J. F. (2005). The fiscal contract: States, taxes, and public services. *World Politics*, 57(4):530–567.

Torgler, B. (2005). Tax morale and direct democracy. *European Journal of Political Economy*, 21(2):525–531.

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

Weigel, J. L. (2020). The participation dividend of taxation: How citizens in congo engage more with the state when it tries to tax them. *The Quarterly Journal of Economics*, 135(4):1849–1903.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, 13:75–98.

Appendices

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.

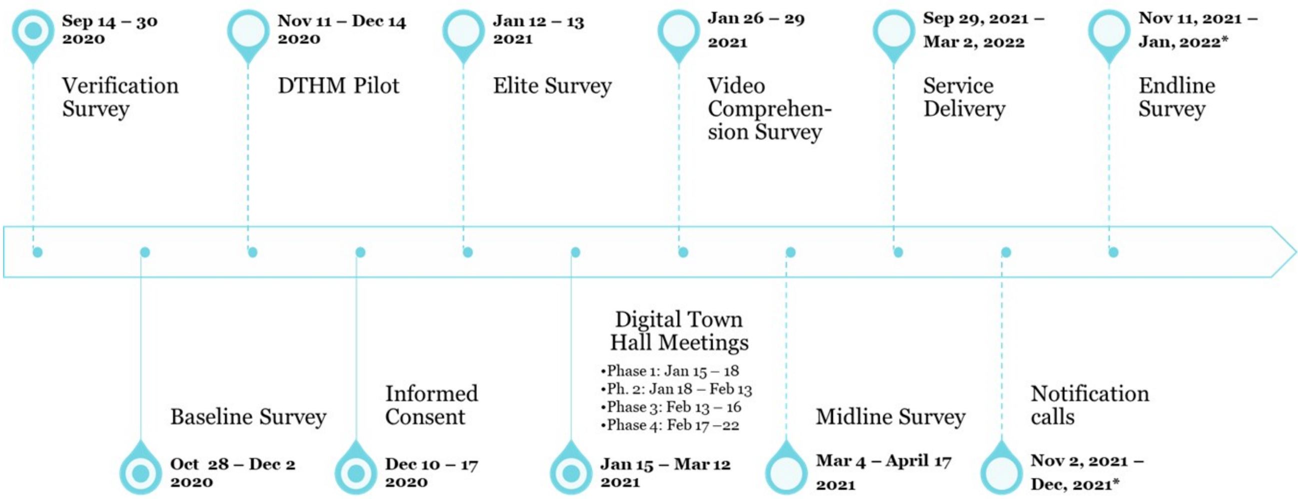
Second, perhaps counter-intuitively, we argue that DTHs hold more deliberative promise: In the Habermasian ideal type of deliberative democracy, participants engage in potentially endless communicative action (an exchange of reasoned arguments) as equals until the best argument prevails (Habermas 1975). 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 Project Timeline

Figure 6: Project timeline



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

B.2 Project Pictures

Figure 7: Project implemented in Ward 418.

B.3 Participation and Experience in DTH

Participation frequency	percent
Never	0.050
Once	0.053
Once per week	0.053
Two or three times per week	0.227
Four to six times per week	0.080
Daily	0.536

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

Table 6: DTH participation

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

Table 7: Voting and messaging behavior

	Perceived responsible actor				
activity	fcc	gov	researchers	citizens	other
organized	0.893	0.019	0.126	0.002	0.014
implement	0.961	0.045	0.018	0.017	0.006
fund	0.842	0.106	0.023	0.115	0.056

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 8: Organization, Implementation, Funding

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 9: DTH experience

C Treatment Assignment Appendix

C.1 Matching

We match property owners using the following covariates:

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

We generated matched-pairs using the *blockTools* package in *R*. We use the Optimal Greedy ("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 predictor to tax compliance, in line with the common use of this variable as proxy for tax compliance behavior. We place equal weight on another set of six measures from our baseline survey. Three of these measures are important factors in the literature on tax compliance: (i) service conditional tax morale, (ii) perceived likelihood of punishment for non-compliance, and (iii) satisfaction with FCC service provision. We also place equal weight on the (iv) gender of the property owner, (v) their awareness and support of the property tax reform,[54] and (vi) the number of these five variables that were imputed.[55]

[54]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.

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

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 10: Summary statistics of matching variables

Table 10 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.[56] The last column displays the number of observations that were imputed for matching. Note that in general, the number of imputed responses is low.

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

C.2 Treatment Assignment Map

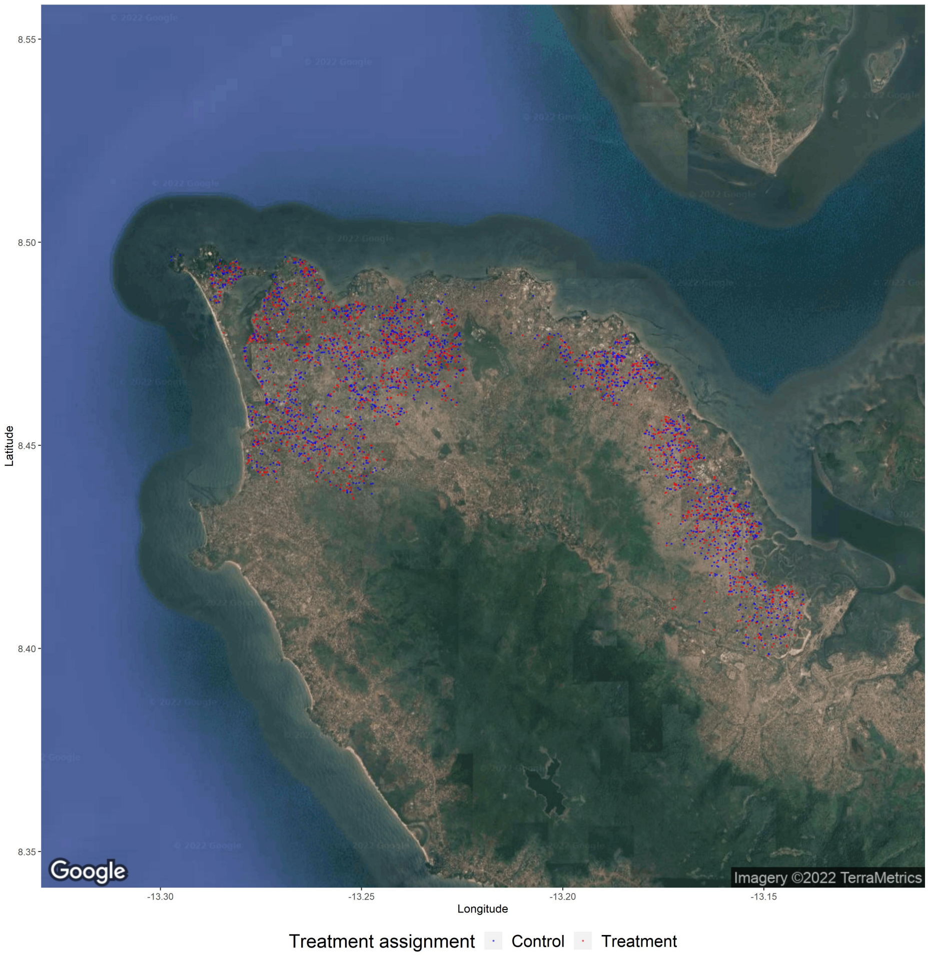


Figure 8: Digital Town Hall treatment assignment in Freetown (red = treatment)

C.3 Distance to Closest Study Property

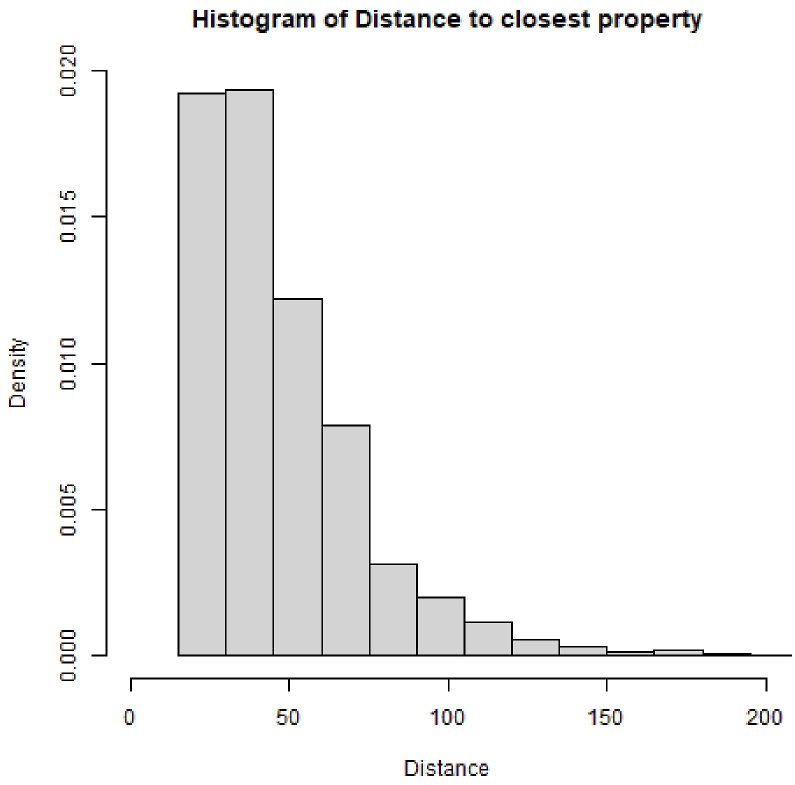


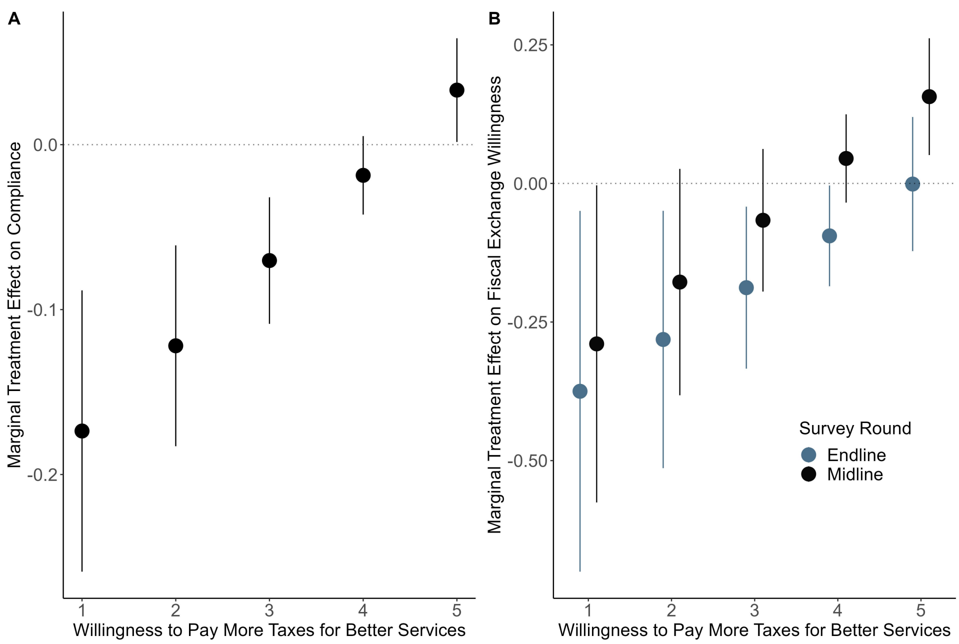
Figure 9: Histogram of minimum distance (in meters) between study properties

D Additional Tables and Figures

D.1 Conditional Treatment Effect

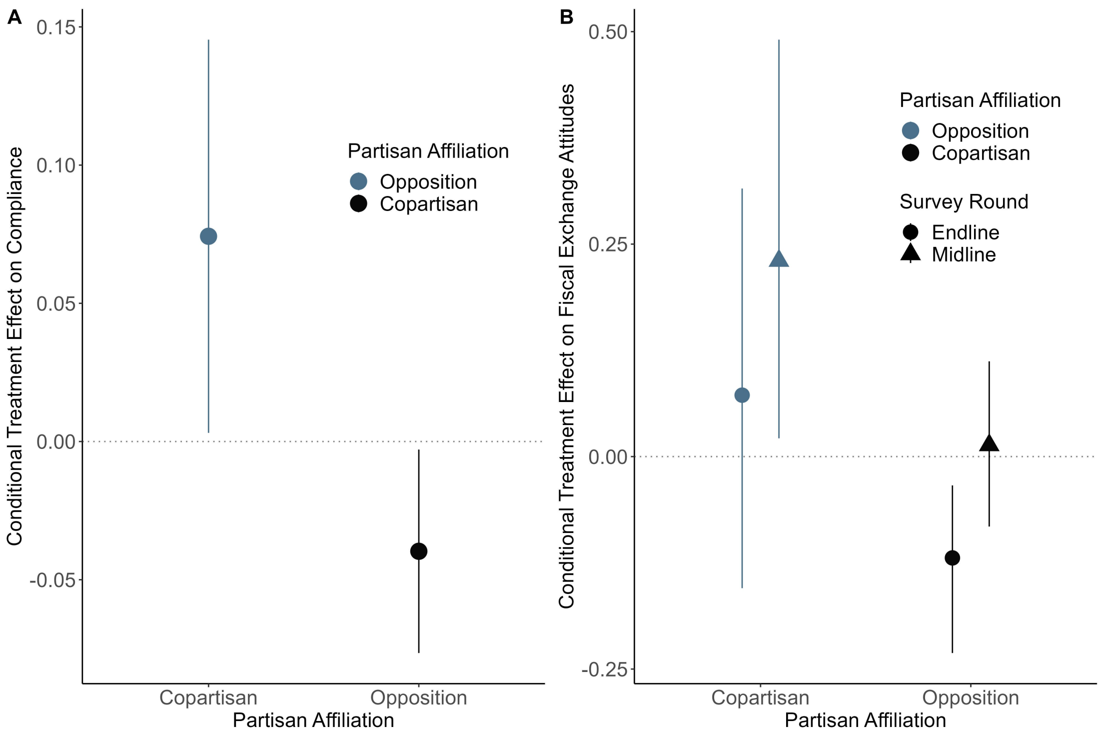
Plot B (Figure 10), presents predicted marginal effects on attitudes towards fiscal exchange from a models that interacts treatment with our five-point measure of (baseline) attitudes towards fiscal exchange, using both midline and endline data. The interaction term is statistically significant (p-value = 0.034; $\beta = 0.11$ SDUs) in the model using midline data and at the threshold of conventional levels of statistical significance when using endline data (p-value = 0.12; $\beta = 0.093$ SDUs). In line with the expectations of motivated reasoning, treated property owners who were initially opposed to fiscal exchange *do not* positively update their attitudes towards fiscal exchange. In fact, in both midline and endline surveys, we find evidence that the intervention hardens the opposition of those initially opposed to fiscal exchange. By contrast, treated respondents who support fiscal become *more* supportive in the mid than their peers in the control group. While this effect disappears at endline, marginal effects of those supportive of fiscal exchange are larger than for those who are not supportive. Note that detecting positive treatment effects for the highly supportive group is difficult due to ceiling effects: for respondents who are fully supportive at baseline, positive treatment effects can only be observed by treated respondents being more likely to *remain* fully supportive.

Figure 10: Marginal Treatment Effects by Attitudes Towards Fiscal Exchange



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

Figure 11: 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."

D.2 Political Engagement

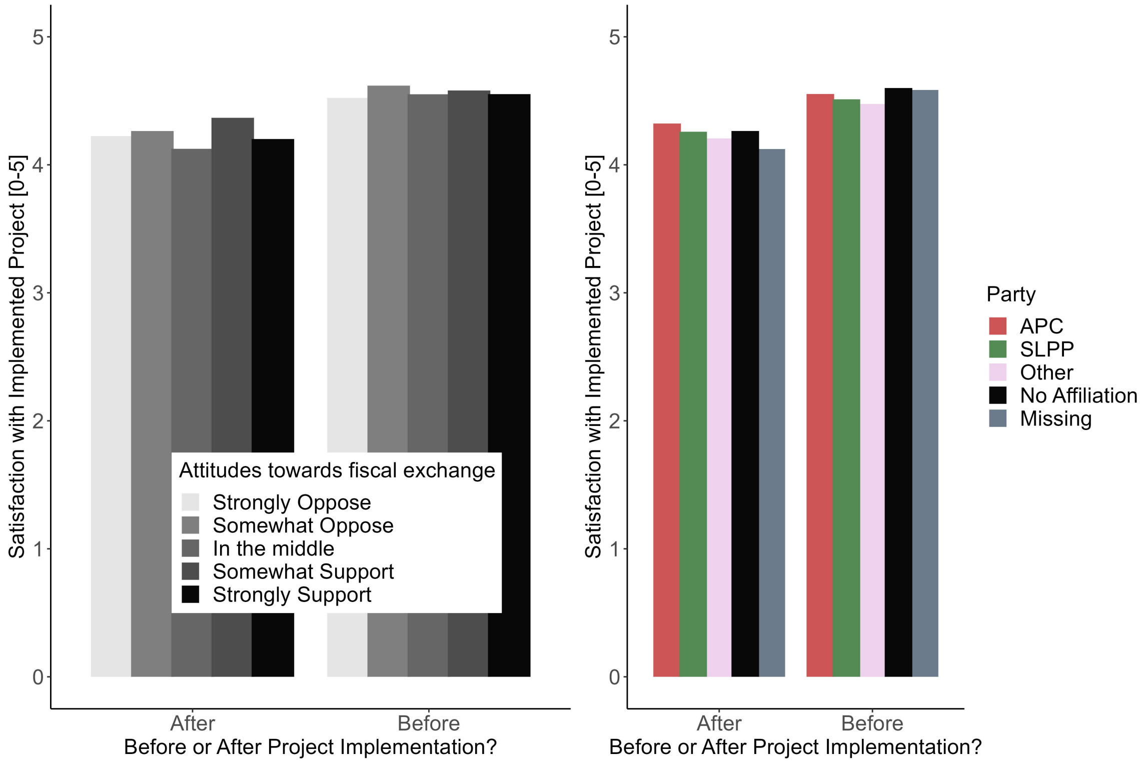
	Baseline	Midline				Endline			
Outcome	Mean	Mean	Effect	N	q-val	Mean	Effect	N	q-val
Knows ward councillor name	0.360 (0.480)	0.383 (0.486)	0.104** (0.040)	3,618	0.007	0.313 (0.463)	0.044 (0.039)	3,618	0.283
Attempted to contact ward councillor	0.193 (0.395)	0.188 (0.391)	0.214*** (0.044)	3,299	0.001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

Table 11 reports the effect of the treatment on political engagement measures. Columns 1, 2, and 6 report the control group mean for each indicator for the baseline, midline, and endline surveys, respectively, with the standard deviation in parentheses. Column 3 presents treatment effects estimates at the midline survey and Column 6 presents treatment effects estimates at the endline survey. Columns 4 and 8 reports the number of non-missing observations in the midline survey and endline survey, respectively. Stars refer to randomization inference p-values. Columns 5 and 9 report corrected q-values, which adjust for multiple hypothesis testing, following Anderson (2008). Reported effects are standardized effects. Attempts to contact MP or Councillor, or attendance at political meeting, are for last six months. $^{*}p < 0.10$; $^{**}p < 0.05$ $^{***}p < 0.01$

Table 11: Political Engagement

D.3 Service Satisfaction by Group

Figure 12: Service Satisfaction by Group



E Spillover

We use a design-based strategy to estimate spillovers that occur due to geographic proximity between properties. For this analysis, we focus on tax compliance spillovers from treated properties to 74,352 properties outside of our study.[57] Our approach compares non-study properties geographically proximate to treated study properties to non-study properties proximate to control study properties.[58] We estimate spillovers with the following equation:

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

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 5, 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.[59] 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.[60]

[57] 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.

[58] 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.

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

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

The top panel A in Table 12 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).[61] 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.

[61] 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.

Threshold Distance	Effect on Compliance 2022	'22 or '23	N observations 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 12 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 12: Spillover Effects