Property tax compliance in Tanzania

Can nudges help?

Matthew Collin
Vincenzo Di Maro
David K. Evans
Fredrick Manang

GLOBAL WORKING PAPER #161
SEPTEMBER 2021

B | Global Economy and Development at BROOKINGS

Property Tax Compliance in Tanzania: Can Nudges Help?

Matthew Collin
Brookings Institution

Vincenzo Di Maro
World Bank

David K. Evans
Center for Global Development

Fredrick Manang
University of Dodoma

September 2021*

Abstract

Low- and middle-income countries around the world struggle with low tax compliance together with limited capacity to enforce compliance. This paper reports the results of a randomly rolled out text-message campaign aimed at promoting compliance among landowners in Dar es Salaam, Tanzania. Landowners were effectively randomly assigned to one of four groups designed to test different aspects of tax morale. They either received a simple text-message reminder to pay their tax (a test of salience), a message highlighting the connection between taxes and public services (reciprocity), a message communicating that non-compliers were not contributing to local or national development (social pressure), or no message (control). Recipients of any message were 11 percent (or 1.2 percentage points) more likely to pay any property tax by the end of the study period. Across treatments, simple reminders and reciprocity messages delivered similar gains in payment rates, whereas social pressure messages delivered lower gains in payment rates. Actual payment amounts were highest for reciprocity messages. The average estimated benefit-cost ratio across treatments is 20:1 due to the low cost of the intervention, with higher cost-effectiveness for reciprocity messages.

Keywords: Tax Compliance, Tax Morale, Public Finance, Nudges
JEL classification: H26, H13, O17

*The authors are grateful for support from the Tanzania Revenue Authority (TRA), especially the Non-tax revenue department at TRA headquarters. We also thank Anne Brockmeyer, Santiago de la Cadena Becerra, Addisu Lashitew, Fabrizio Santoro, Stuti Khemani, Thiago Scot, and participants at the CSAE conference and at World Bank workshops for discussions and comments. We also thank the DIME Analytics team at the World Bank for their careful replication review. The Brookings Institution is a nonprofit organization devoted to independent research and policy solutions. Its mission is to conduct high-quality, independent research and, based on that research, to provide innovative, practical recommendations for policymakers and the public. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not not reflect the views of the Brookings Institution, its management, or its other scholars, nor do they necessarily represent the views of the authors' other employers or their affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent. Corresponding author: Manang (fmanang@udom.ac.tz). Other authors: Collin (mcollin@brookings.edu), Di Maro (vdimaro@worldbank.org), and Evans (devans@cgdev.org).

1 Introduction

Governments need enough revenue to sufficiently fund public goods and services. It is becoming clearer that the source of that revenue also matters (Gadenne 2017). The state's ability to raise tax revenue is thought to invite public scrutiny and strengthen the "social contract" between citizen and state, leading to positive effects on institutional and economic development (Besley and Persson 2013; Ali, Fjeldstad, and Katera 2017; Dincecco and Katz 2014). Despite the purported benefits of higher tax capacity, low- and middle-income countries largely struggle to raise the same levels of tax revenue as their higher income peers. As of 2014, the tax-to-GDP ratio of the median low or lower-middle income country was 15.8 percent, compared to 20.5 percent for higher income countries. Evidence suggests that this may condemn some of these countries to lower levels of economic growth in the long term (Gaspar, Jaramillo, and Wingender 2016).

One area where governments seek to improve tax compliance is the taxation of immovable property, which presents an attractive source of revenue as it is – in theory - both easier to target and easier to tax in a non-distortionary way. But low- and middle-income countries perform even worse in the collection of property tax than they do overall: where OECD countries bring in approximately 2.1 percent of GDP of revenue from property taxes, poorer countries bring in only 0.6 percent (Norregaard 2013).

Tanzania, the context of our study, historically struggled with low rates of tax revenue, both overall (11.8 percent of GDP) and for property tax (0.1 percent of GDP). Furthermore, the government has oscillated between a regime of decentralized revenue collection (where the local authorities are responsible for collecting property tax) and that of centralized collection (where the TRA is responsible) (Fjeldstad, Ali, and Katera 2017). This process has led to unstable and unpredictable levels of property tax revenues over the past decade. Furthermore, the Tanzanian government has struggled with low levels of compliance. This may in part be due to a lack of perceived reciprocity by taxpayers: property owners do not understand how the government will use their money (PO-RALG 2013).[1] It is also the result of a small tax base: while legally every property owner is obligated to pay tax, local authorities have previously prioritized those with larger properties living in the most affluent areas of the city (PO-RALG 2013). It is within this context that the Tanzania Revenue Authority (TRA) is examining new ways to improve property tax compliance in the cities it is responsible for.

This paper investigates the impact of a series of reminders via text message that leverage different aspects of citizens' voluntary motivation to pay property taxes (i.e., their "tax morale"). Working with the TRA we randomly allocated a group of more than 200,000 individuals in Dar es Salaam who were liable to pay property tax - but as of one month prior to the annual deadline had not paid any - into four groups.[2] Three groups received a text message treatment: a simple reminder (increasing the salience of tax paying), a message that emphasized the link between tax revenue and publicly-provided goods (focusing on reciprocity), or a message highlighting the

[1] In Pakistan, even an explicit intervention to collect citizen preferences on public services and then to deliver services based on those preferences had little initial effect on tax payments (Khwaja et al. 2020).

[2] As discussed later, taxpayers were actually assigned to five groups, but two groups received the same treatment due to an implementation error.

non-cooperative nature of tax evasion (focusing on social pressure). The fourth group served as a control.

We find that all three treatments had a positive impact on both the propensity of taxpayers to make payments to the TRA and the total amount paid. Those receiving the reciprocity treatment paid higher amounts of total tax. Those receiving the social pressure message paid at lower rates than those receiving the simple reminder. Similar to what has been found with other "nudge" style interventions, the intervention is highly cost effective, with the increase in revenue driven by treatment exceeding its cost by roughly a factor of twenty.

Finally, we also document two interesting sources of heterogeneity. The first is across the geography of the city: areas that have a higher rate of tax compliance among the control group also had lower treatment effects, suggesting that nudges may be more successful in low-compliance areas when policymakers are able to identify them ex-ante. Second, we find heterogeneity across the amount that taxpayers owed the TRA: average treatment effects appear to be strongest for those who owe around 10,000 TSh, predominantly smaller properties that do not bring in much revenue and so are less likely to be subject to TRA follow up. This indicates that tax nudges may, in some circumstances, be regressive in their impact on compliance.

The rest of the paper is structured as follows. Section 2 discusses prevailing theories of tax compliance and the existing body of evidence on the impact of messages on taxpayer behavior. Section 3 discusses the structure of the experiment and the data we will use to examine its impact. Section 4 presents and discusses the results and we conclude with Section 5.

2 Background

Until recently, economists have viewed tax compliance largely through the lens of enforcement (Allingham and Sandmo 1972), where taxpayers increase their compliance when the perceived probability of detection goes up. There is evidence that letters and electronic forms of communication have the potential to do this: research in many high-income economies suggests that letters containing an implicit or explicit threat of audit increases tax payments (Coleman 1996; Blumenthal et al. 2001; Hasseldine et al. 2007; Kleven et al. 2011; Fellner et al. 2013; Castro and Scartascini 2015; Hallsworth et al. 2017; Pomeranz 2015; Meiselman 2018; Hernandez et al. 2017) although the effect sizes vary across contexts and are not always significant (Ariel 2012). Work in low- and middle-income countries has largely revealed similar results (Ortega and Scartascini 2020; Brockmeyer et al. 2019; Kettle et al. 2016; Brockmeyer et al. 2020), again with results not always significant (Del Carmen, Espinal Hernandez, and Scot 2020). Evidence from Rwanda suggests that less aggressive messages (such as reciprocity or reminder-framed messages) work slightly better than those aimed at deterrence (Mascagni, Nell, and Monkam 2017), whereas evidence from Uganda suggests stronger impacts for enforcement focused messages (Cohen 2020).

In recent years, economists have extended the Allingham and Sandmo model to include the concept of "tax morale," a bundle of mechanisms which explain voluntary tax compliance (Luttmer and Singhal 2014). Recent experiments have attempted to make the components of tax morale more salient through careful messaging, with mixed results. Kettle et al. (2016)

finds that both letters emphasizing national pride and those emphasizing social norms improve compliance in Guatemala, but not significantly more than a letter invoking a heightened probability of audit. In richer countries, randomized studies of letter or e-mail campaigns typically find that attempts to emphasize the social contract or civic duty either have little impact or are marginally effective (Coleman 1996; Blumenthal et al. 2001; Torgler 2004; Ariel 2012; Fellner et al. 2013; Castro and Scartascini 2015; Meiselman 2018) with some exceptions (Hallsworth et al. 2017). Krause (2020) finds that messages that emphasize the social pressure mechanism in Haiti might even have a negative effect on tax compliance. This is consistent with results from psychology research which have shown that in contexts with low rates of pro-social coordination, a mechanism known as antisocial punishment could be at play (Herrmann et al. 2008).

A recent meta-analysis (Antinyan and Asatryan 2020) of studies of nudges for tax compliance finds that deterrence nudges are on average more effective than tax morale nudges. The meta-analysis also finds that nudges seem to work better for the sub-samples of late payers, which is the sample we focus on in this paper.

3 Experiment and data collection

3.1 Baseline data and randomization

The frame for this experiment is a list of 241,200 properties for which, as of June 1st, 2018, no property tax had been paid to the TRA for the 2017/2018 tax year. The deadline for property tax payments to be completed was June 30th. After June 30th had passed, the TRA extended the deadline for another two weeks, although continued to accept payments after this point.

As some taxpayers own multiple parcels, we collapsed these data to the taxpayer level (237,699 taxpayers), as indicated by the taxpayer identification number associated with the property. We use two sources of information in the randomization: the location of the property and whether or not the property had been served a 'demand notice' at the time the data was collected. The location of the property is the lowest level of administration in Dar es Salaam, the sub-ward or 'mtaa' level. We assign taxpayers the same location as their property. When taxpayers have multiple properties that span more than one sub-ward, we pick the modal sub-ward. Where there is no modal sub-ward, we randomly choose one of those sub-wards to assign to the taxpayer. Ultimately, the randomization was conducted within 1,211 different strata, which were defined both by the location of the property (sub-ward) and whether a bill had been issued.

Demand notices are bills issued by the TRA to landowners. Approximately 19 percent of the experimental sample had been issued a bill at the time of the data collection. The TRA issued bills (called "demand notices") in bulk for a specific area of the city (this could be a ward or a set of core streets) and sent them to landowners by manual delivery by TRA officers and temporary interns, typically after seeking the support of street leaders. While the goal is to cover all areas of the city every fiscal year, limited financial and human resources explain why only a subset of the city is covered in practice.

To better understand how our experimental sample compared to the average property in

Table 1: Treatment arms and treatment assignments

Treatment	Type	Message	N
Control	No message		47,555
T1	Reminder	"Dear taxpayer, TRA reminds you to pay your property tax. Pay before 30th June. For more information: dial * 152 * 00 #, visit your nearest TRA office or call 0800780078. Thank you."	47,502
T2	Reciprocity	[Reminder] + "Your tax facilitates access to social services and infrastructure. Together we build our nation."	47,547
T3	Social Pressure	[Reminder] + "Non-taxpayers are not contributing to national development and thus hindering development of their communities."	47,538
T4	Enforcement*	[Reminder] + "Pay your tax early to avoid penalties."	47,553

***Note:** Subjects randomly allocated to T4 were mistakenly sent T1. See sub-section 3.3 for details. Original messages were sent in Swahili.

Dar es Salaam, we matched our experimental data to a set of data comprising every parcel in TRA's database for the city. Out of approximately 830,000 parcels in the city, nearly 30 percent are owned by a taxpayer in our experimental sample. Parcels included in our experiment were only slightly (1.4 percentage points) less likely to have been issued a bill. Of those that were issued bills, only 35 percent of properties in the experimental sample had been valued by the TRA, as opposed to 74 percent in the rest of the city (those that were not valued were charged a TSh 10,000 flat tax). Conditional on being billed and rated, the median value for properties included in the experiment was higher, approximately 31.2 million TSh (13,500 USD) versus 20 million TSh (8,600 USD) for the rest of the city. Properties that were billed and rated faced, on average, an annual tax rate of approximately 0.16% of the property's rated value.

3.2 Treatments

In collaboration with the Tanzania Revenue Authority, we randomized each property owner into one of five groups.[3] The treatments are summarized below in Table 1. The control was not to receive any message from the TRA. Group T1 received a simple message reminding them to pay their property tax, indicating the due date (June 30th) and providing information the taxpayers could use to contact the TRA in case they had any questions. All other treatments included this reminder message.

Treatment 2 added a 'reciprocity' message, where taxpayers were reminded that taxes fund social services and infrastructure and finished with the TRA's slogan "Together we build our Nation." Treatment 3 included the simple reminder as well as a 'social pressure' message in which taxpayers were reminded, in a negative fashion, that non-compliers were not contributing to the development of the country or their own communities. The final planned treatment was an enforcement message which included the simple statement "Pay your tax early to avoid penalties." However, for reasons we discuss below, during the implementation of the experiment no taxpayers were sent this message.

[3]The randomization was conducted using the Stata command `randtreat`, with "misfits" (i.e., observations beyond those that are a multiple of the number of treatment groups) being dealt with using the *strata* method, which randomly allocates misfits across all strata (Carril et al. 2017).

Table 2: Frequencies of actual treatment, by treatment group

Treatment Arm	Received No Message	Received T1	Received T2	Received T3
Control	90.3%	5.1%	2.5%	2.2%
T1 - Reminder	10.4%	81.7%	4.6%	3.8%
T2 - Reciprocity	5.1%	11.5%	77.7%	6.4%
T3 - Social pressure	13.5%	12.2%	7.5%	67.2%

Note: As some some taxpayers received multiple messages, the above frequencies can exceed 100.

3.3 Implementation and challenges

Following the randomization, a list of taxpayers and their phone numbers (included in the original TRA dataset) were provided to the TRA. The majority of messages were sent out after June 20, fewer than ten days before the initial deadline to pay property tax. While there was overlap in the delivery of different treatments, completion of each treatment arm proceeded sequentially, with reminder messages being sent first to group T1, reciprocity messages second to group T2, followed by social pressure messages to group T3.[4]

There were two errors in the message delivery process. First, the firm in charge of sending the messages sent Treatment 1 messages to taxpayers who had been randomly allocate to Treatment 4 (Enforcement), essentially doubling the size of the first treatment arm. Thus no enforcement messages were sent as a part of the experiment (although a few were sent independently by the TRA). Second, instead of using the list of cleaned and prepared phone numbers they were provided with, the firm chose an unformatted list which contained the same numbers, but in some cases were not usable due to missing pre-fixes or mistakenly included county codes. As a result, between 22-33 percent of each treatment arm was not sent the intended message. Using data from the text message delivery, we can account for which taxpayers were or were not sent a message due to this error.

Finally, the randomization was conducted at the taxpayer identification level, but a small subset of taxpayer IDs shared identical phone numbers. This is likely because some taxpayers were issued multiple taxpayer IDs, or households sharing a single number contained multiple taxpayers. This leads to spillovers in actual treatment between the various treatment arms and the control group. The actual frequencies of treatment across these groups are summarized in Table 2. Because we know the actual distribution of messages that were sent, we can instrument for actual message delivery with assignment to treatment.

3.4 Outcome data

Our data on outcomes was retrieved from the TRA at the beginning of August, 2018. We merged the complete record of all property tax payments made for a given taxpayer ID between June and the beginning of August to our original experimental sample. For each taxpayer ID we record, for each date during this period, whether any payments associated with that ID had been made up to that date as well as the cumulative amount of payments made so far. For a

[4]Figure A1 in the Appendix shows the timing of message delivery. One concern is that the differential timing of message delivery affected the impacts. In the results section, we provided suggestive evidence that our pattern of results is robust to this.

subset of approximately 45,000 taxpayers we also have the final demand notice (bill) that was issued by the TRA.[5] This will allow us to also investigate the impact our treatments have had on the proportion of the final tax bill each taxpayer has paid, as well as the probability that they have paid their entire bill.

4 Results

For most of the main results, we display the results from either a single treatment dummy (for any treatment) or a full set of dummies for each treatment:

$$P_i = \alpha + \beta \times treated_i + \gamma_s + \epsilon_i \tag{1}$$

and

$$P_i = \alpha + \beta_1 T1_i + \beta_3 T2_i + \beta_4 T3_i + \gamma_s + \epsilon_i \tag{2}$$

Where P_i is alternatively whether the taxpayer has paid anything to the TRA or the amount the taxpayer has paid. In equation (1), $treated_i$ is an indicator variable equal to one if the taxpayer was randomized into any of the three treatment groups. In equation (2), the indicator variables $T1_i$, $T3_i$ and $T3_i$ are dummy variables equal to one if the taxpayer was randomized into the reminder, reciprocity or social pressure treatments, respectively. Unless otherwise specified, we run both specifications (1) and (2) with strata fixed effects (indicated by γ_s) and cluster the standard errors at the taxpayer level.

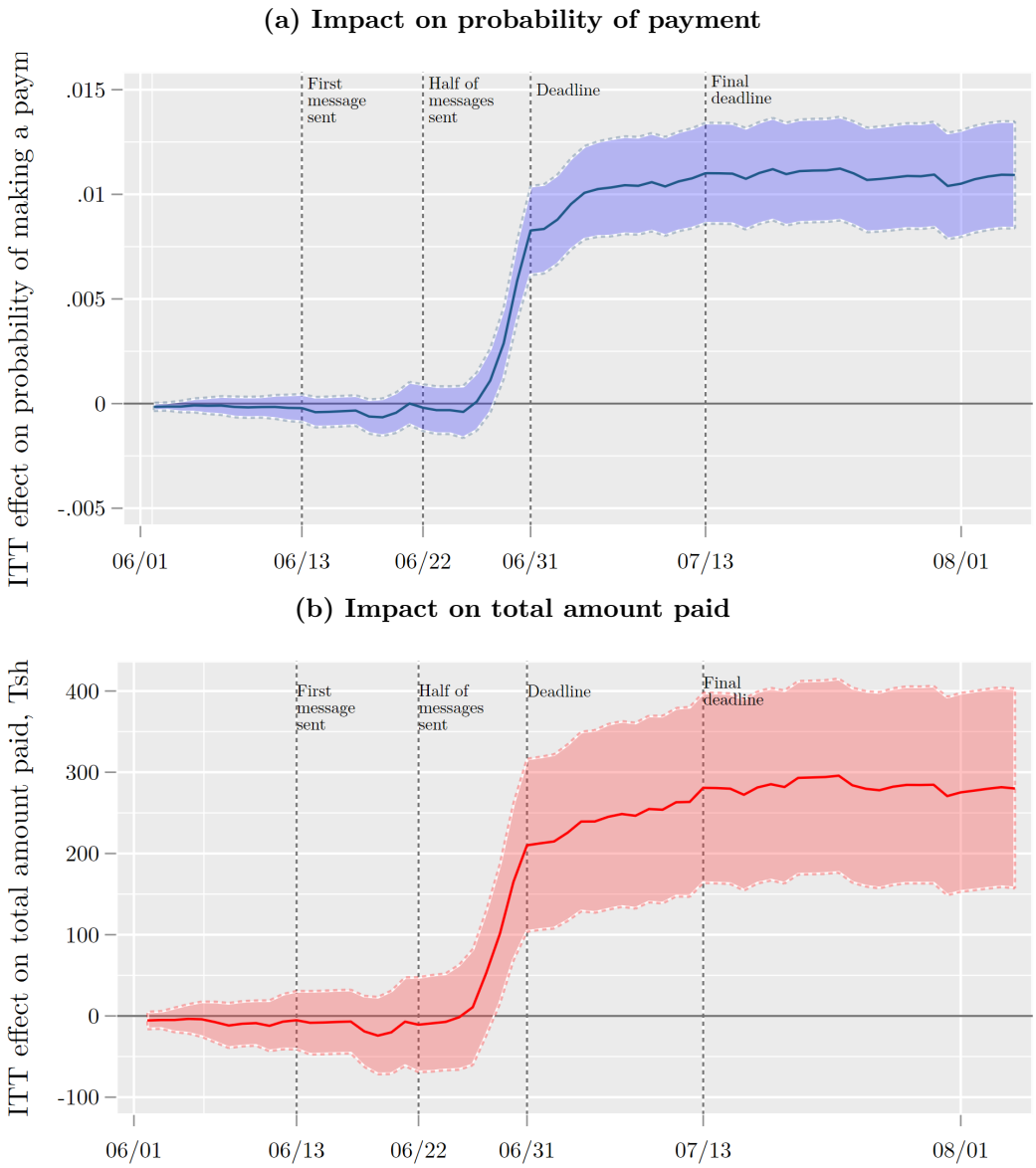
Figure 1 shows the intent-to-treat (ITT) coefficient estimates from specification (1) when the outcome is whether the taxpayer has paid anything, measured at different points of time during the experiment. As can be seen, prior to the introduction of the text messages, treated taxpayers had the same propensity to have made a payment to the TRA as untreated taxpayers. Only following the introduction of the text messages do we see a difference. By the end of the period for which we have administrative data, those randomized into a message treatment were 1.1 percentage points more likely to have made a payment to the TRA, over a baseline of approximately 10 percent.

Figure 2 shows (a) the average payment rate and (b) the average amount paid for the control group and each treatment group over the course of the study relative to the timing of the treatment (indicated by the dark grey-shaded bars). On average, all three treatment messages outperform the control group, with the gap opening up around the time of the first payment deadline on June 31st.

Table 3 shows the results of specification (2) when the outcome is whether the taxpayer has made any payment to the TRA measured at different points of time during the experiment. Columns (1)-(3) show the ITT estimates and (4)-(6) show the 2SLS estimates, where treatment assignment is used to instrument receipt of the correct text message. For the latter, the results indicate that one month after the deadline, those that received a reminder, reciprocity, or a

[5] As we show in Table A1 in the Appendix, there is no imbalance within our experimental sample in the probability of being issued a bill, although for the bill subsample there are some minor imbalances between the reciprocity and social pressure treatment arms for the overall bill amount.

Figure 1: Timeline of ITT effect of all messages on payment rates



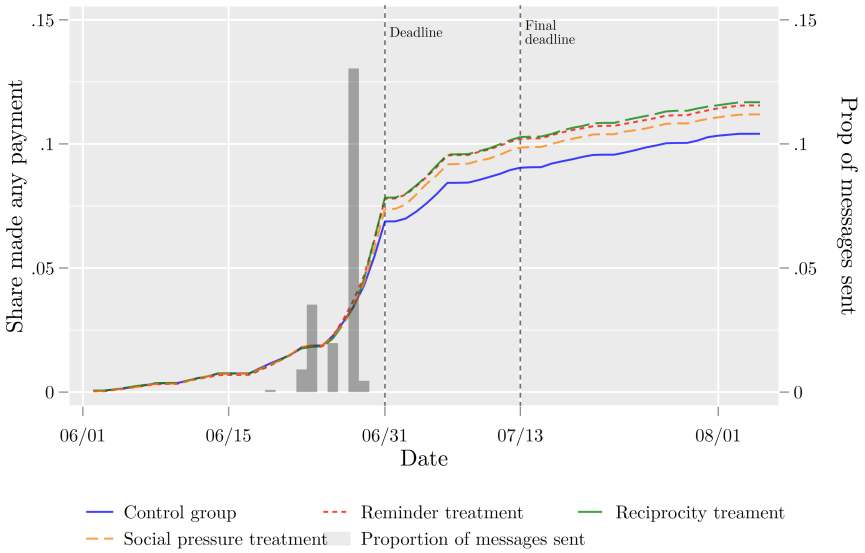
Note: Figure shows the pooled effect of being randomized into one of the treatment groups over time. Subfigure (a) shows the intent-to-treat effect on the probability that the taxpayer has made any payment to the TRA by the date shown. Subfigure (b) shows the intent-to-treat effect on the total amount the taxpayer has paid to the TRA by the date shown. 95% confidence intervals shown.

social pressure message were 1.4, 1.5, or 0.9 percentage points more likely to make a payment, over a control mean of about 10 percentage points. At the bottom of each column we present a test of equality of these coefficients: we find that both the reminder and reciprocity treatments have a larger effect than the social pressure treatment.

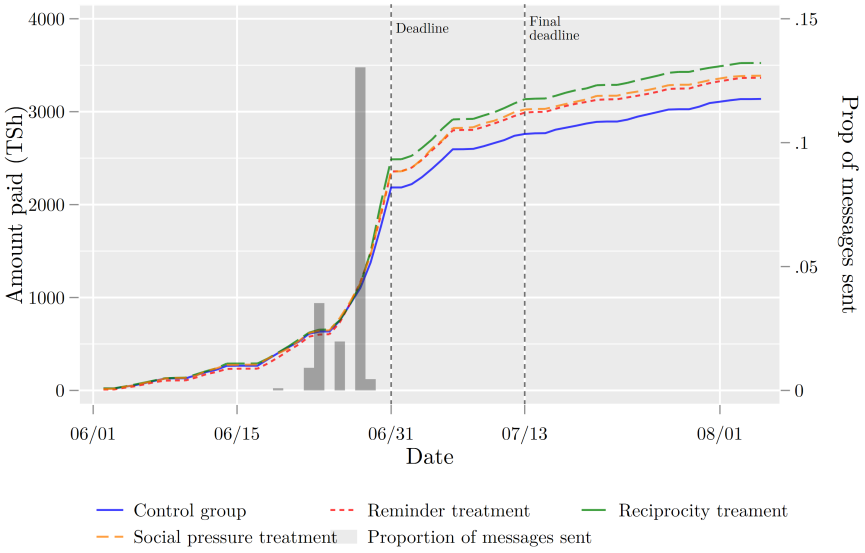
Table 4 follows the same approach, but the outcome is the amount paid in TSh. Recipients of the reminder, reciprocity, and social pressure messages paid an additional 290-480 TSh above a control mean of approximately 3,100 TSh, an increase in about 10 percent. Those receiving the reciprocity message gave significantly more. While these results indicate that those receiving the social pressure message gave more on the intensive margin (as the social pressure message performed worse on convincing taxpayers to pay anything), in our bill subsample (described

Figure 2: Timeline of payment rates by treatment group

(a) Probability of any payment to TRA



(b) Amount paid to TRA



Note: Figures show the average outcomes for each treatment group for (a) the share of taxpayers who made any payment to the TRA and (b) the average amount paid to the TRA in TSh, both over the period we have data for. The shaded bar graph indicates the timing of the treatment: each bar indicates the proportion of messages that were sent to taxpayers on a given day.

in Section 3.4) we find a slight imbalance on the amount that those receiving social pressure messages owe (See table A1 in the Appendix).

In Table 5 we use our subsample of taxpayers for whom we have bill data to unpack what proportion of their final tax bill has been paid. Column (1) replicates column (6) of Table 3, then column (2) restricts the sample to the bill subsample, showing that the effect sizes are similar across the two samples. Columns (3) investigates the impact of the three treatments

Table 3: Impact of message assignment on payment rates

	ITT estimates			2SLS estimates		
	Start of experiment	First tax deadline	One month after first deadline	Start of experiment	First tax deadline	One month after first deadline
Pooled treatment arms						
Treated	-0.000214	0.00828***	0.0109***	-0.000271	0.0105***	0.0139***
	(0.000406)	(0.00129)	(0.00155)	(0.000503)	(0.00160)	(0.00192)
Separate treatment arms						
T1: Reminder	-0.000379	0.00924***	0.0115***	-0.000492	0.0116***	0.0144***
	(0.000442)	(0.00143)	(0.00171)	(0.000561)	(0.00182)	(0.00218)
T2: Reciprocity	-0.000132	0.00962***	0.0128***	-0.000139	0.0115***	0.0153***
	(0.000514)	(0.00167)	(0.00200)	(0.000643)	(0.00209)	(0.00251)
T3: Social pressure	0.0000309	0.00505**	0.00791***	0.000112	0.00562*	0.00942***
	(0.000517)	(0.00164)	(0.00198)	(0.000744)	(0.00237)	(0.00285)
Constant	0.00640***	0.0687***	0.104***			
	(0.000365)	(0.00115)	(0.00138)			
Control mean	0.006	0.069	0.104	0.006	0.069	0.104
First stage f-stat				31,174.8	31,174.8	31,174.8
R^2	0.010	0.037	0.044	0.000	0.000	0.000
Obs	237,591	237,591	237,591	237,591	237,591	237,591
Test: T1 = T2	0.575	0.798	0.455	0.563	0.947	0.713
Test: T1 = T3	0.355	0.004	0.038	0.375	0.007	0.062
Test: T2 = T3	0.752	0.007	0.015	0.752	0.024	0.058

Notes: Outcome is the probability a taxpayer made any payment to the TRA by the date indicated. ITT estimates indicate impact of being assigned to treatment. 2SLS estimates instrument the receipt of each message type with assignment to treatment. Robust standard-errors in parentheses. $^{+}p < 0.10$,$^{*}p < 0.05$,$^{**}p < 0.01$,$^{***}p < 0.001$

on the proportion of the total tax bill (the 2017/18 bill and any arrears) paid by the taxpayer. During the time period we consider the control group has paid roughly 10 percent of their total bill on average, and we find that treated taxpayers pay an additional 1.1-1.8 percent of their bill on average. Only 8.4 percent of control group taxpayers paid off their entire bill. Those receiving a reminder or a reciprocity message were 1.2-1.4 percentage points more likely to fully pay.

To check whether the timing of the messages across groups affected the impacts, we include analysis of our main results – whether citizens paid any part of their bill and the amount that they paid – for the subgroup that received messages on June 26 or later (Appendix Figure A1). Our pattern of results is consistent in this smaller sample. The impact on payment rates is higher for those who received any treatment, and point estimate is highest for the reciprocity group (Appendix Table A2). Likewise, the point estimates on total amount paid are highest for the reciprocity group, although in this subsample, the amount paid is quite high for the social pressure group as well (Appendix Table A3).

The remainder of the results section presents some extensions beyond the main results. This discussion is exploratory in nature and may be extended in future research.

4.1 External validity and the geographic spread of effects

In this subsection, we ask a simple question: had the TRA targeted a specific region of Dar es Salaam with messages, would they have had the same impact observed across the entire city?

Table 4: Impact of message assignment on payment amounts

	ITT estimates			2SLS estimates		
	Start of experiment	First tax deadline	One month after first deadline	Start of experiment	First tax deadline	One month after first deadline
Pooled treatment arms						
Treated	-5.260	210.4***	280.6***	-5.786	271.1***	358.8***
	(18.33)	(54.30)	(62.69)	(22.72)	(67.28)	(77.65)
Separate treatment arms						
T1: Reminder	-24.27	178.4**	235.3***	-32.67	218.2**	287.3***
	(19.62)	(59.39)	(68.57)	(24.86)	(75.37)	(87.02)
T2: Reciprocity	13.28	305.9***	392.9***	19.16	376.1***	480.0***
	(23.85)	(70.12)	(80.66)	(29.95)	(88.06)	(101.2)
T3: Social pressure	14.23	179.0**	258.9**	24.00	222.9*	330.4**
	(23.81)	(69.16)	(79.96)	(34.41)	(99.74)	(115.3)
Constant	223.6***	2178.9***	3130.5***			
	(16.43)	(48.33)	(55.83)			
Control mean	223.436	2184.581	3137.610	223.436	2184.581	3137.610
First stage f-stat				31,174.8	31,174.8	31,174.8
R^2	0.013	0.051	0.060	0.000	0.000	0.000
Obs	237,591	237,591	237,591	237,591	237,591	237,591
Test: T1 = T2	0.065	0.038	0.025	0.066	0.064	0.049
Test: T1 = T3	0.058	0.993	0.734	0.071	0.959	0.688
Test: T2 = T3	0.969	0.073	0.101	0.898	0.160	0.234

Notes: Outcome is the amount in Tanzanian shillings a taxpayer paid to the TRA by the date indicated. ITT estimates indicate impact of being assigned to treatment. 2SLS estimates instrument the receipt of each message type with assignment to treatment. Payment amounts winnorsized at the 99th percentile. Robust standard-errors in parentheses. $^{+}p < 0.10$,* $p < 0.05$,** $p < 0.01$,*** $p < 0.001$

Recent meta-analyses of nudges have revealed a significant amount of heterogeneity across contexts (Antinyan and Asatryan 2020; DellaVigna and Linos 2020). This heterogeneity is important to the policymaker, who will need to know the likelihood a given treatment will be effective in a new setting. If messages sent to taxpayers in different parts of the city have drastically different effects, then a revenue authority may want to consider targeting areas where the intervention is more cost effective, or tailor its messaging to be more effective in those areas.

As mentioned above, the randomization was conducted within 1,211 different strata, which were defined both by the location of the property (sub-ward) and whether a bill had been issued. This allows us to compare effect sizes across different locations to see if any systematic relationships appear.

For this analysis, we drop sub-wards with fewer than 100 observations (roughly 4 percent of the total experimental sample), leaving us with 360 sub-wards. For each sub-ward, we run our standard regressions, where the outcome is an indicator equal to one if the taxpayer made any payment to the TRA during the study period as a function of the reduced form treatment assignment, as well as including a control for whether that taxpayer received a bill not. We then recover both the coefficient β as well as calculate the control group mean μ from each regression. The top half of Figure 3 shows (i) the distribution of those effect sizes across the 360 sub-wards, and (ii) the distribution of average control group compliance. The figures demonstrate a substantial amount of variation in effect sizes across strata. To give a sense of the implications for comparing small-scale experiments to city-wide ones: if the experiment had

Table 5: Payment results for subsample with bills

	Full Sample	Bill Sample		
	Pr(any payment)	Pr(any payment)	Proportion paid	Pr(fully paid)
Pooled treatment arms				
Treated	0.0139***	0.0164***	0.0153***	0.0121**
	(0.00192)	(0.00486)	(0.00436)	(0.00415)
Separate treatment arms				
T1: Reminder	0.0144***	0.0169**	0.0150**	0.0118*
	(0.00218)	(0.00540)	(0.00484)	(0.00460)
T2: Reciprocity	0.0153***	0.0181**	0.0176**	0.0141**
	(0.00251)	(0.00602)	(0.00542)	(0.00515)
T3: Social pressure	0.00942***	0.0111	0.0112^{+}	0.00889
	(0.00285)	(0.00682)	(0.00614)	(0.00583)
Control mean	0.104	0.118	0.102	0.084
First stage f-stat	31,174.8	10,325.3	10,325.3	10,325.3
R^2	0.000	0.001	0.001	0.001
Obs	237,591	44,384	44,384	44,384
Test: T1 = T2	0.713	0.834	0.614	0.620
Test: T1 = T3	0.062	0.330	0.483	0.570
Test: T2 = T3	0.058	0.318	0.319	0.382

Notes: Table shows results for payment outcomes for both entire experimental sample and the subsample for which we have bill data. The outcomes in columns (1) and(2) is an indicator equal to one if the taxpayer has made any paymen to the TRA during the study period. The outcome in column (3) is the share of the total cumulative tax bill the taxpayer made by the end of the study period. The outcome in column (4) is an indicator equal to one if the taxpayer has fully paid their tax bill by the end of the study period. All results are 2SLS results where the actual messages sent to the taxpayer are instrumented using the original treatment assignment. Robust standard-errors in parentheses. $^{+}p < 0.10,^{*}\ p < 0.05,^{**}\ p < 0.01,^{***}\ p < 0.001$

been conducted in a randomly-chosen strata, the chance that the resulting effect size would have been within one percentage points of the "true" effect estimate is only 58 percent. Only 65 percent of sub-wards had estimated effect sizes greater than zero.

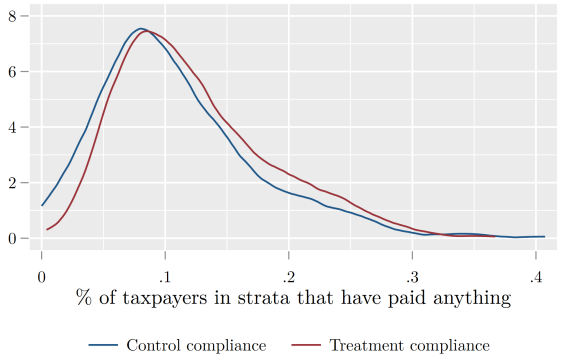
The bottom half of Figure 3 compares the estimated distributions of control group compliance and treatment group compliance across strata. The bottom left graph shows that the distribution of treatment group compliance is shifted up and to the right of control group compliance. The scatterplot in the bottom right shows that the effect sizes are inversely proportional to control level compliance: the treatment seems to be more effective in areas that have a higher "baseline" rate of non-compliance.

4.2 Cost-effectiveness of the impacts and comparisons to other estimated effects

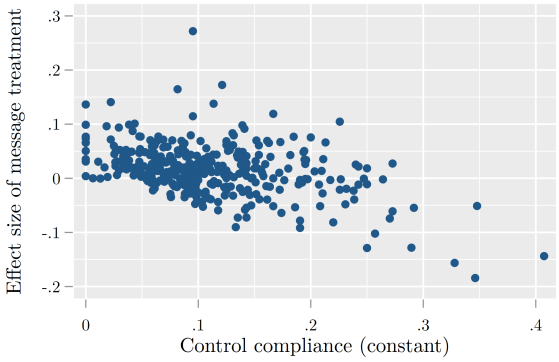
While the messages in our experiment have a relatively small effect on compliance and amount paid, this effect is can still be shown to have substantial practical relevance (cost-effectiveness). Through a straightforward calculation, we estimate that the experiment's benefits are around 20 times their cost. This comes from simply dividing the average increase in the amount paid to TRA due to the intervention (around 300 TSh) with the text-message campaign's cost, which is estimated to be around 15 Tsh per text message. The reciprocity treatment (impact on amount paid estimated at 480 TSh) would be the most cost-effective intervention in our setting with an increase in the amount paid of more than 30 times the messages' cost.

Figure 3: Distribution of effects and end-of-experiment compliance rates across strata

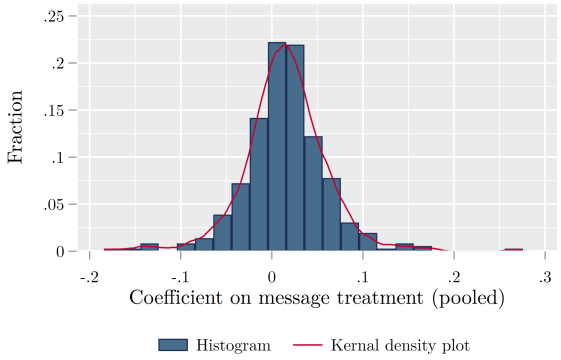
(a) Distribution of average compliance across strata for taxpayers receiving any message versus the control group



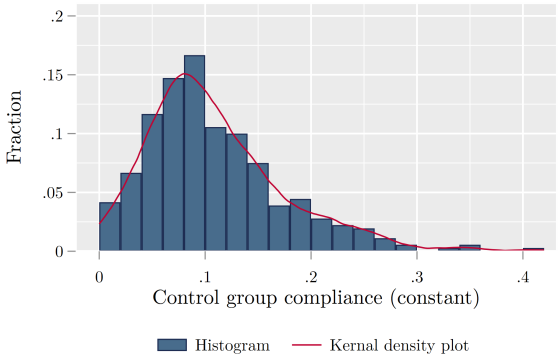
(b) Strata-specific effect size versus the average compliance rate for the control group



(c) Distribution of strata-specific effect sizes



(d) Distribution of control-group compliance rate across strata



Note: Each observation is either the coefficient or the constant from a strata-specific regression of whether a household has made any payments to the TRA during the study period on a dummy = 1 if the household was randomized into a message treatment. Sample includes all strata with at least 100 observations.

There are, of course, several factors that this simple calculation does not consider. First, the cost of the messages could be lower is the government achieves any discount with large numbers of messages. Second, as more taxpayers comply with payment, other taxpayers could start to emulate them, and so the messages could have an additional, spillover effect that we do not consider in this simple cost-effectiveness analysis. On the other hand, we cannot rule out that the messages have only a short term effect, although we suspect this is not the case. Previous work on text message reminders suggest that high frequency messages (e.g., daily reminders to take medicines) may lose efficacy, but not so with occasional reminders (Pop-Eleches et al. 2011).

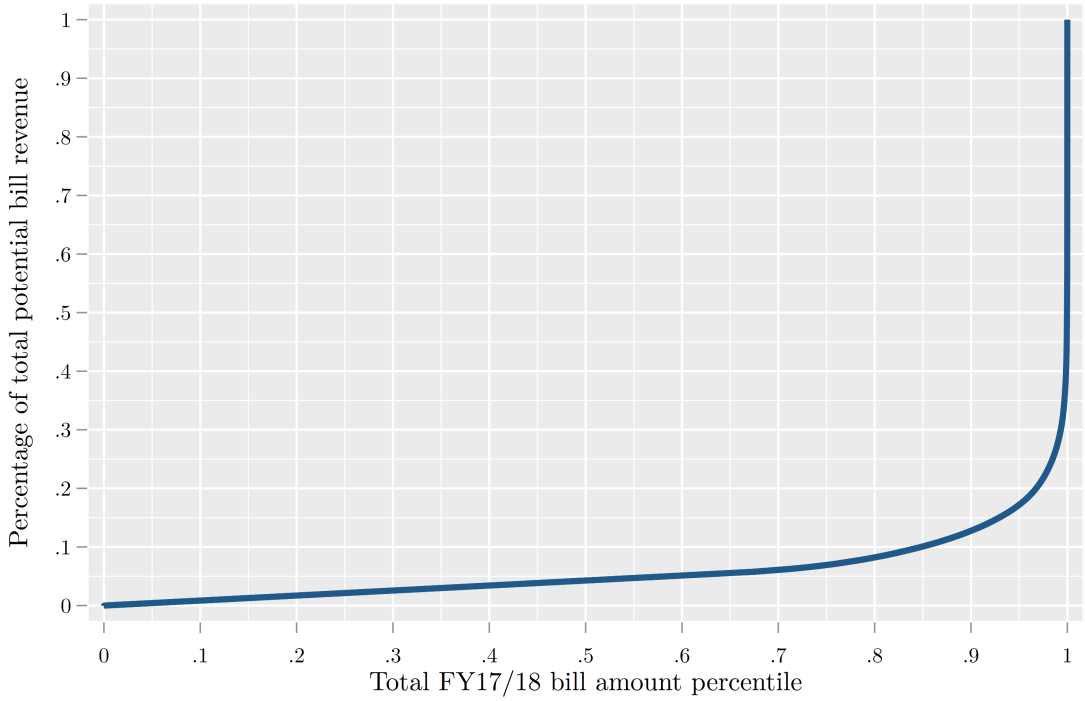
As described earlier in the paper, we did not explicitly test an enforcement message as part of the randomized control trial. However, a subset of taxpayers - not experimentally assigned - did receive an enforcement message from the TRA, indicating that they may face fines if they did not make their payment on time. We include those results in Section A1.2 in the Appendix,

with the caveat that those receiving the enforcement messages differed significantly from other participants across a number covariates.

Related to cost effectiveness, we can also infer some insights on how program implementation affects the program's benefits. From Table 4, we observe that the ITT estimated impact is of around 280 Tsh (see pooled treatment arms) while the 2SLS estimated impact is of around 360Tsh. As the ITT includes potential implementation failures (i.e., text messages not effectively received or opened), it gives an indication of the impact under imperfect implementation. On the other hand, the 2SLS would give an indication of the impact under implementation with greater fidelity.

4.3 Distributional effects

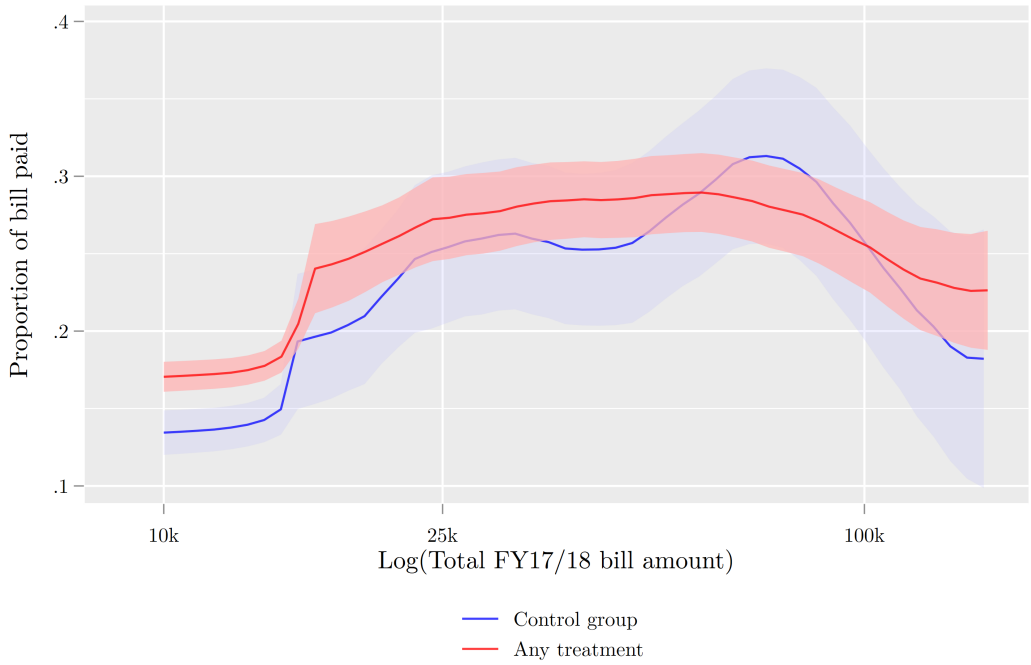
Figure 4: Potential bill revenue and bill amount percentiles



Note: Figure shows (on the y-axis) the cumulative proportion of the total the total 2017/18 fiscal year bill amount charged by the TRA against (on the x-axis) the percentile of bill amounts the taxpayer is in. For example, the first 50% of taxpayers, ranked by their bill amount, account for last than 5% being charged by the TRA.

Another dimension we explore in the data is the distribution of effects across types of properties. For the sample of properties for which a bill was issued, we order the bills by amount owed percentiles (for the 17/18 tax year, not including back taxes) and plot the percentage of the total potential bill revenue for each percentile in Figure 4. The graph indicates that in terms of potential revenues, most of the opportunity for increased revenue would come from the top 10 percent of billed properties, which account for more than 90 percent of the the total potential bill revenue. We then look at treatment heterogeneity across bill amounts in Figure

Figure 5: Proportion of bill paid and bill amount



Note: Figure shows local polynomial estimates of the relationship between (on the y-axis) the proportion of the total total 2017/18 fiscal year bill amount paid by taxpayers in the experimental sample and the log of the total , divided by whether they are (blue) in the control group or (red) received any treatment. 95% confidence intervals shown.

5. The red curve (with 95 percent confidence bands in red) shows the proportion of the bill paid after reception of any treatment, while the blue curve (with 95 percent confidence bands in blue) shows the proportion of the bill paid for the control group (i.e., the space between the lines indicates the impact of the program). The graph suggests that the treatments has a bigger impact for smaller bills, primarily those paying at or just above 10,000 TSh. This result offers insight into why the overall impact of the messages in terms of payment amounts is relatively low: most of the potential tax take would come from the very high valued properties/bills, but the messages have a stronger impact for the smaller bills. Because this heterogeneity is measured relative to the bill amount for a given property rather than household wealth or income, we cannot draw clear conclusions about the progressivity or regressivity of the reminders.[6]

5 Conclusion

In the face of limited tax compliance and limited capacity for enforcement of tax compliance, this study reports the impact of a randomized controlled trial to test different ways of leveraging tax morale in Dar es Salaam, Tanzania. The study tests salience (via a simple reminder), reciprocity (via a message highlighting the link between tax payment and publicly provided services), and

[6]Further complicating inferences on progressivity, we find that the reminders have a larger impact on properties that the TRA has valued, and it is highly likely that TRA has focused on valuing properties that are worth more (Appendix Figure A2).

social pressure (via a message emphasizing that those who do not pay are not contributing to national or local development).

We find positive impacts of all the messages, suggesting an impact of simple salience (since any message boosts salience). But we find statistically significantly higher tax payments with the reciprocity and social pressure messages, suggesting that it is possible for government to leverage these aspects of tax morale. While the absolute gains are not enormous, the interventions are very cheap, such that the benefit to cost ratio is estimated at 20:1 on average across all treatments and 30:1 for the more effective treatments. Text message reminders are one useful tool that governments can draw on to mobilize domestic resources for public services. That said, in absolute terms, tax systems will require a broad range of improvements, including more expansive registration and valuation and more effective enforcement, to substantively raise domestic revenues.

Future work in this area can be designed and statistically powered to test the progressivity of tax morale interventions and a wider range of potential messages to ascertain the sensitivity of tax morale interventions to implementation details around wording, length, and frequency.

References

Ali, M., O.-H. Fjeldstad, and L. Katera (2017). Property taxation in developing countries. *CMI Brief*.

Allingham, M. G. and A. Sandmo (1972). Income tax evasion: a theoretical analysis. *Journal of Public Economics 1*(3-4), 323–338.

Antinyan, A. and Z. Asatryan (2020). Nudging for tax compliance: A meta-analysis. *cesifo Working Papers 8500*.

Ariel, B. (2012). Deterrence and moral persuasion effects on corporate tax compliance: findings from a randomized controlled trial. *Criminology 50*(1), 27–69.

Besley, T. and T. Persson (2013). Taxation and development. In *Handbook of public economics*, Volume 5, pp. 51–110. Elsevier.

Blumenthal, M., C. Christian, J. Slemrod, and M. G. Smith (2001). Do normative appeals affect tax compliance? Evidence from a controlled experiment in Minnesota. *National Tax Journal*, 125–138.

Brockmeyer, A., A. M. Estefan, K. R. Arras, and J. C. S. Serrato (2020). Taxing property in developing countries: Theory and evidence from mexico. *Working Paper*.

Brockmeyer, A., S. Smith, M. Hernandez, and S. Kettle (2019). Casting a wider tax net: Experimental evidence from Costa Rica. *American Economic Journal: Economic Policy 11*(3), 55–87.

Carril, A. et al. (2017). Dealing with misfits in random treatment assignment. *Stata Journal 17*(3), 652–667.

Castro, L. and C. Scartascini (2015). Tax compliance and enforcement in the Pampas evidence from a field experiment. *Journal of Economic Behavior & Organization 116*, 65–82.

Cengiz, D., A. Dube, A. Lindner, and B. Zipperer (2019). The effect of minimum wages on low-wage jobs. *The Quarterly Journal of Economics 134*(3), 1405–1454.

Cohen, I. (2020). Low-cost tax capacity: A randomized evaluation with the Uganda Revenue Authority.

Coleman, S. (1996). The Minnesota income tax compliance experiment.

Del Carmen, G., E. E. Espinal Hernandez, and T. Scot (2020). Targeting in tax compliance interventions: Experimental evidence from honduras. *Working Paper*.

DellaVigna, S. and E. Linos (2020). Rcts to scale: Comprehensive evidence from two nudge units. Technical report, National Bureau of Economic Research.

Dincecco, M. and G. Katz (2014). State capacity and long-run economic performance. *The Economic Journal 126*(590), 189–218.

Fellner, G., R. Sausgruber, and C. Traxler (2013). Testing enforcement strategies in the field: Threat, moral appeal and social information. *Journal of the European Economic Association 11*(3), 634–660.

Fjeldstad, O.-H., M. Ali, and L. Katera (2017). Taxing the urban boom in Tanzania: Central versus local government property tax collection. *CMI Insight*.

Gadenne, L. (2017). Tax me, but spend wisely? sources of public finance and government accountability. *American Economic Journal: Applied Economics*.

Gaspar, V., L. Jaramillo, and M. P. Wingender (2016). *Tax Capacity and Growth: Is there a Tipping Point?* International Monetary Fund.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*.

Hallsworth, M., J. A. List, R. D. Metcalfe, and I. Vlaev (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics 148*, 14–31.

Hasseldine, J., P. Hite, S. James, and M. Toumi (2007). Persuasive communications: Tax compliance enforcement strategies for sole proprietors. *Contemporary Accounting Research 24*(1), 171–194.

Hernandez, M., J. Jamison, E. Korczyc, N. Mazar, and R. Sormani (2017). Applying behavioral insights to improve tax collection: experimental evidence from poland.

Herrmann, B., C. Toni, and S. Gachter (2008). Antisocial punishment across societies. *Science 319*, 1362–167.

Kettle, S., M. Hernandez, S. Ruda, and M. Sanders (2016). *Behavioral interventions in tax compliance: evidence from Guatemala*. The World Bank.

Khwaja, A. I., O. Haq, A. Q. Khan, B. Olken, and M. Shaukat (2020). *Rebuilding the social compact: Urban service delivery and property taxes in Pakistan*. 3ie.

Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. Pedersen, and E. Saez (2011). Unwilling or unable to cheat? evidence from a tax audit experiment in Denmark. *Econometrica 79*(3), 651–692.

Krause, B. (2020). Balancing purse and peace: Tax collection, public goods and protests.

Luttmer, E. F. and M. Singhal (2014). Tax morale. *Journal of Economic Perspectives 28*, 149–168.

Mascagni, G., C. Nell, and N. Monkam (2017). One size does not fit all: a field experiment on the drivers of tax compliance and delivery methods in Rwanda.

Meiselman, B. S. (2018). Ghostbusting in Detroit: Evidence on nonfilers from a controlled field experiment. *Journal of Public Economics 158*, 180–193.

Norregaard, M. J. (2013). *Taxing Immovable Property Revenue Potential and Implementation Challenges*. Number 13-129. International Monetary Fund.

Ortega, D. and C. Scartascini (2020). Don't blame the messenger. the delivery method of a message matters. *Journal of Economic Behavior & Organization 170*, 286–300.

PO-RALG (2013). A study on LGAs own source revenue collection. *PO-RALG - Internal Report*.

Pomeranz, D. (2015). No taxation without information: Deterrence and self-enforcement in the value added tax. *American Economic Review 105*(8), 2539–69.

Pop-Eleches, C., H. Thirumurthy, J. P. Habyarimana, J. G. Zivin, M. P. Goldestein, D. de Walque, L. MacKeen, J. Haberer, S. Kimaiyo, J. Sidle, D. Ngare, and D. R. Bangsberg (2011). Mobile phone technologies improve adherence to antiretroviral treatment in a resource-limited setting: a randomized controlled trial of text message reminders. *AIDS 25*(6), 825–834.

Torgler, B. (2004). Moral suasion: An alternative tax policy strategy? Evidence from a controlled field experiment in Switzerland. *Economics of Governance 5*(3), 235–253.

A1 Appendix

A1.1 Additional Graphs and Tables

Table A1: Balance tests for bill subsample

	Full Sample	**Bill Sample**		
	(1) Probability(bill issued)	(2) Log(total bill)	(3) Log(17/18 bill)	(4) IHS(property value)
T1: Reminder	-0.0000744	-0.000424	0.000955	-0.111
	(0.00219)	(0.0113)	(0.00999)	(0.0775)
T2: Reciprocity	-0.000170	0.000618	0.00814	-0.0450
	(0.00253)	(0.0130)	(0.0115)	(0.0893)
T3: Social pressure	-0.0000395	0.0294*	0.0235*	0.0271
	(0.00253)	(0.0131)	(0.0117)	(0.0896)
Control mean	.1873409736	341,915.137	135,042.359	84,058,493
R^2	0.000	0.452	0.537	0.526
Obs	237,699	44,380	38,107	44,384
Test: T1 = T2	0.965	0.925	0.462	0.393
Test: T1 = T3	0.987	0.008	0.024	0.074
Test: T2 = T3	0.959	0.025	0.180	0.418

Notes: Table shows balance tests for four bill related outcomes. Outcome (1) is a binary outcome = 1 if the taxpayer was issued a bill by the TRA. Outcome (2) is the log of the cumulative bill amount (current and back taxes) across all the taxpayer's properties. Outcome (3) is the log of the 2017/2018 tax owed only. Outcome (4) is the inverse hyperbolic sign transformation of the assessed property value by the TRA. The treatment measures are the original, intent-to-treat indicators. Robust standard-errors in parentheses. $^{+}p < 0.10,^{*}p < 0.05,^{**}p < 0.01,^{***}p < 0.001$

Table A2: Impact of message assignment on payment rates for taxpayers that only received a message four or fewer days before the first deadline

	ITT estimates			2SLS estimates		
	Start of experiment	First tax deadline	One month after first deadline	Start of experiment	First tax deadline	One month after first deadline
Pooled treatment arms						
Treated	-0.000179	0.00623***	0.00982***	-0.000188	0.00801***	0.0125***
	(0.000435)	(0.00137)	(0.00165)	(0.000548)	(0.00173)	(0.00208)
Separate treatment arms						
T1: Reminder	-0.000367	0.00562***	0.00920***	-0.000546	0.00745**	0.0123***
	(0.000507)	(0.00162)	(0.00196)	(0.000725)	(0.00233)	(0.00281)
T2: Reciprocity	-0.0000249	0.00765***	0.0115***	0.00000680	0.00912***	0.0136***
	(0.000588)	(0.00188)	(0.00226)	(0.000752)	(0.00241)	(0.00290)
T3: Social pressure	-0.0000661	0.00594***	0.00934***	-0.0000687	0.00743**	0.0117***
	(0.000537)	(0.00170)	(0.00205)	(0.000723)	(0.00230)	(0.00277)
Constant	0.00648***	0.0681***	0.104***			
	(0.000374)	(0.00116)	(0.00140)			
Control mean	0.006	0.068	0.103	0.006	0.068	0.103
First stage F-stat				49,161.3	49,161.3	49,161.3
R^2	0.012	0.039	0.044	0.000	0.000	0.001
Obs	171,129	171,129	171,129	171,129	171,129	171,129
Test: T1 = T2	0.547	0.276	0.309	0.515	0.549	0.705
Test: T1 = T3	0.560	0.851	0.947	0.547	0.994	0.849
Test: T2 = T3	0.945	0.375	0.356	0.932	0.556	0.589

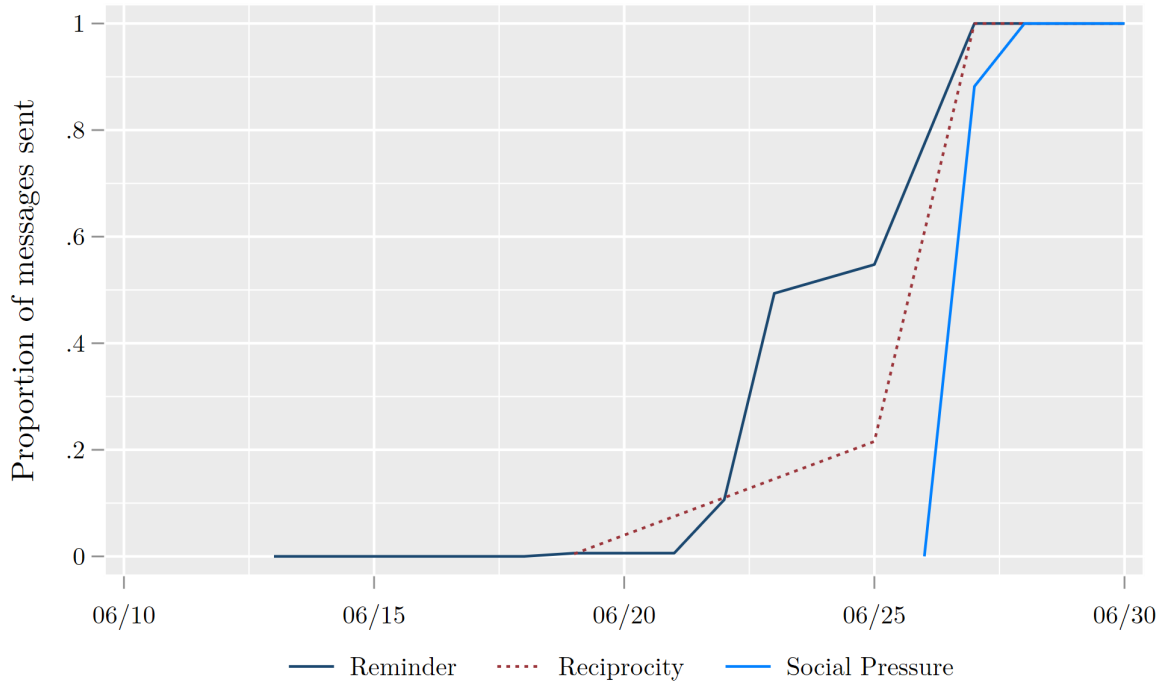
Notes: Outcome is the probability a taxpayer made any payment to the TRA by the date indicated. ITT estimates indicate impact of being assigned to treatment. 2SLS estimates instrument the receipt of each message type with assignment to treatment. Robust standard-errors in parentheses. Sample restricted to taxpayers that received no message or only received a message on June 26th or after. $^{+}p < 0.10$,* $p < 0.05$,** $p < 0.01$,*** $p < 0.001$

Table A3: Impact of message assignment on payment amounts for taxpayers that only received a message four or fewer days before the first deadline

	ITT estimates			2SLS estimates		
	Start of experiment	First tax deadline	One month after first deadline	Start of experiment	First tax deadline	One month after first deadline
Pooled treatment arms						
Treated	-2.896	175.7**	263.0***	1.137	241.7***	353.1***
	(19.65)	(57.09)	(66.10)	(24.89)	(72.07)	(83.39)
Separate treatment arms						
T1: Reminder	-26.35	61.61	130.1+	-41.77	53.70	141.6
	(22.37)	(66.44)	(77.06)	(31.90)	(95.07)	(110.3)
T2: Reciprocity	19.55	272.7***	362.3***	27.70	333.2***	434.8***
	(27.61)	(78.88)	(90.69)	(35.49)	(101.3)	(116.3)
T3: Social pressure	8.728	241.1***	349.0***	12.39	314.9**	455.8***
	(24.70)	(71.72)	(83.13)	(33.35)	(96.86)	(112.3)
Constant	227.2***	2146.8***	3098.3***			
	(16.83)	(48.60)	(56.29)			
Control mean	223.291	2127.732	3075.261	223.291	2127.732	3075.261
First stage F-stat				49,161.3	49,161.3	49,161.3
R^2	0.016	0.052	0.060	0.000	0.000	0.000
Obs	171,129	171,129	171,129	171,129	171,129	171,129
Test: T1 = T2	0.082	0.006	0.009	0.076	0.015	0.027
Test: T1 = T3	0.135	0.010	0.007	0.131	0.014	0.011
Test: T2 = T3	0.704	0.699	0.887	0.717	0.880	0.880

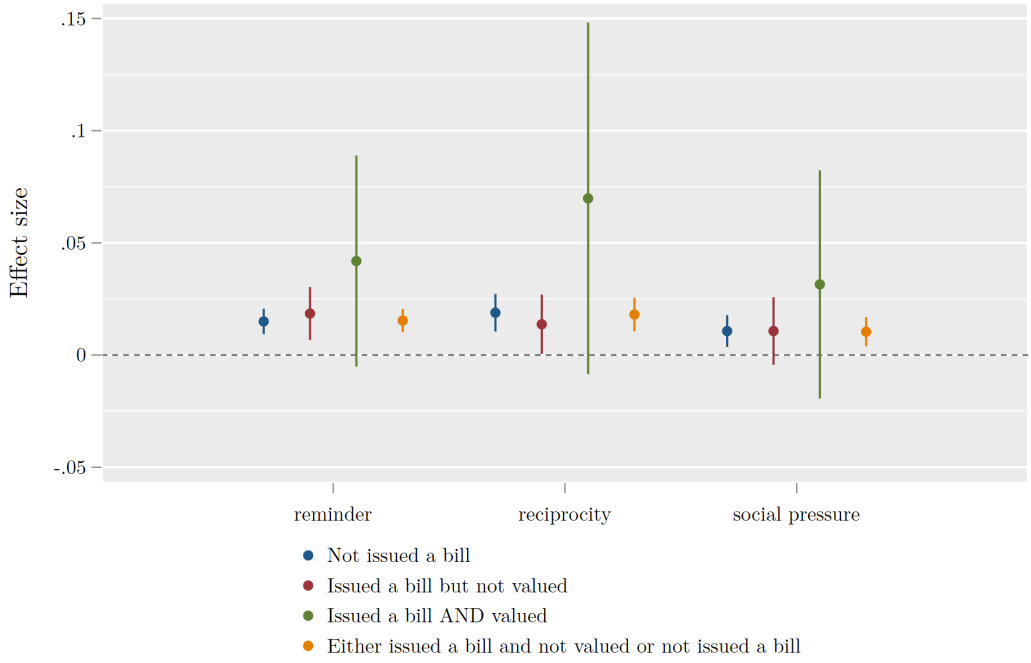
Notes: Outcome is the amount in Tanzanian shillings a taxpayer paid to the TRA by the date indicated. ITT estimates indicate impact of being assigned to treatment. 2SLS estimates instrument the receipt of each message type with assignment to treatment. Sample restricted to taxpayers that received no message or only received a message on June 26th or after. Payment amounts winnorsized at the 99th percentile. Robust standard-errors in parentheses. $^{+}p < 0.10$,* $p < 0.05$,** $p < 0.01$,*** $p < 0.001$

Figure A1: Timeline of message delivery



Note: Graph shows the cumulative proportion of messages sent (out of messages sent to **all** taxpayers, not just those in our experimental sample) over time.

Figure A2: Impact of treatment on probability of any payment by type of property



Note: Figure shows the effect size (estimated using 2SLS) by type of property: (a) those that had not been issued a bill, (b) those that had been issued a bill but had not been valued by the TRA (and thus owed a flat fee of 10,000 TSh), (c) those that had both been issued a bill and positively valued by the TRA and (d) those that had not been issued a bill or had been issued a bill but not valued.

A1.2 Enforcement message difference-in-difference estimates

As described in the main text, those assigned to enforcement messages in our experiment were, due to an error, instead sent reminder messages. However, the Tanzanian Revenue Authority did send its own set of enforcement messages to a subset of all property taxpayers, including nearly 600 taxpayers in our experimental sample.

To investigate how an enforcement message might have induced compliance among this group, we followed an event-study design, to observe whether payment rates increased following the receipt of the first message.

There are several caveats to this approach. The taxpayers that received enforcement messages were not randomly selected. We observe that they were more likely to have received a bill from the TRA and, conditional on receiving a bill, had higher estimated property values. Thus any estimated effect might be driven by a combination of the actual effect of the enforcement message together with the effect of any unobserved characteristics (e.g., propensity to pay, unobserved effort on the part of the TRA) which induced these taxpayers to pay before the initial deadline.

Because different taxpayers received an initial message at different points in time, we account for the fact that this is what is now referred to as a "staggered" difference-in-differences framework. To account for the problems that can develop when units are treated at different times (Goodman-Bacon 2021), there is expected heterogeneity in the treatment and heterogeneity across time, we proceed using a stacked difference-in-differences design (Cengiz et al. 2019).

For a single message type, we retain all taxpayers that received that message and all control taxpayers. We then consider each event-cohort separately: keeping all taxpayers that were treated in a given week w and all control taxpayers. We then stack those event-cohorts in event-time, so that we have a pooled sample consisting of each event-cohort. This means that control units will enter into the sample multiple times, but will be accounted for with fixed effects specific to the event-cohort.

We then run a regression of the following form:

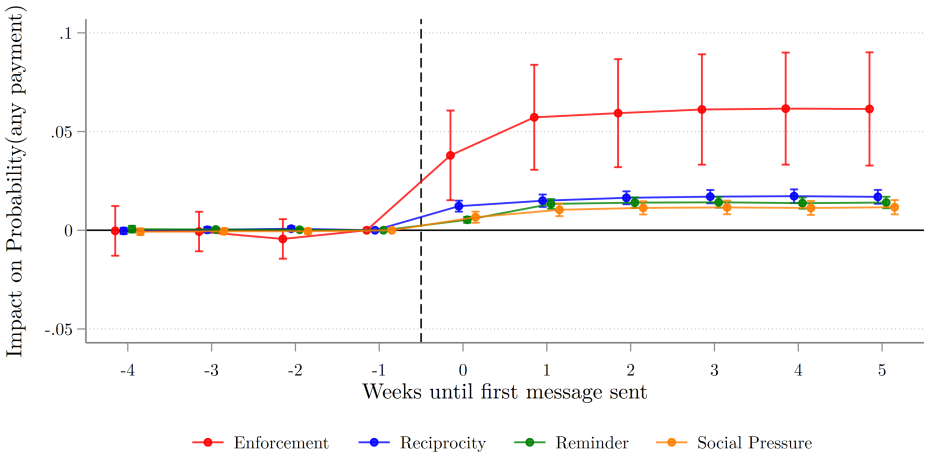
$$P_{icw} = \sum_{j=-4}^{5} \beta_j (message_{ic} \times 1\{w = j\}) + \mu_{ic} + \gamma_{cw} + \epsilon_{icw} \tag{3}$$

In which P_{icw} is the outcome of interest for taxpayer i in week w for event-cohort c. The indicator variable $Message_{ic}$ is $= 1$ if the taxpayer ever received a message, and so the first term in the equation is a series of event dummies that are $= 1$ when the taxpayer is -4,-3... 5 weeks out from receiving his or her first message. Parameters μ_{ic} and γ_{cw} are taxpayer-cohort and cohort-week dummies, respectively. We run Equation (3) separately for each message type (enforcement, reciprocity, reminder and social pressure) separately, so in each instance, the receipt of that type of message is being compared to a pure control group that received no message.

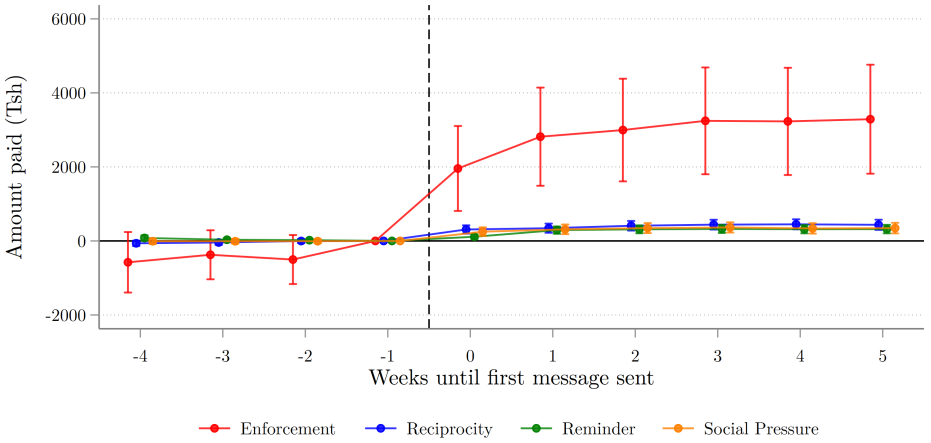
Figure A3 displays the event study coefficients (β_j) for (a) any payment being made and (b) the total amount paid in Tanzanian shillings. For both outcomes, taxpayers that received an enforcement message saw substantially faster growth following the treatment period. Five weeks

Figure A3: Stacked event-study estimates of impact of different message types

(a) Outcome: any payment



(b) Outcome: total paid



Note: Figures graph the event study dummies from equation (3), which chart the impact of receiving a message from the TRA on either (a) whether the taxpayer has made any payment or (b) the total amount paid to the TRA. The sample is the experimental sample described in the main paper. 95% confidence intervals shown.

following the treatment, those sent enforcement messages were roughly 6 percentage points more likely to have made a payment (versus 1.7 percentage points for the reminder treatment) and had paid roughly Tsh 3,200 (versus 430 for the reminder treatment).[7] Given the challenge with non-comparability across groups, this only yields suggestive evidence that enforcement messages may merit further study.

[7]These effects also appear to hold as a percentage of the total amount owed, when we consider the subset of taxpayers with bills (results available upon request).

1775 MASSACHUSETTS AVE NW
WASHINGTON, D.C. 20036

brookings.edu/global

B | Global Economy and Development at BROOKINGS