Online Appendix For: Marketing Taxation? Experimental Evidence on Enforcement and Bargaining in Malawian Markets

Lucy Martin[*], Brigitte Seim[†], Simon Hoellerbauer[‡], and Luis A. Camacho[§]

---

[*]Associate Professor, Department of Political Science, University of North Carolina at Chapel Hill

[†]Associate Professor, Humphrey School of Public Affairs, University of Minnesota

[‡]Lecturer in Data Analytics and Computational Social Science, University of Massachusetts Amherst

[§]Senior Technical Director, Evaluation, Research, and Analytics, Social Impact

Appendix Contents

A   Implementation and Experimental Interventions

This section presents a detailed intervention timeline and supplemental details regarding each treatment component in the bottom-up (BU) treatment bundle (Section A.1) and the top-down (TD) treatment bundle (Section A.2).

This field experiment was conducted as one component of the Local Government Accountability and Performance (LGAP) activity, a program from the United States Agency for International Development (USAID) in Malawi. USAID developed LGAP to support improved democratic accountability and local government capacity to effectively and efficiently deliver public services, for improved government performance. LGAP aimed to rigorously examine this link to support the Government of Malawi in determining the best ways to improve service delivery and democratic practice. It focused primarily on: 1) Supporting citizen engagement and advocacy for accountable local government; 2) Building the capacities of local government to transparently deliver on their mandates; and 3) Supporting decentralization policy and process reforms as required by the Malawian 'Public Sector Reform' agenda. LGAP was implemented by DAI over five years, August 2016-August 2021, with an initial total budget of approximately $15 million. It was implemented in eight districts that were chosen by USAID and the Government of Malawi: Balaka, Blantyre, Kasungu, Lilongwe, Machinga, M'Mbelwa, Mulanje, and Zomba (see Figure A1).

A.1   Bottom-Up Treatment Bundle

Step 1: Facilitate Market Committee Elections and Training
At baseline, not all markets had valid market committees (or, indeed, any market committee at all). Invalid market committees are those that were formed without proper elections — such as committees that were directly installed in markets by the local government — or whose terms had expired. Elections were held in the 54 treatment markets that did not have valid market committees. All newly elected market committees then received a training in which committee members learned about the proper organizational structure and the roles and responsibilities of the market committees, including their role in local governing structures. These trainings, co-run by the District Council's District Capacity Building staff and LGAP district staff, took place in December 2017 and January 2018.

The following flyer (example from Machinga District) was circulated in BU markets as part of Step 1.

LGAP role towards increased market revenue:

- Supports market vendors' committee elections and trainings.
- Supports market infrastructure projects such as water, market sheds, pathways, electricity, garbage bins for use by vendors.
- Supports market vendors on sharing market information and grievances through mHub.

Machinga District Council

Mandate of District Councils

District Councils are charged with the overall development of the districts. In the course of their development work, the Councils are required to provide for local people's (communities) participation in the formulation and implementation of the District Development Plan. It is in this respect that the Councils have been requested to form action committees at Area, Ward or Village level. For District Councils to translate the power and competence into development phenomena, a sound local government financing mechanism is required. They have two main revenue sources:

(a) Locally generated revenues (traditional); and

(b) Central government transfers.

Locally generated revenues include:

(a) Property rates

(b) Ground rent;

(c) Fees and licences;

(d) Commercial undertakings; and

(e) Service charges.

In trying to address the pressing public service delivery issues for its largely impoverished population, the Malawi Government developed a National Decentralization Policy (1998) and passed the Local Government Act (1998), but progress in the implementation of decentralization has been slower than anticipated. At some point, there was a nine-year gap without elected local councillors, resulting in administrative and fiscal decentralization being rolled out without much coordination over the last 16 years. However, an effective local government system has been reinvigorated following the **2014** tripartite elections, with the election of **462** local councillors. The Local Government Performance and Accountability (LGAP) project, thus, is helping to give these new councillors the skills they need to fully represent the people of Malawi.

Why should a vendor pay market fees/taxes?

- Undertake development for communities.
- Construct market structures.
- Support refuse collection.
- Put up street lights.
- Maintain infrastructure.
- Provide health, sanitation and recreation services.
- Pay for bursaries

Role of market committees
- Facilitating common market development agenda.
- Managing Social welfare of members.
- Facilitating and monitoring market ground rules.
- Mediating conflicts.
- Presenting grievances to the council management.

Roles of the council, ticket sellers, market masters and revenue collectors
- Provision of utilities.
- Provision of sanitation.
- Allocation of space.
- Cleaning.
- Maintenance of structures in the markets.
- Ensuring security in markets.
- Collection of revenue.

Role of Chiefs
- Custodians of culture.
- Provide advice on broader issues such as land conflicts, personal conflicts etc. that link to traditional values.
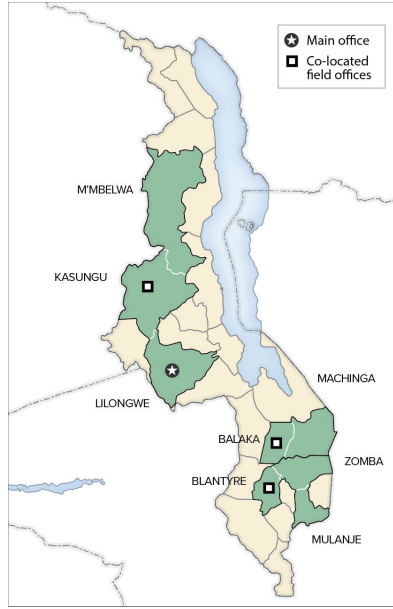
Figure A1: *Intervention Districts.* Main and Field offices refer to the offices of the USAID implementing partner.

Step 2: Facilitate Meetings Between Vendors, Market Committees, and Local Government After ensuring that each treatment market had a valid market committee, local governments held a public meeting in the market to address vendors' sense of exclusion from the taxation system. These meetings included vendors, the market committee, political and bureaucratic local government representatives, and group village headmen. Local government representatives typically in attendance were a representative from the District Finance Office, the market's tax collectors, the market/zone managers, and the ward councilor. A total of 3515 vendors attended these Step 2 meetings in the 64 treatment markets, with a median attendance of 61.5 vendors.[1] These meetings, which took place between January and February 2018, were observed by LGAP and included the following elements:

- A speech by the ward councilor in which they reminded vendors of the connection between taxes and development in the market

- A discussion of the roles and responsibilities of vendors and government officials, including vendors' obligation to pay market fees whether or not they sold any goods.

- A discussion of the perceived problems with the current tax collection system, particularly the barriers faced by vendors. The local government representatives also had the chance to explain how the revenue is used.

- An explanation of the bottom-up intervention and the way it would impact market operations. This included discussing the way the council uses funds from market fees and introducing Step 4 (the SMS system, see below). At the end of the meeting,

---

[1]These statistics exclude M'mbelwa markets; meeting reports are missing for this district.

vendors were able to register for the SMS system (see below).

- Documentation and discussion of the state of market services, including toilets, sanitation, security, and infrastructure. Vendors then selected their market development priorities. This was done using a pairwise ranking method over six options: market shed, borehole, electricity, pathways, a concrete slab, and refuse bins. This list was used to decide which infrastructure project a market would receive (see Step 3). This piece of the intervention in particular was designed to replicate tax bargaining.

Step 3: Jump Start Service Delivery in Markets

While our calculations suggested that market revenue would be sufficient to maintain infrastructure, existing compliance and funds were too low to fund infrastructure improvements. Step 3 therefore aided local governments to implement the top market infrastructure priority chosen in Step 2. Forty-six markets chose a borehole – the other 18 chose a mix of the remaining project types. Projects cost approximately US$5,000, an amount insufficient to completely rehabilitate markets but sufficient to serve as a costly signal of the local government's commitment to improving service provision in markets.

The markets were scoped and visited by infrastructure specialists during the summer of 2018 in order to complete the necessary field assessments. A competitive bidding process between July and September 2018 led to the selection of the appropriate construction firms. Construction began in September 2018 and finished in March 2019. However, almost all markets saw at least some construction progress prior to endline data collection in November-December 2018. Each project was bookended by an opening ceremony and a handover ceremony attended by government officials, market committees, and other vendors. Market committees were responsible for monitoring the state of the projects and, upon their completion, developed a maintenance plan in conjunction with the district council.

Step 4: Increase Transparency in Taxation via an SMS System

To strengthen citizens' trust that their tax funds were being used well, and to facilitate transparency and bottom-up accountability between vendors and local government, Step 4 sought to provide citizens with information about government revenue and spending on an ongoing basis via an SMS messaging system. This system was developed and managed by mHub, an organization in Malawi that works with businesses and other organizations on information and communication technology projects(`http://www.mhubmw.com/`). At the meetings in Step 2, vendors were informed about the SMS system and were able to sign up by sending an SMS message to a market-specific number. A total of 2435 vendors signed up for the SMS system at the Step 2 meetings, with a median of 44 registrations. The median proportion of vendors who attended who signed up for the system was 0.729.[2] Posters and pamphlets were also left at the markets after the meeting so vendors could understand the system better and sign up later.

Vendors then received monthly messages with that month's market revenue, along with information on how the money generally was allocated and spent. The text of these messages was designed to become more specific over the intervention period, as vendors became more comfortable with the system. One of the main advantages of the SMS system was that, once

---

[2]These statistics exclude M'mbelwa markets, due to lack of data.

data on market revenues were collected, the system for passing this information to vendors was centrally managed and required few steps. Messages were first sent in January 2018, in markets where Step 2 meetings had already taken place. The last messages were sent in November 2018. All vendors who signed up received the SMS messages unless or until they opted out.

Vendors were also able to use a related SMS system, also set up and managed by mHub, to report grievances about local government service delivery. During the intervention period, grievances were passed on to local government officials designated by the implementing partner. mHub, in conjunction with the local governments, followed up with complainants when issues had been resolved. The grievance system was designed to give vendors more agency and enable them to take action if revenues were not used in line with their expectations or if market services were lacking.

A.2 Top-Down Treatment Bundle

Step 1: Roll Out Mobile Money Revenue Transfer System
To address the widespread potential for evasion, corruption, and inefficiency at the market level, treatment markets shifted to transferring revenue via mobile money, a phone-based banking system. Airtel Malawi was engaged to transfer revenue to the district council. Tax collectors still collected fees from vendors and then gave the cash to the market manager. Then, instead of bringing the cash to the bank or district headquarters a few times a month, the market manager transferred the cash to an Airtel agent,[3] who then transferred it to the district council bank accounts. Airtel earned 2% of the fees as payment. This streamlined revenue transfer to local governments and improved their ability to reliably track how much each market collects in fees. It also made it easier to see if certain markets were not transferring funds as regularly as they should. Markets started using mobile money in March 2018 and continued using it until December 2018.

Step 2: Provide Accurate and Reliable Market Vendor Counts
One barrier to efficiently collecting market fees was the lack of a reliable estimate of anticipated revenue, which is required to determine collector benchmarks, monitor collector performance, and forecast local government revenue. However, the size of the market (measured in the number of vendors) changes over the course of the week, month, and year. This made a formal registration system infeasible. Instead, the implementing partner hired and trained vendor counters.[4] Counters visited each market at least four times a month - on two market days and two non-market days. These vendor counters systematically walked through the market and recorded the number of vendors by type of business. On each visit, they counted vendors twice at different times of the day in order to obtain a more accurate count. Vendor counting started in February 2018 and continued until October 2018.

Step 3: Forecast Revenue and Generate Revenue Targets Based on Vendor Numbers

---

[3]No market vendors were recruited as Airtel agents.
[4]These individuals could not be market vendors or government staff.

The figures produced in Step 2 were used to determine tax collector compensation schemes, forecast local government revenue, and track LGAP performance. The counts, once provided to the local government, were fed into a revenue target calculator that adjusted targets based on the previous month's revenues collected for each market and the number of market days a week that market had in order to create monthly estimates of the expected revenue for each market. These targets were then communicated to market masters and tax collectors. They were also used to evaluate market performance in Step 4. Targets were first sent to markets starting in April 2018. The last targets were communicated in November 2018.

Step 4: Introduce Incentives for Tax Collectors
Prior to the intervention, tax collectors reported low motivation due to low incomes - on average less than US$1 per day. This was true regardless of whether tax collectors received a fixed wage or commission pay.[5] This led to vendor perceptions of collector corruption. Salary paychecks were also often late, reducing the incentive to work hard.[6]

Step 4 addressed these issues with an incentive system using the revenue targets created in Step 3. These incentives were non-monetary in nature and were applied at two levels: market and individual. If a market met or surpassed its monthly revenue target, the market received either wheelbarrows, rakes, hoes, or shovels - valuable supplies that make management of the market easier. In addition, if the market met its target, each tax collector also received an individual incentive, which could have been a bicycle, fertilizer, certificates of excellence, mattresses, or work suites. A tax collector whose market kept meeting its targets was able to choose to alternate incentive goods. These incentives were designed to inspire tax collectors to perform their jobs without having to resort to bribery.

B   Deviations from Research Design and Intervention Plan

The treatments are described in their idealized form in Sections A.1 and A.2. Despite extensive buy-in from the implementing partners, local governments, and the national government, some markets saw significant deviations from the originally-planned interventions. This section catalogs the major deviations and addresses how these deviations may have affected our ability to detect results from the interventions. Because increasing taxation is almost always politically sensitive, the deviations we observe are also informative about the types of interventions that are most politically feasible in similar contexts.

B.1   Project and Intervention Component Delays

The project as a whole was delayed several times. Baseline data collection began in July 2017. The goal at that point was to implement the interventions over the course of the next

---

[5]Six districts reported at least some tax collectors paid based on commission for at least some of the implementation period.

[6]All districts except for Lilongwe experienced tax collector salary paycheck delays in at least some markets for some of the implementation period.

year, so that endline collection could take place in summer 2018 so that it would not overlap with the May 2019 elections or preceding campaign period. However, implementation delays meant that most of the top-down treatment components had only been in the field for 6-8 months during endline data collection. The fact that markets and local governments were only exposed to these components for approximately half a year means that the effects we observe are likely the lower bound of potential treatment impact.

In the the bottom-up treatment arm specifically, the infrastructure component was significantly delayed for logistical reasons. Ultimately, only eight of the 64 projects started before October 30, when endline data collection began. Ultimately, three markets were visited for endline data collection before any construction had taken place. Nineteen markets were visited after the handover ceremony had taken place and all construction had been completed.

B.2   Treatment Implementation Issues

In addition to general delays touched on in the previous section, there were a series of specific issues with treatment implementation that have the potential to weaken the interventions.

Service Delivery in Markets (BU):
Due to budget constraints, about half of bottom-up markets ended up not receiving their first choice infrastructure project. The majority of them got their second choice. Additionally, despite scoping visits by hydrologists, no water was found after drilling in fourteen of the forty-six markets that selected a borehole. These markets received alternative projects, typically wheelbarrows and other cleaning supplies, or mobile refuse bins.

Mobile Money Revenue Transfer System (TD):
In Balaka district, a dispute with the district's bank led to a temporary halt to the mobile money program for two months during the intervention period.

Incentives for Tax Collectors (TD):
The top-down treatment included individual incentives for tax collectors who met targets. However, after the first month of incentives, local governments switched the incentives to be market-level instead of individual. Following this change, markets that met their targets received market-level rewards of wheelbarrows and bicycles. Additionally, several districts did not consistently communicate revenue targets, limiting their impact. Finally, there was often a significant lag before incentives were actually received by markets that met their targets.

B.3   Protests, Boycotts, and Strikes

Throughout the course of the intervention period, a number of markets saw vendor protests and vendor boycotts of fee payments. These are common in the study context and, to the

best of our knowledge, were not the result of any intervention components.[7] Often, these protests had to do with lack of adequate services in markets, underscoring the importance of the bottom-up treatment (particularly Step 3). Other protests were due to other market concerns unrelated to service provision.

C   Data Collection Strategy

The data in the main paper come from baseline and endline surveys of market vendors and tax collectors. All surveys were implemented by Innovations for Poverty Action (IPA). Government revenue data, discussed below, was provided by our implementing partner, along with monthly reports on implementation. Finally, IPA conducted monitoring visits to each of the sample markets, providing further data during the course of implementation. This all resulted in a rich tapestry of information upon which we draw in our analysis.

C.1   Baseline and Endline Surveys

C.1.1   Market Vendor Survey

Market vendor surveys were conducted among 100 vendors from each of 128 markets; 20 for the full survey and 80 for the short version focusing only on tax compliance. Different individuals were surveyed at baseline and endline, unless the same individual was chosen by chance. Vendors were selected using a modified random walk, details of which are available upon request. Markets were visited on their market days to ensure that the sample represented the market when the largest number of vendors was present. Vendors received a small airtime voucher in return for completing the survey, valued at USD0.28 or USD0.42-USD0.84[8] for the short and long surveys, respectively.

C.1.2   Tax Collectors Survey

Enumerators administered a 20-30 minute survey to up to seven tax collectors in each market. The survey included questions on knowledge of tax law; knowledge of customer service practices; number of points of contact with market vendors and businesses; rejection rate in tax collection attempts; perceived proportion of market vendors paying taxes per day; amount collected in local taxes; and perceived barriers to tax compliance. At baseline, 302 tax collector surveys were completed, with an average of 2.44 per market. At endline, 264 tax collector surveys were completed, with an average of 2.06 per market.

---

[7]There was one claim that the endline data collection team had set off protests in two markets; an independent investigation found that this was not the case, and the fee boycott was unrelated to the interventions.

[8]A delayed gratification experiment embedded in the long survey affected the amount provided.

C.2 Monitoring Data

Some of our robustness checks investigate treatment noncompliance. These estimates rely on monitoring data from the intervention period between baseline and endline data collection. LGAP provided us with information on the status of the interventions on a monthly basis. They also collected government records related to tax collection on a monthly basis for the entire period between baseline and endline data collection. In addition, we carried out periodic focus groups with vendors and interviews with tax collectors, market committee members, and market managers to collect nuanced qualitative data on intervention implementation.

C.2.1 Data Exchange

The implementing partner, LGAP, provided key information to help assess intervention roll-out. In particular, this information helped us assess spillover and violations of intervention design.[9] LGAP also served as the intermediary for data from the local governments, including market revenues. They also collected monthly market-level data on estimated numbers of vendors (in TD markets); numbers of tax collectors; the revenue targets (for TD markets); total revenue for each month; and LGAP activities in and around the market. More aggregated data were provided on how market revenues had been allocated/spent. In addition, LGAP provided district-level information on any treatment compliance issues, changes in tax collector and market manager employment, and potential spillovers.

C.2.2 Market Visits

IPA also carried out unannounced market visits throughout the intervention period to supplement the quantitative analysis, allowing us to assess local perception of the interventions, provide an additional check of treatment compliance, and identify which mechanisms were being affected by the interventions. On average, 25 percent of the study markets were visited during every two-month period. Over the course of eight months, all 128 were visited. Each visit lasted about three hours and included an anonymous walk in which the observer recorded observations, an interview with the market manager, an interview with the market committee chairperson, an interview with the tax collectors, and a one-hour focus group discussion with a small group of market vendors. These data provided us additional information on treatment compliance, spillovers, and how the interventions were perceived by different market actors.

---

[9]In reality, this information was often delayed, which meant that we were often not able to react as quickly to issues as we would have liked. It was, however, invaluable to our analysis.

C.3 Measures

Table C1 presents the variables used to test hypotheses H1–H11, the hypotheses laid out in the main text. All measures are drawn from the data sources discussed in Section C.

Table C1: Key Outcome Measures

Outcome	Question Wording	Answer Options
Receipt measure	Enumerators verified whether respondent could show a tax receipt and recorded the date on the receipt. Date used to confirm whether receipt dated from within past 7 days.	1 / 0.
Self-reported compliance	Now, I am going to put 5 tokens on the table here. Think about the last 5 days you sold goods or services in this market. Please put a token here [indicate location] for each time you happened to be able to pay your K100 fee in the last 5 days	0 - 5
Perceived group compliance	Similar to self-reported compliance, but asked to allocate 10 tokens to represent other vendors paying fees.	0 - 10
Trust in local government	In your opinion or based on what you have heard, would you say the district government is trustworthy?	4-point scale
Trust in ward councilors	In your opinion or based on what you have heard,would you say your ward councilor for this market is trustworthy?	4-point scale
District manages funds well	Do you strongly agree, agree, disagree or strongly disagree with how your district government is managing the following: Managing public funds effectively on behalf of citizens in this area	4-point scale
District spending transparent	Do you strongly agree, agree, disagree or strongly disagree with how your district government is managing the following: Transparency in how it uses funds collected from market revenues	4-point scale
District tax collection transparent	Do you strongly agree, agree, disagree or strongly disagree with how your district government is managing the following: Transparency in how much it is collecting from market revenues in this area	4-point scale
Satisfaction with services (combined)	In general, how satisfied are you with the developments in THIS market provided by the district government?	4-point scale
Satisfaction with water access	Now I am going to ask you about different developments in the market provided by the district government. For each, please tell me how satisfied you are with them. Access to clean water	4-point scale
Perception of relative amount spent on services	For every 1,000 kwacha the government collects from this market, how much do you think goes towards developments in the market provided by the district government and paying market staff?	0-1000
Paying taxes is a duty	Whether agree/disagree: "Paying taxes is a duty of all citizens, even when you do not approve of how elected officials spend money."	4-point scale
Vendors should pay taxes	Which of these statements do you agree with more? Statement 1: Vendors should always pay tax even if they disagree with local government. Statement 2: Vendors should only pay tax if they agree with local government.	1/0
Vote	Do you intend to vote in the next national election?	1 / 0
Sign Petition (Anon)	Enum explains it's a hypothetical, then asks "Would you be willing to sign an anonymous petition to the District Finance Committee telling them that you want the District to improve funding for markets? An anonymous petition is one that does not have your name on it, just how many people have signed."	1/0
Sign petition (Name)	Would you be willing to use your real name to sign a petition to the District Finance Committee telling them that you want the District to improve funding for markets? The Councilors who receive the petition will have the names of all the people who have signed it.	1/0
Reduce Gov Overreach	Whether sent an SMS saying "We would like to demand a reduction in the over-reach of government power in collecting revenue in this district"	1 / 0
Demand Increased Effectiveness	Whether sent an SMS saying "We would like to demand an increase in government effectiveness in spending government revenue collected in this district."	1 / 0
Ind'l Evasion Possible	Do you Strongly Agree, Somewhat agree, somewhat disagree, or strongly disagree with the following statements: If I wanted, I could refuse to pay my market fee.	4-point scale
Group Evasion Possible	Do you Strongly Agree, Somewhat agree, somewhat disagree, or strongly disagree with the following statements: If all vendors together decided to stop paying market fees, we could do so successfully.	4-point scale
Pay Because Consequences	Do you Strongly Agree, Somewhat agree, neither agree nor disagree [**Note:sic** in question but not offered as an answer option], somewhat disagree, or strongly disagree with the following statements: I pay market fees because I'll get in trouble if I don't	4-point scale
Perception of Relative Amount of Taxes Reaching Gov't	For every MK 1,000 fee collectors in this market collect, how much do you think reaches the district government?	0-1000
Hours working in market per day	START: Each day you work at THIS market, what time do you start work at this market? END: Each day you work at THIS market, what time do you end work at this market?	Calculated time difference between end time and start time (numeric)
Vendors visited in market per day	Each day that you work at THIS market, about how many vendors do you visit?	Open-ended numeric

D Survey Descriptive Statistics

The tables in this section provide summary statistics for key demographics and outcomes in the baseline and endline surveys.

Table D1: Summary Stats for Demographic Variables Vendor Survey - Baseline. Superscripts in column names identify groups. Superscripts in cells indicate that a value is significantly different from the value for the superscripted group

Variable	Overall Mean	BOTH[1]	BU[2]	Control[3]	TD[4]	SD	Min	Max	N
Age	33.609	33.803	33.495	33.537	33.603	10.427	18.000	87.000	12356
Female	0.334	0.342	0.355	0.302	0.337	0.472	0.000	1.000	12388
Education (ordinal)	8.268	8.117	8.387	8.23	8.336	3.436	0.000	15.000	12383
Literate	0.781	0.757[4]	0.787	0.761[4]	0.816[1,3]	0.414	0.000	1.000	2494
Household income (Malawian kwacha)	71513	68903	73051	72963	71156	86942	100	600000	11943
Vendor sells services	0.102	0.114	0.086	0.102	0.106	0.303	0.000	1.000	12388
Vendor sells daily	0.283	0.282	0.297	0.269	0.283	0.450	0.000	1.000	12386
Years selling in market	6.402	6.5	6.199	6.354	6.561	6.383	0.000	47.000	2522

Table D2: Summary Stats for Demographic Variables Vendor Survey - Endline. Superscripts in column names identify groups. Superscripts in cells indicate that a value is significantly different from the value for the superscripted group

Variable	Overall Mean	BOTH[1]	BU[2]	Control[3]	TD[4]	SD	Min	Max	N
Age	33.974	34.165	33.828	34.334[4]	33.57[3]	10.140	18.000	86.000	12351
Female	0.348	0.344	0.376	0.337	0.333	0.476	0.000	1.000	12370
Education (ordinal)	8.184	8.016	8.186	8.116	8.421	3.460	0.000	17.000	12358
Literate	0.845	0.854	0.864	0.838	0.825	0.362	0.000	1.000	2516
Household income (Malawian kwacha)	62909	60600	63105	60990	66955	77482	1	600000	12159
Vendor sells services	0.087	0.093[2]	0.059[1,3,4]	0.094[2]	0.104[2]	0.282	0.000	1.000	12370
Vendor sells daily	0.290	0.283	0.299	0.279	0.299	0.454	0.000	1.000	12369
Years selling in market	6.581	6.602	6.285	6.718	6.726	6.335	0.000	50.000	2525

Table D3: Summary Stats for Demographic Variables Tax Collector Survey - Baseline. Superscripts in column names identify groups. Superscripts in cells indicate that a value is significantly different from the value for the superscripted group

Variable	Overall Mean	BOTH[1]	BU[2]	Control[3]	TD[4]	SD	Min	Max	N
Age	41.523	41.905	41.932	43.838[4]	39.046[3]	11.895	20.000	88.000	302
Female	0.265	0.23	0.274	0.221	0.322	0.442	0.000	1.000	302
Education (ordinal)	10.046	9.878	10.493[3]	9.338[2,4]	10.368[3]	2.453	0.000	14.000	302
Literate	0.973	0.959	0.973	0.971	0.988	0.161	0.000	1.000	301
Household income (Malawian kwacha)	40264	39777	40947	39294	40863	37121	4000	350000	302
Days worked in market	3.722	3.554[2]	4.301[1,4]	3.559	3.506[2]	2.251	1.000	7.000	302
Does not speak English	0.358	0.338	0.397	0.294	0.391	0.480	0.000	1.000	302

Table D4: Summary Stats for Demographic Variables Tax Collector Survey - Endline. Superscripts in column names identify groups. Superscripts in cells indicate that a value is significantly different from the value for the superscripted group

Variable	Overall Mean	BOTH[1]	BU[2]	Control[3]	TD[4]	SD	Min	Max	N
Age	41.038	42.322	38.97[3]	43.141[2]	40.068	10.914	19.000	71.000	264
Female	0.311	0.305	0.373	0.25	0.311	0.464	0.000	1.000	264
Education (ordinal)	10.163	10.254	10.06	10.016	10.311	2.232	3.000	13.000	264
Literate	0.924	0.915	0.91	0.906	0.959	0.265	0.000	1.000	264
Household income (Malawian kwacha)	41329	46892	38743	40391	40046	31936	6000	250000	264
Days worked in market	4.045	4.475[3]	4.254[3]	3.266[1,2,4]	4.189[3]	2.475	1.000	7.000	264
Does not speak English	0.473	0.424	0.478	0.406	0.568	0.500	0.000	1.000	264

Table D5: Summary Stats for Outcome Variables Vendor Survey - Baseline. Superscripts in column names identify groups. Superscripts in cells indicate that a value is significantly different from the value for the superscripted group

Variable	Overall Mean	BOTH[1]	BU[2]	Control[3]	TD[4]	SD	Min	Max	N
Self-reported compliance	3.808	3.786	3.902	3.746	3.796	1.676	0.000	5.000	12359
Perceived group compliance	6.689	6.7	6.811	6.57	6.672	2.506	0.000	10.000	12221
Ability to produce receipt	0.257	0.255	0.253	0.246	0.272	0.437	0.000	1.000	12371
No receipt when paying	1.357	1.337	1.344	1.367	1.383	0.783	1.000	5.000	2496
Trust in Local Gov.	2.735	2.693	2.773	2.75	2.723	0.944	1.000	4.000	2463
Trust in Ward Cllr.	2.674	2.619[2]	2.794[1,4]	2.669	2.609[2]	0.986	1.000	4.000	2413
District Manages Funds Well	2.716	2.717	2.722	2.76	2.665	1.259	1.000	4.000	2495
Satisfaction w/ Water Access	2.000	1.963	2.05	1.861[4]	2.125[3]	1.241	1.000	4.000	2511
" " Garbage Collection	2.277	2.349	2.289	2.202	2.27	1.232	1.000	4.000	2506
" " Pathways Conditions	2.486	2.57	2.502	2.418	2.456	1.185	1.000	4.000	2511
" " Stall Conditions	2.265	2.302	2.265	2.214	2.279	1.222	1.000	4.000	2509
" " Security	2.597	2.585	2.694[3]	2.514[2]	2.589	1.272	1.000	4.000	2506
" " Developments in Mkt	2.046	2.082	2.111	1.961	2.03	1.143	1.000	4.000	12339
Respondent Avg. Services Satisf.	2.328	2.359	2.363	2.243	2.347	0.867	1.000	4.000	2467
Perception of Relative Amount Spent on Services (Malawian kwacha)	301.970	303.375	308.407	284.094	311.771	216.008	0.000	1000.000	2361
Vendors Should Pay Taxes	1.544	1.556	1.541	1.536	1.542	0.498	1.000	2.000	12343
Paying Taxes is a Duty	3.691	3.727	3.654	3.689	3.695	0.694	1.000	4.000	2519
Ind'l Evasion Possible	1.483	1.476	1.45	1.537	1.471	1.006	1.000	4.000	2499
Group Evasion Possible	1.749	1.71	1.722	1.792	1.771	1.151	1.000	4.000	2512
Pay Because Consequences	3.825	3.834	3.818	3.839	3.807	0.506	1.000	4.000	2508
Sign Petition Anon.	0.708	0.707	0.709	0.724	0.692	0.455	0.000	1.000	2512
Sign Petition w/ Name	0.508	0.507	0.522	0.526	0.475	0.500	0.000	1.000	2518

Table D6: Summary Stats for Outcome Variables Vendor Survey - Endline. Superscripts in column names identify groups. Superscripts in cells indicate that a value is significantly different from the value for the superscripted group

Variable	Overall Mean	BOTH[1]	BU[2]	Control[3]	TD[4]	SD	Min	Max	N
Self-reported compliance	3.945	3.89	4.009	3.844[4]	4.034[3]	1.437	0.000	5.000	11822
Perceived group compliance	6.508	6.508	6.661	6.392	6.473	2.357	0.000	10.000	12294
Ability to produce receipt	0.328	0.325	0.386[3]	0.265[2]	0.336	0.470	0.000	1.000	12365
No receipt when paying	1.477	1.414[3,4]	1.417[3,4]	1.533[1,2]	1.544[1,2]	0.828	1.000	5.000	2516
Trust in Local Gov.	2.692	2.734	2.761	2.609	2.664	0.978	1.000	4.000	2509
Trust in Ward Cllr.	2.600	2.657[4]	2.715[3,4]	2.555[2]	2.47[1,2]	1.006	1.000	4.000	2447
District Manages Funds Well	2.516	2.477	2.457	2.545	2.587	1.128	1.000	4.000	2521
District Spending Transparent	2.382	2.363	2.316	2.409	2.441	1.151	1.000	4.000	2518
District Tax Collection Transparent	2.362	2.323	2.307	2.401	2.419	1.162	1.000	4.000	2510
Satisfaction w/ Water Access	2.225	2.239[2]	2.584[1,3,4]	1.968[2]	2.103[2]	1.277	1.000	4.000	2517
" " Garbage Collection	2.356	2.282	2.461	2.359	2.323	1.230	1.000	4.000	2520
" " Pathways Conditions	2.441	2.389	2.467	2.411	2.498	1.138	1.000	4.000	2520
" " Stall Conditions	2.188	2.141	2.144	2.142	2.325	1.153	1.000	4.000	2517
" " Security	2.414	2.396	2.391	2.449	2.42	1.221	1.000	4.000	2525
" " Developments in Mkt	2.149	2.17	2.284[3]	1.999[2]	2.143	1.100	1.000	4.000	12365
Respondent Avg. Services Satisf.	2.336	2.297	2.426	2.278	2.341	0.853	1.000	4.000	2478
Perception of Relative Amount Spent on Services (Malawian kwacha)	371.816	365.629	388.952	361.847	370.776	264.198	0.000	1000.000	2411
Perception of Relative Amount of Taxes Reaching Gov't (Malawian kwacha)	724.017	737.372	717.399	722.463	718.709	260.398	0.000	1000.000	2463
Vendors Should Pay Taxes	0.604	0.612	0.605	0.595	0.602	0.489	0.000	1.000	12355
Paying Taxes is a Duty	3.680	3.68	3.727[3]	3.638[2]	3.674	0.658	1.000	4.000	2531
Ind'l Evasion Possible	1.530	1.521	1.489	1.57	1.541	0.981	1.000	4.000	2514
Group Evasion Possible	1.802	1.736	1.809	1.79	1.873	1.141	1.000	4.000	2524
Pay Because Consequences (Unclustered t-Test)	3.665	3.668	3.669	3.642	3.681	0.677	1.000	4.000	2518
Sign Petition Anon.	0.747	0.781[4]	0.771[4]	0.736	0.701[1,2]	0.435	0.000	1.000	2514
Sign Petition w/ Name	0.541	0.583[4]	0.573[4]	0.518	0.49[1,2]	0.498	0.000	1.000	2514
Reduce Gov. Overreach (Agree)	0.689	0.696	0.716	0.671	0.674	0.463	0.000	1.000	2531
Reduce Gov. Overreach (SMS)	0.328	0.359[4]	0.365[4]	0.314	0.272[1,2]	0.470	0.000	1.000	2531
Increase Gov. Effectiv. (Agree)	0.739	0.746	0.759	0.727	0.723	0.439	0.000	1.000	2531
Increase Gov. Effectiv. (SMS)	0.366	0.402	0.389	0.362	0.309	0.482	0.000	1.000	2531

Table D7: Summary Stats for Outcome Variables Tax Collector Survey - Baseline. Superscripts in column names identify groups. Superscripts in cells indicate that a value is significantly different from the value for the superscripted group

Variable	Overall Mean	BOTH[1]	BU[2]	Control[3]	TD[4]	SD	Min	Max	N
Hours in Market	9.405	9.617	9.71	8.996	9.289	2.595	1.000	14.000	302
Vendors Visited	93.282	78.716	100.443	67.552	119.724	187.634	10.000	3000.000	298

Table D8: Summary Stats for Outcome Variables Tax Collector Survey - Endline. Superscripts in column names identify groups. Superscripts in cells indicate that a value is significantly different from the value for the superscripted group

Variable	Overall Mean	BOTH[1]	BU[2]	Control[3]	TD[4]	SD	Min	Max	N
Hours in Market	9.841	10.359[3]	9.276[4]	9.218[1,4]	10.495[2,3]	3.002	0.500	20.250	262
Vendors Visited	113.923	178.966	77.621	72.922	130.347	434.360	12.000	6000.000	261

17

E Revenue Results

Hypothesis 2 in the PAP posited that each treatment would also increase the tax payments that reached local governments. We anticipated analyzing this outcome using market-level government revenue data, which was to be provided by the implementing partner on a monthly basis. We pre-specified that our analysis would include a difference-in-means estimate using November 2018 (endline) revenue data, as well as a diff-in-diff approach using November 2017 as the pre-treatment month.

However, there were significant issues with the revenue data. Due in part to low capacity for record-keeping among local governments at baseline, we did not receive the "baseline" November 2017 revenue numbers until May 2018, well after treatment was implemented. We therefore cannot confirm that these numbers actually represented baseline revenues. Furthermore, for 17 disproportionately control group markets, we never received any baseline revenue data at all. In contrast, later months were received much more promptly. This suggests that baseline revenue data were especially prone to error, and that baseline revenue data issues may not be constant across treatment groups.

Figure E1 shows monthly market revenue estimates for each treatment group. For this analysis, the raw amount of revenue was divided by the daily fee payment, to create a more standardized measure of "market fee units per month." While there were noticeable differences at endline in the predicted direction, those differences were also present at baseline.
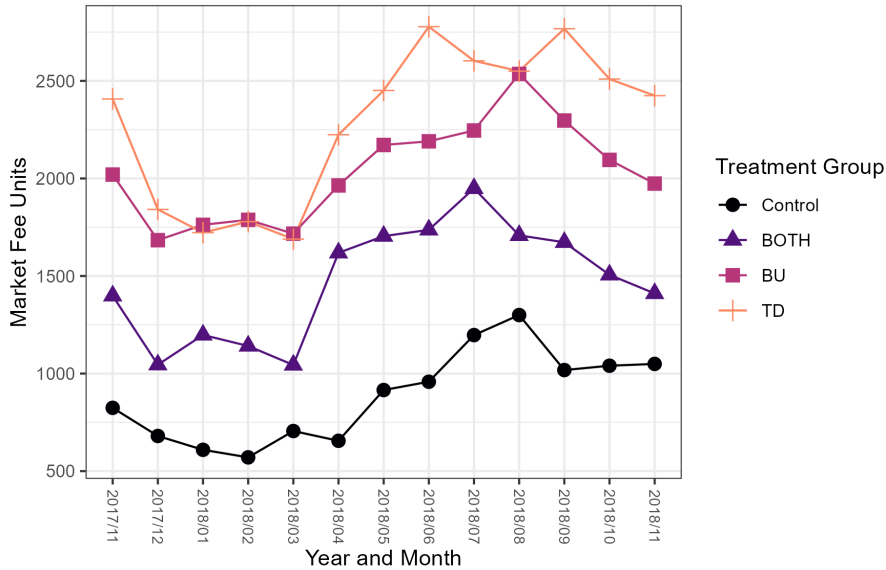
Figure E1: Market Fee Units, Average by Treatment Group

Table E1 presents the DIM and DID estimates for the estimated number of monthly fee payments at the market level, as pre-specified. The DIM regression shows a significant, positive effect of the TD treatment. However, as discussed above, the DID analysis, which uses November 2017 as a baseline, finds no significant treatment effects for any condition.[10]

---

[10]Using December 2017 as the baseline produces similar results.

Table E1: Treatment Effects on Market Fee Units

	Market DIM	Market DID
BU	961.004	−94.445
	(617.697)	(366.193)
TD	1,251.142*	−154.090
	(598.921)	(348.530)
BOTH	341.216	−159.338
	(614.391)	(363.802)
Observations	123	108
Adjusted R^2	0.137	−0.093

Notes:
  (1) Models include block fixed-effects.
  (2) *p<0.05; **p<0.01; ***p<0.001.

These results have two possible interpretations. First, it is possible that the treatments really did increase government revenues, and that the baseline numbers are not accurately capturing pre-treatment revenue levels for the reasons described above. Second, it is possible that the treatments did not affect market revenues at all, and the DIM results reflect pre-existing differences in revenue. Note that the estimates also use slightly different samples, as we are missing baseline data for a number of markets. Ultimately, due to low confidence in the data, we have omitted these results from the main paper, focusing instead on measures where data quality is higher.

F   Understanding Differentiated Effects for BU & TD Treatment

In the results presented above, effects in the markets that received both the BU and the TD interventions (the BOTH group) are often either substantively smaller or statistically insignificant compared to the groups that received only one bundle of interventions (BU or TD). This pattern runs contrary to our expectation, which was that the two bundles would complement each other and, in combination, would have the greatest effect.

We posit four explanations for this pattern in the BOTH markets:

1. **Crowding Out Explanation:** In the BOTH markets, it is possible vendors were more inclined to pay taxes voluntarily due to the BU components, but that this effect on voluntary tax compliance was counteracted ("crowded out") by the focus on consequences and monitoring in the TD bundle. This explanation is supported by academic work from diverse fields (Agrawal, Chhatre and Gerber, 2015; Frey and Jegen, 2001;

Ostrom, 2000).

2. **Vendor Capacity Explanation:** In the BOTH markets, it is possible that having eight intervention components rolled out in one year was overwhelming for vendors, and that their response to this overwhelm was to ignore some of the components.

3. **State Capacity Explanation:** Planning, staffing, and managing all of the BU and TD components in the BOTH markets was resource-intensive, and it is possible local governments delivered weaker versions of the treatments as a result.

4. **Intervention Timing Explanation:** The timing of the intervention rollout was such that some of the BU components (elections, meetings, SMS transparency campaign) rolled out before the TD interventions rolled out, and that many of the market infrastructure projects had not been completed prior to endline data collection. This means that, in the BOTH markets, vendors learned about their rights and responsibilities surrounding government revenue collection, then government focus on revenue collection was ramped up, but without the corresponding service improvements vendors were promised. This experience may have been particularly demoralizing for the vendors in the BOTH markets, especially in light of the Crowding Out Explanation (explanation (1)).

Without additional research, it is not possible to definitively determine which of these explanations is correct. Further, it is highly likely that all of these explanations are at play to some extent. Future work can disentangle these explanations in order to determine the most effective way to implement two-pronged interventions of this kind.

G Mechanism Treatment Effects as Percent Increase over Control

Table G1 contextualizes the substantive significance of the treatment effects by presenting each as the percent increase over the control group mean.

Table G1: Treatment Effects as a Percent of Control Group

Outcome	Treatment Group	Treatment Effect	Percent Increase	Control Mean
Trust in local government	BU	0.176	6.745	2.609
Trust in local government	BOTH	0.142	5.432	2.609
Trust in ward councilors	BU	0.168	6.591	2.555
Satisfaction with services (Combined)	BU	0.293	14.672	1.999
Satisfaction with water access	BU	0.654	33.220	1.968
Satisfaction with water access	BOTH	0.315	16.024	1.968
Paying taxes is a duty	BU	0.072	1.983	3.638
Pay because of consequences	TD	0.056	1.526	3.642
Perception of Relative Amount of Taxes Reaching Gov't	BOTH	26.126	3.616	722.463
Hours [TC] working in market a day	TD	1.154	12.521	9.218

H Trust and Engagement: Additional Results

This appendix reports the additional analysis referenced in Bottom-Up Treatment Downstream Effects in the main paper. Note that Table H1 includes two unrelated models, combined in one table to save space.

The table shows that vendors in BU treatment markets who agreed to send a message against government over-reach had a lower level of trust in the local government than those who did not (a 7.55% decrease). In addition, the interaction between levels of trust in the local government and the BU only treatment group is statistically significant in the regression on whether vendors agreed with statement 1 in the whole sample, showing that individuals in this subgroup were actually slightly less likely to agree with statement 1, compared to vendors who had no trust at all in the local government.

Table H1: Political Engagement Outcomes

	Dependent variable:	
	Trust in Local Gov	Reduce Gov. Over-Reach
Agreed to Send Message	-0.213^{***} (0.051)	
BU		0.122^* (0.057)
TD		0.014 (0.058)
BOTH		0.048 (0.061)
BU * Local Gov - Not Very Trustworthy		-0.103 (0.096)
BU * Local Gov - Somewhat Trustworthy		-0.023 (0.065)
BU * Local Gov - Very Trustworth		-0.153^* (0.072)
TD * Local Gov - Not Very Trustworthy		-0.054 (0.093)
TD * Local Gov - Somewhat Trustworthy		0.036 (0.069)
TD * Local Gov - Very Trustworth		-0.103 (0.073)
BOTH * Local Gov - Not Very Trustworthy		-0.062 (0.101)
BOTH * Local Gov - Somewhat Trustworthy		0.051 (0.068)
BOTH * Local Gov - Very Trustworth		-0.018 (0.072)
Observations	1,263	2,509
Adjusted R^2	0.011	0.318

Notes: *p<0.05; **p<0.01; ***p<0.001

Models include enumerator and block fixed-effects.

Models have SEs clustered on market.

I Spillovers

I.1 Introduction

Spillovers are a possibility in all experiments. In order to assess the extent to which treatment spillover is enhancing or diminishing the effects of the interventions, we employ two approaches: an inverse probability weighting (IPW) approach and a treatment externalities approach based on Miguel and Kremer (2004).

We use two approaches because the IPW approach, while canonical and useful, is somewhat of a poor fit for our situation, as our treated units (markets) are a level higher than the observed units (vendors). Even when we use the individual level data on other markets in which vendors sell, we can only get an endline market-level measure of spillover potential, not an individual-level one, because we do not have panel data (see the next section for a more in-depth explanation). The treatment externalities approach allows us to take into account market size and how that may be affecting spillover effects (termed "treatment externalities" by Miguel and Kremer (2004), hence the name).

I.2 IPW Approach

With inverse probability weighting, "units are weighted by the inverse of the probability of being in the condition that they are in."[11] It requires making an assumption about where spillovers occur. In our case, we think about spillovers occurring geographically. If two markets are close to one another, it is possible the vendors from those markets actually visit or work in both markets. If the two markets have been assigned to different treatments, then those treatments may have "spilled over" between the two markets. For example, a vendor in a control market who also sells in a bottom-up market may observe the infrastructure project there and may then change their attitudes and behavior in a similar way to a vendor in the bottom-up market.

We assume that spillovers will only occur within a certain distance around each market. We then use the distance between markets to create an adjacency matrix. An adjacency matrix allows us to state mathematically whether individuals (or treated units) are connected (geographically, in our case) to another treated unit. We use the adjacency matrix to determine the actual treatment condition of a market - which is a composite of assigned and spillover conditions. There are 32 possible conditions - eight for each "pure" condition. For example, a market could be assigned to the bottom-up treatment, but it could be within x km of another market that was assigned to the top-down treatment. This market would then be in the "Bottom-Up_Top-Down" spillover condition group. We then simulate treatment assignment 10,000 times and calculate the number of times each market falls into each possible treatment condition. This gets us an estimate of the probability a market is in each possible treatment condition.

---

[11]https://egap.org/methods-guides/10-things-you-need-know-about-spillovers

We use multiple adjacency matrices, which get us different probabilities and therefore different weights. A traditional adjacency matrix is an NxN indicator matrix, where N is the number of units, and where the cell [i,j] is 1 if unit i is adjacent to unit j and 0 otherwise.

We know the distance between each of our markets (except for Linjidzi, for which we are missing GPS coordinates). We also have information on where a portion (approx. 20%) of our sample sold in addition to the market in which they were interviewed.[12] We use three different versions of the adjacency matrix for IPW, combining these two data sources:

1. Distance only: a NxJ matrix, where N is the number of respondents and J is the number of markets – 1 if market j is within d distance of respondent's market, 0 otherwise.

2. Other Market Selling: a n X J matrix, where n is the number of respondents in our subsample (vendors who completed the long survey), and J is the number of markets. Cell [i,j] is 1 if respondent i says they sell in market j.

3. Distance + Other Market Selling: this is once again an n x J matrix. We first add together the adjacency matrices for 1. and 2. If $A_{i,j_1} + A_{i,j_2} > 0$, cell [i,j] in this adjacency matrix takes a value of 1. If 0, it remains 0.

We use distances of two km, five km, and ten km. In each case, our results are only accurate if there is no spillover outside of that distance. In effect, this results in a sensitivity analysis: What happens when the spillover radius increases? With three distances and three different types of adjacency matrices, we end up with seven different adjacency matrices (as 2. above does not depend on distance).

We create 2. and 3. using baseline survey responses. We do this because we were concerned that responses to the question might have been affected by the intervention itself. To then incorporate this information into the endline analysis, we average the probabilities of being in the **modal** treatment condition among the market's respondents. This results in a market average. This means that for all models, all individuals within a market receive the same weights. We do this because we do not have a panel.

In our context, the IPW approach has some significant limitations. When we only use the distance between markets, we assume that *all* vendors are equally likely to go sell in nearby markets. Our data tell us this is very likely not the case. However, because we do not have a panel survey, when we incorporate individual responses, we are still forced to consider all individuals as having equal probability of being in the condition in which they are.

To account for spillovers in our main analysis, we first drop all markets that are currently in a spillover condition, and then weight individuals by the inverse of the probability that their market is in the pure treatment condition. We repeat this with the various probabilities calculated using our different adjacency matrices.

We do this for our main outcomes, with results shown in tables I1, I2, and I3.

---

[12]Vendors who completed the *long* survey were asked "Do you sell in any markets other than this one?" Those who responded *yes*, were then asked "what are the names of those markets?" An individual who noted a market within our sample was then "connected" to that market.

Table I1: Spillover Analyses, Self-Reported Compliance

	Dependent variable:						
				Self-Rep. Comp.			
	D2	D2 - Mixed	D5	D5 - Mixed	D10	D10 - Mixed	Ind. Only
BU	0.103	0.095	0.076	0.080	0.106	0.098	0.097
	(0.079)	(0.093)	(0.079)	(0.093)	(0.115)	(0.113)	(0.092)
TD	0.161*	0.193*	0.158*	0.194*	0.266*	0.314*	0.180*
	(0.077)	(0.085)	(0.079)	(0.086)	(0.128)	(0.126)	(0.084)
BOTH	0.021	0.064	−0.038	0.028	−0.030	−0.002	0.063
	(0.097)	(0.097)	(0.099)	(0.099)	(0.122)	(0.121)	(0.096)
Observations	11,568	10,835	10,906	10,317	5,804	5,606	10,990
Adjusted R^2	0.116	0.125	0.111	0.123	0.103	0.102	0.123

Notes:

*p<0.05; **p<0.01; ***p<0.001
Individual-level models include enumerator
and block fixed effects.
Individual-level models have SEs clustered
on market.

Table I2: Spillover Analyses, Perceived Group Tax Compliance

	Dependent variable:						
				Per. Group Comp.			
	D2	D2 - Mixed	D5	D5 - Mixed	D10	D10 - Mixed	Ind. Only
BU	0.154	0.049	0.127	0.024	−0.139	−0.043	0.048
	(0.130)	(0.156)	(0.139)	(0.158)	(0.160)	(0.177)	(0.154)
TD	0.024	0.023	0.021	0.018	0.100	0.145	0.011
	(0.115)	(0.128)	(0.118)	(0.130)	(0.108)	(0.111)	(0.127)
BOTH	0.023	0.051	−0.038	0.019	−0.255	−0.189	0.047
	(0.143)	(0.145)	(0.150)	(0.152)	(0.165)	(0.153)	(0.144)
Observations	12,037	11,280	11,354	10,754	6,022	5,821	11,438
Adjusted R^2	0.116	0.121	0.121	0.124	0.145	0.132	0.121

Notes:

*p<0.05; **p<0.01; ***p<0.001
Individual-level models include enumerator
and block fixed effects.
Individual-level models have SEs clustered
on market.

Table I3: Spillover Analyses, Evidence of Recent Receipt

				Dependent variable:			
				Evidence of Recent Receipt			
	D2	D2 - Mixed	D5	D5 - Mixed	D10	D10 - Mixed	Ind. Only
BU	0.104***	0.094**	0.106**	0.092**	0.028	−0.019	0.092**
	(0.031)	(0.031)	(0.032)	(0.032)	(0.047)	(0.042)	(0.031)
TD	0.071*	0.097**	0.078*	0.103***	−0.008	0.025	0.099**
	(0.031)	(0.031)	(0.032)	(0.031)	(0.032)	(0.023)	(0.031)
BOTH	0.051	0.044	0.030	0.029	−0.017	−0.001	0.046
	(0.032)	(0.032)	(0.030)	(0.032)	(0.036)	(0.039)	(0.032)
Observations	12,108	11,348	11,422	10,820	6,037	5,836	11,506
Adjusted R^2	0.262	0.260	0.286	0.278	0.314	0.284	0.264

Notes:

*p<0.05; **p<0.01; ***p<0.001
Individual-level models include enumerator
and block fixed-effects.
Individual-level models have SEs clustered
on market.

I.3 Treatment Externalities Approach

This approach is described in more depth in Miguel and Kremer (2004). We assume that spillovers are a function of the number of vendors or number of markets of a certain treatment condition within a certain distance from each market; the more vendors there are at nearby markets or, more roughly, the more nearby markets there are, the more likely it is that the respondent will have heard about the treatment.

This amounts to fitting the following model:

$$Y_{ijkl} = \beta_0 + \beta_1 * BU_j + \beta_2 * TD_j + \beta_3 * BOTH_j +$$
$$\sum_d (\gamma_d * N_{dj}^{BU}) + \sum_d (\xi_d * N_{dj}^{TD}) + \sum_d (\zeta_d * N_{dj}^{BOTH}) + \sum_d (\phi_d * N_{dj}) +$$
$$\beta_k * ENUM_k + \beta_l * Block_l + \epsilon_{ijkl}$$

where N_{dj} is the total number in markets at distance d from market j, including market j itself, and N_{dj}^{BU}, N_{dj}^{TD}, and N_{dj}^{BOTH} are the numbers in markets assigned to the BU, TD, and BOTH treatments at distance d from market j, respectively. To create the various N_{dj}, we add up

1. A daily average of the number of vendors who sell in a market

2. The maximum number of vendors who sell in a market during a week

3. The number of markets itself

We use the same distances as we do in the IPW approach: two km, five km, and ten km.

Table I4: Treatment Externalities, Self-Reported Tax Compliance

| | Dependent variable: | | |
| | Self-Rep. Compliance | | |
	Avg. Vend. pr. Day	Max Num. Vendors	Num. Mkts.
BU	0.109 (0.091)	0.106 (0.089)	0.201 (0.116)
TD	0.127 (0.089)	0.124 (0.092)	0.075 (0.118)
BOTH	-0.031 (0.106)	-0.036 (0.108)	0.017 (0.144)
N-2-BU			
N-5-BU	-0.002^{***} (0.002)	-0.001 (0.001)	-0.178 (0.174)
N-10-BU	0.00004 (0.0001)	0.00000 (0.00003)	-0.102 (0.081)
N-2-TD	-0.002 (0.002)	-0.001 (0.001)	-0.340^{***} (0.083)
N-5-TD	0.005^{***} (0.003)	0.003 (0.001)	0.043 (0.219)
N-10-TD	0.0001 (0.0003)	0.00005 (0.0001)	0.096 (0.089)
N-2-BOTH	0.001 (0.001)	0.0004 (0.0005)	0.185 (0.152)
N-5-BOTH	-0.0001 (0.001)	0.0001 (0.0005)	0.378^{*} (0.185)
N-10-BOTH	0.0001 (0.0001)	0.0001 (0.00003)	-0.042 (0.092)
N-2-All	-0.001 (0.001)	-0.0004 (0.0004)	
N-5-All	0.001^{***} (0.001)	0.0004 (0.0004)	0.119 (0.130)
N-10-All	-0.0002 (0.0001)	-0.0001^{**} (0.00002)	-0.037 (0.068)
Observations	11,723	11,723	11,723
Adjusted R^2	0.117	0.117	0.119

Notes:
*p<0.05; **p<0.01; ***p<0.001
Individual-level models include enumerator
and block fixed-effects.
Individual-level models have SEs clustered
on market.

Table I5: Treatment Externalities, Group-Perceived Tax Compliance

	Dependent variable:		
	Group-Per. Compliance		
	Avg. Vend. pr. Day	Max Num. Vendors	Num. Mkts.
BU	0.116 (0.163)	0.117 (0.157)	0.078 (0.168)
TD	0.061 (0.134)	0.055 (0.143)	0.029 (0.167)
BOTH	−0.007 (0.162)	−0.021 (0.167)	−0.087 (0.201)
N-2-BU			
N-5-BU	−0.001*** (0.003)	−0.0002 (0.001)	−0.307 (0.269)
N-10-BU	0.0001 (0.0001)	0.00005 (0.00004)	0.106 (0.112)
N-2-TD	0.001 (0.004)	−0.0002 (0.001)	−0.268 (0.139)
N-5-TD	0.010*** (0.005)	0.005* (0.002)	0.198 (0.318)
N-10-TD	−0.0003 (0.0004)	−0.0001 (0.0002)	0.00005 (0.118)
N-2-BOTH	0.003 (0.002)	0.001 (0.001)	0.121 (0.275)
N-5-BOTH	−0.001*** (0.002)	−0.0002 (0.001)	0.366 (0.285)
N-10-BOTH	0.00002 (0.0001)	0.00002 (0.00005)	0.077 (0.116)
N-2-All	−0.003 (0.002)	−0.001 (0.001)	
N-5-All	0.003*** (0.002)	0.001 (0.001)	0.311 (0.195)
N-10-All	−0.00003 (0.0001)	−0.00001 (0.00002)	−0.048 (0.083)
Observations	12,195	12,195	12,195
Adjusted R^2	0.116	0.116	0.117

Notes: *p<0.05; **p<0.01; ***p<0.001
Individual-level models include enumerator and block fixed-effects.
Individual-level models have SEs clustered on market.

28

Table I6: Treatment Externalities, Evidence of Recent Receipt

	Dependent variable:		
	Evidence of Recent Receipt		
	Avg. Vend. pr. Day	Max Num. Vendors	Num. Mkts.
BU	0.074* (0.034)	0.075* (0.034)	0.076 (0.044)
TD	0.034 (0.032)	0.033 (0.032)	0.003 (0.042)
BOTH	0.018 (0.031)	0.018 (0.031)	0.011 (0.045)
N-2-BU			
N-5-BU	−0.001*** (0.001)	−0.0004 (0.0002)	−0.230*** (0.065)
N-10-BU	0.0002 (0.00003)	0.0001*** (0.00001)	0.050 (0.037)
N-2-TD	0.002 (0.001)	0.001** (0.0004)	0.165*** (0.038)
N-5-TD	−0.002*** (0.001)	−0.001 (0.0005)	−0.336*** (0.091)
N-10-TD	0.0002 (0.0001)	0.0001 (0.0001)	0.094* (0.039)
N-2-BOTH	0.001* (0.001)	0.0005* (0.0002)	0.269*** (0.056)
N-5-BOTH	−0.001*** (0.001)	−0.0002 (0.0002)	0.072 (0.086)
N-10-BOTH	0.0001 (0.00004)	0.00004** (0.00001)	0.028 (0.038)
N-2-All	−0.001 (0.001)	−0.0004* (0.0002)	
N-5-All	0.001*** (0.001)	0.0004* (0.0002)	0.201*** (0.053)
N-10-All	−0.0001 (0.00002)	−0.00004*** (0.00001)	−0.037 (0.029)
Observations	12,266	12,266	12,266
Adjusted R^2	0.277	0.276	0.275

Notes: *p<0.05; **p<0.01; ***p<0.001
Individual-level models include enumerator
and block fixed-effects.
Individual-level models have SEs clustered
on market.

J   Compliance Analysis

In this section, we estimate the so-called local average treatment effect (LATE), also known as the effect on compliers, using an instrumental variables strategy. We use treatment assignment as an instrument for treatment compliance. We operationalize treatment compliance in two ways, using the same set of compliance variables.

We say that a bottom-up treatment market had a compliance issue if it had one of three possible problems: 1) if endline data collection occurred before mobilization for its infrastructure project had started; 2) if a vendor from a given market sent in a grievance message that was not responded to; and 3) if a market did not receive the infrastructure project it had been promised after mobilization began (there are multiple reasons for this, the most prevalent being that a borehole was drilled but no water was found).

We say that a top-down treatment market had a compliance issue if it had one of three possible problems: 1) if a market met its target but did not receive an incentive; 2) if the incentive a market was supposed to receive arrived delayed; and 3) if mobile money was not active in any given month after May 2018.

We consider a market as having had a compliance issue under the *strict* operationalization when it had at least one of the three issues. We consider a market as having had a compliance issue under the *relaxed* operationalization only when it had all three issues. The tables below present results for these models, focusing on our main outcome of the receipt measure.

Table J1: Compliance IV Regression 2nd-Stage Treatment Group Approach

	Dependent variable:					
	Self-Rep. Compl		Per. Group Compl.		Recent Rcpt.	
	Strict	Relaxed	Strict	Relaxed	Strict	Relaxed
BU - Str.	0.320		0.566		0.307*	
	(0.246)		(0.400)		(0.127)	
TD - Str.	0.407		0.131		0.202	
	(0.233)		(0.332)		(0.103)	
BOTH - Str.	−1.202		−0.897		−0.508	
	(1.055)		(1.512)		(0.434)	
BU - Rel.		0.130		0.214		0.116**
		(0.086)		(0.144)		(0.036)
TD - Rel.		0.179*		0.056		0.084*
		(0.085)		(0.126)		(0.034)
BOTH - Rel.		0.011		0.052		0.050
		(0.113)		(0.173)		(0.037)
Observations	11,822	11,822	12,294	12,294	12,365	12,365
Adjusted R^2	0.097	0.112	0.110	0.115	0.196	0.261

Notes: *p<0.05; **p<0.01; ***p<0.001
Individual-level models include enumerator
and block fixed-effects.
Individual-level models have SEs clustered
on market.

Table J2: Compliance IV Regression 2-Stage Treatment Group Approach Market Level DIM

| | Self-Rep. Compl | | Per. Group Compl. | | Recent Rcpt. | |
	Strict	Relaxed	Strict	Relaxed	Strict	Relaxed
BU - Str.	0.233		0.586		0.309	
	(0.354)		(0.489)		(0.156)	
TD - Str.	0.456		0.125		0.146	
	(0.324)		(0.448)		(0.143)	
BOTH - Str.	−1.075		−0.507		−0.416	
	(1.187)		(1.640)		(0.524)	
BU - Rel.		0.088		0.222		0.117*
		(0.123)		(0.178)		(0.047)
TD - Rel.		0.189		0.052		0.060
		(0.123)		(0.178)		(0.047)
BOTH - Rel.		0.012		0.100		0.050
		(0.129)		(0.188)		(0.050)
Observations	128	128	128	128	128	128
Adjusted R^2	0.087	0.235	0.058	0.132	0.292	0.550

Dependent variable:

Notes: *p<0.05; **p<0.01; ***p<0.001
Models include block fixed effects.

Table J3: Compliance IV Regression 2nd-Stage Treatment Group Approach Market Level DID

	Self-Rep. Compl		Per. Group Compl.		Recent Rcpt.	
	Strict	Relaxed	Strict	Relaxed	Strict	Relaxed
BU - Str.	−0.220 (0.489)		−0.047 (0.655)		0.316 (0.176)	
TD - Str.	0.303 (0.449)		−0.149 (0.600)		0.089 (0.162)	
BOTH - Str.	−0.334 (1.642)		−0.166 (2.196)		−0.388 (0.591)	
BU - Rel.		−0.084 (0.170)		−0.018 (0.238)		0.120* (0.055)
TD - Rel.		0.125 (0.170)		−0.062 (0.238)		0.037 (0.055)
BOTH - Rel.		−0.034 (0.179)		−0.059 (0.251)		0.045 (0.058)
Observations	256	256	256	256	256	256
Adjusted R^2	0.080	0.227	0.163	0.233	0.408	0.594

Dependent variable:

Notes: $^*p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$
Models include block fixed effects.

K Multilevel Modeling Analysis

Table K1: Multilevel Models

	Dependent variable:		
	Self-Rep. Compl	Per. Group Compl.	Recent Rcpt.
BU	0.073	0.139	0.100**
	(0.108)	(0.166)	(0.038)
TD	0.163	0.050	0.063
	(0.108)	(0.166)	(0.038)
BOTH	0.037	0.057	0.056
	(0.108)	(0.166)	(0.038)
Observations	11,822	12,294	12,365

Notes: $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$
Models include random intercepts by enumerators and
by markets nested in blocks nested in districts.

L Heterogeneous Treatment Effects Analysis

This section presents analysis of potential heterogeneous treatment effects at the individual and market levels.

L.1 Vendor-Level Heterogeneity

In this section, we fit only individual endline DiM models for the main outcomes, as vendor covariates are at the individual level. For binary variables (gender, service vs. goods stall type, selling daily), we fit subgroup models as well as interaction models.

Table L1: Subgroup Analysis by Gender

	Self-Reported Compliance			Perceived Group Compliance			Evidence of Recent Receipt		
	Male	Female	Int.	Male	Female	Int.	Male	Female	Int.
Female			−0.093			0.060			0.087***
			(0.064)			(0.104)			(0.021)
BU	0.144	0.105	0.137	0.140	0.303	0.143	0.099**	0.096**	0.102**
	(0.081)	(0.097)	(0.083)	(0.141)	(0.155)	(0.143)	(0.031)	(0.034)	(0.033)
TD	0.150*	0.182	0.145	0.036	0.092	0.015	0.094**	0.038	0.093**
	(0.073)	(0.102)	(0.075)	(0.111)	(0.147)	(0.118)	(0.030)	(0.035)	(0.031)
BOTH	0.035	0.047	0.031	0.029	0.111	0.021	0.059	0.042	0.061
	(0.090)	(0.123)	(0.092)	(0.131)	(0.194)	(0.136)	(0.030)	(0.034)	(0.031)
Female * BU			−0.034			0.129			−0.003
			(0.090)			(0.139)			(0.030)
Female * TD			0.038			0.110			−0.049
			(0.085)			(0.145)			(0.029)
Female * BOTH			0.020			0.122			−0.015
			(0.089)			(0.157)			(0.030)
Observations	7,698	4,124	11,822	8,029	4,265	12,294	8,068	4,297	12,365
Adjusted R^2	0.101	0.140	0.114	0.120	0.113	0.116	0.262	0.287	0.271

Note: $^*p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Table L2: Subgroup Analysis by Services vs Goods

	Self-Reported Compliance			Perceived Group Compliance			Evidence of Recent Receipt		
	Goods	Services	Int.	Goods	Services	Int.	Goods	Services	Int.
Service			−0.461***			−0.148			−0.041
			(0.130)			(0.139)			(0.035)
BU	0.088	0.465*	0.086	0.161	0.688**	0.160	0.105***	0.012	0.105***
	(0.079)	(0.219)	(0.080)	(0.134)	(0.232)	(0.134)	(0.031)	(0.051)	(0.031)
TD	0.163*	0.128	0.156*	0.062	0.059	0.059	0.078*	0.013	0.079**
	(0.073)	(0.204)	(0.073)	(0.115)	(0.208)	(0.115)	(0.031)	(0.040)	(0.030)
BOTH	0.033	0.042	0.029	0.075	−0.262	0.076	0.060	−0.010	0.060*
	(0.093)	(0.201)	(0.092)	(0.144)	(0.213)	(0.145)	(0.031)	(0.042)	(0.031)
Service * BU			0.353			0.503*			−0.024
			(0.189)			(0.218)			(0.049)
Service * TD			0.066			−0.070			−0.029
			(0.164)			(0.226)			(0.044)
Service * BOTH			0.085			−0.131			−0.032
			(0.177)			(0.201)			(0.047)
Observations	10,874	948	11,822	11,223	1,071	12,294	11,285	1,080	12,365
Adjusted R^2	0.113	0.160	0.118	0.116	0.130	0.116	0.275	0.172	0.267

Note: $^*p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Table L3: Subgroup Analysis by Selling Daily or Not

	Self-Reported Compliance			Perceived Group Compliance			Evidence of Recent Receipt		
	Not Daily	Daily	Int.	Not Daily	Daily	Int.	Not Daily	Daily	Int.
Sell Daily			−0.196			0.184			0.043
			(0.113)			(0.108)			(0.035)
BU	0.108	0.083	0.088	0.222	−0.034	0.222	0.112***	0.047	0.114***
	(0.075)	(0.150)	(0.080)	(0.127)	(0.199)	(0.133)	(0.032)	(0.040)	(0.034)
TD	0.098	0.332*	0.079	0.045	0.024	0.034	0.076*	0.068	0.074*
	(0.067)	(0.145)	(0.071)	(0.115)	(0.170)	(0.122)	(0.033)	(0.038)	(0.034)
BOTH	0.034	0.024	0.017	0.098	−0.107	0.093	0.053	0.030	0.057
	(0.079)	(0.189)	(0.082)	(0.132)	(0.223)	(0.132)	(0.031)	(0.039)	(0.031)
Sell Daily * BU			0.126			−0.113			−0.038
			(0.147)			(0.150)			(0.047)
Sell Daily * TD			0.288*			0.040			0.001
			(0.138)			(0.166)			(0.046)
Sell Daily * BOTH			0.078			−0.110			0.0004
			(0.166)			(0.173)			(0.045)
Observations	8,491	3,330	11,821	8,719	3,574	12,293	8,780	3,584	12,364
Adjusted R^2	0.119	0.150	0.114	0.128	0.098	0.116	0.285	0.248	0.266

Note: *p<0.05; **p<0.01; ***p<0.001

Table L4: Wealth Heterogeneous Effects Analysis

	Self-Reported Compliance	Perceived Group Compliance	Evidence of Recent Receipt
HH Income	0.00000	0.00000	−0.000
	(0.00000)	(0.00000)	(0.00000)
BU	0.080	0.193	0.106**
	(0.083)	(0.140)	(0.034)
TD	0.138	0.077	0.084**
	(0.077)	(0.126)	(0.032)
BOTH	−0.019	0.095	0.071*
	(0.100)	(0.164)	(0.033)
HH Income * BU	0.00000	−0.00000	−0.000
	(0.00000)	(0.00000)	(0.00000)
HH Income * TD	0.00000	−0.00000	−0.00000
	(0.00000)	(0.00000)	(0.00000)
HH Income * BOTH	0.00000	−0.00000	−0.00000
	(0.00000)	(0.00000)	(0.00000)
Observations	11,713	12,180	12,250
Adjusted R^2	0.117	0.115	0.267

Note: *p<0.05; **p<0.01; ***p<0.001

L.2 Market-Level Heterogeneity

At the market level, we analyze heterogeneous effects by market size (using baseline data) and vendors' collective action propensity. We measure market-level collective action propensity by taking the market-level average of an endline survey question that asked: "Do you agree or disagree with the following statement? When there is a problem in this market, we work together to solve it." Responses were on four-point scale that we then normalized; higher numbers indicate more agreement with the question - a higher propensity for collective action.

Table L5: Het. Treatment Effects by Market Size

| | Self-Reported Compliance | | | Perceived Group Compliance | | | Evidence of Recent Receipt | | |
	Mkt. DID	Mkt DIM	Ind. DIM	Mkt. DID	Mkt DIM	Ind. DIM	Mkt. DID	Mkt DIM	Ind. DIM
Market Size	0.00001	0.00003	0.00004	−0.00002	−0.00002	0.000000	0.000003	0.00001	0.00001
	(0.0001)	(0.0001)	(0.00004)	(0.0002)	(0.0001)	(0.0001)	(0.00003)	(0.00003)	(0.00003)
BU	−0.1167	0.0491	0.0992	−0.1380	0.1331	0.1577	0.0952	0.1000*	0.1118**
	(0.1848)	(0.1337)	(0.0904)	(0.2679)	(0.1976)	(0.1717)	(0.0524)	(0.0497)	(0.0368)
TD	0.1455	0.1434	0.1558	−0.1095	0.0371	0.0761	−0.0043	0.0007	0.0588
	(0.1935)	(0.1400)	(0.0932)	(0.2805)	(0.2069)	(0.1399)	(0.0549)	(0.0520)	(0.0360)
BOTH	0.0197	0.1431	0.1718	−0.0436	0.1986	0.1905	0.0196	0.0314	0.0760*
	(0.1803)	(0.1304)	(0.1037)	(0.2613)	(0.1927)	(0.1629)	(0.0511)	(0.0485)	(0.0360)
Market Size * BU	0.00005	0.00003	0.000004	0.0001	0.0001	0.00003	0.00001	0.00001	−0.00001
	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.00004)	(0.00004)	(0.00003)
Market Size * TD	−0.00004	0.00003	0.000003	0.0001	0.00002	−0.00002	0.00005	0.0001	0.00002
	(0.0002)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.00005)	(0.00005)	(0.00003)
Market Size * BOTH	−0.0001	−0.0002	−0.0002**	−0.00002	−0.0002	−0.0002*	0.0001	0.00004	−0.00003
	(0.0002)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.00005)	(0.00004)	(0.00003)
Observations	128	128	11,822	128	128	12,294	128	128	12,365
Adjusted R^2	0.0151	0.2515	0.1173	−0.0053	0.1172	0.1166	0.2482	0.5882	0.2670

Notes:

*p<0.05; **p<0.01; ***p<0.001
Individual-level models include enumerator and block fixed-effects.
Individual-level models have SEs clustered on market.
Market-level models include block fixed-effects.

Table L6: Het. Treatment Effects by Collective Action Propensity

	Self-Reported Compliance	Perceived Group Compliance	Evidence of Recent Receipt
Collective Action Propensity	−0.046	−0.024	−0.037
	(0.118)	(0.177)	(0.037)
BU	−0.186	−0.151	0.107*
	(0.149)	(0.222)	(0.046)
TD	0.076	−0.104	0.044
	(0.145)	(0.218)	(0.045)
BOTH	−0.031	−0.120	0.047
	(0.148)	(0.221)	(0.046)
Collective Action Propensity * BU	0.465**	0.504*	0.079
	(0.160)	(0.240)	(0.050)
Collective Action Propensity * TD	−0.193	−0.236	0.034
	(0.156)	(0.233)	(0.049)
Collective Action Propensity * BOTH	0.145	0.348	0.098
	(0.184)	(0.276)	(0.058)
Observations	128	128	128
Adjusted R^2	0.177	0.105	0.236

Note: *p<0.05; **p<0.01; ***p<0.001

All models market-level DID models.

All models include block fixed-effects.

M Linear Hypothesis Tests for Main Paper Models

To compare the treatment effect sizes between treatment groups and to test for the additive nature of the BU and TD treatments in the context of the BOTH treatment group, we carried out individual Wald tests for the corresponding linear restraints (BU = TD, BU = BOTH, TD = BOTH, BU + TD = BOTH), reported in Tables M1 through M5. Note that the p-values from Table M1 are shown in Panel B of Table 2 in the main paper.

Table M1: p-Values for Linear Hypothesis (Wald) Tests for Receipt Outcome Models in Table 2

Linear Hypothesis	Receipt Outcome (Individual DIM)	Receipt Outcome (Market DID)
BU = TD	0.326	0.089
BU = BOTH	0.117	0.179
TD = BOTH	0.524	0.715
BU + TD = BOTH	0.004	0.137

Table M2: p-Values for Linear Hypothesis (Wald) Tests for Bottom-Up Mechanism Models in Table 3

Linear Hypothesis	Trust in Local Government	Trust in Ward Councilors	District Manages Funds Well	District Spending Transparent	District Tax Collection Transparent
BU = TD	0.005	0.000	0.216	0.473	0.623
BU = BOTH	0.497	0.353	0.645	0.514	0.829
TD = BOTH	0.014	0.001	0.389	0.817	0.382
BU + TD = BOTH	0.673	0.579	0.671	0.480	0.871

Table M3: p-Values for Linear Hypothesis (Wald) Tests for Bottom-Up Mechanism Models in Table 4

Linear Hypothesis	Satisfaction with Services	Satisfaction with Water Access	Perception of Relative Amount Spent on Services	Paying Taxes Is Duty	Vendors Should Pay Taxes
BU = TD	0.030	0.002	0.192	0.321	0.656
BU = BOTH	0.195	0.056	0.063	0.413	0.103
TD = BOTH	0.455	0.312	0.638	0.921	0.184
BU + TD = BOTH	0.088	0.024	0.097	0.130	0.423

Table M4: Linear Hypothesis (Wald) Tests for Political Engagement Outcome Models in Table 5

Linear Hypothesis	Vote	Sign Petition Anonymously	Sign Petition with Name	Reduce Gov. Over-reach	Increase Gov. Effectiveness
BU = TD	0.798	0.001	0.003	0.008	0.070
BU = BOTH	0.172	0.665	0.658	0.878	0.428
TD = BOTH	0.271	0.001	0.002	0.007	0.008
BU + TD = BOTH	0.416	0.523	0.627	0.986	0.358

Table M5: p-Values for Linear Hypothesis (Wald) Tests for Top-Down Mechanism Models in Tables 6-7

Linear Hypothesis	Ind'l Evasion Possible	Group Evasion Possible	Pay Because Consequences	Perception of Relative Amount of Taxes Reaching Gov't	Hours Working in Market per Day	Vendors Visited in Market per Day
BU = TD	0.953	0.432	0.383	0.157	0.113	0.557
BU = BOTH	0.920	0.176	0.840	0.576	0.587	0.232
TD = BOTH	0.859	0.034	0.547	0.018	0.263	0.406
BU + TD = BOTH	0.410	0.089	0.158	0.560	0.245	0.843

N General Robustness Models

N.1 Main Outcomes

Figure N1: Evidence of Recent Receipt: Difference between Baseline and Endline by each market in each treatment group.
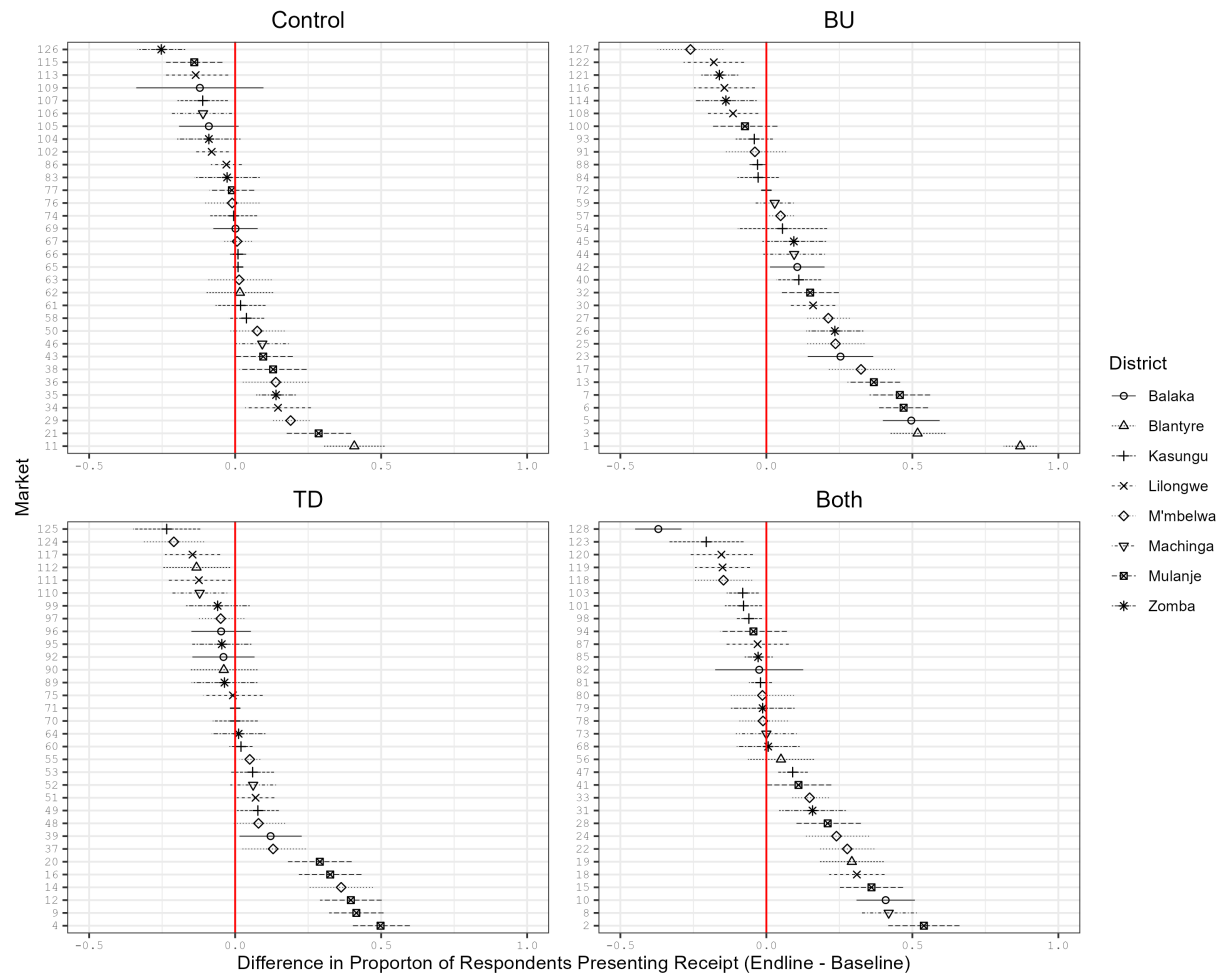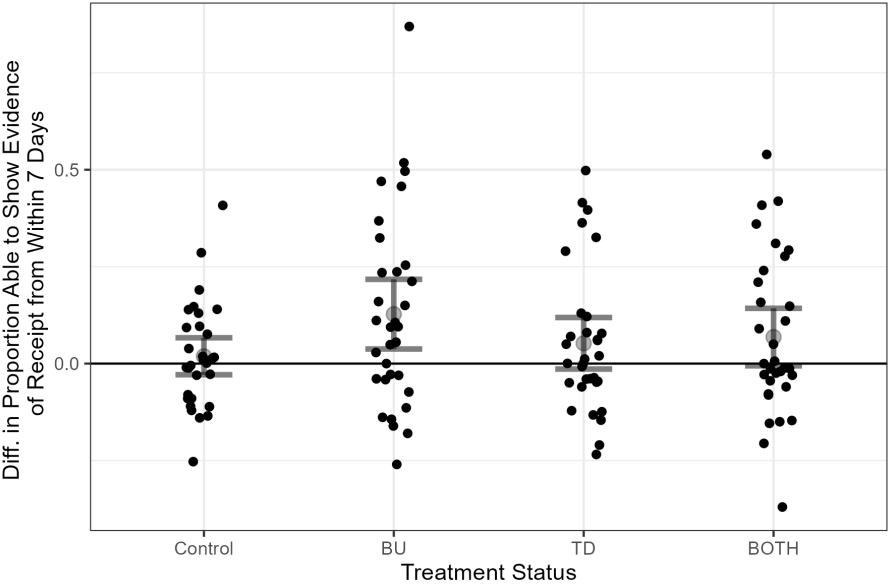
Figure N2: Evidence of Recent Receipt: Difference between Baseline and Endline

N.1.1 Main Outcomes - Self-Reported and Perceived Group

Table N1: Hypothesis 1 Results Table - Individual-Level DIM and Market-Level DID

Panel A: Individual-Level DIM Models

	Self-Reported Full Tax Compliance	Perception of Others' Always Complying	Evidence of Receipt from Past 7 Days
BU	0.119	0.194	0.104***
	(0.079)	(0.132)	(0.031)
TD	0.158*	0.050	0.076*
	(0.075)	(0.114)	(0.030)
BOTH	0.037	0.064	0.057
	(0.094)	(0.142)	(0.031)
Observations	11,822	12,294	12,365
Adjusted R^2	0.113	0.115	0.266

Panel B: Market-Level DID Models

	Self-Reported Full Tax Compliance	Perception of Others' Always Complying	Evidence of Receipt from Past 7 Days
BU	−0.076	−0.016	0.109*
	(0.150)	(0.217)	(0.044)
TD	0.114	−0.056	0.033
	(0.150)	(0.217)	(0.044)
BOTH	−0.023	−0.054	0.049
	(0.150)	(0.217)	(0.044)
Observations	128	128	128
Adjusted R^2	0.049	0.024	0.231

Notes

*p<0.05; **p<0.01; ***p<0.001

Individual-level models include enumerator and block fixed effects

Individual-level models have SEs clustered on market.

Market-level models include block fixed effects.

N.1.2 Other Main Outcome Specifications

Table N2: Ability to Produce a Receipt Robustness Models

| | Evidence of Receipt from Past 7 Days | | | |
	Market DIM	Market DID	Individual DID	Individual DID Incl. Lagged DV
BU	0.106*	−0.002	−0.003	0.107***
	(0.042)	(0.035)	(0.025)	(0.031)
TD	0.055	0.021	0.021	0.064*
	(0.042)	(0.035)	(0.023)	(0.029)
BOTH	0.056	0.007	0.005	0.053
	(0.042)	(0.035)	(0.019)	(0.030)
recent_receipt_7_bl_avg				0.426**
				(0.139)
Endline:BU		0.109*	0.111*	
		(0.049)	(0.050)	
Endline:TD		0.033	0.037	
		(0.049)	(0.040)	
Endline:BOTH		0.049	0.049	
		(0.049)	(0.042)	
Observations	128	256	24,736	12,365
Adjusted R²	0.568	0.607	0.155	0.269

Notes:
*p<0.05; **p<0.01; ***p<0.001
Individual-level models include enumerator
and block fixed-effects.
Individual-level models have SEs clustered
on market.
Market-level models include block fixed-effects.

Table N3: Self-Reported Tax Compliance Robustness Models

| | Self-Reported Full Compliance | | | |
	Market DIM	Market DID	Individual DID	Individual DID Incl. Lagged DV
BU	0.080	0.156	0.144	0.104
	(0.111)	(0.108)	(0.088)	(0.076)
TD	0.171	0.057	0.045	0.145*
	(0.111)	(0.108)	(0.103)	(0.073)
BOTH	0.036	0.058	0.041	0.033
	(0.111)	(0.108)	(0.080)	(0.093)
fee1_full_bl_avg				0.133
				(0.090)
Endline:BU		−0.076	−0.007	
		(0.153)	(0.118)	
Endline:TD		0.114	0.120	
		(0.153)	(0.124)	
Endline:BOTH		−0.023	−0.002	
		(0.153)	(0.133)	
Observations	128	256	24,181	11,822
Adjusted R^2	0.242	0.236	0.032	0.114

Notes:

*p<0.05; **p<0.01; ***p<0.001
Individual-level models include enumerator
and block fixed-effects.
Individual-level models have SEs clustered
on market.
Market-level models include block fixed-effects.

Table N4: Perceived Group Tax Compliance Robustness Models

| | | Group Perceived as Always Complying | | |
	Market DIM	Market DID	Individual DID	Individual DID Incl. Lagged DV
BU	0.202	0.218	0.225*	0.207
	(0.162)	(0.153)	(0.112)	(0.128)
TD	0.047	0.103	0.091	0.060
	(0.162)	(0.153)	(0.141)	(0.113)
BOTH	0.104	0.158	0.127	0.071
	(0.162)	(0.153)	(0.110)	(0.141)
fee2_always_bl_avg				−0.055
				(0.083)
Endline:BU		−0.016	0.013	
		(0.216)	(0.181)	
Endline:TD		−0.056	−0.047	
		(0.216)	(0.182)	
Endline:BOTH		−0.054	−0.020	
		(0.216)	(0.212)	
Observations	128	256	24,515	12,294
Adjusted R²	0.129	0.231	0.025	0.115

Notes:
*p<0.05; **p<0.01; ***p<0.001

Individual-level models include enumerator and block fixed-effects.

Individual-level models have SEs clustered on market.

Market-level models include block fixed-effects.

N.1.3 Interacting BU and TD Treatment Assignment

Table N5: Analysis as Factorial Design (w/ Int., no Int.)

	Self-Reported Full Compliance		Self-Reported Always Complying		Evidence of Receipt from Past 7 Days	
BU	0.119	−0.002	0.194	0.103	0.104***	0.042
	(0.079)	(0.061)	(0.132)	(0.094)	(0.031)	(0.022)
TD	0.158*	0.038	0.050	−0.040	0.076*	0.014
	(0.075)	(0.060)	(0.114)	(0.094)	(0.030)	(0.022)
BU:TD	−0.240*		−0.180		−0.123**	
	(0.119)		(0.195)		(0.043)	
Observations	11,822	11,822	12,294	12,294	12,365	12,365
Adjusted R^2	0.113	0.112	0.115	0.115	0.266	0.261

Notes: *p<0.05; **p<0.01; ***p<0.001
Individual-level models include enumerator
and block fixed-effects.
Individual-level models have SEs clustered
on market.

N.1.4 0s as 0s for Self-Reported and Perceived Group Tax Compliance

For the self-reported and perceived group tax compliance measures, individuals were supposed to allocate 5 and 10 tokens, respectively, into three groups. The survey software was then supposed to check that all tokens had been allocated - enumerators should not have been able to proceed if allocations added up to less than 5 or 10. However, in some instances, the survey software seemingly malfunctioned, allowing respondents to report totals of more than 5 or 10 for a single category or to report 0 for all categories. This was a larger problem for the self-reported compliance question, with ∼560 respondents dropping out of the data due to this issue. For perceived group compliance, the dropped number of respondents was smaller, at ∼100. In the main models, all of these were treated as NAs (that is, if an individual seemingly allocated none of their tokens or if they allocated more than 5 or 10 to a single category, that category's information was considered missing). To see what impact this may have had on our results, we reran the data where only nonsensical (greater than 5 or 10 or less than 0) are treated as NAs, and *all 0* outcomes are retained.

Results are very similar. Tables omitted here to save space but are available upon request.

N.1.5 Alternative Outcomes

Table N6: H1: Evidence of Receipt from Past 10 Days

| | Evidence of Receipt from Past 10 Days | | |
	Individual DIM	Market DIM	Market DID
BU	0.105***	0.107*	−0.002
	(0.031)	(0.042)	(0.035)
TD	0.076*	0.055	0.023
	(0.030)	(0.042)	(0.035)
BOTH	0.057	0.056	0.007
	(0.031)	(0.042)	(0.035)
Endline:BU			0.109*
			(0.049)
Endline:TD			0.031
			(0.049)
Endline:BOTH			0.048
			(0.049)
Observations	12,365	128	256
Adjusted R^2	0.265	0.567	0.607

Notes: *p<0.05; **p<0.01; ***p<0.001
Individual-level models include enumerator
and block fixed-effects.
Individual-level models have SEs clustered
on market.
Market-level models include block fixed-effects.

Table N7: H1: Outcome 3 - Tax Collector Does **Not** Give You A Receipt When You Pay Fee

| | No Receipt When Paying | | |
	EL DIM	BL-EL DID	EL DID (Lagged DV)
Endline		0.052**	
		(0.020)	
BU	−0.059**	−0.014	−0.059**
	(0.020)	(0.016)	(0.020)
TD	−0.004	0.010	−0.003
	(0.019)	(0.017)	(0.019)
Both	−0.054**	−0.017	−0.055**
	(0.019)	(0.017)	(0.019)
Endline:BU		−0.036	
		(0.032)	
Endline:TD		−0.020	
		(0.030)	
Endline:Both		−0.037	
		(0.029)	
Observations	2,516	5,012	2,516
Adjusted R²	0.105	0.009	0.105

Notes	*p<0.05; **p<0.01; ***p<0.001
	All models have individuals as unit of analysis.
	All include block fixed-effects.
	Endline only models include enumerator fixed-effects as well.
	All models have SEs clustered on market.
	Lagged DV model includes baseline market average of DV.
	Outcome is binary.

Measure is based on a survey question where respondents were asked how often they paid the tax, but did not get a receipt. Those who said "it happens", "it happens a lot", or "always" were coded as a 1; others were coded as zero. We view this as a rougher measure than that in the next table. Note that while some effects are statistically significant, the point estimates are smaller than the overall compliance effects.

Table N8: H1: Outcome 3 - Additional Recent Receipt Variable Robustness Models

	Age of Actual Receipt	Received Receipt Last Time Paid	Number of Days Receipt Usually Kept
BU	−1.429	0.013	−3.363
	(1.337)	(0.012)	(3.471)
TD	−0.383	−0.001	−1.361
	(1.665)	(0.013)	(3.431)
Both	−1.574	0.026*	−5.037
	(1.650)	(0.012)	(2.965)
Observations	4,180	2,498	2,413
Adjusted R^2	0.049	0.062	0.040

Notes: *p<0.05; **p<0.01; ***p<0.001
(1) All models are estimated using ordinary least squares.
(2) Models include enumerator and block fixed-effects and market-level clustered standard errors.
(3) Outcome for models 1 and 3 are days. Outcome 2 is binary.

N.2 Mechanism Outcomes

Table N9: Bottom-Up Causal Mechanism Outcomes: H4 - H5 - Individual-Level DID Results

	Dependent variable:					
	Trust Local Gov.		Trust in Ward Cllr.		DC Manages Funds Well	
	OLS		*OLS*		*OLS*	
	BL-EL DID	EL DID (Lagged DV)	BL-EL DID	EL DID (Lagged DV)	BL-EL DID	EL DID (Lagged DV)
Endline	−0.139 (0.075)		−0.109 (0.060)		−0.211* (0.088)	
BU	0.027 (0.058)	0.176** (0.063)	0.134* (0.066)	0.144* (0.069)	−0.033 (0.067)	−0.081 (0.058)
TD	−0.025 (0.056)	−0.0004 (0.067)	−0.055 (0.065)	−0.108 (0.061)	−0.091 (0.073)	0.004 (0.060)
BOTH	−0.057 (0.056)	0.137* (0.060)	−0.043 (0.079)	0.109 (0.066)	−0.039 (0.070)	−0.054 (0.055)
tr1_bl_avg		−0.077 (0.104)				
tr2_bl_avg				0.211* (0.094)		
tr9e_bl_avg						0.121 (0.078)
Endline:BU	0.124 (0.102)		0.027 (0.088)		−0.055 (0.114)	
Endline:TD	0.074 (0.112)		−0.033 (0.094)		0.134 (0.126)	
Endline:BOTH	0.184 (0.103)		0.147 (0.107)		−0.029 (0.113)	
Observations	4,972	2,509	4,860	2,447	5,016	2,521
Adjusted R²	0.017	0.182	0.018	0.115	0.009	0.332

Notes:

*p<0.05; **p<0.01; ***p<0.001
All models include block fixed-effects.
Endline models include enumerator fixed-effects.
All models have SEs clustered on market.
Lagged DV models include market baseline average for DV.
All outcomes are on a 4-point scale.

Table N10: Bottom-Up Causal Mechanism Outcomes: H6 - Individual-Level DID Results

	Dependent variable:			
	Services Satisfaction		Percep. of Sp. on Services	
	OLS		*OLS*	
	BL-EL DID	EL DID (Lagged DV)	BL-EL DID	EL DID (Lagged DV)
Endline	0.037		78.395***	
	(0.047)		(18.109)	
BU	0.146	0.214**	23.733	24.396
	(0.093)	(0.071)	(16.199)	(14.631)
TD	0.067	0.069	28.196	3.165
	(0.069)	(0.074)	(14.768)	(15.426)
BOTH	0.121	0.091	19.923	−2.189
	(0.073)	(0.075)	(14.956)	(13.931)
tc2_10_bl_avg				0.185
				(0.115)
Endline:BU	0.133		2.329	
	(0.078)		(24.635)	
Endline:TD	0.073		−20.490	
	(0.084)		(22.971)	
Endline:BOTH	0.047		−16.253	
	(0.075)		(23.116)	
Observations	24,704	12,365	4,772	2,411
Adjusted R^2	0.033	0.196	0.034	0.291

Notes:
*p<0.05; **p<0.01; ***p<0.001
All models include block fixed-effects.
Endline models include enumerator fixed-effects as well.
All models have SEs clustered on market.
Lagged DV models include market baseline average of DV.
Outcome 1 is on a 4-point scale. Outcome 2 is a number out of 1000.

Table N11: Bottom-Up Causal Mechanism Outcomes: H6 - Satisfaction with Specific Services

	Clean Water Access	Garbage Collection	Condition of Paths	Condition of Stalls	Security
	OLS	OLS	OLS	OLS	OLS
BU	0.654***	0.149	0.073	0.022	−0.035
	(0.160)	(0.090)	(0.079)	(0.097)	(0.097)
TD	0.161	−0.041	0.073	0.150	−0.060
	(0.129)	(0.087)	(0.072)	(0.080)	(0.087)
BOTH	0.315*	−0.045	0.004	−0.010	−0.034
	(0.148)	(0.094)	(0.075)	(0.088)	(0.085)
Observations	2,517	2,520	2,520	2,517	2,525
Adjusted R²	0.140	0.221	0.208	0.190	0.181

Dependent variable:

Notes:

*p<0.05; **p<0.01; ***p<0.001
Individual-level models include enumerator
and block fixed effects.
Individual-level models have SEs clustered
on market.
All outcomes are on a 4-point scale.

Table N12: Bottom-Up Causal Mechanism Outcomes: H6 - Satisfaction with Specific Services (Water Through Paths)

| | Dependent variable: | | | | | |
| | Clean Water Access | | Garbage Collection | | Condition of Paths | |
	BL-EL DID	OLS EL DID (Lagged DV)	BL-EL DID	OLS EL DID (Lagged DV)	BL-EL DID	OLS EL DID (Lagged DV)
Endline	0.113 (0.102)		0.159* (0.081)		0.002 (0.067)	
BU	0.190 (0.135)	0.575*** (0.155)	0.090 (0.101)	0.129 (0.083)	0.087 (0.096)	0.056 (0.074)
TD	0.272* (0.125)	0.045 (0.124)	0.068 (0.099)	−0.058 (0.084)	0.043 (0.092)	0.068 (0.071)
BOTH	0.114 (0.126)	0.263 (0.135)	0.152 (0.087)	−0.089 (0.087)	0.162* (0.080)	−0.034 (0.070)
ms1_bl.avg		0.429*** (0.103)				
ms3_bl.avg				0.309*** (0.081)		
ms4_bl.avg						0.271*** (0.073)
Endline:BU	0.425* (0.185)		0.010 (0.124)		−0.036 (0.120)	
Endline:TD	−0.127 (0.143)		−0.113 (0.122)		0.030 (0.117)	
Endline:BOTH	0.157 (0.165)		−0.229 (0.125)		−0.188 (0.100)	
Observations	5,028	2,517	5,026	2,520	5,031	2,520
Adjusted R²	0.052	0.168	0.043	0.229	0.036	0.214

Notes:

Table N13: Bottom-Up Causal Mechanism Outcomes: H6 - Satisfaction with Specific Services (Stall Condition and Security)

	Dependent variable:			
	Condition of Stalls		Security	
	OLS		*OLS*	
	BL-EL DID	EL DID (Lagged DV)	BL-EL DID	EL DID (Lagged DV)
Endline	−0.063		−0.063	
	(0.066)		(0.091)	
BU	0.049	−0.0002	0.186*	−0.086
	(0.087)	(0.084)	(0.080)	(0.090)
TD	0.069	0.127	0.075	−0.080
	(0.082)	(0.076)	(0.090)	(0.085)
BOTH	0.099	−0.047	0.070	−0.054
	(0.077)	(0.080)	(0.091)	(0.080)
ms5_bl_avg		0.422***		
		(0.098)		
ms6_bl_avg				0.288**
				(0.106)
Endline:BU	−0.057		−0.243*	
	(0.100)		(0.122)	
Endline:TD	0.103		−0.113	
	(0.117)		(0.125)	
Endline:BOTH	−0.111		−0.123	
	(0.098)		(0.123)	
Observations	5,026	2,517	5,031	2,525
Adjusted R^2	0.026	0.204	0.015	0.186

Notes:
*p<0.05; **p<0.01; ***p<0.001
All models include block fixed-effects.
Endline models include enumerator fixed-effects as well.
All models have SEs clustered on market.
Lagged DV models include market baseline average of DV.
All outcomes are on a 4-point scale.

Table N14: Bottom-Up Causal Mechanism Outcomes: H7 - Individual-Level DID Results

	Dependent variable:			
	Paying Tax as Duty		Pay Tax Even if Disag. w. Gov.	
	OLS		*OLS*	
	BL-EL DID	EL DID (Lagged DV)	BL-EL DID	EL DID (Lagged DV)
Endline	−0.053		0.060**	
	(0.054)		(0.022)	
BU	−0.036	0.069*	0.005	0.001
	(0.047)	(0.034)	(0.016)	(0.012)
TD	0.005	0.043	0.007	0.006
	(0.045)	(0.030)	(0.017)	(0.011)
BOTH	0.035	0.048	0.021	0.020
	(0.046)	(0.033)	(0.016)	(0.013)
tc2_4b_bl_avg		−0.151*		
		(0.063)		
pay_even_disagree_bl_avg				0.130
				(0.074)
Endline:BU	0.126		0.005	
	(0.072)		(0.030)	
Endline:TD	0.031		−0.001	
	(0.069)		(0.027)	
Endline:BOTH	0.009		−0.003	
	(0.067)		(0.027)	
Observations	5,050	2,531	24,698	12,355
Adjusted R^2	0.004	0.112	0.006	0.082

Notes:

*p<0.05; **p<0.01; ***p<0.001
All models include block fixed-effects.
Endline models include enumerator fixed-effects as well.
All models have SEs clustered on market.
Lagged DV models include market baseline average of DV.
Outcome 1 is on a 4-point scale. Outcome 2 is dichotomous.

Table N15: Top-Down Causal Mechanisms Outcomes, Vendor Survey - Individual-Level DID Results

	Dependent variable:					
	Could Refuse to Pay		Group Non-Comp. Poss.		Pay Because Consequences	
	BL-EL DID	EL DID (Lagged DV)	BL-EL DID	EL DID (Lagged DV)	BL-EL DID	EL DID (Lagged DV)
Endline	0.030 (0.070)		−0.003 (0.064)		−0.197*** (0.039)	
BU	−0.091 (0.056)	−0.043 (0.057)	−0.071 (0.066)	0.026 (0.057)	−0.020 (0.031)	0.035 (0.026)
TD	−0.068 (0.059)	−0.046 (0.051)	−0.022 (0.068)	0.065 (0.056)	−0.033 (0.029)	0.052* (0.025)
BOTH	−0.065 (0.060)	−0.039 (0.057)	−0.085 (0.073)	−0.045 (0.061)	−0.006 (0.029)	0.041 (0.028)
tc5a_bl_avg		0.137 (0.086)				
tc5b_bl_avg				0.148 (0.079)		
tc2_15b_bl_avg						−0.121 (0.117)
Endline:BU	0.010 (0.091)		0.091 (0.087)		0.048 (0.064)	
Endline:TD	0.040 (0.085)		0.104 (0.092)		0.069 (0.053)	
Endline:Both	0.018 (0.093)		0.032 (0.104)		0.033 (0.055)	
Observations	5,013	2,514	5,036	2,524	5,026	2,518
Adjusted R^2	0.004	0.123	0.007	0.145	0.031	0.308

Notes:

*p<0.05; **p<0.01; ***p<0.001
All models include block fixed-effects.
Endline models include enumerator fixed-effects as well.
All models have SEs clustered on market.
Lagged DV models include market baseline average of DV.
All outcomes are on a 4-point scale.

Table N16: Top-Down Causal Mechanisms Outcomes, Tax Collector Survey

	Dependent variable:			
	Hours Working in Market A Day		Vendors Visited Per Day	
	BL-EL DID	EL DID (Lagged DV)	BL-EL DID	EL DID (Lagged DV)
Endline	0.271 (0.377)		−9.468 (15.163)	
BU	0.607 (0.427)	0.089 (0.587)	34.680 (23.926)	35.928 (59.231)
TD	0.319 (0.572)	0.921 (0.505)	44.153 (39.753)	103.900 (64.511)
Both	0.537 (0.386)	0.309 (0.573)	18.215 (21.026)	189.129 (122.715)
hrs_in_mkt_bl_avg		0.297** (0.098)		
vendors_visited_trim_99_bl_avg				−0.370 (0.402)
Endline:BU	−0.495 (0.633)		−11.485 (20.328)	
Endline:TD	0.781 (0.841)		13.731 (49.906)	
Endline:Both	0.148 (0.629)		103.736 (108.582)	
Observations	566	260	559	257
Adjusted R^2	0.235	0.379	0.089	0.308

Notes: *p<0.05; **p<0.01; ***p<0.001

All models include block fixed-effects.

Endline models include enumerator fixed-effects as well.

All models have SEs clustered on market.

Lagged DV models include market baseline average of DV.

Outcome 1 is hours in a day. Outcome 2 is a positive integer.

O Multiple Hypothesis Tests Correction

O.1 Main Paper Results

As specified in the PAP, we perform multiple hypothesis tests corrections to assess the robustness of our findings. Table O1 shows the p-values for the primary outcome models with significant treatment effects presented in the main paper. We present the Holm correction, which controls the family-wise error rate (FWER).

We correct for running tests using multiple outcomes for each hypothesis. For the main hypothesis (H1), we also correct for the fact that there are three comparisons against the control (evaluating the effect of the three treatment groups). For the BU mechanism and downstream hypotheses (H4 - H8) and TD mechanism hypotheses (H9 - H11), we correct only for the tests of the intervention effect relevant to each hypothesis (BU and BOTH, then TD and BOTH, respectively).

Table O1: Multiple Hypothesis Tests Correction

Level	Outcome	Term	Hypothesis	p	p Holm	Survives Holm
Individual	Self	TD	H1	0.034	0.504	No
Individual	Receipt	BU	H1	0.001	0.014	Yes
Individual	Receipt	TD	H1	0.013	0.214	No
Market	Receipt	BU	H1	0.015	0.234	No
Individual	Trust in Local Gov.	BU	H4	0.005	0.021	Yes
Individual	Trust in Local Gov.	BOTH	H4	0.017	0.047	Yes
Individual	Trust in Ward Counc.	BU	H4	0.016	0.047	Yes
Individual	Services Satisfaction	BU	H6	0.002	0.012	Yes
Individual	Paying Tax as Duty	BU	H7	0.038	0.152	No
Individual	Petition Anon.	BU	H8	0.007	0.047	Yes
Individual	Petition Anon.	BOTH	H8	0.004	0.042	Yes
Individual	Petition w. Name	BU	H8	0.012	0.060	No
Individual	Petition w. Name	BOTH	H8	0.007	0.047	Yes
Individual	Agree St. 1	BU	H8	0.006	0.047	Yes
Individual	Agree St. 1	BOTH	H8	0.004	0.042	Yes
Individual	Agree St. 2	BOTH	H8	0.022	0.088	No
Individual	Pay Because Consequences	TD	H9	0.026	0.155	No
Individual	Perception of Relative Amount of Taxes Reaching Gov't	BOTH	H10	0.020	0.040	Yes
Individual	Hours Working in Market A Day	TD	H11	0.020	0.078	No

O.2 Holm-Corrected Wald Tests

Below, we also apply MHTC to the Wald tests reported in Appendix M (Tables O2 through O6).

As with the main results, to control the family-wise error rate, we apply the Holm-correction to the p-values for all four linear hypotheses per model in each main-paper hypothesis. In other words, for H1, we are correcting for the fact that we are doing eight tests: four in the individual DIM receipt model and four in the market DID receipt model.

Table O2: Holm-corrected p-Values for Linear Hypothesis Tests for Receipt Outcome Models in Table 2

Linear Hypothesis	Receipt Outcome (Individual DIM)	Receipt Outcome (Market DID)
BU = TD	0.979	0.620
BU = BOTH	0.699	0.715
TD = BOTH	1.000	1.000
BU + TD = BOTH	0.035	0.699

Table O3: Holm-corrected (Within Hypotheses H4 and H5) p-Values for Linear Hypothesis (Wald) Tests for Bottom-Up Mechanism Models in Table 3

Linear Hypothesis	Trust in Local Government	Trust in Ward Councilors	District Manages Funds Well	District Spending Transparent	District Tax Collection Transparent
BU = TD	0.032	0.000	1.000	1.000	1.000
BU = BOTH	1.000	1.000	1.000	1.000	1.000
TD = BOTH	0.072	0.004	1.000	1.000	1.000
BU + TD = BOTH	1.000	1.000	1.000	1.000	1.000

Table O4: Holm-corrected (Within Hypotheses H6 and H7) p-Values for Linear Hypothesis (Wald) Tests for Bottom-Up Mechanism Models in Table 4

Linear Hypothesis	Satisfaction with Services	Satisfaction with Water Access	Perception of Relative Amount Spent on Services	Paying Taxes Is Duty	Vendors Should Pay Taxes
BU = TD	0.304	0.026	0.959	1.000	1.000
BU = BOTH	0.959	0.502	0.505	1.000	0.824
TD = BOTH	0.959	0.959	0.959	1.000	1.000
BU + TD = BOTH	0.617	0.265	0.617	0.910	1.000

Table O5: Holm-corrected p-Values for Linear Hypothesis (Wald) Tests for Political Engagement Outcome Models in Table 5

Linear Hypothesis	Vote	Sign Petition Anonymously	Sign Petition with Name	Reduce Gov. Over-reach	Increase Gov. Effectiveness
BU = TD	1.000	0.027	0.053	0.124	0.907
BU = BOTH	1.000	1.000	1.000	1.000	1.000
TD = BOTH	1.000	0.018	0.042	0.115	0.124
BU + TD = BOTH	1.000	1.000	1.000	1.000	1.000

Table O6: Holm-corrected (Within Hypotheses H8, H9, H10) p-Values for Linear Hypothesis (Wald) Tests for Top-Down Mechanism Models in Table 6-7

Linear Hypothesis	Ind'l Evasion Possible	Group Evasion Possible	Pay Because Consequences	Perception of Relative Amount of Taxes Reaching Gov't	Hours Working in Market per Day	Vendors Visited in Market per Day
BU = TD	1.000	1.000	1.000	0.471	0.903	1.000
BU = BOTH	1.000	1.000	1.000	1.000	1.000	1.000
TD = BOTH	1.000	0.407	1.000	0.073	1.000	1.000
BU + TD = BOTH	1.000	0.978	1.000	1.000	1.000	1.000

P   Explanation of Deviations from PAP

We delineate four deviations from the pre-analysis plan. First, we are treating the bottom-up and top-down treatment combination as its own treatment, for reasons already explained in the main text.

Second, the PAP specified that we would estimate corruption by comparing estimated market-level tax compliance to the actual revenues reported by the local government. Revenue data quality issues (see discussion in Appendix E) compelled us to drop this measure of corruption.

Third, the PAP specified that we would use a spillover radius of the largest distance a vendor seemed to travel, based on baseline responses. However, this turned out to be unreasonable: several vendors traveled more than thousand kilometers, according to our data (although this could be due to similar market names). The majority, however, sold only at one market, and so the mean distance traveled by a vendor was closer to zero. We decided to use distances of 2 km, 5km, and 10 km, before we did any spillover analysis.

Finally, the PAP specified that we would include district fixed effects. However, because we used block randomization, we used block fixed effects instead.

Q   Ethics

This appendix contains information on the ways in which this paper and the associated project comply with the ethical and transparency obligations described in APSA's A Guide to Professional Ethics in Political Science and in the APSA-approved Principles and Guidance for Human Subjects Research. Some of this information can also be found in prior appendices but is compiled here for ease of access.

Q.1 Human Subjects Research

This project was evaluated and declared exempt from further review by the University of North Carolina - Chapel Hill IRB (IRB Number 17-1043) and by the IRB at NORC at the University of Chicago (project number: 7554.030.01; IRB protocol number: 17.06.18). It was also evaluated and approved by the National Commission on Research in the Social Sciences and Humanities (which serves as the Malawi country-level IRB) as protocol P03/17/161.

We affirm that the study is in compliance with APSA's Principles and Guidance for Human Subjects Research. See below for more details.

Consent and Confidentiality. All interventions were conducted with the approval and participation of local government authorities. For the baseline and endline surveys, approval was given by local government as well as by the market master. Enumerators asked each survey participant for their consent verbally before proceeding with the survey. During training, the necessity and importance of obtaining consent was repeatedly reinforced. The consent statement, which was approved by all relevant IRBs, stated that the survey was being conducted by Innovations for Poverty Action, and that the survey was about markets in Malawi. The voluntary nature of the study was stressed, as was confidentiality, and respondents were given contact information for IPA and the IRBs in case they had questions or concerns. We avoided mentioning the intervention in the consent process in order to avoid priming respondents about the intervention itself and to avoid additional social desirability bias when assessing tax compliance.

To promote confidentiality, enumerators were instructed to take respondents to a quiet place to complete the survey so that respondents' answers could not be overheard. Responses were stored on tablets during enumeration. At the end of each day, they were uploaded to IPA's server on the Box platform, where all data were encrypted using BoxCryptor. Only the PIs, their co-authors, and select IPA staff had access to the keys for the encrypted data. The data were also uploaded to a secure SFTP-protected server by NORC, to facilitate analysis. Only the PIs and their co-authors had access to this server. All publicly released data has been de-identified.

Compensation. At both baseline and endline, study participants who completed the vendor survey received a small airtime voucher in return for completing the survey (MWK200 for the short survey and either MWK300 or MWK600 for the long survey, depending on a delayed gratification experiment embedded in the long survey).

Participation in the surveys was not associated with any particular risks. Although several questions asked about tax compliance, which is technically required to operate in the market, the questions were phrased to avoid asking directly about compliance or requiring the respondent to report on particular instances of compliance (or noncompliance).

Risk and ethical issues. The interventions themselves were not assessed as causing undue risk by any of the IRBs that approved the project. It was technically possible that the top-down intervention would increase repression on behalf of the local authorities, although

this would be an unintended side effect, as none of the top-down interventions were designed to give the authorities more power vis-à-vis market vendors. Checking for unintended consequences such as these was part of the ongoing monitoring assessments during the intervention period.

The implementing partners did not report any ethical issues over the course of the study. There was one claim that the endline data collection team had set off protests in two markets; an independent investigation found that this was not the case.

Data Collection Procedures, Data, and Code. See Appendix C for information on data collection procedures. All quantitative data and code necessary to replicate the results can be found in the APSR Dataverse. The raw data was also supposed to be accessible through USAID's Development Data Library (https://data.usaid.gov/). The authors submitted the data to USAID through NORC at the University of Chicago in April 2022. At the time of publishing, the Development Data Library was offline.

References

Agrawal, Arun, Ashwini Chhatre and Elisabeth Gerber. 2015. "Motivational Crowding in Sustainable Development Interventions." *American Political Science Review* 109(3):470–487.

Frey, Bruno and Reto Jegen. 2001. "Motivation Crowding Theory." *Journal of Economic Surveys* 15(5):589–611.

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

Ostrom, Elinor. 2000. "Crowding Out Citizenship." *Scandinavian Political Studies* 23(1):3–16.