Building Fiscal Capacity with Traditional Political Institutions: Experimental and Qualitative Evidence from Sierra Leone*

Kevin Grieco†

March 25, 2025

Abstract

I argue that governments in weak states can build fiscal capacity by collaborating with non-state, traditional political institutions (TPIs). To study the impact of collaboration, I partnered with the local government in Kono District, Sierra Leone (the KDC), and embedded an experiment within their awareness campaign for a new rural property tax. Property owners in 118 villages were shown videos with varying content. Those in the treatment group viewed an additional segment where their paramount chief discussed the collaboration between the chiefdom government and the KDC in the tax effort. Priming collaboration significantly increased tax compliance and strengthened property owners' belief in their obligation to pay taxes. To assess mechanisms, I developed additional video segments where paramount chiefs emphasized either their coercive capacity or their accountability to constituents. The experimental findings, reinforced by qualitative evidence from 300 interviews, demonstrate that both coercion and accountability are crucial sources of TPIs' authority.

*This study received IRB approval from UCLA (IRB #21-000453) and in Sierra Leone (approved on 21 April 2021). The project received generous funding for data collection from the International Centre for Tax and Development, with additional funding for scoping research provided by the International Growth Centre. I am grateful for feedback from Kate Baldwin, Graeme Blair, Darin Christensen, Cesi Cruz, Barbara Geddes, Ellen Lust, Stephen Monroe, Colette Nyirakamana, Daniel Posner, Wilson Prichard, Sridhar Telidevara, Frank Wyer, and Kerem Yildirim. I would like to thank participants at the Politics of Order and Development Lab and the COMPASS workshop at UCLA. Seminar participants at the 6^{th} GLD Conference, the Sierra Leone Tax for Development Conference, the GLD Work-in-Progress session, WGAPE 2024 (Accra), and the APCG Online Colloquium provided helpful comments and suggestions. I gratefully acknowledge the hardworking enumeration team for data collection. Tamba Moitenga, Henry Mondeh, and Aiah Morsay provided excellent research assistance throughout this project. Of course, this project would not have been possible without the support and collaboration of the Kono District Council, the Kono Council of Paramount Chiefs, and the Knowledge for Community Empowerment Organization. The pre-analysis plan for this project is registered at osf.io/g7xwv.

†Postdoctoral Fellow, International Centre for Tax and Development. e: kgrieco@ucla.edu

Tax collection provides the resources necessary to carry out government activities, including the provision of basic public goods (Schumpeter 1918). The state's capacity to collect taxes contributes to both political order and economic development (Besley and Persson 2011). However, many governments lack the capacity to secure citizens' compliance with their tax demands or their policies more broadly (Lee and Gordon 2005; Migdal 1988).[1] As a result, weak states can become trapped in a pernicious, low-capacity equilibrium: they lack the capacity to collect taxes and, therefore lack the tax revenues to invest in greater capacity.

How can weak states escape this trap? In this paper, I investigate a contemporary case of collaboration between a weak state and local intermediaries in Sierra Leone. Specifically, I examine whether partnering with traditional political institutions (TPIs) can enhance the state's fiscal capacity by increasing citizens' compliance with government taxes. TPIs coexist alongside governments throughout the developing world (Holzinger et al. 2016), governing important parts of day-to-day life (Baldwin 2016; Baldwin and Raffler 2019). Indeed, Baldwin and Holzinger (2019) estimate that 83% of the population of sub-Saharan Africa is governed, at least in part, by TPIs.

Throughout history, governments have relied on local intermediaries to secure compliance in peripheral areas where state control was weak. For example, during the Scramble for Africa colonial powers claimed vast inland territories (Robinson and Gallagher 1961), but were often unwilling to invest in the infrastructure and human capital needed for direct administration (Herbst 2014; Michalopoulos and Papaioannou 2020). Unable to implement state policies directly, "African intermediaries had to be called upon" (Curtin et al. 1995, pg. 425) to accomplish key objectives of the colonial state, such as suppressing the slave trade, securing trade routes, and collecting taxes (Crowder 1968; Migdal 1988).

Governing through intermediaries, often referred to as "indirect rule," has a long history and encompasses a variety of practices (Naseemullah and Staniland 2016). Although this method is perhaps most frequently associated British colonialism (Crowder 1964), it has been employed by various states throughout history. For example, the Mughal Empire governed Bengal through

[1] This is especially true in sub-Saharan Africa, where state capacity has been lower than in any other region of the world since at least the 1960s, the dawn of African independence (Hanson and Sigman 2021).

local intermediaries (Van Schendel 2009, pg. 49-56), as did the Sokoto Caliphate in what is now northern Nigeria (Paden 1970). Moreover, governing through intermediaries has "persisted and continue[s] to structure contemporary state-society relations" (Naseemullah and Staniland 2016). Recent qualitative research highlights ongoing collaborations between weak states and TPIs in contemporary sub-Sahara Africa (Buur and Kyed 2007).

Although research suggests that TPIs' can influence certain behaviors of their constituents (e.g., Baldwin 2013), it is unclear whether this influence can be harnessed to increase citizens' compliance with state policy. In fact, there are several reasons why collaboration with TPIs might fail to achieve this. First, indirect rule has often enabled intermediaries to exploit local populations (Mamdani 1996). If constituents fear that contemporary collaboration will be similarly exploitative, partnering with TPIs may erode citizens' support for the state, ultimately reducing compliance with state policy. Second, by collaborating with TPIs, the state may signal to citizens that it cannot achieve its goals independently, which could lower their perceptions of state capacity and undermine compliance. Third, recent research suggests that TPIs' authority over their constituents' behavior is limited to activities within chiefs' geographic jurisdiction and area of expertise (Baldwin et al. 2023). If citizens believe a state policy—in this case, tax policy—falls outside these domains, they may be less willing to follow traditional leaders' directives.[2]

To study the impact of collaboration and the sources of TPIs' authority, I partnered with the local government in Kono District, Sierra Leone (KDC) during a recent tax reform. This reform aimed to systematically collect taxes in rural villages for the first time since the post-war reintroduction of district councils in 2004. To achieve this, the KDC enlisted the support of the district's 14 paramount chiefs, who are leaders of non-state TPIs. I embedded an experiment in the KDC's campaign to collect property taxes. During an awareness campaign that preceded tax collection, property owners were shown videos that varied in content, particularly whether and how their local paramount chief characterized his involvement in tax collection.[3] All respondents watched a video that began with a local government official explaining the property tax, which serves as the

[2]I use the terms "traditional leaders" and "chiefs" interchangeably to refer to the political leaders of TPIs.

[3]In Kono, where the study takes place, the 14 highest level chiefs (paramount chiefs) are all men.

control condition. I then randomly assigned some respondents to view a follow-up segment where their paramount chief discussed the collaboration between the chiefdom government and the KDC. Video versions were randomized among 1,752 property owners in 118 villages across five chiefdoms. I measured tax compliance with two survey-based indicators that capture propensity to pay and a behavioral game that captures voluntary compliance.

Using a preregistered specification, I estimated the effects of informing property owners about their TPI's collaboration in the tax effort. I find this significantly increases a preregistered index of tax compliance. The magnitude of this effect translates roughly to a three percentage point increase in property owners' propensity to pay property tax. Collaboration also increases property owners' belief that they *should* pay taxes (i.e., "tax morale").

How do traditional leaders obtain compliance from their constituents? When state officials collaborate with TPIs, they aim to leverage the authority of traditional leaders over their constituents. Some scholarly accounts suggest that TPIs use coercion to influence constituents' behavior, such as imposing fines on those who disobey directives or threatening biased decisions in land allocation or dispute resolution (Goldstein and Udry 2008; Mokuwa et al. 2011). Conversely, other accounts depict TPIs as inclusive and accountable political institutions whose leaders advance their constituents' interests (Baldwin and Holzinger 2019; Baldwin 2016). In this view, constituents expect to benefit from directives issued by chiefs and (quasi)-voluntarily comply (Levi 1988, 1997).

To tease out the mechanisms behind TPIs' authority, I created additional video versions where the paramount chief made statements that primed either their coercive capacity or their accountability to constituents. To prime coercion, the paramount chief stated that he would discuss with other chiefs "what to do" with noncompliers and emphasized that chiefs "will not be merciful." To prime accountability, the paramount chief first mentioned that he would hold a meeting with all his sub-chiefs to "discuss and map out" how the tax revenue would be spent, underscoring the inclusive and transparent decision-making often associated with TPIs. He then acknowledged that if tax revenue was not used for development, constituents "will be annoyed," reflecting research that suggests leaders are held accountable through the threat of sanctions from their constituents.

The coercion video increases tax compliance by 0.13 SDUs compared to control (p-value < 0.001), which is twice the effect of the collaboration video. To isolate the impact of the chief's coercion prime, I compare the effect of the coercion video to the effect of the collaboration video. This difference is statistically significant at the 90% level (p-value = 0.09) and provides support for the argument that coercive capacity is a key source of TPIs' authority. To dig deeper into mechanisms, I collected 300 semi-structured interviews across 29 villages in four of the five chiefdoms targeted by the awareness campaign. These interviews provide detailed insights into TPIs' enforcement mechanisms and offer additional evidence supporting the coercion hypothesis. A majority of respondents reported at least one strategy that traditional leaders use to monitor (68%) and punish (78%) individuals who try to evade taxes.

Turning to the accountability hypothesis, I find that the accountability video increases tax compliance by 0.124 SDUs, relative to control (p-value < 0.01). While this effect is larger than the effect of T1 by 0.056 SDUs, the difference is not statistically significant (p-value = 0.18). Further investigation of preregistered secondary outcomes yields similarly suggestive but inconclusive results. Given these inconclusive experimental results, the qualitative findings are particularly instructive. The qualitative data suggest local law-making is participatory and inclusive: nearly all respondents (97%) report that local laws are developed and enacted in open meetings, rather than behind closed doors, and 94% indicate that village representatives are invited to attend chiefdom byelaw meetings held outside their village. These meetings are described as forums for discussion between citizens and their leaders, rather than simply opportunities for leaders to inform citizens about decision. Moreover, qualitative data provides indirect evidence that leaders are held accountable. Although chiefs can call for mandatory labor contributions for any reason, over 80% of respondents report that this "communal labor" in their town is devoted exclusively to public projects, with only 7% indicating that it is occasionally used for projects that do not benefit the public. Finally, in an exploratory analysis where I combine experimental and qualitative data, I find that the effect of collaboration is stronger where chiefs are perceived as more accountable. I interpret this bundle of evidence—derived from a combination of preregistered hypothesis testing,

qualitative data collection, and exploratory heterogeneous effects analyses—as suggestive support for the accountability hypothesis.

This paper makes two central contributions. First, this paper contributes to the literature on fiscal capacity, building on recent work that examines whether state bureaucracies should incorporate or engage with *non-state* actors to improve tax collection and fiscal capacity. In finding that non-state actors can be effectively engaged by the state for tax collection, my results are similar to Balán et al. (2022). This paper differs from their work by focusing on a set of distinctly political non-state institutions that impact citizens' compliance behavior through different mechanisms. Whereas Balán et al. (2022) find that local elites collect more revenue because they have better *information* than state agents, I argue TPIs possess greater coercive capacity and, in some cases, are perceived as more accountable. My findings are more optimistic about collaboration between state and non-state actors than Gottlieb et al. (2024), who find that, "tax appeals by trusted social intermediaries *on behalf of the state* have no impact on citizens' compliance." More generally, this study adds to a growing body of policy experiments that explore whether collaboration with non-state actors can enhance state functions, such as targeting beneficiaries for social assistance programs (Basurto et al. 2020), distributing development aid (Carlson and Seim 2020), and implementing and coordinating development projects (Casey et al. 2018).

Second, this paper improves our understanding of how TPIs exercise their authority. Directives or endorsements from traditional leaders can influence constituents' behavior on important outcomes such as vote choice (Brierley and Ofosu 2023; De Kadt and Larreguy 2018; Kramon 2019), contributions to public projects (Baldwin et al. 2023), and compliance with health regulations (Kao et al. 2021; Van der Windt and Voors 2020). However, few studies have carefully considered the mechanisms through which traditional leaders exert this influence. Where previous work has investigated the sources of TPIs' authority, it has found little support for the coercion mechanism (e.g., Brierley and Ofosu 2023; Baldwin et al. 2023) and reaches contradictory conclusions about the accountability mechanism (Baldwin and Mvukiyehe 2015; Baldwin et al. 2022; Brierley and Ofosu 2023). Indeed, a recent review of the literature on TPIs calls for scholars to

more precisely unpack how "key characteristics of traditional authorities affect governance and politics" (Baldwin 2025).

I contribute to the literature by providing evidence that TPIs obtain citizens' compliance through a combination of coercion and accountability. My findings further suggest that coercion and accountability are complementary aspects of chiefs' authority, rather than substitutes. Existing theories propose that chiefs derive their authority from *either* coercion or accountability. Some argue that despotic, unaccountable chiefs rely solely on coercion to assert their power (Mamdani 1996; Richards 1996; Ntsebeza 2005). In other accounts, coercion is unnecessary because accountable chiefs and citizens share the goal of community development (Baldwin 2013, 2016), with aligned incentives leading to citizens' compliance. However, my findings indicate that neither mechanism alone is sufficient to shape citizens' behavior. On the one hand, coercion alone fails because unaccountable chiefs who abuse their coercive power lose favor with citizens and struggle to secure compliance. On the other hand, accountability alone is also inadequate, as chiefs often need to direct individuals to participate in collective actions that produce public goods. In these cases, coercion is necessary because aligned incentives alone are unlikely to overcome the freerider problem and ensure compliance.

The Sources of TPIs' Authority: Coercion and Accountability

Arguments about the source of TPIs' authority can be divided into two bins. A first strand of the literature focuses TPIs' coercive capacity. In these accounts, citizens comply with chiefs' directives because noncompliance can be detected and punished. Indeed, recent survey research document that citizens expect traditional leaders to sanction noncompliance. On the Ghana-Togo borderland, Wilfahrt and Letsa (2023) find that 85% of respondents would be worried about punishment if they failed to follow village chiefs' directives. Surveys conducted in Malawi, Kenya, and Zambia reveal that 43% of respondents are motivated to make contributions (e.g., money and labor) to community water and sanitation projects because they fear being fined by chiefs (Lust et al. 2023).

Another version of the coercion hypothesis suggests that traditional leaders influence constituents' behavior by exploiting their unchecked authority within local governance systems. Mam-

dani (1996) influentially argues that colonial governments undermined existing accountability mechanisms within indigenous political institutions and enabled traditional leaders to become "decentralized despots" (see Richards 1996, 2005, for similar accounts in Sierra Leone). Citizens comply with TPIs' directives out of fear that noncompliance will result in biased governance decisions, particularly in areas like dispute resolution or land allocation. Supporting this view, Ntsebeza (2005) argues in his study of TPIs in South Africa that "control of the land allocation process [is] central to. . . how [traditional leaders] derive their authority" (pg. 295). This perspective aligns with political economy research that documents elite control of land rights. For example, Goldstein and Udry (2008) show that in Ghana, individuals who hold an office of social or political power have more secure property rights. Similarly in Sierra Leone, Acemoglu et al. (2014) find that less accountable chiefs "have more authority to influence whether or not people can farm or sell a piece of land" (pg. 323). Related research highlights similar biases in local court systems (Mokuwa et al. 2011). Given that TPIs influence land allocation and dispute resolution across much of sub-Saharan Africa (Baldwin and Raffler 2019), governance biases may serve as a general mechanism of coercive social control.

Despite the prominence of the coercion hypothesis, recent studies find no evidence that TPIs influence their constituents' behavior through coercion (Brierley and Ofosu 2023; Baldwin et al. 2023). Additionally, while TPIs need to detect noncompliance in order to impose penalties, Kao et al. (2021) find that TPIs are not more effective at detecting noncompliance than state officials.

A second strand in the literature portrays TPIs as accountable political institutions whose leaders advance their constituents' interests. Several mechanisms facilitate accountability. First, TPIs are highly participatory, featuring elements of direct democracy and transparent decision-making processes (Baldwin and Holzinger 2019; Skalník 1996). Because key decisions are made with public input, citizens have a benchmark for evaluating leaders' performance and can hold them accountable if they deviate from expectations. Second, traditional leaders may be held accountable through their social proximity to the populations they govern. According to Baldwin (2016), unelected but socially embedded chiefs are motivated to act in the community's interest because

they must promote community development to extract rents and because they face social sanctions if they perform poorly. Third, some chiefs—especially those at lower levels of administration—face elections that may discipline leaders to act in the community's interest (Acemoglu et al. 2014; Barro 1973; Baldwin and Holzinger 2019). These accountability mechanisms may encourage leaders to perform well and therefore foster quasi-voluntary compliance from citizens (Levi 1988, 1997). Additionally, public participation in political affairs can legitimate leaders (Pateman 1970) and thus "bolster willing obedience" (Levi et al. 2009; also see Bó et al. 2010).

The existing literature does not directly test whether TPIs' accountability mechanisms lead to increased citizen compliance with chiefs' directives. The few studies that partially address this question provide mixed evidence. On one hand, Baldwin et al. (2022) find that expanding participatory decision-making processes in TPIs enhances governance performance, consistent with the accountability hypothesis. However, they do not investigate whether this improvement translates into greater citizen compliance. Similarly, Brierley and Ofosu (2023) find that in Ghana, citizens believe their chiefs consider community interests when endorsing political candidates, but the study does not show that this belief drives voting behavior and does not explain *why* citizens hold this belief. On the other hand, contrary to the accountability hypothesis, Acemoglu et al. (2014) find in their study of TPIs in Sierra Leone that increased accountability actually *lowers* citizens' compliance with chiefs' directives.Similarly, Baldwin and Mvukiyehe (2015) find that in villages where chiefs are selected through more participatory methods (elections rather than elite selection), individuals contribute less to "collective endeavors," as measured by a public goods game.

In summary, while the literature suggests multiple mechanisms that may drive TPIs' authority, the evidence supporting these mechanisms is limited. Moreover, even if TPIs can influence certain behaviors of their constituents, it remains unclear whether this influence can be leveraged to increase citizens' compliance with state policy. In this paper, I address these questions by testing the following preregistered hypotheses:

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

H2: TPIs use (the threat of) coercion to obtain citizens' compliance with state tax demands.

H3: TPIs' accountability enables them to obtain citizens' compliance with state tax demands.

State Weakness, TPIs and Property Taxation in Sierra Leone

Jeffrey Herbst (2014) argues that the challenge of "project[ing] authority over inhospitable territories that contain relatively low densities of people" has long impeded state-building efforts in Africa. This assessment would ring true for any visitor to Kono District in eastern Sierra Leone, where the median village contains 17 buildings.[4] The potholed, dirt roads that connect the district's sparsely populated villages make administration cumbersome and expensive for government officials: traveling from the median village to the district's capital costs more than the minimum wage (Grieco 2024). In this context the government has certainly struggled to project its authority: villages have no connection to the electrical grid or piped water, schools and health facilities are chronically understaffed and unequipped, and the justice system is largely left in the hands of traditional leaders who enact and enforce laws and resolve disputes.

Perhaps unsurprisingly, the Kono District Council (KDC)—the local government with jurisdiction over the rural parts of the district—collects very little taxes. In-person tax collection is both logistically challenging and resource-intensive. Moreover, since the KDC has a poor track record of delivering services, citizens are likely reluctant to pay. This creates a vicious cycle: without tax revenue the KDC cannot invest in the physical or administrative infrastructure to efficiently demand taxes; cannot invest in the enforcement infrastructure to compel payments; and cannot improve service delivery to encourage citizens compliance. To escape this cycle, KDC initiated a tax reform in 2018, aiming to systematically collect taxes in rural areas for the first time since the post-war reintroduction of local council in 2004 (Zhou 2009). Central to this reform was a property tax that applied to all residential and commercial structures.[5]

Aware of their limited capacity, the KDC partnered with the district's chiefs to collect this

[4]Calculated by author from the 2015 National Census.

[5]Buildings were categorized into tax rate bands based on their size and construction materials, with all structures in the same band taxed at a uniform rate. Tax collectors were assigned mutually exclusive areas of the district and were compensated with a share of the revenue they collected. In 2019, 95% of the building structures in the district had an assessed rate of US $1.50 or less.

newly introduced property tax. Most of Sierra Leone, except for the peninsula that includes the capital city, is divided into 190 chiefdoms. Each chiefdom, though also under the jurisdiction of the state (both central and local governments), is governed by a chiefdom council and a hierarchy of chiefs. The top traditional political authority in each chiefdom is the paramount chief, who is elected for life by the chiefdom's elite, with candidates drawn from a restricted set of ruling families (see Reed and Robinson 2013). Chiefdoms are further subdivided into sections, each containing several villages, with section chiefs and village chiefs overseeing these areas, respectively.

The chiefs agreed to assist the KDC in collecting the new property tax. In return, half of the collected revenue would be earmarked for spending within the chiefdom where it was collected. By April 2018, the KDC had signed a memorandum of understanding with all the district's paramount chiefs, formalizing this revenue-sharing agreement between the KDC and the chiefdom councils. The chiefdom councils were responsible for recruiting and supervising tax collectors, enforcing tax compliance, and overseeing local bank accounts where the revenue was stored.

The tax reform got off to a rocky start, with collected revenue amounting to only about two percent of the potential revenue in 2019. This low revenue was largely due to poor tax compliance: data from 2019 indicated that compliance rates were below 10% in the villages visited by tax collectors. Interviews with tax collectors revealed that property owners were often reluctant to pay the new tax because they were unaware of the chiefs' involvement and support. One tax collector noted that his efforts were successful in some villages because "the traditional authorities passed the message [of TPI involvement] to the people, and the people have respect for the traditional authorities." However, in other villages, collection was less successful because "the message [of collaboration] didn't reach them soon enough."Another tax collector suggested that compliance might improve the following year if "the paramount chief calls a meeting. When the chiefs are more strongly backing [the new tax], that's going to make people pay."

Chiefs and local government officials recognized that the involvement of TPIs had not been effectively communicated to the district's residents. Although plans were made for chiefdom meetings to discuss the tax reform, these meetings never materialized because neither the chiefs nor the

district council could marshal the funding to hold them. To address this lack of awareness before the 2021 tax collection season and improve compliance, I collaborated with the KDC and the district's paramount chiefs to design a tax awareness campaign. The tax awareness video central to this campaign is the intervention under investigation in this study.[6]

Research Design and Data Collection

Working with the KDC and the district's paramount chiefs, I led a door-to-door tax awareness campaign in the summer of 2021. The campaign was centered around a tax awareness video and targeted five chiefdoms.[7] In each village, a team hired through a local civil society organization met with property owners to share tax awareness videos about the new property tax. I embedded an experiment in this campaign, randomly assigning different versions of the tax awareness video to property owners, with each version priming a concept of theoretical interest. To measure outcomes and test hypotheses, enumerators conducted a survey immediately after participants watched the video. Additionally, in villages not visited by the campaign, I collected qualitative data to gain deeper insights into chiefdom governance and the sources of TPIs' authority.

Interventions: Tax Awareness Videos

In March 2021, I worked with local government officials and the district's paramount chiefs to record four distinct video segments, which were combined into different versions of the awareness video. Segments featuring a paramount chief were chiefdom-specific, so while speakers followed a script to ensure message consistency, the precise wording varied across chiefdoms. Below, I use the paramount chief's segments from Gbane Chiefdom as an example.

Segment 1: Local politician provides information about tax collection

- First, he introduces himself: *"Greetings my people! This is your son Solomon Sahr Gbondo who is heading the Kono District Council*."

[6] See Appendix A.1 for more details on the development of this awareness campaign.

[7] We created tax awareness videos in five additional chiefdoms but did not include them in the door-to-door campaign for several reasons. Three were excluded because the video featured the paramount chief's representative instead of the chief himself; the most rural and sparsely populated chiefdom was excluded due to the high expected cost of a door-to-door campaign; and one chiefdom was omitted because the paramount chief's language in the video deviated significantly from the agreed script. See Appendix A.2 for more details on the selection of chiefdoms into the campaign and evaluation.

- Second, he provides information about the tax rates: *". . . Stick house, with local [palm tree leaves] roof: You pay 20,000 Leones . . ."*
- He concludes with an appeal to pay: *"Please, let us pay our taxes in order for us to able to carry out development projects in the district . . ."*

Segment 2: Paramount chief mentions collaboration with local government

- First, the paramount chief introduces himself: *"My Gbane people, I greet you all. This is your paramount chief Aiah Bindi Faefankongor the* 2^{nd}*."*
- Second, he explains the collaboration: *"Gbane Chiefdom Council and Kono District Council are working in unity to collect property taxes . . ."*

Segment 3: Paramount chief primes coercion

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

Segment 4: Paramount chief primes accountability

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

Experimental treatment conditions are different tax awareness videos that combine different video segments and are designed to test different hypotheses (see Table 1). Property owners assigned to the control condition see only the first video segment where the local government official provides information about tax collection. For property owners assigned to the collaboration treatment

condition (T1), the government information segment is followed by the second segment where the chief mentions their collaboration with local government. I expect T1 to increase property owners' perception that TPIs are collaborating with the local government on the property tax, relative to control. I test my first hypothesis by comparing tax compliance outcomes between T1 and C.

In the coercion treatment condition (T2), the paramount chief's statement is expanded to include the third video segment where the chief focuses on punishment for non-compliers. The goal is to prime punitive actions that can be taken against non-compliers. I assess my second hypothesis by comparing tax compliance outcomes between T1 and T2. In the accountability treatment condition (T3) the additional messaging from the paramount chief focuses on accountability, rather than coercion. The goal of T3 is to prime aspects of TPI's accountability that are highlighted in the literature—namely that important decisions will have the input of additional actors beyond the paramount chief and his close inner circle, that revenue spending decisions must be justified in public, and that poor governance will anger constituents.[8] I can address my third hypothesis—TPIs' accountability enables them to obtain citizens' compliance with state tax demands—by comparing tax compliance outcomes between T1 and T3.

Table 1: Summary of Treatment Conditions

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

Randomization and Balance

I randomly assigned property owners to treatment conditions with equal probability using simple randomization. As respondents watched the tax awareness video on the tablets that enumerators used to conduct the survey, I programed the treatment randomization into the tablet-based survey.

[8]The paramount chief's promise to discuss tax revenue allocation is credible, as an MoU mandated that 50% of collected revenue fund local development through the chiefdom council. As its head, the paramount chief is well-positioned to promote citizen inclusion.

Treatment groups are balanced on nearly all immutable covariates, and the number of observed differences is no more than expected (Appendix Table A1 presents balance).[9] While I do observe imbalance on two variables—gender and education—this would only be a concern if these covariates predict my primary outcome of interest, the compliance index. The bivariate relationship between gender and the tax compliance index is small and statistically insignificant. However, the education variable is positively associated with the compliance index. As the control group is more educated than any of the treatment groups, this imbalance could introduce downward bias into the estimates, though the magnitude of that bias is likely to be small.[10] I account for this when estimating treatment effects, as I preregistered the education variable as a control variable in my main specification.

Table A1 also allows us to characterize the sample. The average respondent is about 46 years old, uneducated ($\approx 65\%$), Kono speaking ($\approx 81\%$), male ($\approx 72\%$), and married ($\approx 79\%$). Property owners in my sample do not appear to be wealthy. In rural Kono District wealth is largely held in animal stocks, and the average respondent has animal stocks with a market value of \$225.[11] The sample also captures a mix of elite and non-elite respondents, with roughly a quarter of the sample holding a community position of social or political importance (e.g., chief, mammy queen, religious leader, youth leader). Finally, respondents are primarily engaged in small-scale agriculture. Roughly 56% of the sample works exclusively on their own farm, while an additional third of the sample mixes work on their personal farm with outside employment (e.g., trader or miner).

Survey Data

Between May and June 2021, I worked with a team of 33 enumerators to conduct 1,752 surveys across 118 villages in the five chiefdoms where the door-to-door tax awareness campaign was implemented. With villages in Sierra Leone's 2015 national census as a sampling frame, I used

[9] Given that I run 44 balance tests, we would expect 4.4 tests to appear significant at the 90% confidence level.

[10] According to a bivariate regression (using only control data), moving from no education to some education increases the tax compliance index by 0.18 standard deviations. Respondents in control are seven percentage points more likely to have received some form of education than respondents in T2. Therefore, if left unadjusted, we should expect bias in the order of 0.0125 standard deviations. Regarding the relationship between gender and compliance, the p-value on the regression coefficient of this estimated bivariate relationship is 0.81 (control group data only).

[11] Animal stock value is calculated based on market values in the district capital at the time of data collection.

geographic cluster random sampling to select villages for the study from a set of 434 eligible villages in the five chiefdoms. Enumerators used a random walk procedure to select households for interviews. See Appendix A.2 for more details on sampling.

For the majority of survey questions, respondents were asked to gauge their expectations or perceptions on a ten-point scale. To make this scale more concrete to survey respondents, all enumerators were given ten beans and a plastic plate, which served as a visual aid regarding the ten-point scale (Delavande 2014; see Appendix A.4). Respondents were asked to allocate some, none, or all of the ten beans to the plastic plate to represent their perceptions and expectations. Before entering the main modules of the survey, enumerators guided respondents through several sample questions to familiarize respondents with this scale. The response patterns to these practice questions were encouraging: average responses were low to unlikely events ("chance that the president visits this community tomorrow"), high for highly likely events ("chance that you will drink water this month"), and were in keeping with basic laws of probability (Appendix Table A2).

Tax compliance is measured through two survey questions and a behavioral game, all administered immediately after the tax awareness video. The first survey question, *self-reported propensity to pay*, directly elicits compliance by asking respondents how likely they would be to pay the full tax rate if a tax collector visited their home the next day. A potential limitation of this measure is that respondents may view paying taxes as socially desirable and, as a result, over-report their true compliance propensity.[12] To address this concern, I collected a second survey outcome that indirectly measures compliance by eliciting expectations about others' behavior (Fisher 1993; Haire 1950). The second question, *perceived neighbors' propensity to pay*, asked respondents to estimate the proportion of other property owners in their village they believed would pay the new tax.

A limitation of survey-based compliance measures is that respondents do not face the material consequences of their answers. To address this, my third measure uses a real-money behavioral game to capture the voluntary channels of tax compliance. Specifically, respondents participated in a dictator game where they were given a small sum of money to allocate between themselves and

[12]Social desirability bias would only impact the *levels* of this measure, but would not bias treatment effect estimates.

the local government's property tax fund, referred to as the "house money fund." Each respondent received five 500 Leone coins (each worth approximately $0.05) and was instructed to decide how to split these coins between themselves and the house fund. They were informed that the coins allocated to the house fund would go into the same government bank account as property tax revenue, and that the allocation was entirely at their discretion. The number of coins given to the house fund was recorded by the enumerators. The contributions were anonymous and could not be tracked by political authorities, implying that contributions were purely voluntary.

The preregistered primary outcome of interest is an additive index (compliance index) that comprises these three measures. To construct the summary index of the three compliance measures, I follow Kling et al. (2007) and standardize each sub-indicator relative to the control group and combine them in an equally weighted index that averages across standardized sub-indicators. I impute missing sub-indicators using the group mean.

There are two reasons why I do not use a measure of real tax compliance behavior. First, at the time of the study, the Kono District Council collected paper tax receipts from property owners but lacked administrative records on compliance. The absence of identifying information on the receipts prevented linking tax payment data to individuals in my study. Secondly, even if compliance data from administrative records were available, using it would have introduced a time lag between the treatment (the tax awareness campaign) and measurement, since compliance behavior would only be observable in administrative records months later, after tax collectors visited villages. This time lag would have likely created spillover between experimental units.[13]

One way to assess whether the compliance index is a valid proxy for tax compliance is to evaluate its ability to confirm established relationships between tax compliance and other variables, a process known as nomological validation (Adcock and Collier 2001). Using administrative data from Freetown, Sierra Leone's capital, Appendix Table A4 (Panel A) shows that tax compliance behavior is positively associated with education and wealth, patterns also found in the literature.[14]

[13]Moreover, the KDC ruled out collecting taxes *during* the tax awareness campaign because they wanted to inform property owners about the tax before trying to collect it.

[14]For education see Besley 2020 and D'Arcy 2011; for wealth see Ali et al. 2014.

The tax compliance index confirms these relationships in my study: Panel B (Appendix Table A4) shows that the compliance index is positively associated with measures of education and wealth.

A second way to assess measurement validity is by examining the relationships between the different indicators of tax compliance. In addition to the indicators included in the compliance index, I preregistered a survey-based measure of voluntary compliance (*tax morale*) as a secondary outcome.[15] Since all four indicators aim to measure tax compliance, convergent validation suggests that they should be positively correlated (Adcock and Collier 2001). Appendix Table A3 confirms that the four compliance indicators are positively correlated. However, these indicators capture two distinct compliance constructs: the direct and indirect survey-based measures reflect compliance *propensity*, while the behavioral game and tax morale indicator measure *voluntary* compliance. Appendix Table A3 also supports divergent validation—the expectation that correlations should be stronger between measures of the same construct than between different constructs (Adcock and Collier 2001). Specifically, the behavioral indicator of voluntary compliance (*coins given to KDC's house fund*) is twice as strongly correlated with the survey-based measure of voluntary compliance (*tax morale*) as with either survey-based indicator of compliance propensity.

Estimation

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

$$Y_i = \alpha + \beta_1 T1_i + \beta_2 T2_i + \beta_3 T3_i + \theta \mathbf{X}_i + \gamma C_c + \delta N_k + \epsilon_i \quad (1)$$

where Y_i is the outcome of property owner *i* and T1, T2, and T3 are dummy variables for each treatment condition. Following Lin (2013), $\mathbf{X}_i$ is a set of preregistered control variables, centered and interacted with each treatment condition. Prespecified control variables include: (i) educational level (a dummy indicating whether the respondent received any schooling); (ii) a set of dummy variables for community positions of social or political importance; (iii) expected likelihood of

[15] Respondents were asked to imagine a situation where they would not face fines or penalties for not paying their property tax and then indicate whether they believed it was (morally) right to pay the tax. This tax morale measure is the only preregistered secondary compliance outcome.

travel to the district and country capital; (iv) expected occurrence of an unlikely event; (v) the village level literacy rate; (vi) the percentage of households in the village that own a radio; and (vii) percentage of village residents born in the chiefdom.[16] I also include chiefdom and enumerator fixed effects as C_c and N_k, respectively. ϵ_i is the idiosyncratic error term. As randomization occurs at the level of the observation (respondent), I do not cluster standard errors.

Qualitative Interviews

In the fall of 2022, just over a year after the tax awareness campaign, I worked with a team of six research assistants to conduct interviews with 300 respondents across 29 villages in four of the five study chiefdoms (Gbane, Soa, Lei, and Nimikoro). Interviews focused on key elements of chiefdom governance, such as taxation, mandatory communal labor, law-making and enforcement, and citizens' perceptions of chiefs' performance. To select villages for interviews, I first randomly chose sub-chiefdom administrative units called sections within each chiefdom. From each selected section, I then chose the section headquarter town and one other large town for interviews. The interviews were conducted in Krio, Kono, or a combination of both, based on the respondent's preference. Each interview was recorded and lasted approximately 20 minutes. Interviewers underwent a five-day training workshop before data collection. For more details on the qualitative data collection process, see Appendix D.1.

Does State Collaboration with TPIs Increase Compliance?

Before reporting the impact of the TPI collaboration treatment (T1) on compliance, I first present the results of a manipulation check. T1 attempts to manipulate respondents' perceptions about the collaboration between state actors and TPIs. To measure perceptions of involvement, enumerators presented respondents with a laminated paper divided into four squares, where each square represented one of the four actors: (i) Kono District Council, (ii) TPIs, (iii) the central government, and (iv) NGOs and civil society organizations. Each respondent was then given 10 beans and asked to

[16] I include the respondent's perceived likelihood of an unlikely event—the president visiting the respondent's village on the following day—because it tells us how the respondent is using the 10-point scale and is therefore prognostic (see Appendix A.4). I selected prognostic variables for covariate adjustment using a LASSO model that predicted my outcomes of interest.

allocate the beans across the four squares, placing more beans on actors they thought were more involved in and responsible for the property tax (see Appendix Figure B1).

Table 2 reports the average number of beans allocated to each actor by treatment condition. The collaboration treatment (T1) successfully increased respondents' belief that TPIs were involved in the property tax. Table 2 shows that T1 increases the perceived involvement of TPIs by 0.56 beans, equivalent to 24% of the baseline mean (0.27 SDUs; p-value < 0.001).[17] While treated participants are unambiguously more likely to believe TPIs are involved, control group respondents still perceive chiefs as somewhat involved in the tax. This implies that my study design influences respondents' beliefs about TPIs' involvement at the margin: when interpreting subsequent experimental results, note that treatment effects estimates leverage this marginal, rather than completely distinct, difference in perceptions.[18]

Table 2: T1 Increases Perceived Involvement of TPIs

	Perceived Involvement			
Actor	**Control**	**Treatment**	**Difference**	***p*-value**
TPIs	2.35	2.91	0.56***	0.00
Kono District Council	3.96	3.91	0.05	0.75
Central Govt.	2.83	2.47	-0.36**	0.04
NGOs	0.87	0.72	-0.15*	0.08

Table 2 reports the effect of the collaboration treatment (T1) on citizens' perceived involvement of several actors in the property tax. Respondents allocated ten beans across actors, such that more beans represents greater perceived involvement. Columns 1 and 2 show the number of beans allocated to each actor by the control and treatment group, respectively. Column 3 reports the difference in perceived involvement and Column 4 reports the p-value of that difference. Differences are estimated using OLS with preregistered specifications. The reported treatment group levels are predicted values, rather than raw group levels.
Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3 presents the effect of the TPI collaboration treatment (T1) on tax compliance. Treatment effects are reported in standard deviation units (SDUs). The TPI collaboration treatment

[17]Note that the measurement strategy forces respondents to allocate a finite number of beans. Therefore, an increase for one actor must lead to a decrease for one or more of the other actors. Correspondingly, T1 led to a decrease in the perceived involvement of the central government and NGOs, but not the local government.

[18]In addition to this manipulation check, I conducted several attention checks. Nearly all respondents correctly recalled the number of different speakers in the awareness video (94%) and accurately identified those speakers (93%). Furthermore, respondents can recall theoretically important messages delivered in the video they watched. Details of these attention checks are presented and discussed further in Appendix B.

(T1) increases the compliance index by 0.068 SDUs, relative to the control, a difference that is statistically significant. This effect is primarily driven by changes in self-reported propensity to pay and perceptions of neighbors' likelihood to pay. To better understand the effect size, consider the impact on the *self-reported propensity to pay* indicator. The effect size is 0.09 SDUs, which corresponds to 0.27 beans on the ten-bean scale. Since each bean represents ten percentage points, we can interpret this effect as a 2.7 percentage point increase in the respondent's self-reported propensity to pay the property tax. Interpreted in this way, this effect is similar in magnitude to the effect in Balán et al. (2022), who show that using local elites (city chiefs) as tax collectors increases compliance rates by 3.3 percentage points.

Table 3: Effect of Collaboration (T1) on Compliance

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

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

The last row of Table 3 shows that TPI collaboration (T1) has a large and statistically significant impact on *tax morale*, the respondent's belief that paying the property tax is morally right. The magnitude of this effect corresponds to 0.44 beans on the ten-point scale, an increase of 5.9% over the baseline level. By finding positive, statistically significant treatment effects on both my primary and secondary measures of tax compliance, these results provide strong evidence that in-

dividuals are more willing to comply with the newly introduced property tax when they know their local government is collaborating with TPIs.

Sources of TPIs' Authority: Coercion and Accountability

Why does government collaboration with TPIs increase citizens' compliance with property tax? This section examines two arguments for why TPIs are able to generate citizen compliance: *coercion* and *accountability*. Table 4 reports the effect of the Coercion treatment (T2) (Columns 2-3) and Accountability treatment (T3) (Columns 4-5) on the compliance index.

Table 4: Effects of Mechanism Treatments (T2/T3) on Compliance Outcomes

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

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

Coercion

Relative to control, the effect of the Coercion treatment (T2) on the compliance index is 0.134 SDUs (Column 2; p-value < 0.001), an effect that is twice as large as the increase generated by T1. Comparing the effect of T2 to the effect of T1 allows us to isolate the impact of the coercion prime. This difference is statistically significant at the 90% level (Column 3; p-value = 0.09). The positive effect on the compliance index is primarily driven by an increase in the direct measure of tax compliance, *self-reported propensity to pay tax*. The Coercion treatment has no impact on the

voluntary compliance measure (*coins give to KDC's house fund*), intuitively suggesting that T2 impacts compliance through non-voluntary channels. While T2 increases respondents' tax morale by 0.219 SDUs relative to control (Column 2; p-value < 0.001), T2 does not provide a statistically significant boost to tax morale, relative to the Collaboration treatment (Column 3).

These experimental results are buttressed by evidence from qualitative data that show that TPIs have enforcement mechanisms in place to detect, and punish, noncompliance with a widespread existing poll tax (called the local tax) that is collected by chiefdom authorities (see Appendix Table D1). Traditional leaders commonly used roadblocks (mentioned by 32% of respondents) to monitor compliance with the local tax, erected either inside the village or at key junctions on the road network.[19] Another common monitoring strategy, noted by one village chief, is for authorities to "go house-to-house to check for tax payers" (24%).[20] Respondents also reported that authorities keep records of who has paid (24%). While village chiefs can monitor compliance directly by, for example, making lists of compliant community members,[21] chiefdom authorities can monitor *villages* by tracking the number of tax receipts and associated revenue turned in by a given village.[22] In total, 68% of respondents described at least one strategy that authorities used to monitor compliance with the local tax, at either the village or chiefdom level.

The majority of respondents (55%) report that individuals found to have not paid their local tax will be issued a fine by authorities: "either you buy the tax, or you pay a fine."[23] Other respondents note that non-compliers can be taken to higher authorities (45%), a situation also likely to end with a fine.[24] The most commonly mentioned non-fine form of punishment is for village authorities to prevent noncompliant community members from accessing their farms (9%), thus cutting off a major source of income. Most respondents (78%) believe non-compliant individuals will face some

[19] One respondent explained that chiefdom authorities "erect check points in collaboration with the chiefdom police, especially when the compliance rate is low" (Interview: 405; also see Interview 404).

[20] Interview: 404

[21] Interview: 406

[22] Chiefdom authorities distribute receipt books to village chiefs based on village population. Village chiefs must return completed books with the collected tax revenue, allowing authorities to identify low-compliance villages by unreturned books. To enforce compliance, chiefdom authorities may fine village chiefs in under-performing villages, or require them to "buy" additional tax receipts to resell to villagers (Interviews: 4; 27; 53; 207).

[23] Interview: 402

[24] Interview: 404

consequences at the hand of either village or chiefdom authorities. Given this existing enforcement infrastructure, it is plausible that property owners believe that TPIs can enforce compliance with the new property tax. Taking together, the experimental and qualitative data provide strong evidence in favor of the hypothesis that TPIs' authority stems from their coercive capacity.

Accountability

Relative to control, the effect of the Accountability treatment (T3) on the compliance index is 0.124 SDUs (Column 4; p-value = 0.004). Comparing T3 to T1 allows us to isolate the impact of the accountability prime. The effect of T3 is larger than the effect of T1 by 0.056 SDUs, but this difference is not statistically significant (p-value = 0.18). However, it is worth noting that all three sub-indicators move in the expected positive direction. Further, the positive T3 point estimate is driven by an increase in the measure of voluntary compliance (*coins given to KDC's development fund*), in accordance with theoretical expectations that accountability should lead to consent-based compliance. T3 also increases tax morale relative to control (Column 4; p-value = 0.011), but this effect is not statistically distinguishable from the T1 effect (Column 5).

To further evaluate the accountability hypothesis, I examined a set of preregistered secondary outcomes that measured whether respondents expected to benefit from taxation and whether they believed tax revenue would be spent efficiently and transparently. If chiefs are accountable, the Accountability treatment (T3) should increase these indicators. Conversely, if the T3 has no real effect—and the positive point estimate in Column 5 (Table 4) is simply a product of noise—we would expect no significant effects on these secondary outcomes. In Appendix Table C1, I find that T3 increases an index of secondary outcomes by 0.06 SDUs, but that this effect is, again, just outside the threshold of statistical significance (p-value of 0.15). When these results are considered alongside the borderline significant effect on the main compliance index, they offer suggestive but ultimately inconclusive evidence for the accountability hypothesis.

Given these inconclusive experimental results, the qualitative evidence is particularly instructive. In line with the argument that TPIs have participatory and inclusive decision-making processes (e.g., Baldwin and Holzinger 2019), the qualitative data reveals that local laws are created

with direct and indirect citizen participation. Specifically, I find that (1) byelaws are developed and enacted at meetings, rather than behind closed doors; (2) village representatives are invited to attend meetings for chiefdom byelaws, which are held outside the village; and (3) participants in these policy-making meetings have space to actively engage (see Appendix Table D3).

Respondents highlighted that authorities "made laws in consultation with the people" and that the law making process was open to all community members: "whether you have a [leadership] position in the town or not...it is us all that sit and make [the laws]."[25] Byelaws are developed during open meetings, rather than behind closed doors. To discuss village byelaws, authorities "invite the entire community" or "the whole town" to meetings.[26] Similarly, chiefdom authorities call a "general meeting" to formulate chiefdom byelaws.[27] Nearly all respondents (97%) explicitly mentioned that meetings are called when byelaws are created, either within the village or more broadly within the chiefdom. When meetings are called to discuss chiefdom byelaws, all villages are represented by local chiefs and community leaders. Chiefdom authorities invite "town chiefs, section chiefs, youth chairmen, and the mammy queen [i.e, women leaders]" from each village.[28] Of respondents who reported a meeting would be called, 94% said that a representative from their village or section would be invited to attend these meetings and only 1.8% say that they would not.

These policy-making meetings are spaces where attendees are "given a chance to talk" about "burning issues."[29] Participants can "ask questions and make suggestions" about byelaws that chiefdom authorities or other participants are putting forward.[30] At the village level, 84% of respondents describe meetings as forums for discussion between village authorities and villagers, compared to 8.6% who report that these meetings are only a space for village authorities to *inform* the village's residents about a byelaw. For chiefdom meetings, 78% of respondents describe these meetings as containing active participation from attendees; only 3.7% of respondents report that

[25]Interview 52; Interview: 404

[26]Interview: 400; Interview 404

[27]Interview: 401

[28]Interview: 404. Another respondent describes the universal representation from villages in their area: "all the nineteen villages are invited. No one is left out" (Interview: 10).

[29]Interview: 120; Interview: 404

[30]Interview: 32

these meetings are not open for active participation.

The qualitative data also provides indirect evidence that chiefs are held accountable by citizens. In rural areas, many public works are organized locally and involve communal labor—a form of mandatory labor that chiefs can demand. If chiefs were unaccountable, we might expect them to misuse this authority by demanding labor for projects that serve private interests rather than the public good. However, I find little evidence of this. Instead, the data shows that communal labor is predominantly directed towards projects with broad social benefits, such as clearing vegetation from roadways ("road brushing," 68%) and road maintenance (39%) (see Appendix Table D4). Moreover, respondents themselves report that communal labor is directed towards public benefit projects. For example, over 80% of respondents describe town-level communal labor as being devoted exclusively to public projects, while only 7% indicate that it is sometimes or often used for projects that do not benefit the public (see Appendix Table D5).

Coercion and Accountability as Complementary Sources of Authority

What do these findings tell us about TPIs' authority? Existing theories propose that traditional leaders derive their authority from *either* coercion or accountability—that is, these mechanisms are substitutes. In contrast to the existing literature, I argue in this section that coercion and accountability are best viewed as complementary sources of TPIs' authority.

Is TPIs' Authority Rooted in Coercion Alone?

Some scholars argue that despotic and unaccountable chiefs rely exclusively on coercion to assert their power. In the most widely cited study of traditional political institutions in Africa, Mahmood Mamdani characterizes chiefs as "decentralized despots" whose authority rests on their unchecked coercive capacity (Mamdani 1996). Similar conclusions are drawn in influential studies of TPIs in Sierra Leone (Richards 1996) and South Africa (Ntsebeza 2005). According to Jesse Ribot and his coauthors, "many of the 'indigenous' governance systems. . .[could] be labelled totalitarian, despotic, oppressive, patriarchal, gender biased or gerontocratic" (Ribot et al. 2008).

While I find good evidence that TPIs' use (the threat of) coercion to influence their con-

stituents' behavior, my findings are not consistent with the interpretation that TPIs' authority lies *solely* in their coercive capacity. First, note that the experimental evidence for the accountability mechanism (T3) is not much weaker than the evidence for the coercion mechanisms (T2). While I rely on conventional thresholds of statistical significance to draw inferences about the effects of these treatments, the point estimate of the Accountability treatment (0.124 SDUs) is only 7.5% smaller than the point estimate of the Coercion treatment (0.134).

Second, the perspective that only coercion matters implies that accountable chiefs are not better at obtaining compliance than unaccountable ones. In the data, I find clear variation in perceptions of chiefs' accountability. At one extreme, chiefdom authorities in Lei Chiefdom are perceived as unaccountable and biased in law enforcement, particularly in planter-herder conflicts that are common in the chiefdom (see Appendix E.1 for a case study of law enforcement in Lei Chiefdom). Specifically, interview respondents consistently accuse chiefdom leaders as unfairly favoring livestock farmers in these disputes. "There is no equity in [chief's] judgment. . . Cattle rearers are favored against crop farmers."[31] Indeed, while respondents in other chiefdoms frequently cite law enforcement as something their chiefdom authorities do well (31% of respondents), not a single respondent in Lei Chiefdom (n = 48) mentioned law enforcement when asked about the strengths of their traditional leaders (Table D6, Column 1). Gbane Chiefdom presents a contrasting picture, where 42% of respondents cite law enforcement as something their traditional leaders do well. Relative to respondents in Gbane Chiefdom, respondents in Lei Chiefdom are are also three times more likely to name a particular law they dislike, and twice as likely to say their chiefdom leaders are performing worse than leaders in other chiefdoms. In a case study of law enforcement in Gbane Chiefdom, I find no systematic complaints about local law enforcement or chiefs' performance more generally (see Appendix E.2).

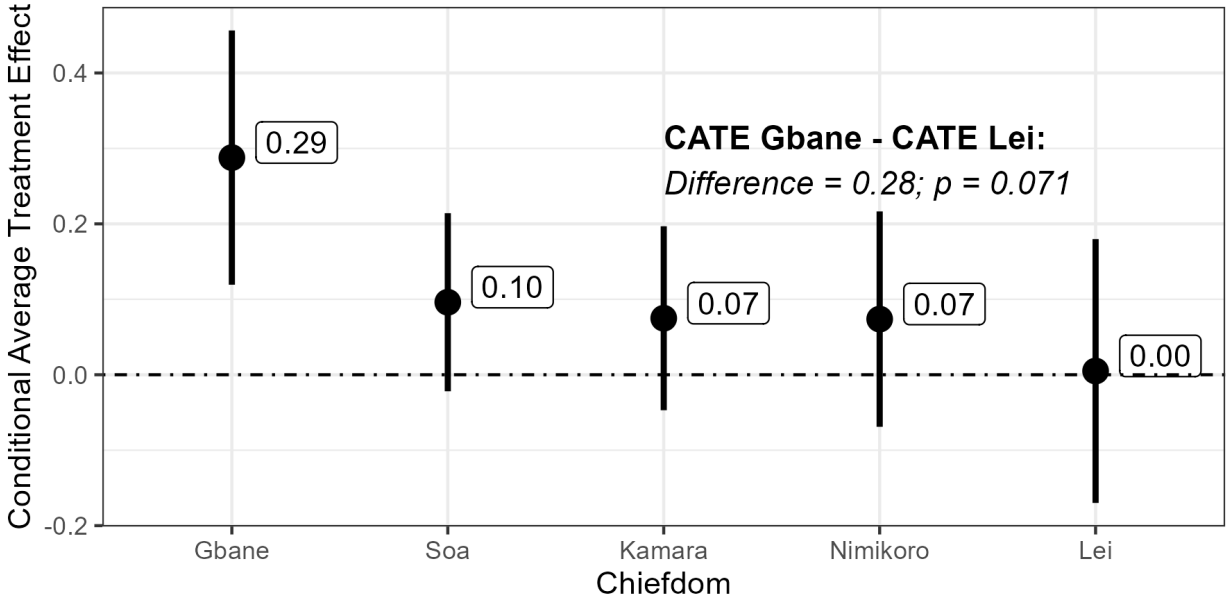
In contrast with the perspective that chiefs' power lies in their despotic coercion, I find evidence that chiefs' have more authority where they are more accountable. Figure 1 plots the joint treatment effect by chiefdom, pooling across all treatment conditions to buy statistical power.[32] The

[31] Interview: 100

[32] While this strategy bundles the effects of collaboration with the effects of coercion and accountability appeals,

treatment effect in Gbane Chiefdom (0.288 SDUs), where chiefs are perceived as accountable, is much larger than the effect observed in Lei Chiefdom (0.005 SDUs), where chiefs are perceived as least accountable. This difference in the magnitude of the collaboration effect between these two chiefdoms is statistically significant (p-value = 0.071). While these analyses are exploratory, they suggest that TPIs' have more authority where they are perceived as more accountable, at odds with theories that suggest TPIs rely on unaccountable coercion alone.

Figure 1: Combined Treatment Effects by Chiefdom



Why Even Accountable Chiefs Need Coercion

In stark contrast to these accounts, other scholars argue that coercion is unnecessary because accountability mechanisms align the interests of traditional leaders with those of their constituents. According to these views, chiefs' authority is legitimized by their pursuit of community welfare (Williams 2010), and citizens follow their chiefs' directives because they trust that these actions will benefit the community (Baldwin 2013, 2016). Likewise, coercion is absent in accounts that attribute chiefs' authority to their popular support (Logan 2013).

this is appropriate because the objective is to explore variation in the authority of TPIs across chiefdom. I use the same model specification for these chiefdom-level estimates, but I omit enumerator fixed effects. Observations per enumerator are not balanced across chiefdoms and including enumerator fixed effects is likely to control away real chiefdom effects. Lines in Figure 1 represent 90% confidence intervals.

While accountability may be an important source of authority for some chiefs, my qualitative data suggests that even accountable chiefs rely on (the threat of) coercion to secure compliance with their directives. To see why coercion is important even for accountable chiefs, consider the case of communal labor. Respondents overwhelmingly reported that chiefs called communal labor for public projects that benefit the community broadly, rather than private projects that benefit a narrow elite (see Appendix Tables D4 and D5). Theories where citizens follow chiefs' directives when these directives are in the community's interest would predict that citizens comply with these demands for communal labor *in the absence of* enforcement mechanisms.

In contrast, I find that TPIs have well-established enforcement mechanisms in place to detect and punish noncompliance with communal labor. According to respondents, the Youth Leader is commonly responsible for monitoring attendance and participation in communal labor and reporting to the authorities (mentioned by 67%).[33] Authorities also monitor attendance themselves, either relying on their knowledge of community members to identify who has failed to show up (46%) or keeping an attendance list (33%). Says one respondent, "if you failed to go, your town chief will know because the town chief knows everyone."[34] Taken together, 90% of respondents described at least one monitoring mechanisms at either the town or chiefdom level. And nearly all respondents agreed that individuals will be punished if they are caught missing community labor (Appendix Table D2).

The perspective that TPIs' coercive capacity and accountability are complementary elements of their authority builds on Margaret Levi's theory of quasi-voluntary compliance. According to Levi, while citizens may be more inclined to comply with a government that is accountable and trustworthy, they will only do so if the government also has the enforcement capacity to punish noncompliance (Levi 1988). The complementarity of these mechanisms also has implications for *how* chiefs use coercion. If chiefs are accountable political authorities, they should use coercion to advance public interests. In doing so, they play the role of the Hobbesian state, enhancing public welfare by enforcing the rule of law and maintaining social order. The use of coercion by

[33]Notes one respondent, "we have the youth leader, he reports to the chief" (Interview: 32).
[34]Interview: 401

TPIs to facilitate the production of public goods reflects Bates' observation that violence can be a "productive resource" for economic development (Bates 2001).

This perspective may explain why my findings support the coercion hypothesis, while two recent studies do not. If chiefs use coercion accountably, they should enforce compliance with laws, such as the property tax I study. In contrast to my findings, Brierley and Ofosu (2023) find no evidence that paramount chiefs in Ghana influence vote choice through coercive mechanisms. Similarly, Baldwin et al. (2023) find no evidence that coercion drives traditional leaders' influence over presidential vote choice or voluntary contributions to burial or education funds in Kenya, Malawi, and Zambia. If chiefs use coercion accountably, these null effects are understandable: individuals who do not vote for the chief's preferred candidate or fail to contribute to a *voluntary* fund do not violate any laws and therefore should not be punished.

Conclusion

Against the predictions of modernization theorists (e.g., Huntington 1968), traditional political institutions are "resurgent" across contemporary sub-Saharan Africa (Englebert 2005), with increasing legal protection (Holzinger et al. 2020; Baldwin 2016, chpt. 3) and high levels of public support (Logan and Amakoh 2022). In this paper, I show that governments in weak states can, and do, rely on contemporary TPIs to secure citizens' compliance with state policy. Specifically, using a field experiment in Sierra Leone, I find evidence that governments can increase citizens' tax compliance by collaborating with TPIs. This finding is robust: using a preregistered specification, all three treatments that mention collaboration increase both a preregistered tax compliance index and respondents' belief that they ought to pay taxes. This finding expands our understanding of the domains where traditional leaders can influence constituents' behavior (Baldwin et al. 2023) and the ways local intermediaries can be leveraged by the state to achieve its goals (Balán et al. 2022).

What implications does this finding have for state-building? In the short term, collaboration addresses state weakness by increasing citizens' compliance. However, collaboration can only resolve state weakness in the long run if the revenue from tax compliance is invested in enhancing state capacity. Future work should investigate politicians' incentives to make these investments

(see Christensen and Garfias 2021).

Another important consideration is how collaboration impacts citizens' perceptions and attitudes toward the government, a question this paper does not directly address. On one hand, if collaboration builds citizens' trust in the state, it may foster long-term state-building. On the other hand, if collaboration diminishes trust in the state, it could hinder it. The literature does not offer a clear prediction, as recent work suggests that whether citizens' attitudes toward TPIs and the government are complements or substitutes depends on the relationship between these actors (Henn 2023; Van der Windt et al. 2019). Investigating the impacts of collaboration on citizens' attitudes toward the government and TPIs is a promising path for future research.

The polarizing literature on TPIs contains competing accounts of these institutions, often dichotomously casting them as either coercive and despotic or accountable and legitimate. Combining experimental and qualitative evidence, I argue that both of these mechanisms are important sources of TPIs' authority. My conclusion that accountability is an important source of TPIs' authority hinges, in part, on exploratory analyses that suggest chiefs are less influential where they are less accountable. In acknowledging that not all chiefs are accountable, my findings do not contradict the many case studies that document traditional leaders abusing of power.

While I acknowledge that some chiefs govern better than others, this study does not attempt to explain this variation. We know little about what explains variation in governance quality across TPIs, though Acemoglu et al. (2014) are a notable exception with their argument about electoral competition. Future research should explore this variation, possibly through more systematic data collection on the coercive powers, institutional checks, and governance performance of TPIs.

Finally, my results speak to the fragmented nature of political authority in contemporary Sierra Leone. My findings are in agreement with Joel Migdal's post-independence assessment that in Sierra Leone, "social control . . . has remained in social organizations apart from the state" (Migdal 1988, pg. 137). Understanding the conditions under which weak states consolidate political authority, and the consequences of that consolidation, remains a central challenge for scholars of contemporary African politics.

References

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

Adcock, R. and Collier, D. (2001). Measurement Validity: A Shared Standard for Qualitative and Quantitative Research. *American political science review*, 95(3):529–546.

Ali, M., Fjeldstad, O.-H., and Sjursen, I. H. (2014). To pay or not to pay? citizens' attitudes toward taxation in kenya, tanzania, uganda, and south africa. *World development*, 64:828–842.

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

Baldwin, K. (2013). Why Vote with the Chief? Political Connections and Public Goods Provision in Zambia. *American Journal of Political Science*, 57(4):794–809.

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

Baldwin, K. (2025). Traditional Authorities Around the World: Toward a Better Understanding of their De Facto Power, Downward Accountability, and State Recognition. *Annual Review of Political Science*, 28.

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

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

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

Baldwin, K. and Mvukiyehe, E. (2015). Elections and collective action: Evidence from changes in traditional institutions in liberia. *World Politics*, 67(4):690–725.

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

Barro, R. J. (1973). The control of politicians: an economic model. *Public choice*, pages 19–42.

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

Bates, R. H. (2001). *Prosperity and Violence: The Political Economy of Development*. W. W. Norton & Company.

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

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

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.

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

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

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

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

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.

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

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

Curtin, P., Feierman, S., Thompson, L., and Vansina, J. (1995). *African History: From Earliest Times to Independence*. Pearson, second edition.

D'Arcy, M. (2011). Why do citizens assent to pay tax? legitimacy, taxation and the african state. *Afrobarometer Working Paper No. 126*.

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

Delavande, A. (2014). Probabilistic expectations in developing countries. *Annu. Rev. Econ.*, 6(1):1–20.

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

Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of consumer research*, 20(2):303–315.

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

Gottlieb, J., LeBas, A., and Magat, J. (2024). Can Social Intermediaries Build the State? Evidence from an Experiment in Lagos, Nigeria.

Grieco, K. (2024). *Can Rural Property Tax Generate Revenue?: A Simple Accounting Exercise in Sierra Leone*. International Centre for Tax and Development.

Haire, M. (1950). Projective techniques in marketing research. *Journal of marketing*, 14(5):649–656.

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

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

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

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

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

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

Kao, K., Lust, E., Dulani, B., Ferree, K. E., Harris, A. S., and Metheney, E. (2021). The abcs of covid-19 prevention in malawi: Authority, benefits, and costs of compliance. *World development*, 137:105167.

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

Kramon, E. (2019). Ethnic group institutions and electoral clientelism. *Party Politics*, 25(3):435–447.

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.

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

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

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

Lust, E., Kao, K., Landry, P. F., Harris, A., Dulani, B., Metheney, E. A., Nickel, S., Carlitz, R., Gatua, J. G., Jöst, P., Mechkova, V., Maxim, F., Tengatenga, J., Grimes, M., Jansson, C. A., Alfonso, W., Nyasente, D., Brahim, N. B., Jordan, J., Bauhr, M., Boräng, F., Ferree, K., Hartmann, F., and Lueders, H. (2023). The Local Governance Performance Index (LGPI) Household Survey 2019: Kenya, Malawi, Zambia.

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

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

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

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

Naseemullah, A. and Staniland, P. (2016). Indirect rule and varieties of governance. *Governance*, 29(1):13–30.

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

Paden, J. N. (1970). Aspects of Emirship in Kano. In Crowder, M. and Ikime, O., editors, *West African Chiefs: Their Changing Status Under Colonial Rule and Independence*. Africana Pub. Corp.

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

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

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

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

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

Robinson, R. and Gallagher, J. A. (1961). *Africa and the Victorians: The Official Mind of Imperialism*. Macmillan and Company.

Schumpeter, J. (1918). The Crisis of the Tax State. *Zeiqragen azu dem Gebiet der Soziologie*.

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

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

Van der Windt, P. and Voors, M. (2020). Traditional leaders and the 2014–2015 ebola epidemic. *The Journal of Politics*, 82(4):1607–1611.

Van Schendel, W. (2009). *A history of Bangladesh*. Cambridge University Press.

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

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

Zhou, Y., editor (2009). *Decentralization, Democracy and Development: Recent Experience from Sierra Leone*. The World Bank.

Online Appendix

Building Fiscal Capacity with Traditional Political Institutions:
Experimental and Qualitative Evidence from Sierra Leone

Following text to be published online.

Table of Contents

A Research Design

A.1 Intervention Development

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

At a November 2020 (Zoom) meeting, I presented a proposal for a video-based property tax awareness campaign, involving both KDC officials and traditional leaders. This proposal was met with general approval, and I was directed to continue developing plans for a tax awareness video that contained local government officials and traditional leaders. At a January 2021 meeting in Kono, the Revenue Mobilization Team agreed on the basic contours of a tax awareness video, including the three key messages that Traditional Leaders should emphasize (collaboration with KDC, accountability, and enforcement). I then met with senior paramount chiefs to workshop the video script.

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

We shot tax awareness videos with traditional leaders in 10 chiefdoms, and I conducted this study in five of those chiefdoms. In four of these chiefdoms, we filmed both Kono and Krio versions. In one (Nimikoro), we only filmed a Kono version, so property owners who did not speak Kono were excluded.[35]

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

A.2 Sampling

i. Selecting Study Chiefdoms: The research design relies on the development of chiefdom-specific tax awareness videos. Therefore, it was only possible to conduct the study in chiefdoms where we were able to create a tax awareness video with the paramount chief. I reached out to paramount chiefs or senior chiefdom authorities in all 14 chiefdoms and successfully shot videos with senior chiefs in 10 of those 14 chiefdoms. For the four chiefdoms where I did not shoot a video, I was unable to schedule a recording session in the three-day period for which I had hired a professional filmmaker.[36] In three chiefdoms, we shot videos not with the paramount chief, but with his representative: I excluded these chiefdoms.[37] I excluded one chiefdom (Gorama Kono) because I judged that the language used by the paramount chief in the video differed too much from the agreed script. Finally, I decided to exclude Toli Chiefdom for practical budgetary reasons. Toli contains less than 2% of the villages in Kono and is the most sparsely populated and least accessible chiefdom in the district. I determined that enumeration costs in Toli would be too high to warrant inclusion in the study. This left five chiefdoms included: Soa, Lei, Gbane, Nimikoro, and Kamara.

ii. Sampling Villages: Using villages from Sierra Leone's 2015 national census as a sampling frame, I excluded certain villages based on three criteria. First, the research design depends on the ability to influence respondents' beliefs about the involvement of TPIs in the new property tax collection. I reasoned that this belief would be more flexible in villages where property tax had not previously been collected, so I excluded all villages visited by tax collectors in 2019. Second, to enhance enumeration efficiency, I excluded villages listed in the 2015 census as having fewer than three building structures. Finally, I excluded chiefdom headquarters towns from the sample. This process left me with a sampling frame of 434 eligible villages.

Due to the poor road infrastructure in Kono District, traveling between villages is time-consuming and resource-intensive. To minimize transportation costs during data collection, I employed geographic cluster sampling to select 123 villages from the set of 434 eligible villages. The enumeration team was unable to locate five of the 123 sampled villages, likely because the sample frame was based on the 2015 National Census, which may not have accurately reflected the existing set of village in Kono at the time of surveying. Errors may have occurred during the census, and some villages may have been abandoned since then.

I first grouped the 434 eligible villages into 155 geographical clusters, dropping three isolated

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

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

villages. Clustering was done within each chiefdom, so that villages were not clustered across chiefdom boundaries.[38] I then sampled clusters, and within sampled clusters, I sampled villages.

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

1. Clusters near the chiefdom's headquarters town that contain a large village.
2. Clusters near the chiefdom's headquarters town that do not contain a large village.
3. Clusters far the from chiefdom's headquarters town that contain a large village.
4. Clusters far the chiefdom's headquarters town that do not contain a large village.[39]

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

- First, I drew two clusters in each strata. (There are two strata that contain only one cluster of villages—in these I drew one cluster).[40]
- Second, I selected two village in each stratum. In strata that contain large villages, I selected one large and one small village.
- Third, I checked whether the number of potential observations in each stratum was at least 100. As a proxy for the number of potential observations in each village, I used the number of structures recorded in the 2015 census.
- Fourth, in strata where the target number of potential observations was not been met, I drew an additional village from the set of sampled clusters.[41]
- It remains possible that the maximum number of potential observations in a given strata did not reach 100. In this case, I drew an additional cluster from the appropriate cluster stratum.

[38] After the initial clustering, 25 villages were in clusters of their own. I placed these villages in the closest cluster. In three instances, these one-cluster villages were more than three kilometers from the closest village in their new cluster; I dropped these three villages.

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

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

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

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

- The enumeration team arrived in the village in the morning and went directly to the house of the village chief (or another village authority if the village chief was not present that day). A letter had been dropped off to village authorities within the previous three days specifying the date of the enumeration team's arrival. Enumeration teams ranged between two and six people, depending on the size of the village.
- From the house of the village chief, the enumerators agreed to walk in separate directions. After agreeing which directions they would each travel, enumerators used their tablets to select a distance, which told enumerators whether to interview a respondent at the first, second, third, or fourth house in their chosen direction. If the enumerator found no one home at the relevant house, the enumerator proceeded to the next house in that direction.
- The enumerators asked to speak to the person "most responsible and influential" for making decisions related to the property. If that person was home, the enumerator began the informed consent process to start the interview. If that person was not home, the enumerator asked if he or she would return later that day. If so, the enumerator scheduled a time to return to interview that person. If not, the enumerator asked if there is "someone else who is involved in decision-making related to this property." If so, the enumerator asked to interview that person. If not, the enumerator attempted to schedule an appointment for later. If that was not possible, the enumerator moved on to a different property.
- After completing an interview, the enumerator used the tablet to select the direction and distance of the next house. Previously interviewed houses (marked by a sticker) were not included in the count.
- If an enumerator walked past the last structure of the village in a given direction, he or she turned around and finished the count, walking back in the direction they came.
- If an enumeration team completed interviews with all available respondents before the end of the day, they proceeded to their next scheduled village. Otherwise, the enumerators left for their next scheduled village in the morning.[42]

iv. Informed Consent

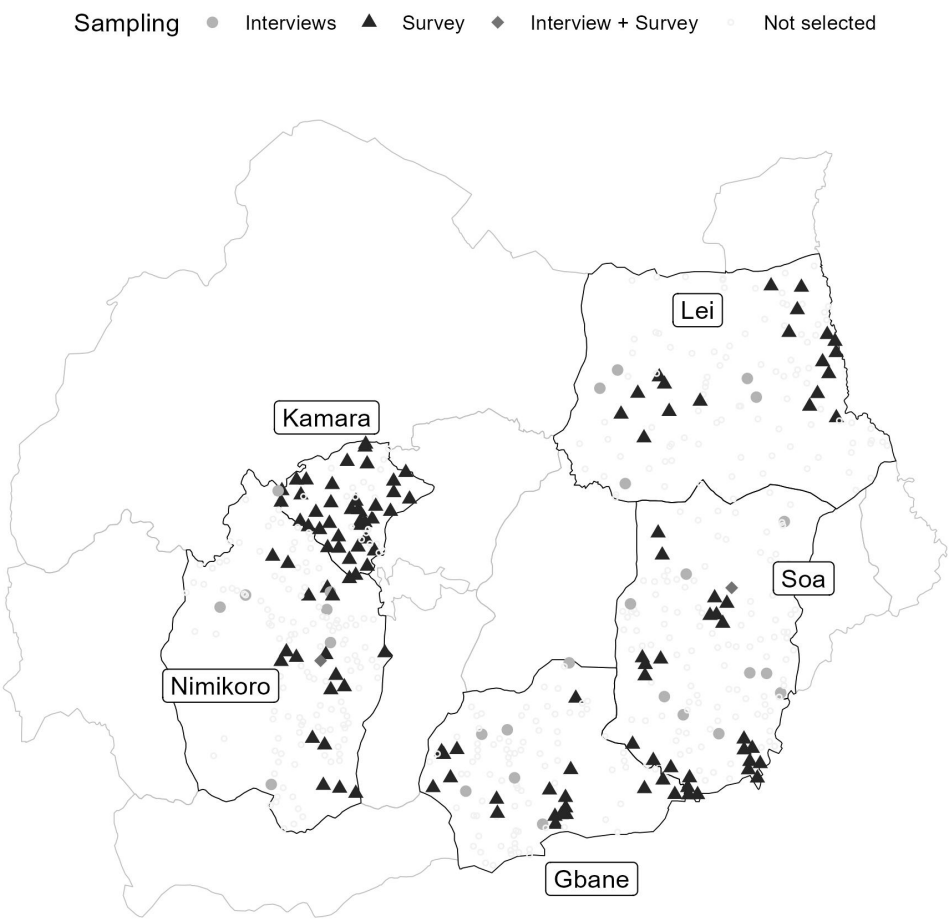
Enumerators explained their role to respondents in the following way: "I work for an organization that is between the people and the government, which is called KoCEPO. This organization

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

is doing some research to find out ways to improve conditions in Sierra Leone." They also clarified that the research was being carried out by an academic institution, not a political entity. Enumerators assured respondents that their answers would be recorded on a tablet but would remain confidential, that they could skip any questions they did not wish to answer, and that the survey would take approximately 45 minutes to complete. They emphasized that choosing not to participate would have no negative consequences.

Respondents were informed that while they would not receive direct compensation for participating in the survey, they would receive three Maggi spice cubes as a token of appreciation. Additionally, although it was not discussed in advance, respondents kept their proceeds from a modified dictator game. For the control group, the average amount retained by respondents was 1,670 Sierra Leonean Leone (approximately US $0.15). After providing this information, enumerators asked if the respondent would like to participate in the survey.

Figure A1: Sampling Map



A.3 Balance

Table A1: Balance Table

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

Table A1 reports balance across immutable covariates. Columns 1-4 report treatment group means; Column 5 reports the control group standard deviation; Columns 6-8 report differences standardized relative to the control group standard deviation; Column 9 reports the *F*-statistic for the joint null hypothesis.
Significance: * $p < 0.10$

Table A1 presents balance statistics. Columns 1-4 display group means for each covariate and column 5 presents the control group standard deviation. Columns 6-8 present differences between each treatment group's mean and the control group mean, standardized by the control group standard deviation. For metrics to gauge the magnitude of these differences, I provide two test statistics from a model that regresses a given covariate on the three treatment indicators. First, where a treatment group mean for a given covariate is statistically different than the control mean ($\alpha < 0.1$), I star the corresponding standardized difference in columns 6-8. Second, Column 9 presents the *F*-statistic for the joint null hypotheses—a significant result here implies that the treatment indicators collectively have predicative power (i.e., treatment group means are different than the control group mean).

Given the 33 tests I run in Columns 6-8, under the null hypothesis of no differences between groups, we would expect 3.3 tests to appear significant at the 90% confidence level; I find only one significant difference on the *education* covariate. In Column 9, I run 11 tests and therefore expect 1.1 to appear significant at the 90% confidence level; I find one significant difference on the *gender* variable.

A.4 Measurement Scale

Table A2 reports (control group) responses to four key sample question that respondents were asked to familiarize them with the 10-point bean scale. Average responses were low to unlikely events ("chance that the president visits this community tomorrow") and high for highly likely events ("chance that you will drink water this month"). In addition, the response patterns were in keeping with basic laws of probability—respondents overwhelmingly reported that they had an equal or greater likelihood of visiting the district headquarters town in the next 30 days than in the next seven days: just 4% of respondents report they are more likely to travel to the district capital over the next seven days than over the next thirty days.

Table A2: Responses to Practice Questions (Control Group)

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

This exercise also provides insight into how respondents interpreted the levels of my measurement scale. While our enumeration team coached respondents that each bean represented 10 percentage points of probability ("each bean is one chance out of 10"), it seems more likely that respondents understood each bean as an increase (or decrease) in relative likelihood, rather than representing exactly 10 percentage points. This means that between-respondent differences in measured outcomes may represent differences in the way respondents map perceived probabilities to the 10-point scale—in addition, of course, to representing real differences in beliefs.[43] Therefore, responses to these practice questions might predict responses to other survey questions, a relationship that can be leveraged to reduce noise when estimating treatment effects. In include several of these measures in the pre-specified covariate adjustment.

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

Figure A2: Using Beans to Measure Perceived Probability

A.5 Measurement Validity

Table A3: Correlation Matrix of Compliance Indicators

	Compliance Propensity		*Voluntary Compliance*	
	Direct	Indirect	Game	Tax Morale
Direct (*self-reported propensity to pay*)	1.00	0.42	0.08	0.41
Indirect (*perceived neighbors' propensity to pay*)	0.42	1.00	0.085	0.38
Game (*coins given to KDC's house fund*)	0.08	0.085	1.00	0.16
Tax Morale	0.41	0.38	0.16	1.00

Table A3 reports the Pearson Correlation Coefficient between each indicator of tax compliance.

Table A4: Nomological Validity

	Outcome = compliance		
	(1)	**(2)**	**(3)**
Panel A: Correlates of Compliance Behavior in Freetown (2023)			
Wealth (property value)	0.003***		0.003***
	(0.000)		(0.000)
Education (attended inst. of higher education)		0.066***	0.040***
		(0.015)	(0.015)
N	3,615	3,379	3,377
Panel B: Correlates of the Compliance Index (this study)			
Wealth (animal wealth index)	0.002***		0.002***
	(0.001)		(0.001)
Education (received any schooling)		0.102***	0.093***
		(0.034)	(0.034)
N	1,752	1,752	1,752

Table A4 summarizes the results from OLS regressions that examine the relationship between tax compliance (the dependent variable) and wealth and education (explanatory variables). In Panel A, the dependent variable is property tax compliance behavior in Freetown (the capital of Sierra Leone), as observed from administrative records. Compliance is coded as 1 if a property owner paid any tax in 2023, and 0 otherwise. Columns 1-3 present regression coefficients (standard errors in parentheses) from models that include measures of wealth (Column 1), education (Column 2), and both (Column 3). The *Wealth* variable reflects the property's assessable value, where the variable score is the property's percentile rank in the sample (e.g., the median value property in the sample has a wealth score of 50). The *Education* variable is an indicator equal to 1 if the property owner attended an institution of higher education (mean value = 0.40). In Panel B, the dependent variable is the tax compliance index, which serves as the primary outcome of interest in this study. Columns 1-3 present regression coefficients from models that include measures of wealth (Column 1), education (Column 2), and both (Column 3). In Panel B, the *Wealth* variable captures the market value of the respondent's animal stock, where the variable score is the respondent's percentile rank within this stock. The *Education* variable is a dummy variable equal to 1 if the respondent has attended any formal school (mean value = 0.35). **Significance:** * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B Manipulation and Attention Checks

B.1 Manipulation Check

Figure B1: Manipulation Check Measurement

B.2 Attention Check 1 - Recalling Video Speakers

The tax awareness videos contain information that I expect to modify respondents' beliefs in theoretically important ways. First, I check whether respondents can correctly recall the number and identity of the speakers in the video they watched. Respondents who see the control video see only one speaker, District Council Chairman Solomon Bundo. Respondents who see treatment videos see two speakers—Chairman Bundo and the paramount chief of their chiefdom. 94% of respondents correctly state the number of speakers and 93% correctly identify the speakers. If there were two speakers, this meant correctly naming both. Enumerators asked these questions directly after a respondent watched the video.

B.3 Attention Check 2 - Recalling Video Content

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

1. A property tax that will be collected on houses. ["Tax"]
2. The Chiefdom Council is working with Kono District Council on this property tax. ["Collaboration"]
3. After taxes are collected chiefdom authorities will call a meeting to discuss how to spend the money collected. ["Spend"]

4. After taxes are collected chiefdom authorities will call a meeting to discuss how to punish non-compliers. ["Punish"]
5. Tax collectors will be paid 10% of the money they collect. ["Salary"]
6. All tax collectors have an identification card with their name and picture. ["ID Card"]

Table B1 reports property owner responses by treatment condition. Column 2 ("n") refers to the number of observations in each treatment group.[44] The value in each of the remaining six columns is the percent of respondents that affirmed a given message was given in the video. First, let's consider a set of three questions that all respondents should answer in a similar way, regardless of treatment condition. Of course, the central messaging of the video is around a house/property tax. Column 3 ("Tax") tells us that across treatment and control groups 95 to 98 percent of respondents correctly state that the video contained messaging about a house tax.

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

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

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

Table B1: Attention Check - Recalling Video Content

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

Next, let's consider responses to three questions that we do expect to change with the respondent's treatment condition. First, recall that the treatment videos attempt to manipulate respondents' perceptions about the collaboration between the state and TPIs. As reference to this collaboration only appears in treatment versions (and not in control), we should see respondents in T1, T2, and

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

T3 more likely to agree that the collaboration between chiefs and state was discussed in the video (compared to control video). Indeed, that is what we see. Respondents in T1, T2, and T3 are respectively 18 percentage points, 22 percentage points, and 21 percentage points more likely to state that collaboration between KDC and chiefdom authorities was mentioned in the video.

In the Coercion treatment (T2) the paramount chief said that he would call a meeting with chiefdom authorities to discuss how to punish non-compliers. Respondents who viewed the T2 video are 33 percentage points more likely to state that their video contained this message, compared to respondents who watched the T1 video.[45] The goal of the Accountability treatment (T3) is to prime respondents to accountability mechanisms in TPIs and the T3 video the paramount chief says he will call a meeting to discuss with his sub-chiefs on how the collected revenue will be spent. Respondents who watched T3 videos are 25 percentage points more likely to affirm that their videos referred to these meetings, compared to respondents who watched the T1 video.[46]

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

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

[45] This increase jumps to 39 percentage points when comparing T2 to the pure control.

[46] This increase jumps to 34 percentage points when comparing T3 to the pure control.

messaging about TPIs. It is possible that respondents infer that the subjects they were asked about in the survey (ie., TPIs) are likely to have been addressed in the video. Third, placing the recall questions at the end of the survey creates a time lapse between the video and the recall questions, which may lower recall.

If the high false positive rate is driven mainly by the placement of the recall questions at the end of the survey (rather than general lack of comprehension), we should see much lower false positives if the recall questions were asked directly after the video. Prior to undertaking my primary data collection, I conducted a pilot survey where we did ask recall questions directly after the video. Table B2 shows results from that pilot. False positive rates in the control group plummet. Only 30% of respondents who watch the control video incorrectly state their video discussed collaboration, down from the 68% we saw in our true study. Rates of false positives drop across each of the other four comprehension check measures.

Table B2: Attention Checks (Pilot)

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

C Experimental Analyses

C.1 Secondary Accountability Outcomes

To further evaluate the accountability hypothesis, I examined a set of preregistered secondary outcomes. If chiefs are indeed accountable, the involvement of TPIs in the property tax should increase respondents' expectations of benefiting from the tax, as well as enhance their perceptions of spending transparency and efficiency. I measured four secondary outcomes. The first indicator reflects respondents' perceptions that their own village will benefit from taxation, while the second measures their belief that other villages in their chiefdom will benefit. The third and fourth indicators assess expected benefits indirectly, capturing respondents' views on whether tax revenue will be spent transparently and efficiently. For these outcomes, the appropriate comparison group is the pure control group (C; Tax information). Using T1 (the TPI collaboration treatment) as a comparison is problematic because questions about expected benefits and spending efficiency and transparency may prime respondents to focus on TPIs' accountability. For instance, after seeing their paramount chief in a video, T1 respondents might evaluate these survey questions in the context of their chief's governing performance, which could act as an accountability prime. This would undermine the impact of the accountability treatment (T3).

Table C1 presents the impact of the Accountability treatment (T3) on an index composed of these four survey measures. Column 2 shows that the accountability treatment (T3) increases the accountability index, relative to the control group, but this effect is not statistically significant (p-value = 0.15).

Table C1: Effects of Accountability Treatment (T3) on Secondary Outcomes

Secondary Outcome	**Mean** (1)	**T3-C** (2)	**N** (3)
Accountability Index	**0.000**	**0.060**	**1,752**
	(0.698)	(0.042)	
Own village will benefit from tax	7.114	0.088	1,722
	(2.853)	(0.061)	
Other villages will benefit from tax	6.710	0.023	1,646
	(2.838)	(0.068)	
Ease of discovering how tax revenue was spent	4.460	0.103	1,742
	(3.094)	(0.064)	
Proportion of revenue towards development	6.180	0.021	1,709
	(2.614)	(0.066)	

Table C1 reports the effect of Accountability treatment (T3) on the secondary outcomes. Column 1 reports the control group mean for each indicator, with the standard deviation in parentheses. Column 2 present standardized treatment effects for T3, relative to control. Models are estimated using OLS with preregistered specifications. Column 3 reports the number of non-missing observations.

C.2 Variable Definitions

Table C2: Description of Outcome Variables

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

C.3 Deviations from PAP

In this section I report deviations from the preanalysis plan:

- For ease of interpretation, I report treatment effects on index sub-indicators with p-values, rather than 95% CIs.
- Treatment effects for the Accountability treatment (T3) on secondary outcomes (Appendix Table C1) are reported relative to control (as preregistered); however, they are not also reported relative to T1, as I believe this comparison to be inappropriate (see Appendix Section C.1).

D Qualitative Data and Analyses

D.1 Qualitative Data Collection

Qualitative data collection was motivated by one descriptive question and one causal question. First, I wanted a richer understanding of the mechanisms tested in the experiment. Specifically, how do TPIs deploy their coercive capacity? How might leaders of TPIs be held accountable? Second, I wanted to understand why the effect of collaboration varied across chiefdoms. To answer these questions, in fall of 2022, I worked with a team of six research assistants to conduct interviews with 300 respondents across 29 villages in four chiefdoms in Kono district (Gbane, Soa, Lei, and Nimikoro). Chiefdoms were selected to maximize treatment effect variation (see Figure 1). Figure A1 shows the villages visited for qualitative interviews, marked with light circles. In two of these villages, we had already conducted surveys, indicated by diamonds. Respondents were randomly selected using a random walk procedure, and interviewers followed the same consent process as during the surveys (see Appendix Section A.2).

With respect to my causal question, my prior was that this variation may be explained by differences in the way chiefs governed their chiefdoms. Therefore, interviews focused on key elements of chiefdom governance, such as taxation, mandatory communal labor, law-making and enforcement, and citizens' perceptions of chiefs' performance. Before conducting data collection, interviewers were trained during a five-day workshop to follow an interview guide which was structured to cover the following topics:

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

D.2 Qualitative Data Processing

To summarize qualitative data, I developed a coding scheme to capture respondents' (i) perceptions of enforcement mechanisms for local tax and communal labor, (ii) participation in byelaw creation and attitudes towards byelaws, (iii) judgments of leaders' performance, (iv) descriptions of and attitudes towards projects undertaken with communal labor. A team of three research assistants applied this scheme to code 261 interviews.

To investigate why the effect of collaboration varied across chiefdoms, my approach was to listen to a random set of (Krio language) interviews, with an ear towards cross-chiefdom variation in governance outcomes that could explain the observed variation in the effect of collaboration. This preliminary investigation of the interview data revealed substantial variation in citizens' perceptions of traditional leaders' enforcement of local laws: a significant proportion of respondents complained that chiefs unfairly enforced local laws.

I also considered two other potential drivers of chiefs' effectiveness as collaborators: (1) public participation in law-making and (2) TPIs' enforcement capacity. I found little evidence of variation along these dimensions; in general, respondents reported overall high levels of participation and enforcement capacity. Therefore, I ruled these out as factors explaining the observed variation in chiefs' effectiveness as collaborators.

Focusing on complaints about law enforcement, I went back and listened to all Krio language interviews, village-by-village, in the chiefdoms where I observed the largest (Gbane Chiefdom) and smallest (Lei Chiefdom) treatment effects. I took notes on each interview, and then wrote village-level summaries of these notes. Based on these summaries, I wrote the law enforcement vignettes for these chiefdoms (Appendix E).

D.3 Qualitative Tables

Table D1: Local Tax: Monitoring and Punishment Strategies

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

Table D1 presents qualitative evidence of TPIs' monitoring and punishment strategies for Local Tax. Chiefdom authorities retain most of the revenue from this tax; a small percentage is transferred to the local government. Monitoring interview prompt: "Did village (chiefdom) leaders do anything to check if people had paid Local Tax this year (2022)? Or do they not do anything like that?" Punishment interview prompt: "When village (chiefdom) leaders found out that someone had not paid, did they anything about it, or did they not do anything?" Percentages are rounded to the nearest integer.

Table D2: Communal Labor: Monitoring and Punishment strategies

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

Table D2 presents qualitative evidence of TPIs' monitoring and punishment strategies for Communal Labor. Interviewers' prompt: "What happens if someone who was supposed to participate in Community Labor does not? Would the Village (chiefdom) Leaders find out? Would they do anything?" Percentages are rounded to the nearest integer.

Table D3: Citizens' Participation in Local Law-making

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

Table D3 describes respondents' perceptions of participation in local policy making. Interviewers asked respondents about the creation of town, section and chiefdom level byelaws, without explicitly mentioning meetings. For ease of exposition, I combine responses about section and chiefdom level meetings. For the first outcome, respondents coded as "agree" ("disagree") explicitly mention that a meeting was (not) called. For the second outcome respondents are coded as "agree" if they say that representatives would be called to attend either section or chiefdom level meetings. For the third outcome, research assistants coded interviews for evidence of active participation ("agree"), evidence of lack of participation ("disagree"), or no evidence for either ("unclear"). Throughout, respondents are coded as "unclear" for a given outcome when their response is ambiguous or when they don't answer a given prompt. If the interviewer did not ask the question the respondent is removed for that outcome.

Table D4: Common Communal Labor Projects

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

Table D4 reports projects to which communal labor is devoted, according to respondents in semi-structured interviews. Respondents were asked to name up to three recent projects carried out with communal labor. This table report the percent of respondents that name a given type of project. This table presents a non-exhaustive list. Percentages are rounded to the nearest integer.

Table D5: Who Benefits from Communal Labor?

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

Table D5 presents respondents' perceptions regarding the benefits of mandatory communal labor. Interviewers prompted "Is the Communal Labor called for by the Town (Section; Chiefdom) Leaders usually used in a fair way that benefits the community or is it used in an abusive way that benefits only a few people?" Responses are coded as "agree" respondents say community labor is used for broad community benefits, "disagree" if respondents say community labor is sometimes or often used for narrow benefits, and "unclear" is the response is ambiguous or there was no direct response to the question.

Table D6: Perceptions of Law Enforcement

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

Table D6 reports citizens' perceptions of laws and law enforcement, and traditional leader performance, by chiefdom. Column 1 reports the percent of respondents that mentioned law enforcement when prompted: *"Please tell me something the leaders of this section (chiefdom) are doing well?"*. Column 2 reports the percent of respondents who reported there is a law in their chiefdom they do not like. Column 3 reports the percent of respondents who state the performance of traditional leaders in their chiefdom is *worse* than the performance of traditional leaders in other chiefdoms. Column 4 shows the number of coded interviews.

E Chiefdom Vignettes on Law Enforcement

E.1 Lei Chiefdom

In Lei Chiefdom a local law has been passed that outlines compensation for crop farmers when cattle damage their crops and compensation due to herders if a farmer attacks a cow.[47] A respondent in one village takes issue not with the law itself but with its implementation.

> I have a problem with one [law] that has not been implemented fairly. This is occurring during the process of adjudicating on matters where a livestock farmer's animal has eaten a farmer's crop. In matters like that, the crop farmer's complaint is not treated seriously or followed through on according to the byelaw and most times unreasonable [i.e., very low] compensation is made. On the other hand, if a crop farmer kills a cow of a livestock farmer, [that crop farmer] will be beaten, molested, and treated poorly. There is no equity in [chief's] judgment of this byelaw. Cattle rearers are favored against crop farmers.[48]

Another respondent in the same village agrees, "The laws between the cattle owners and the crops farmers are very fine in writing and when reading them, but its implementation is very bad."[49] A third respondent from the same village takes issue with the perceived difference in standards applied by chiefdom authorities to crop farmers and cattle rearers, "If a cow eats the rice you've planted, they eat the money that you would need to pay the children's school fees. If you complain nothing happens. . .But if you kill one cow. . . that's an issue."[50]

This perception that crop farmers are getting the short end of the stick turns up in other villages. Says a respondent in a different village, "If a cow ruins someone's farm, [the authorities] should summon that person [to court]. At times it can take the chief a month to do so, as they are avoiding the case. But if something happens to a cattle, within 30 minutes or an hour, an arrest is made and someone is detained."[51] Another respondent in the same village has similar frustrations with inaction from chiefdom authorities: "If a cow eats my rice, and I make a report to [the section chief] take action! ...[the authorities] should take action, but they don't."[52] In a third village, there are similar complaints, "As a man of the country, I haven't see anything good yet that [chiefdom authorities] have done. Like when those cows ruin our rice, we cry. The money! But when the cattle herder comes [inaudible] he doesn't have money [for us]. The authorities don't do anything."[53] Even Chiefdom leaders admit that this is a problem. In a section headquarter town in Lei, when

[47] Interview: 96
[48] Interview: 100
[49] Interview: 130
[50] Interview: 10
[51] Interview: 1
[52] Interview: 31
[53] Interview: 106

asked what chiefdom authorities could improve, the first topic discussed by the Section Chief is the "settling of dispute among farmers and cattle rearers."[54]

E.2 Gbane Chiefdom

In Gbane Chiefdom, I fail to find similar systematic complaints about local law enforcement. While many of citizens' complaints in Lei Chiefdom focused on planter herder conflict, there were no such issues that cut across the six villages where we conducted interviews in Gbane Chiefdom. In fact, only in one village did respondents' complaints converge on a topic: the role of chiefdom leaders in resolving a boundary dispute with a neighboring village. Three of the eight interviews I reviewed in this village mention the boundary dispute and place negative judgment on chiefs' role in this dispute. Across the 30 interviews I reviewed in the remaining five villages, I fail to document strong criticism of local laws (or their implementation). In two of these villages, the strongest criticism I can find against chiefs is that the paramount chief does not live in the chiefdom headquarter town, but the district headquarter. For reference, nearly all paramount chiefs reside most of the time in the district headquarter town (Koidu).[55] In the remaining three villages, the biggest complaints against chiefs are fairly normal demands for development (e.g., improve water access, improve roads), demands that are also commonplace in other chiefdoms.[56] I cannot find a complaint against a law or implementation of a law.

[54]Interview: 76

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

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