Technology and Tax Capacity: Evidence from Local Governments in Ghana*

James Dzansi — *IGC*
Anders Jensen — *HKS and NBER*
David Lagakos — *BU and NBER*
Henry Telli — *IGC*

September 2022

Abstract

This paper studies the role of technology in local-government tax collection capacity in the developing world. We first conduct a new census of all local governments in Ghana to document a strong association between technology use and property tax billing, collection and enforcement. We then randomize the use of a new revenue collection technology within one large municipal government. Revenue collectors using the new technology delivered 27 percent more bills and collected 103 percent more tax revenues than control collectors. Collectors using the new technology learned faster about which households in their assigned areas were willing and able to make payments. We reconcile these experimental findings in a simple Beckerian time-use model in which technology allows revenue collectors to better allocate their time towards households that are the most likely to comply with taxpaying duties. The model's predictions are consistent with experimental evidence showing that treatment collectors are more likely to target households with greater liquidity, income, awareness of taxpaying duties, and satisfaction with local public goods provision.

*james.dzansi@theigc.org, anders_jensen@hks.harvard.edu, lagakos@bu.edu, henry.telli@theigc.org.
For helpful comments on this project we thank Augustin Bergeron, Anne Brockmeyer, Ray Fisman, Andrew Foster, Rema Hanna, Gordon Hanson, Asim Khwaja, Gabriel Kreindler, Craig McIntosh, Karthik Muralidharan, Dani Rodrik, Nii Sowa, Chris Udry, Silvia Vannutelli, Mazhar Waseem, Jaya Wen as well as seminar participants at Brown, BU, Dartmouth, Harvard, NYU Abu Dhabi, Peking HSBC Business School, USC, Virginia, Williams and Yale. For outstanding research assistance we thank Manon Delvaux, Soala Ekine, Radhika Goyal, Mary Nyarkpoh, Isaac Otoo and Cynthia Zindam. For help implementing a pilot version of the experiment we thank IPA Ghana. For financial support we thank the International Growth Centre and J-PAL. All potential errors are our own. Harvard IRB approval: IRB17-1310. AEA RCT Registry ID: AEARCTR-0007267.

1 Introduction

A common feature of lower-income countries is a government that collects little tax revenue and provides few public goods. The literature on state capacity argues that the inability to collect taxes is at the heart of why low-income countries are as poor as they are (e.g. Besley, Ilzetzki, and Persson, 2013; Dincecco and Katz, 2016; Mayshar, Moav, and Pascali, 2022). This literature suggests that the path to development may begin with investing in the capacity to collect taxes to finance productivity-enhancing public goods.

This paper studies the role of technology in improving government tax capacity. The setting is local governments in Ghana, which are in charge of collecting property taxes but collect very little in practice. As we detail below, the technology in question consists of a geospatial database of properties embedded into an electronic tablet with GPS capabilities. Similar technologies have been implemented in a number of other developing countries in recent years with a goal of increasing tax revenues (Fish and Prichard, 2017). However, technology investments may ultimately have limited impact if they do not address the most important constraints on local tax capacity; moreover, the implementation of technology in the field may fail due to logistical and technical challenges or subversion by officials. To our knowledge, our paper is the first to randomize the presence of a new technology for tax collection to causally study its impacts.[1]

We set the stage by establishing facts about local tax capacity, based on a new census we conducted in every local government in Ghana. The census data highlight how poor infrastructure constrains tax collection practices. In a typical district fewer than half of property tax bills are actually delivered. Limited delivery stems from the fact that only around one in four properties has a physical address, and a minority of streets have names. As a result, location information on bills is imprecise and revenue collectors struggle to navigate in the field. Household compliance is low due to citizens' limited awareness of taxpaying duties, low satisfaction with public services, and limited enforcement. The majority of tax payments are paid in cash directly to revenue collectors, creating opportunities for "leakages." Our census reveals that a minority of local governments have adopted revenue management software and electronic databases of properties, and these governments have significantly better outcomes at every stage in the tax collection process. In particular, they deliver more bills, have higher payment rates and collect revenues with lower cost than governments without technology.

[1]Several prior studies have leveraged policy reforms to create non-experimental variation in technology usage, focusing on technologies which digitize third-party transactions between taxpayers, including Eissa and Zeitlin (2014); Brockmeyer and Somarriba (2022) and Fan, Liu, Qian, and Wen (2021); see Okunogbe and Santoro (2021) for a review of other related studies.

The tight empirical link between technology use and tax collection outcomes at the local-government level naturally invites questions about the direction of causality. To address these issues, we partnered with one large municipal government in Ghana and a private technology firm to randomize the use of technology within the government's jurisdiction. In particular, we randomized the use of a new revenue collection software and geospatial database of properties at the level of a revenue collector. In the experiment, both treatment and control collectors were given a stack of around 135 bills of similar value in a randomly assigned area and tasked with collecting as much revenue as possible in six weeks. The treatment group was given an electronic tablet that uses the geospatial data and improves navigation and locating properties. Otherwise the two groups of collectors, and their assigned areas, were observationally similar.

Revenue collectors using new technology delivered 27 percent more bills than the control collectors by the end of the study. We view this result as reflecting the navigational advantage that the technology provides in locating taxpayers more efficiently in an environment with scant property addressing. The time series of cumulative bills delivered exhibits a concave pattern over time, as collectors shift emphasis from delivering bills to following up with the households that already received a bill in order to collect payment from them. Revenue collections were 103 percent higher among the collectors assigned to the technology group, on average, implying a treatment effect on revenue collections around four times as large as the effect on bills delivered. Moreover, we find that the treatment effect on collections grows over time, leading to a rising average amount collected per bill delivered throughout the experiment.

We explore several potential hypotheses for why the treatment effect on collections is so much larger than the treatment effect on bill deliveries. One story is that households in treatment areas change attitudes toward payment after being visited by collectors with a modern technology. Yet households in treatment and control areas surveyed right after the experiment report statistically similar levels of perceived integrity and ability to enforce tax payments among local government officials. A second hypothesis is that the technology helps reduce leakages, e.g. in the form of payments made by households but diverted by revenue collectors before reaching the local government's coffers. However, several types of household survey questions about the preponderance of bribe payments point to more – rather than less – bribe activity in treatment areas than in control areas.

Our preferred explanation, which is new to this literature, is that technology allows collectors to directly gather knowledge about which households are most likely to make tax payments, and to better target those households in their collection efforts. By reducing navigational challenges, technology frees up scarce time for collectors which they

use to learn, through repeat visits or longer visits, about the hard-to-observe household characteristics that are most relevant in determining a household's propensity to pay. Using surveys of collector behavior and strategies, we show that treatment collectors report fewer navigational challenges and better knowledge over time about which households in their assigned areas are willing and able to pay property taxes. In addition, we show that treatment collectors switch more over time into collection strategies focused on households with greater liquidity, income, awareness of taxpaying duties, and satisfaction with local public goods. Importantly, none of these household characteristics would have been known to the collectors at the start of the study period, highlighting the importance of technology for acquiring information about the `soft' characteristics of taxpayers that would not be readily apparent on a tax bill or in a property database. [2]

We formalize this mechanism by embedding time-savings and learning in a dynamic Beckerian time use model in which forward-looking revenue collectors maximize collections subject to a time constraint each period. Households have a high or low payment probability, and the type is initially unknown to the collectors. Treatment collectors have higher probabilities of delivering a bill, which captures technology's navigational advantage in locating households. Treatment collectors also have a higher probability of learning a household's type, which captures how treatment collectors leverage the technology to learn more about household types. The probability of collecting a tax payment from a household of each type is identical across treatment and control collectors.

We calibrate the model to match the treatment effects on bill deliveries and on collector focus on hard-to-observe household characteristics. The calibrated model reproduces the concave time series pattern of treatment effects on bill deliveries and the convex effect on collections. The model's implied treatment effect on collections is roughly twice as large as its treatment effect on deliveries. Counterfactual simulations reveal that without faster learning about household types in the treatment group, the treatment effect on bill deliveries and collections would be similar in magnitude. Thus, our model shows that the differential learning and targeting mechanism explains roughly half the gap between delivery and collection effects observed in our experiment, making it about equally important as the direct effect of reduced navigational challenges in delivering bills.

Improved learning through technology has important distributional impacts. Due to increased knowledge and subsequent targeting of higher-income households in treatment areas, the tax system becomes more progressive: we nd that tax payments as

[2] Our nding that technology impacts collector strategies is related to other experiments with tax collectors, including on performance-based postings and nancial incentives (Khan, Khwaja, and Olken, 2015, 2019) and group-work assignments (Bergeron, Bessone, Kabeya, Tourek, and Weigel, 2021).

a share of taxes due increases in the top quartiles of the income-asset distribution but remains unchanged in the bottom quartile. However, technology appears to be a double-edged sword, as the treatment effect on bribes is concentrated in the bottom quartiles.

Our experimental ndings shed novel light on the promises and pitfalls of using technology to build tax capacity.[3] Our results support theories of government arguing that technology investments lead to growth in government size, including due to ef - ciency improvements (Brennan and Buchanan, 1980; Kau and Rubin, 1981; Becker and Mulligan, 2003). At the same time, our bribe results suggest that early expansions of government size may be accompanied by increased corruption (as in Daunton, 2001; Carpenter, 2020; Cui, 2022). The technology we study improved tax outcomes by alleviating constraints on collection which arose from incomplete property addressing. The United Nations estimates that 4 billion people live in places without physical addresses (web link); in this context, our paper provides some of the rst evidence about how technology may help overcome constraints on public service delivery stemming from incomplete property addressing infrastructure.

Our results suggest that the positive effects on tax outcomes are only partly due to the presence itself of technology. By reducing navigational challenges, technology freed up time for collectors in the eld which they used to directly build 'soft' information about taxpayers' propensity to pay. Our ndings therefore relate to papers which show how pre-existinginformation, mainly from third-parties including employers and nancial institutions, can be leveraged to improve collection (Kleven, Knudsen, Kreiner, Pedersen, and Saez, 2011; Pomeranz, 2015; Naritomi, 2019; Balan, Bergeron, Tourek, and Weigel, 2022; Londoño-Vélez and Ávila-Mahecha, 2021). Most prior studies place third-party 'hard' information at the center of governments' informational capacity (Gordon and Li, 2009; Kleven, Kreiner, and Saez, 2016). Our work shows how, when such hard information sources are largely non-existent, the state can still strengthen its informational capacity by directly building 'soft' information on taxpayers. The importance of 'soft' information for tax capacity is most relevant in developing countries where third-party information often remains limited (Almunia, Hjort, Knebelmann, and Tian, 2022; Jensen, 2022; Waseem, 2022).[4]

[3]Experimental evidence on technology exists in other governance areas, including social transfers (Muralidharan, Niehaus, and Sukhtankar, 2016) and monitoring (Callen, Gulzar, Hasanain, Khan, and Rezaee, 2020; Dal Bo, Finan, Li, and Schechter, 2021; Vannutelli, 2022). Studies in public nance have indirectly highlighted technology's value by providing taxpayers with incentives or information made available due to its presence (Carillo, Pomeranz, and Singhal, 2017; Okunogbe and Pouliquen, 2022).

[4]Our work relates to recent studies on property taxation in developing countries, including those cited in footnote 2 and Best, Gerard, Kresch, and Naritomi (2020) and Brockmeyer, Estefan, Suárez Serrato, and Ramírez (2020). For a historical analysis of the US property tax, see Dray, Landais, and Stantcheva (2022).

2 Census of Tax Collection Capacity in Local Government

There is limited systematic evidence about the process of tax collection by local governments in the developing world. To better understand this process, and the constraints on collection, we conducted a census of all local governments in Ghana in 2017, focusing on taxation. In this section we summarize the main ndings from this census.

2.1 Census of Local Government Tax Capacity

We conducted the census of local governments in fall 2017 in collaboration with several national ministries and all the local governments in Ghana's 216 districts (one local government per district). The aim of the census was to collect data on every relevant dimension of tax collection in each local government. Three sets of respondents were interviewed: local of cials; locally elected assembly members; and, citizens. Within the rst set – which is the most extensive – survey responses were collected from every of cial that participated in the tax collection process. These included the chief executive (the political head), the coordinating director (the bureaucratic head), the nance of cer, the budget of cer, the physical planning of cer, the revenue accountants, and the revenue collectors. Survey modules for of cials and assembly members captured information on the tax collection process and demographics and experience. Surveys of the citizens measured tax morale, knowledge about local taxes and demand for public goods.

The census contains 5,375 citizen responses (approximately 25 per district) and 2,785 local government of cials and assembly members (13 per district). In addition to the survey data, we digitized and harmonized administrative records to measure all sources of local tax collection and all types of local public expenditure across the 216 districts.

2.2 Local Governments have Limited Tax Collection and Information

Our census data allows us to document facts on local tax capacity and its constraints, which are reported in Table 1. As can be seen in Panel A, the average local taxes collected per person is only 4.2 Ghanaian currency GHC ($0.67), with a median of 2.6 GHC. This is a small amount relatively to per capita income in Ghana, and a disappointingly low amount in the eyes of the Ghanian Government (Government of Ghana, 2014).

It is useful to consider that tax collections are determined both by the probability of bill delivery (the delivery margin) and the amount paid conditional on delivery (the payment margin). On the delivery margin, we nd that in the average local government only 43 percent of property tax bills are delivered. The delivery margin is thus an important determinant of low tax collection. Many studies in public nance and development

abstract from bill delivery, and focus instead on the payment margin. Of cials in our survey indeed report challenges on the payment margin. In the typical district, the likelihood of a property owner paying a bill after receiving it is just 30.2 percent.

Government of cials cite limited information as a key constraint on tax collection. Indeed, in Panel A of Figure 1, the absence of data on residential and commercial properties is the most frequently cited constraint on tax collection amongst chief executives, other local bureaucrats, and assembly members. In relation to the delivery margin, absence of data on property owners and challenges in locating them are cited as two of the three most important constraints on bill delivery (Panel B, Figure 1). The lack of information starts with the simple absence of precise street addressing: Panel B of Table 1 shows that, in the average district, only 26.7 percent of properties have an of cial address. The property tax registry inherits this limited address information, which leads the location on many property tax bills to be imprecise. Figure A1 provides an illustration of an actual tax bill in Madina (the local government where our experiment takes place). The only information on the location of the property is ”Opposite Presec School” (a secondary boarding school). Across all of Ghana's districts, 74 percent of the revenue collectors we surveyed report that it is common not to be able to locate the property and/or the owner (Panel B, Table 1).

By constraining tax collector activities, limited information contributes to a high cost of collecting taxes (Panel C of Table 1). We proxy for this cost with the average monthly salary of revenue collectors as a percent of average monthly collections. Note that this is likely to be an under-estimate of the total cost, which may include other indirect expenditures such as for vehicles, fuel, and of ce staff time. Nevertheless, in the typical local government, the cost of collection is an astonishing 64.1 percentIn other words, for every 100 Ghanaian Cedi in taxes collected, the local government retains only 35.9 Cedis after its tax collectors are paid. By contrast, the U.S. Internal Revenue Service estimates that it retains 99.7 out of 100 dollars collected (web link).

Finally, three other constraints on tax capacity are important to highlight in this study's context. First, only 17.1 percent of properties have of cial valuations in the average district. In the absence of of cial valuations, local governments are forced to tax properties according to a presumptive schedule, where the tax liability is based on easily observable proxies, such as number of oors, proximity to the city center, or the type of business, in the case of a bill to a business.[5] Second, in the average local government, an

[5]To feasibly implement market valuation methods requires continuously updated property and market information, including from third-parties such as banks and mortgage providers. Market valuation property taxes are more common in developed countries, while presumptive tax schedules are more common in developing countries with limited information and administrative constraints.

estimated 72.1 percent of property tax payments are made in cash directly to collectors, rather than by check or via electronic transfers directly to the local government nance of ce. Cash payments provide collectors with discretion to capture some of the household taxes and reduce the payments to the local government. Third, local governments face signi cant constraints to enforcing taxes due. Court action (whereby the delinquent property can be con scated), or the threat of such action, is in practice the main enforcement tool that effectively has monetary consequences for tax delinquents. Yet only 22 percent of local governments report taking any tax defaulters to court in the previous year. Survey responses indicate that the reasons for limited court action lie outside the tax administration's immediate scope, and are due to legal constraints or political costs.

2.3 Technology and Taxation Outcomes

Some of the districts in the census report using computers, software and databases to help them distribute bills and collect revenues. We summarize the use of technology in each district by whether the local government has either a digital database of properties or a revenue software. Using this de nition, technology is only adopted by 17 percent of local governments in the country (Panel B, Table 1). Of these, 12 percent have both technology components, while 5 percent has only one of the two. Conditional on adopting, the average district has been using technology for nearly a decade (8.75 years).

Adoption of technology is at the discretion of each local government, and the variation in adoption across the country re ects individual governments' choices. Appendix Table A1 provides cross-district correlates of adoption choices. We nd that local governments are more likely to adopt technology when their district has a larger share of properties with of cial street addresses and property valuations and when legal capacity is stronger. Adoption is also positively correlated with district population size and its urban share. These results may suggest that technology investment is complementary to other characteristics that permit higher tax collection (Besley and Persson, 2009).

These patterns of selection contextualize the impacts of technology adoption on tax capacity. While we provide experimental evidence on these impacts within one local government in Section 3 onward, here we leverage the variation in adoption across the country to investigate the cross-district association between technology and tax outcomes at the level of entire local governments. In this exercise, the key identi cation concern is that adoption of the technology may be correlated with other district characteristics that also determine tax collection. We make some headway on this concern. First, we include the district covariates that are found to statistically predict adoption (Table A1). Second, we include the (district-speci c) share of geographically adjacent districts that

have adopted technology, to capture the in uence of neighboring policies on local governments' decisions. Third, we include ten region xed effects to narrow the comparison between adopters and non-adopters within each region.

The results are presented in Table 2. Technology is associated with improved outcomes at each step of the collection process: tax collection per capita (Panel A); share of bills delivered (Panel B); taxes paid per bill delivered (Panel C); and cost of collection (Panel D). When extensive controls are included (column 5), technology adoption is associated with a 78 percent increase in taxes collected. With these same controls, technology adoption is associated with a 21 percent increase in the share of bills delivered and a 37 percent increase in taxes paid per bill delivered. Finally, by improving both the delivery and payment margins, technology adoption is associated with a 13 percent decrease in the cost of collection (as a percent of taxes collected, Panel D).

3 Experimental Evidence

The census results of the previous section showed a strong association between technology use and tax outcomes in the cross-section of local governments in Ghana. This association suggests a potentially important role for technology in alleviating some constraints on tax capacity. Yet there is also evidence that districts which are more urban and have better administrative infrastructure are more likely to adopt technology in the rst place. Moreover, there are other stated constraints on tax collections that may or may not be relaxed through technology use, such as political will to collect or legal constraints on enforcement. In this section we describe an experiment that we conducted with one large local government in urban Ghana. The goal of the experiment is to causally estimate the impacts of technology on tax outcomes and understand the mechanisms behind how technology affects taxation at different stages of the collection process.

3.1 Setting

We conducted the experiment in 2021 in La Nkwantanang Madina Municipal Assembly (henceforth, Madina). Madina is part of the Greater Accra region, and is more af uent and urban than the average district. We collaborated with the municipal government of Madina and a private technology rm called Melchia Investments which had developed a new technology aimed at increasing property tax revenues. The technology features the two components described in Section 2.3: a geospatial database of properties and a revenue software that integrates this database and assists in bill delivery and enforcement. The database of properties was created by combining high-resolution aerial photographs

with digital registry maps. [6] At the “last mile” of the taxation process, the technology consists of a tablet that assists collectors who work in the eld to deliver bills and collect payments. The tablet provides navigational assistance to help the collector go from an initial point to the location of a designated property (Figure A1). As we detail below, what we vary across treatment and control groups is the presence of the tablet.

During a scal year, the local government assigns collectors to designated geographical areas for approximately six weeks at a time (a 'campaign'). The designated areas are called 'collection units' and are de ned with geographical boundaries that create a cluster of physically adjacent properties (Figure A2). During each six-week campaign, collectors are responsible for delivering bills and collecting payments from assigned property taxpayers in their collection unit. After each campaign, the collector is assigned to a new collection unit. Each area of Madina is only covered once during a scal year, due to the large number of properties relative to the limited number of collectors. Property owners are legally required to pay within four weeks of receiving the tax bill. Pay stations do exist, but in practice virtually all payments are made directly to the collector.

Our experiment was speci cally embedded in the six-week campaign between March 15th and April 25 th in 2021. Before the campaign, collectors received training from both municipal of cers and employees of the technology rm. The main training sessions, common to all collectors, described the rules for property tax collection in Madina and the protocols to follow during interactions with property owners. In addition, the collectors assigned to the treatment group received training in how to use the handheld tablets. The compensation scheme (chosen by the municipal government and technology rm) was constant across treatment and control groups. Each collector received an 8 percent commission rate on taxes collected from their assigned bills. Collectors also received a daily transportation allowance and base salary.

3.2 Experimental Design

We trained 56 collectors and randomly assigned 28 to the treatment group and 28 to the control group. Of the 56 collectors, 39 had previously worked with the rm and 17 were hired shortly before the experiment. Of the 39 collectors with previous experience, 11 were designated as `high performing' by the private rm. Collectors worked individually in their assigned collection unit, where they were assigned to approximately 135 bills each. Each collector had a supervisor randomly assigned to them during the campaign. Supervisors were in charge of monitoring the revenue collectors and assist-

[6] Casaburi and Troiano (2016) study the impacts of an enforcement program in Italy which detected property tax evasion by overlaying aerial photographs and property registry data.

ing them with challenges in the eld. [7] All supervisors were randomly assigned to be in charge of both treatment and control collectors.

At the beginning of the experiment, all collectors in the treatment group were given the tablet for use during the six-week tax campaign. As described above, the database of properties was embedded into the tablet in order to help collectors navigate in the eld and locate properties. Other than the tablet, the treatment group was not provided with any additional advantages. At the beginning of the campaign both groups were provided with their set of physical bills (see Figure A1 for an example). Apart from the electronic map, the information provided to collectors was thus constant across groups.

Randomization and Balance Our randomization proceeded in two steps. First, we randomly assigned each collector to a collection unit. Second, we randomly assigned the collector-collection unit pair to the treatment or control group. We strati ed on the share of properties in the collection unit that were businesses (rather than residential). To avoid chance imbalances, we ran the full randomization 100 times and selected the run with the minimum t-statistic from balance checks on six variables (as in Banerjee, Chassang, Montero, and Snowberg, 2020). Two of these variables were speci c to collectors: a dummy for previous work experience with the rm, and a dummy for high-performance rating (by the rm). The other four variables were speci c to the collection unit: total bills to deliver; total taxes (current due and arrears); average current amount due per bill; and average previous pay status per bill (unpaid, partially paid, fully paid).

Table A3 summarizes a series of balance checks. In Panel A, we consider a set of characteristics at the tax bill level, based on administrative registry data. In Panel B we consider characteristics at the collector-unit level. None of the variables are statistically signi cantly different between groups at the 10 percent level or lower. For example, control and treatment collectors have statistically insigni cant differences in the number of bills assigned to them, the average bill amount and the fraction of bills that are residential (rather than for a business). The fraction of collectors with previous experience in Madina, or that received a high performance rating, also have statistically insigni cant differences across the two groups. In Panel C, we compare characteristics of households in the treatment and control areas, nding statistically insigni cant differences in indices of income, liquidity and taxpayer awareness. At the bottom of each panel, we report the F-test from the null hypothesis that the difference in characteristics across variables are

[7] One potential concern is that the supervisors provide more assistance to the control group, due to greater navigational challenges, or to the treatment group, in order to improve the perceived performance of the technology. However, we nd no differences by treatment status in the amount of supervisor support or supervisor monitoring reported by collectors (Appendix Table A2).

all zero. We fail to reject the null at the tax bill level ($F = 0.71, p = 0.66$); the collector-unit level ($F = 0.16, p = 0.95$), and the household level ($F = 1.07, p = 0.38$).

3.3 Experimental Data and Estimation

In this section, we describe the data sources that we make use of in our analyses and our estimation. We use administrative data at the property level, covering 7,560 residential and business properties, which contain information on owner names, property location, current tax due and arrears. This data set served to create the collection units for all collectors and to issue all the bills that were to be delivered during the tax campaign.

Our research team collected daily data from each collector on the number of bills delivered and the amount of revenue collected. These data allow us to study the activity of revenue collectors in the treatment and control groups at a high frequency. Non-compliance with reporting was low overall (and uncorrelated with treatment), though our small sample of 56 collectors raises concerns about the role of idiosyncratic measurement error at the daily level. For this reason, our main results winsorize the administrative outcomes at the 95th percentile, separately by group and day.

In addition to the daily data, enumerators working for the research team conducted three rounds of surveys with all 56 collectors – at the beginning, middle, and end of the tax campaign. The rst round was conducted during the initial week of the campaign; the mid-line during the third and fourth weeks; and, the end-line at the end of the sixth week. Topics covered include challenges in the eld, strategies used for bill delivery and collection, and self-assessed knowledge about households, among other topics.[8]

Finally, the enumerators administered end-line surveys with 4,353 randomly selected households in April and May of 2021. A random sample of equal size was drawn from each of the 56 collection units. Whenever an initially selected property could not be located or contacted, the enumerator would randomly pick an adjacent property within the same collection unit. The end-line survey covered household characteristics, interactions with and views of collectors, taxation, and beliefs about enforcement and governance.

Given the random treatment assignment, we use OLS to estimate the causal impacts of technology. The econometric speci cation varies slightly depending on the unit of observation. For outcomes that vary at the day and collector level, we estimate:

$$y_{cd} = b_d \; 1(\text{Tech})_c + q_d + W \; X_c + e_{cd}, \tag{1}$$

[8]Attrition in the collector surveys is 17 percent but is uncorrelated with treatment. For both the collector surveys and the daily collector data, the results in the paper are based on the full sample; in both cases, estimates based on the balanced panel data-sets are nearly identical (results available upon request).

where y_{cd} is the outcome for collector-collection unit c on day d, q_d are campaign-day xed effects, and X_c contains time-invariant controls. In the main analysis, X_c only includes strata xed effects for the share of businesses in total properties. In robustness checks, we include additional controls for previous work experience in Madina, a dummy for high quality collector rating, total number of bills to deliver, and the average tax due per bill. The dummy $1(Tech)_c$ takes a value of 1 for all collector-units randomly assigned to the technology treatment. The treatment coef cient, b_d, is indexed by day because we estimate dynamic treatment effects by interacting the treatment dummy with the individual campaign-day xed effects. In a robustness check, we leverage the panel-nature of the daily collector data and include xed effects for each collector-collection unit pair in equation (1). In this case, the identifying variation is the treatment effect that varies within a collector-unit over time, relative to the initial impact on day 1, b_1 (our chosen omitted category). Standard errors are clustered at the collector-unit level.

For outcomes at the household level, we estimate:

$$y_{hc} = b \quad 1(Tech)_c + W \quad X_{hc} + e_{hc}, \tag{2}$$

where h indexes households and c collector-units. Standard errors are clustered by collector-unit. X_{hc} always includes strata xed effects. In robustness checks, we also include the controls at the collector-unit level in (1), as well as previous pay status and property category at the household level. Previous pay status measures if the property tax bill in the past year was fully paid, partly paid or not paid at all.

3.4 Experimental Effects on Tax Outcomes

We begin by studying the impacts of technology on bill delivery and tax collection using the collector daily reports. In Figure 2, we show the impacts on bills delivered. Panel A shows the averages by group and day, while Panel B reports the daily treatment coef - cients b_d (equation 1). The treatment group delivers more bills than the control group. This difference initially builds up and peaks by the 24 th day, where treatment collectors have delivered 34 more bills than the control group (a 58 percent increase). The gap narrows in the second half of the campaign, where the stock of bills delivered in the treatment group steadies while control collectors continue to hand out bills. The con - dence interval around the treatment coef cients is meaningfully wide, likely owing to the limited sample size and number of clusters; notwithstanding, the effect is statistically signi cant at the 5 percent level in all campaign-days beyond the 10 th day. At the end of the campaign, the treatment collectors have delivered 21.5 more bills on average,

representing a 27 percent increase over the 80.7 bills in the control group.

In Figure 3, we nd that technology causes a large increase in total taxes collected. There are no differences in tax performance during the rst week, in which most collectors focus on bill delivery. However, from the second week onward, the treatment group collects at a higher rate; the treatment effect is statistically signi cant at the 5 percent level on all subsequent days and grows over time. At the end of the campaign, the treatment group has collected an additional 856 GHC on average, representing a 103 percent increase over the 829 GHC collected on average in the control group.

We can infer from these results that the treatment group collects more taxes per bill delivered. Appendix Figure A3 shows that this outcome grows over time; at the end of the campaign, the treatment group has collected 118 percent more taxes per bill delivered than the control group. This result implies that the tax collection impact is not only driven mechanically by the increase in bills delivered. The higher collection rate from each delivered bill motivates our investigation of mechanisms in Section 4.

Robustness In Figures A4 and A5, we explore the robustness of our experimental estimates from equation (1) for bills delivered and taxes collected. First, we nd that the estimates are similar when using non-winsorized outcomes. [9] Second, the results are also similar, but more precisely estimated, upon including additional covariates. Third, we include collector-unit xed effects, which implies that estimating b_d in equation (1) relies on the existence of a time-varying component in the treatment effect. In other words, b_d will re ect the treatment effect based on changes within collector-unit over time (relative to the initial impact b_1 on day 1). The presence of a dynamic treatment effect within collector is consistent with our mechanism evidence on learning in the eld over the course of the tax campaign (Section 4.3).[10] Tellingly, the estimated impacts at the end of the campaign are comparable with and without the collector-unit xed effects.

Another important concern is whether the context of the COVID-19 pandemic impacted the results. We conducted a pilot experiment in early 2019 in the same location, using the same technology and the same research protocol (though with a smaller sample of collectors). In that pilot we found similar effects as in the main experiment (Appendix Figure A7). Qualitatively, both the pilot and main experiment produce an effect on bill delivery that is larger in the middle-periods of the intervention rather than at the

[9] Winsorizing is partly motivated by the small sample size of collectors. Figure A6 shows that the results are almost identical across all sub-samples which leave out one collector at a time. This alleviates concerns that the average effects are unduly in uenced by outlier performances of any individual collector.

[10] The xed effect technically captures variation within each collector-collection unit. However, we interpret it as re ecting a treatment effect over time within collector, since we found no evidence suggesting there are time-varying effects within collection units unrelated to changes in collectors' behavior.

end. On the quantitative side, at the end of the interventions, the impact on bills delivered was 32 percent in the pilot versus 27 percent in the main experiment; the impact on taxes collected was 79 percent in the pilot versus 103 percent in the main experiment. This suggests that the results of the main experiment were not somehow an artifact of abnormal conditions during the pandemic.

Complementary evidence from household surveys Independent evidence about the effects of the technology on tax outcomes are available from our household surveys. Table 3 reports the treatment effects on key tax outcomes based on estimating equation (2). Households in the treatment group are more likely to report having been visited by a tax collector, get more total visits from tax collectors, and are more likely to have a bill delivered. The impacts on visit probability and total visits are statistically signi cant, whereas the impact on receiving a bill is positive but insigni cant. [11] One potential explanation is households make excuses for their lack of payment (just 16 percent of these households report actually making any tax payment). A second is that the lack of bill delivery in spite of successful visits re ects collusive bribes (Section 4.2). In terms of magnitudes, the impact of technology on bill delivery is smaller in the household surveys than in the daily collector reports, though we cannot reject the null hypothesis that the two effects are the same in percentage terms (p-value 0.30).

Households in treatment areas are more likely to report making a tax payment and report higher payments than in control areas. Treatment areas also exhibit higher reported payments conditional on bill delivery. While the magnitudes of the effects on collections also differ between the household surveys and the collector reports, we fail to reject the null hypothesis that the effects are similar across sources (p-value 0.27).[12]

4 Mechanisms Behind Experimental Effects on Taxes

The technology was designed largely with the goal of improving the bill delivery margin. In the experiment, the treatment effect on bill deliveries was 27 percent, consistent with this goal. If the technology had only worked through the delivery margin, then the

[11] In Appendix Table A4, we show that the household-level results are robust to the removal of all controls and to the inclusion of more extensive controls, speci cally the variables in Panel C of Table A3.

[12] We can also compare magnitudes by estimating equation (2) but using end-line administrative bill-level data. This bill-level data is unique at the household level and underpins the collector report data. The percentage impacts of technology on bill delivery and amount paid (both statistically signi cant at 5 percent) are moderately larger in the administrative data than in the household survey, but we cannot reject the null hypothesis of equal magnitudes across sources (p-values of 0.28 and 0.29, respectively). These additional results are available upon request.

experiment would have showed a similarly sized effect on revenue collections (as our model demonstrates in Section 5 below). Yet the experiment showed a treatment effect on collections of 103 percent, or four times the size of the treatment effect on deliveries. In other words the technology had a disproportionately large effect on the payment margin. But how did the technology allow collectors to improve so much on the payment margin relative to the delivery margin?

In this section, we investigate three potential mechanisms for why the technology had such a large impact on collections relative to bill deliveries. The rst is that the presence of technology improves citizens' tax morale or strengthens perceived tax enforcement capacity. The second is that the technology reduces the potential for leakages by collectors. The third is that the technology makes it easier for collectors to directly learn about and target the households that have higher propensity to pay. We argue that evidence points most strongly to the third channel.

4.1 Tax Morale and Perceived Enforcement Capabilities

The rst mechanism we consider is that technology improved households' tax morale or increased households' perceived enforcement capabilities of local government. Tax morale is broadly de ned as the non-pecuniary motivations for tax compliance (Luttmer and Singhal, 2014). For instance, the presence of technology may improve households' views that the government is making efforts to collect taxes in more ef cient and equitable ways, or to improve service delivery. Household perceptions of government enforcement capabilities may change if seeing a revenue collector with a new technology raises their expected pecuniary costs of non-compliance.

We use our household survey to create three indices for tax morale: government efforts to collect taxes in equitable and ef cient ways; satisfaction with government services; and government governance capacity and integrity. We also create an index for information-enforcement, which tracks households' perceptions of government informational capacity and enforcement strength. Each index is based on several individual questions, which are detailed in Data Appendix B.2.

In Table 4, we study the impact of technology on these indices, by estimating equation (2). We nd null effects on all outcomes. [13] In Table A5, we nd null effects on 12 of the 13 individual underlying questions used to build the indices. [14] For example, there are

[13] Prior studies have also found that stronger tax collection effectiveness increases tax payments and bribes but does not impact households' morale or beliefs (Khan et al., 2015; Balan et al., 2022).

[14] Out of the 13 individual outcomes, the only one that is statistically impacted shows a decrease in the perception that everyone pays their fair share of taxes. If anything, this effect should lower morale.

null effects on enforcement-related questions such as ”Next time the tax collectors come to collect, what percent of households do you think will pay their taxes?” and ”Imagine someone refuses to pay taxes – how likely do you think it is that the local government will pursue and enforce sanctions?” There are also null effects on questions about satisfaction with government, including ”In your opinion, what has been the overall quality of services offered by the local tax department of Madina?” and ”Overall, how would you rate the competency of the local government of Madina?” In Appendix Figure A8, we investigate the possibility that the average null effects mask heterogeneity along the asset-income distribution. For example, it is possible that the presence of technology stimulates tax morale but only amongst more well-off households that are more likely to have paid taxes in the past. We nd null effects across the income-asset distribution. [15]

4.2 Bribes

The second mechanism we consider is that technology may have improved the payment margin by reducing bribe activity. Bribes can come in the form of a “collusive bribe,” where the household and collector agree on a payment made to the collector in exchange for a cessation of follow-up visits. They can also take the form of a “coercive bribe,” in which the collector pockets tax payments made by the household in combination with a threat of retaliation against whistle-blowing. [16]

Technology can reduce these two types of bribe activities of collectors through better monitoring by supervisors, or easier reporting of bribe taking by households. Yet the effect of technology on bribes is ambiguous ex-ante. Technology may increase households' perception of collectors' enforcement capacity and raise collectors' bargaining power, which could increase bribe taking. Technology could also free up more time for the collectors to do all of their previous activities, including attempting to take bribes. [17]

To investigate how technology affects bribes, we estimate equation (2) with various outcome measures of bribe activity from the household survey (Data Appendix B.3). Importantly, due to the illegal and culturally sensitive nature of bribes, some of the measures came from indirect questions: for example, we ask if it is likely that collectors in the household's area are likely to ask for bribes. In Table 5, we nd positive and statistically

[15] These null effects do not necessarily imply that technology investments cannot increase households' tax morale or perceived enforcement capacity. Such views may be shaped in the longer run, while our experiment captures short run impacts. We discuss long run effects of technology in Section 6.

[16] Other studies have found that coercive and collusive forms of private capture often co-exist within the same setting (Djankov and Sequeira, 2014; Okunogbe and Pouliquen, 2022).

[17] Other technologies, such as electronic ling of tax returns, may reduce private capture by limiting the extent of in-person interactions between of cials and taxpayers (Okunogbe and Pouliquen, 2022).

signi cant effects of the treatment on bribes. While technology causes a meaningful increase in the likelihood of coercive or collusive bribes (column 1), the treatment effects on bribe amounts are smaller (columns 2-5). For example, technology causes a 1 percentage point increase in the amount of collusive bribes, expressed as a percent of the household's tax bill, and a 4 percentage point increase in the amount of coercive bribe, expressed as a percent of taxes collected. The treatment effect on collusive bribe amount in GHC (column 5) is approximately 4.5 times smaller than the treatment effect on tax amount paid (Table 3). Though caution is in order when interpreting the actual magnitudes, since these questions relate to a sensitive topic that is hard to measure directly.[18]

Overall, the fact that in all speci cations we nd positive, rather than negative, effects of the technology on bribe activity suggests that the technology's substantial impact on revenue collections does not work through a decrease in leakage by collectors. To the contrary: the experiment highlights how there is serious potential downside to technology in tax collection efforts that local governments should be aware of.

4.3 Learning and Differential Targeting

The third mechanism we consider is that collectors leverage the time savings from better navigation to directly build knowledge about households' propensity to make a tax payment. This greater knowledge about taxpayers then allows the collectors to better target their collection efforts to the households with higher propensity to pay. To our knowledge, this channel is new to the literature.

Three sets of observations motivate this mechanism. First, bill delivery in the eld is characterized by signi cant navigational challenges that arise from imperfect addressing (Section 2.2 and Figure A1). As a result, 71 percent of control collectors at the beginning of the campaign reported nding it challenging to locate their assigned taxpayers (Panel B of Figure 4). Self-reported time-use data suggests that the average control collector would require 10.4 weeks to deliver all the assigned bills (while the campaign only last 6 weeks).[19] The time constraint on collectors' ability to deliver bills, let alone collect payments (which often requires repeat-visits), therefore appears to be binding.

Second, household propensity to pay is an important determinant of tax payment. This point was raised by numerous local government of cials we interacted with. We use the household survey to proxy for willingness to pay with the household's awareness

[18]The results are robust to different speci cations (Table A4) and measures of bribe (Figure A13).

[19]At the beginning of the campaign, control collectors report that the average weekly time devoted to work in the eld is 19.5 hours, and the average time required to deliver a single bill is 1.5 hours. Thus, to deliver the average assigned 135 bills would in principle require 10.4 weeks: $(135 \quad 1.5)/19.5 = 10.4$.

of taxation, and for ability to pay with income and liquidity (Appendix B.4). We nd that the propensity to pay index (which combines the measures of awareness, income and liquidity) strongly predicts actual compliance outside of the experiment (Table A6).

Third, propensity to pay is heterogeneous across households and collectors have limited knowledge ex-ante about which households have higher propensity. Indeed, our survey data reveal that 71 percent of collectors at the beginning of the campaign report not having a good understanding of which households are more able and willing to pay (Panel A of Figure 5). Household propensity to pay is hard to know in this setting for several reasons. Ability to pay depends on income and liquidity, which are hard to observe and partly transitory. Moreover, propensity to pay is only weakly correlated with characteristics that are more easily observable to the collector, including the value of property taxes indicated on the tax bill. [20] Consistent with this, in Figure A9 we nd that variation across properties in the value of the tax bill accounts for approximately 1 percent of the variation in household income and the index of propensity to pay.

Collector Behavior and Strategy To investigate this mechanism, we start by examining technology's impact on challenges in the eld using the collector surveys. From the outset of the campaign, treatment collectors report less navigational challenges and less challenges in locating taxpayers (Figure 4). These gaps in reported challenges are statistically signi cant in all survey rounds, despite the small sample size. The gaps do decrease at the end of the campaign, suggesting that control collectors also improve their navigation over time. In Appendix Table A2, we nd no strong evidence that other challenges in the eld signi cantly differ between groups (e.g. resistance by property owners, wrong bill information, and lack of supervisor support). Due to improved navigation, treatment collectors spend 48 percent less time per bill delivered, but do not work less hours per week than the control group (Appendix Table A7).

What do treatment collectors spend their freed-up time on? We nd they make more return-visits to property owners (Table 3, column 2). By conducting more return-visits, and possibly by extending the length of each visit, the collectors likely interact with households and learn about their propensity to pay. Consistent with this interpretation, the collector surveys show that treatment collectors increase their knowledge about households' propensity to pay over time. Panel A of Figure 5 shows that at the begin-

[20] This is because the property tax in Madina is calculated on a presumptive schedule (Section 2). In the presumptive tax, coarse proxies for capital value (e.g. number of oors and area-size) are used to calculate taxes owed – thereby weakening the link between property tax value and income or wealth.

ning of the campaign, there were no differences in collectors' knowledge. [21] Over time, a positive knowledge gap opens up (signi cant at 5 percent), as treatment collectors gather information about households' propensity to pay while doing their return-visits. [22]

The collector surveys reveal that the treatment group uses this additional information to target those households with higher propensity to pay. Mirroring the result on knowledge in Panel A, Panel B of Figure 5 shows that there were no differences in collection strategies at the beginning of the campaign experiment, but, over time, treatment collectors increasingly make use of the strategy to visit areas on speci c days where property owners are more likely to be able to pay. Figure 6 shows that treatment collectors also increasingly make use of collection strategies that target property owners that are more willing to pay – by visiting households that have a stronger awareness of taxation (Panel A; p-value = 0.07) and that are more satis ed with public goods (Panel B; p-value = 0.08).

It is useful to divide collection strategies into two broad types: those that focus on hard-to-observe household characteristics, and those that focus on easy-to-observe characteristics. We de ne hard-to-observe characteristics as those that make up propensity to pay: ability to pay taxes (income and liquidity), awareness of taxation, and satisfaction with public goods. As argued above, these characteristics cannot be observed readily in the eld nor can they be inferred via any information directly available to the collector (including on the tax bill). We de ne easy-to-observe characteristics as those that can be directly observed from the tax bill (tax bill amount and amount of previous payment - see Panel A of Figure A1) or easily observed in the eld (greater proximity to the main road; government headquarters; collector's own house). Table 6 shows the treatment effects on the use of these two broad strategy types, by estimating equation (1) on the three rounds of collector surveys.[23] The table shows no differences in strategies at the beginning of the experiment. Over time, however, treatment collectors make disproportionately more use of the hard-to-observe strategies (signi cant at the 10 percent level, column 5), consistent with learning. In fact, the inclusion of collector-unit xed effects makes the disproportionate reliance on hard-to-observe strategies more pronounced (and signi cant at the 5 percent level, column 6). Since these xed effects isolate the part of the treatment impact

[21]Knowledge is a dummy variable which takes a value of 1 if the collector chooses the statement "I think I have a good understanding of which properties are more able and willing to pay" rather than the statement "I put a lot of effort to get my job done, but it remains unclear to me which exact properties are more likely or more willing to pay their property rates". See Data Appendix B.5 for more details.

[22]Part of the knowledge gap in the middle and nal rounds is also due to a decrease in reported knowledge amongst control collectors, which could re ect updated accuracy-beliefs while in the eld.

[23]For each characteristic, strategy use takes a value of 1 if the collector reports using 'all the time' or 'often' the collection strategy which focuses on this household characteristic, and 0 otherwise (Data Appendix B.5). Use of the hard-to-observe and easy-to-observe collection strategies are constructed as the average use over all characteristic strategies in each set.

that varies within collector over time (relative to the initial impact at the beginning of the experiment), this result is strongly consistent with learning and differential targeting over the course of the experimental period.

Which Household Types Get Targeted with the Technology? As a second piece of evidence related to this mechanism, we investigate how targeted households (those that make payment) differ from non-targeted households within a collection unit and, importantly, how technology causes this to differ between treatment and control areas.

The mechanism predicts positive selection under technology on proxies for propensity to pay. We estimate selection using the following speci cation on the household survey

$$y_{hc} = q \quad 1(Pay)_h + b \quad [1(Pay)_h \quad 1(Tech)_c] + W \quad X_h + m_c + e_{hc} \tag{3}$$

y_{hc} is a xed household/property characteristic and $1(Pay)_h$ is a dummy for making any positive tax payment. Since $1(Pay)_h$ is endogenous, q indicates whether there is a statistical (non-identi ed) difference in a xed characteristic between targeted (paying) and non-targeted (non-paying) households in the control group. The treatment coef - cient b shows how the difference in characteristic between targeted and non-targeted households causally changes in treatment versus control areas; any non-zero b would indicate differential selection. We can include collection area xed effects (m_c) since we focus on differences in characteristics between households within collection areas.

We focus on the three xed household characteristics of propensity to pay: income, liquidity, and taxation awareness. As argued above, these are hard-to-observe characteristics which local of cials identi ed as determinants of compliance. [24] We also consider characteristics that are more easily observable: tax bill value, previous tax payment, and observable assets. The rst two are directly observable on the tax bill (see the bill example in Figure A1). The third characteristic is derived from the household survey and measures assets that can more readily be observed outside the property (e.g. car, truck, electric generator). Targeting households with these observable characteristics may be useful, in particular for collectors with less knowledge about propensity to pay.

The results in Figure 7 show the level of selection in the control group (q) and the

[24] The construction of these variables is described in detail in Data Appendix B.4. Even though these proxies are based on end-line household surveys, we think they are plausibly not impacted by the treatment. It is unlikely that technology-induced payment of taxes affects households' earnings choices within the six-week span of the tax campaign. The questions on liquidity refer to a 'typical' month rather than the speci c past month during the campaign. Finally, no property owner from the areas of the experiment was neither summoned to court nor had their property con scated during the tax campaign.

treatment group (q + b); differential selection (b) can visually be inferred as the difference in levels. In Panel A, we focus on the hard-to-observe characteristics. We nd that targeted households in the treatment group have higher liquidity, income, and taxation awareness than non-targeted households (all signi cant at 5%) In contrast, targeted and non-targeted households in the control group are precisely estimated to have no differences in liquidity and income; targeted households have slightly more awareness than non-targeted, but the difference is not signi cant. As a result, the hard-to-observe propensity index (which combines awareness, income and liquidity) reveals strongly positive selection in the treatment group and almost null selection in the control group. [25] Panel B studies selection on more easily observable characteristics. There is little targeting on current tax amount due in both groups – suggesting it may not be a useful predictor of payment. We observe positive selection both on previous tax payment and on assets but of similar magnitude in treatment and control groups.

The strong differential selection between treatment and control on hard-to-observe characteristics (Panel A) thus contrasts with the absence of differential selection on easy-to-observe characteristics (Panel B). This contrasting selection result across hard versus easy to observe characteristics in Figure 7 mirrors the result on disproportionate use of hard-to-observe collection strategies by treatment collectors in Table 6 – though Figure 7 is based on household surveys while Table 6 is based on collector surveys.

Taken together, the results in this sub-section paint a picture consistent with the proposed mechanism: technology reduces navigational challenges and frees up scarce time; collectors use this extra time to learn about the hard-to-observe household characteristics of propensity to pay and subsequently target those with higher propensity. [26]

Finally, additional results suggest treatment collectors may also have learned about households' amenability to bribes. In Figure A11, we nd that treatment households exposed to bribes have higher taxpayer awareness than those not exposed; speci cally, they are more likely to have witnessed or heard about court actions and property con-

[25] An important concern is that these payment patterns re ect heterogeneous effects of technology on tax morale or perceived enforcement. However, Table A8 nds no differential impacts of technology on tax morale and perceived enforcement outcomes by income, liquidity or taxpayer awareness. Moreover, Figure A10 shows that the selection patterns on bill delivery are similar to the selection patterns on tax payment. Bill delivery may be a dimension of targeting if collectors learn during their very rst encounter and selectively choose who to deliver to (Balan et al., 2022). Both of these results suggest that heterogeneous payment effects are unlikely to confound the targeting interpretation of the selection patterns in Figure 7.

[26] Our results are not consistent with a strategy where propensity to pay is in fact observable and collectors focus initially on those with highest propensity and then 'move down the curve'. Since the treatment collectors deliver more bills (Figure 2), this strategy would generate negative selection on proxies for propensity to pay such as income and liquidity (while we nd positive selection in Figure 7). Moreover, this strategy would generate decreasing collection per bill delivered over time (while we nd it increases).

scations by of cials. This increased perception of effective enforcement may raise the credibility of collectors' threat of retribution in a setting of collusive or coercive bribes.

5 Model

The experimental results uncovered larger effects of technology on collections than on bill deliveries. We argued in the previous section that the data best support a mechanism that combines time-savings due to navigational improvement with the fact that collectors use the extra time to learn about taxpayer types. In this section, we formalize this mechanism by embedding time-savings and learning in a dynamic Beckerian time use model of collectors. We use the model to assess how much of the larger treatment effect on collections is plausibly due to learning, relative to the initial navigational advantage. [27]

5.1 Environment

The experiment lasts W weeks. Collectors are endowed with one unit of time each week and split time between delivering bills and trying to collect revenues. In week one each collector is endowed with a large number of bills. Each bill has a face value of one local currency unit. Neither the number of bills or variation in face value across bills play an important role in our experimental results, so we abstract of these in the model.

Collectors come in two types: treatment (T) and control (C). The two types differ exogenously in the number of bills they can distribute in a given amount of time. In one unit of time a treatment collector can distribute q_T bills, and a control collector can distribute q_C bills. We assume that q_T q_C, which captures the greater ef ciency in delivering bills in the treatment group. In relation to our empirical results, the assumption that q_T q_C is motivated by the technology's reduction in navigational challenges (Figure 4) and reduction in time spent to deliver a given bill (Table A7).

Households also come in two types: “high” and “low,” referring to their probability of paying a bill. For simplicity we assume that high types have a positive probability of payment while low types have probability zeroof making a payment after each visit. The household type is not known to the collectors until after they deliver a bill to that household; that is, collectors learn about household types only after bill delivery.

[27] A setting where collectors use the time-savings induced by technology to indiscriminately make return-visits at every property they delivered bills to could generate the larger treatment effect on collections than deliveries in Section 3. However, this indiscriminate time-use would not generate the results on collector knowledge and strategies in Section 4. The full set of observed results therefore motivate our choice to model both time-savings and learning and to investigate the quantitative importance of learning.

The treatment collectors have an advantage in learning about whether households are high types or not. For each bill delivered, treatment collectors have a probability h_T of discovering that the household is a high type, and a probability $1 - h_T$ of learning that the household is a low type. Control collectors have probabilities h_C and $1 - h_C$ of learning that the household is high type and low type. We assume that $h_T > h_C$, which captures the better opportunities for learning afforded in the treatment group.[28] In relation to our empirical results, the assumption that $h_T > h_C$ re ects the improved knowledge gathered by treatment collectors about households' propensity to pay (Figure 5) and their increased use of collection strategies that focus on hard-to-observe household characteristics related to propensity to pay (Figure 5 and Figure 6).

The collection technology is exactly the same for treatment and control collectors. Each week, collectors devote time, c, to collecting from households that have been delivered bills. We assume that collectors attempt to collect only from households they know to be high-type households, which makes sense given that low-types have a probability zero of paying. We assume that each unit of time has diminishing marginal value in collecting from each household in a given week. This could be because repeat visits in the same week reduce the household's willingness to comply, or because the constraints keeping the household from paying earlier in a week are still binding later in the week (e.g. the household is waiting for a paycheck).

We model diminishing returns to collection activity at the weekly level as follows. Spending c units of time per bill trying to collect from h bills yields the following probability of collection per bill: $l c^m$, where $l > 0$ and $0 < m < 1$. Hence, the total collections in the week are the following: $l c^m h$. As a simple example, suppose the collector has identi ed measure $h = 5$ high-type bills. If they spend $c = 1$ on the full measure 5 of bills then they can collect from each of those bills with probability l. They then have measure $5(1 - l)$ bills left over in the next period.

The collector's choice variables each week are time spent distributing bills, b, and time spent on each bill trying to collect, c. Note that a collector would never spend a different amount of time on different bills because of the concavity of the collection probability in time spent trying to collect. Since $m < 1$, the highest returns are for the rst units of time spent trying to collect from each bill. So optimality implies that all bills should get equal time devoted to collection efforts.

The Collector's Dynamic Problem The goal of a collector is to maximize tax rev-

[28] In a richer spatial model one could endogenize the increased learning that comes from the navigational advantage and time-savings of the technology. We abstract from spatial considerations, and other micro-founded channels of learning, for simplicity.

enues. Collectors have the following state variables each week: h, the number of bills that have been delivered to a household identi ed as a high type, and w, the week of the experiment. The collectors' choice variables for each week are b, the time spent trying to deliver bills, and c, the time spent trying to collect from each high-type bill. The time constraint for the collector is that $b + ch = 1$ in each week.

The dynamic trade-off for a collector is that spending more time trying to collect from households this week means spending less time delivering bills that can be collected from next week. Intuitively, this will mean that in the later weeks, collection is a larger priority, while in the earlier weeks, delivering bills is more important.

State variables evolve each week according to the collectors' time allocation choices. After spending b units of time delivering bills, the stock of high types learned about increases by $q_j h_j b$ for collector type $j \in \{C, T\}$. After spending c units of time per bill trying to collect, a fraction λc^m get collected from, leaving $h(1 - \lambda c^m)$ remaining high-type bills for the next week. So the law of motion for known high types with a bill delivered becomes $h^0 = q_j h_j b + h(1 - \lambda c^m)$.

Let $V(h, w)$ be the present discounted value of having h high-type bills delivered by day w. The collector's dynamic problem is therefore:

$$V(h, w) = \max_{\{b,c\}} \left\{ \lambda c^m h + E\left[V(h^0, w^0) \right] \right\} \tag{4}$$

where $w^0 = w + 1$, subject to the time constraint, $b + ch = 1$, and the law of motion for high-type bills discovered, $h^0 = q_j h_j b + h(1 - \lambda c^m)$.

In the last week, collectors spend the maximum time trying to collect. This means that $b = 0$ and c is the maximum amount of time spent on each bill that uses up the collector's full time endowment. If the collector has measure h high-type bills, then the collector can spend $c = 1/h$ units of time collecting from each bill. The result is $\lambda (1/h)^m h$ revenues collected, meaning that $V(h, W) = \lambda (1/h)^m h$.

The remaining periods can be solved by backwards iteration. The optimal dynamic program is to allocate time so that the marginal bene t of trying to collect this week equals the marginal bene t of delivering more bills to collect from next week. If a collector spends too much time collecting now, the bene t of collecting this week will be very low at the margin relative to the value of having more bills ready to collect from next week. If the collector spends too little time collecting this week, the marginal bene t of collecting now will be very high compared to the value of having more bills delivered.

5.2 Quantitative Predictions of Model

We now calibrate the model and use it to simulate the effects of counterfactual changes to the environment. We begin by setting the number of weeks to be $W = 6$ as in the experiment. We then set $q_C = 1$, which is a normalization on bill delivery ef ciency in the control group. We set $h_C = 0.1$ meaning that control collectors have a 10 percent change of nding a high type after delivering a bill. We set the parameters of the collection function to be $l = m = 0.5$. The former controls the average level of collections, and the latter controls the degree of curvature in collection efforts. Appendix Table A9 presents results for alternative parameterizations showing that our results are not particularly sensitive to these parameter choices.

The most important two parameters in our calibration are q_T and h_T, which govern the treatment group's advantage in delivering bills and nding high-type households. Our strategy is to choose values of these two parameters to match two moments from our experiment: (i) the treatment effect on bill deliveries, and (ii) the treatment effect on collectors' focus on strategies that target hard-to-observe household characteristics. The former is 27 percent, as shown in Figure 2. The latter is 22 percent, as shown in column (6) of Panel A of Table 6, corresponding to the disproportionate reliance on collector strategies targeting hard-to-observe household characteristics relative to strategies targeting easy-to-observe characteristics. Matching these two moments requires setting $q_T = 1.38$ and $h_T = 0.13$.

Figure 8 plots the calibrated model's predictions for bills delivered, taxes collected, the stock of high-type bills delivered but not collected from, and the fraction of time spent on collections in each group. The model does well in reproducing the concave time pattern of bill deliveries (top left panel), and by construction gets the 27 percent treatment effect on deliveries correct. The model also gets the convex pattern of collections, with the largest treatment effects coming at the end of the experiment period, as in the data. The model's treatment effect on collections is 51 percent (top right panel), or about twice the treatment effect on deliveries.

To get more intuition about how the model works, the bottom panels of Figure 8 plot the stock of high-type bills delivered (but not collected from) and the time spent on collections each week. In the bottom left panel one can see that the treatment group accumulates a larger stock of bills delivered to known high-types, with a difference that peaks during during the middle and later weeks of the experiment. Equipped with the large stock of high-type bills delivered, the return from allocating time to collection increases relative to the return from allocating time to bill delivery. As a result, in the bottom right panel, one can see that the treatment group spends more time on

collections than the control group, particularly in weeks four and five of the experiment. These model predictions are broadly consistent with the empirical observations that the treatment group is better aware of which households are willing and able to pay during the experiment (Figure 5, Panel A), and more focused on collection strategies targeting these households (Figures 5 and 6), particularly in the later parts of the experiment.

As a frame of reference, we compute a counterfactual simulation of the model in which we shut down the learning and differential targeting channel. In particular we set $h_T = h_C$, and leave all other parameters as in the benchmark calibration. Note this leaves in place the treatment group's advantage in delivering bills, but gives them no additional advantage in learning whether households are high types or not.

Figure 9 plots the model's predictions in this counterfactual simulation. Now the treatment effect on bill deliveries is modestly larger than before, at 34 percent. The treatment effect on collections is substantially smaller than before, at 26 percent. The bottom panels help illustrate why. The treatment group's stock of high-type bills is still larger than the control group in this counterfactual, but basically only because of the advantage in delivering bills to begin with. Relative to the setting with the additional learning advantage, the slower rate of discovery of high-type bills in the middle weeks of the campaign leads to a smaller increase in the relative benefit of allocating time to collection versus delivery. As a result, the treatment group spends only slightly more time each period on collection than the control group. The result is a treatment effect on collections that is similar in magnitude to the treatment effect on bills delivered.

In summary, when the model is calibrated to match the treatment effect on targeting hard-to-observe household characteristics that signal a high payment probability, the model predicts a treatment effect on collections that is about twice the treatment effect on deliveries. In a counterfactual simulation in which the treatment group has no such learning advantage, the treatment effect on collections and deliveries are roughly equal in magnitude. Since the treatment effect on collections was around four times larger than on deliveries in the experiment (Section 3.4), the model's learning channel seems to explain around half the gap in the magnitudes of the two treatment effects. More generally, we conclude that the learning channel is roughly equally important as the navigational direct effect of higher bill deliveries on revenue collections. [29]

[29] In Appendix C, we conduct a regression-based analysis which is complementary to the model analysis. We use data to proxy for the navigational advantage by measuring days since bill delivery at the property level. We find that after controlling for days since delivery in our equation for tax outcomes (equation 2), which can be interpreted as 'shutting down' the navigational advantage in the model, there remains a large treatment effect of technology on tax outcomes. Though based on a different research design, the regression results yield a conclusion that is consistent with the model results and that suggests learning is an important channel in determining the total treatment effect on collections.

6 Distributional Implications of Technology and Broader Relevance

One important focus of our study thus far is how technology leads to differential targeting of households that are more likely to make payment. One aspect of this targeting is that households with higher income are more likely under technology to receive a bill and make tax payments, while households with more assets are equally likely to be targeted with and without technology (Figure 7). These targeting results have implications for the distributional impacts of technology.

To see this, we create four quartiles of the joint income-asset distribution. This distribution is an aggregate index which captures information on both the household's income and assets (Data Appendix B.4). We estimate distributional effects in the household survey by allowing the treatment to vary across the income-asset quartiles (q):

$$y_{hqc} = b_q \ 1(\text{Tech})_c \ 1[\text{Quartile} = q] + W \ X_{hc} + e_{hc} \tag{5}$$

The results from estimating equation (5) on the likelihood of making a positive tax payment are presented in Panel A of Figure 10. Due to the targeting of higher income, we nd that technology strongly improves the equity of the local tax system: treated households in the top two quartiles are 9.8 to 10.1 percentage points more likely to pay than control households, representing a 65 percent increase. In contrast, technology has a somewhat precise null effect on tax payments in the bottom income-asset quartile. In Appendix Figure A12, we show that technology raises taxes paid as a percent of taxes due in the top quartiles. Since technology does not plausibly impact household income, this result implies that technology improves the progressivity of the tax system (it makes taxes paid, as a share of household income, more positively correlated with income).

By estimating equation (5) for bribes, we nd that technology has a positive impact on bribe likelihood which is concentrated among households with lower income-assets. This result is presented in Panel B of Figure 10 which, by visual comparison with Panel A, highlights that technology's impacts on tax and bribe payments affect different segments of households. These results parallel the conclusions of Khan et al. (2015) and Gauthier and Goyette (2014) that tax and bribe payments are substitutes. In Figure A13, we show that the percent increase in bribe amount is concentrated in the bottom quartiles – implying that technology makes the bribe system more regressive. [30]

Thus, technology has starkly contrasting incidence impacts on taxes and bribes: tax payments become more progressive and concentrated amongst well-off households,

[30] Previous work has also found that 'street-level' bribes affect poorer households to a larger extent (Fried et al., 2010; Khan et al., 2015; Peiffer and Rose, 2018).

while bribes become more regressive and concentrated amongst less well-off households.

Broader Relevance The experiment in our study focuses on the effects of technology in one large municipal government in Ghana over a six week horizon. How might the results extend to other settings and longer time horizons?

Our results show that technology positively impacted tax collection by alleviating constraints that arose largely from having incomplete property addressing. Technology likely has smaller effects in settings with developed addressing systems. The United Nations estimates that approximately 4 billion people live in places without physical addresses, and most of these are in the developing world. This gives reason to think that technology will have an impact on property tax collections in other developing nations more broadly. The impacts of technology will likely be smaller in settings where other policies help alleviate the challenges associated with collecting property taxes, such as hiring more collectors or extending the duration of each tax campaign. Yet each of these alternatives are expensive, suggesting that technology may still be valuable from a policy perspective, to the extent that it can help economize on manpower costs.

Finally, this technology permitted the gathering of soft information on taxpayers' propensity to pay, which is most relevant in settings with limited coverage of hard information (such as third-party reports on income) and enforcement. This technology may be less relevant at higher levels of development, where governments are characterized by broad third-party coverage and enforcement (Kleven et al., 2016; Jensen, 2022).

One important question left unanswered is whether technology, such as the one in this study, would have a larger or smaller effect on tax collections at greater time horizons. If the knowledge gleaned about speci c taxpayers' propensity to pay is permanent, then the returns to learning induced by technology would diminish over time. To the extent that knowledge about taxpayers is transitory, though, such as for liquidity or income shocks, then the longer-run effects of technology may also be large. The longer-run effects would likely be lower if governments implement other policies over time that alleviate collectors' time constraints or expand third-party coverage. At the same time, while we found no short-run impacts on households' tax morale or enforcement beliefs, the sustained use of technology may positively affect these views over the longer run and increase compliance.[31] In our cross-sectional census data, we found that local govern-

[31]The increase in bribes under technology, if persistent in the longer-run, may however counter these positive tax morale impacts. Re ecting this ambiguity, 88 percent of treatment households report a preference for the technology-based collection system over the manual system but treated households also report a stronger and statistically signi cant dis-interest in engaging with the state – concentrated amongst those that were subject to increased bribes (results available upon request)

ments which have had technology in place for an average of 8.75 years were associated with a 78 percent increase in tax collection (Section 2 and Table 2). Compared to the short-run causal tax increase of 103 percent we estimate over one tax campaign, these numbers suggest somewhat smaller returns to technology at greater time horizons. [32]

7 Conclusions

This paper has studied the role of technology in improving tax capacity. We focus on local governments in Ghana, which are in charge of property taxes but in practice collect very little. Both cross-sectional census data and experimental evidence point to a close connection between use of technology and positive outcomes at every stage of the tax collection process. In our experiment, we found that technology had a direct impact on deliveries, as intended. But then it allowed collectors to directly gather information on taxpayers' propensity to pay and, as a result, change their collection strategies to focus on those most likely to pay. This novel channel highlights the important ways in which government of cials change their behavior in the presence of a new technology, and how this behavioral change is a key component of the large treatment effect on tax collections.

Our ndings contribute to the literature on informational capacity and state 'legibility' – the breadth and depth of government's knowledge about its citizens and their activities (Scott, 1998; Lee and Zhang, 2017). Prior work shows that pre-existing information can improve collection, including from employers, nancial institutions and local chiefs. We relate to these studies by showing how governments can directly build 'soft' information on taxpayers' propensity to pay through feasible policy investments. Directly building 'soft' information is most likely to strengthen state legibility in developing countries where 'hard' information, including from third-parties, is constrained.

Use of technology for local taxation is limited but growing in Africa and other areas of the world (Knebelmann, 2022). More work is needed to understand the impacts and barriers to tax collection technologies in diverse settings and over the longer run. More work is also required to rigorously establish the extent to which technology in practice leads to changes in public expenditures, since one of the ultimate goals of technology investments in taxation is to fund improved provisions of useful public goods.

[32] Relative to the tablet's running costs, the technology in our experiment was cost-effective and delivered a 96 percent increase in taxes collected net of cost (Appendix Table A10). However, cost as a percent of taxes collected did not markedly decrease in the treatment group relative to control. By contrast, the cross-sectional regressions show a large reduction in cost associated with technology adoption over nearly a decade (Panel D of Table 2) – possibly suggesting that ef ciency gains materialize over the longer run.

References

Almunia, M., J. Hjort, J. Knebelmann, and L. Tian (2022): "Strategic or Confused Firms? Evidence from "Missing" Transactions in Uganda," Review of Economics and Statistics, forthcoming.

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

Banerjee, A. V., S. Chassang, S. Montero, and E. Snowberg (2020): "A Theory of Experimenters: Robustness, Randomization, and Balance," American Economic Review, 110, 1206–30.

Becker, G. and C. Mulligan (2003): "Deadweight Costs and the Size of Government," The Journal of Law and Economics, 46.

Bergeron, A., P. Bessone, J. Kabeya, G. Tourek, and J. Weigel (2021): "Optimal Assignment of Bureaucrats: Evidence from Randomly Assigned Tax Collectors in the DRC," Unpublished Working Paper.

Besley, T., E. Ilzetzki, and T. Persson (2013): "Weak States and Steady States: The Dynamics of Fiscal Capacity," American Economic Journal: Macroeconomics, 5, 203–235.

Besley, T. and T. Persson (2009): "The Origins of State Capacity: Property Rights, Taxation, and Politics," American Economic Review, 99, 1218–1244.

Best, M., F. Gerard, E. Kresch, and J. Naritomi (2020): "Greener on the Other Side? Spatial Discontinuities in Property Tax Rates and their Effects on Tax Morale," Unpublished Working Paper.

Brennan, G. and J. Buchanan (1980): The Power to Tax, Cambridge University Press.

Brockmeyer, A., A. Estefan, J. C. Suárez Serrato, and K. Ramírez (2020): "Taxing Property in Developing Countries: Theory and Evidence from Mexico," Working Paper.

Brockmeyer, A. and M. S. Somarriba (2022): "Electronic Payment Technology and Tax Compliance: Evidence from Uruguay's Financial Inclusion Reform," World Bank Policy Research Working Paper.

Callen, M., S. Gulzar, S. A. Hasanain, M. Y. Khan, and A. Rezaee (2020): "Data and Policy Decisions: Experimental Evidence from Pakistan," NBER Working Paper Series.

Carillo, P., D. Pomeranz, and M. Singhal (2017): "Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement," American Economic Journal: Applied Economics, 9, 144–64.

Carpenter, D. (2020): The Forging of Bureaucratic Autonomy, Princeton University Press.

Casaburi, L. and U. Troiano (2016): "Ghost-House Busters: The Electoral Response to

a Large Anti-Tax Evasion Program," Quarterly Journal of Economics, 131, 273–314.

Cui, W. (2022): The Administrative Foundations of the Chinese Fiscal State, Cambridge University Press.

Dal Bo, E., F. Finan, N. Li, and L. Schechter (2021): "Information Technology and Government Decentralization: Experimental Evidence from Paraguay," Econometrica, 89, 677–701.

Daunton, M. (2001): Trusting Leviathan: The Politics of Taxation in Britain 1799-1914, Cambridge University Press.

Dincecco, M. and G. Katz (2016): "State Capacity and Long-run Economic Performance," The Economic Journal, 126, 189–218.

Djankov, S. and S. Sequeira (2014): "Corruption and Firm Behavior: Evidence from African Ports," Journal of International Economics, 94, 277–94.

Dray, S., C. Landais, and S. Stantcheva (2022): "Wealth and Property Taxation in the United States," Working Paper.

Eissa, N. and A. Zeitlin (2014): "Using Mobile Technologies to Increase VAT Compliance in Rwanda," Unpublished Working Paper.

Fan, H., Y. Liu, N. Qian, and J. Wen (2021): "Computerizing VAT Invoices in China," NBER Working Paper Series.

Fish, P. and W. Prichard (2017): "Strengthening IT Systems for Property Tax Reform," African Property Tax Initiative Brief, 1–8.

Fried, B., P. Lagunes, and A. V. and (2010): "Corruption and Inequality at the Crossroad," Latin American Research Review, 45, 76–97.

Gauthier, B. and J. Goyette (2014): "Taxation and corruption: Theory and firm-level evidence from Uganda," Applied Economics, 45, 2755–2765.

Gordon, R. and W. Li (2009): "Tax Structures in Developing Countries: Many Puzzles and a Possible Explanation," Journal of Public Economics, 93, 855–866.

Government of Ghana (2014): "Internally Generated Revenue Strategy and Guidelines: Maximizing Internally Generated Revenue Potentials for Improved Local Level Service Delivery," Ghana Ministry of Finance.

Jensen, A. (2022): "Employment Structure and the Rise of the Modern Tax System," American Economic Review, 112, 213–234.

Kau, J. and P. Rubin (1981): "The Size of of Government," Public Choice, 37.

Khan, A., A. Khwaja, and B. Olken (2019): "Making Moves Matter: Experimental Evidence on Incentivizing Bureaucrats through Performance-Based Postings," American Economic Review, 109, 237–270.

Khan, A. Q., A. I. Khwaja, and B. A. Olken (2015): "Tax Farming Redux: Experimental

Evidence on Performance Pay for Tax Collectors," Quarterly Journal of Economics, 131, 219–271.

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, 651–692.

Kleven, H. J., C. T. Kreiner, and E. Saez (2016): "Why can Modern Governments Tax so Much? An Agency Model of Firms as Fiscal Intermediaries," Economica, 83, 219–246.

Knebelmann, J. (2022): "Digitalisation of property taxation in developing countries: Recent advances and remaining challenges," ODI Report Series.

Lee, M. and N. Zhang (2017): "Legibility and the Informational Foundations of State Capacity," Journal of Politics, 79, 118–32.

Londoño-Vélez, J. and J. Ávila-Mahecha (2021): "Enforcing Wealth Taxes in the Developing World: Quasi-Experimental Evidence from Colombia," American Economic Review: Insights, 3, 131–148.

Luttmer, E. F. P. and M. Singhal (2014): "Tax Morale," Journal of Economic Perspectives, 28, 149–68.

Mayshar, J., O. Moav, and L. Pascali (2022): "The Origin of the State: Land Productivity or Appropriability?" Journal of Political Economy, 130.

Muralidharan, K., P. Niehaus, and S. Sukhtankar (2016): "Building State Capacity: Evidence from Biometric Smartcards in India," American Economic Review, 106.

Naritomi, J. (2019): "Consumers as Tax Auditors," American Economic Review, 109.

Okunogbe, O. and V. Pouliquen (2022): "Technology, Taxation and Corruption: Evidence from the Introduction of Electronic Tax Filing," American Economic Journal: Economic Policy, Forthcoming.

Okunogbe, O. and F. Santoro (2021): "The Promise and Limitations of Information Technology for Tax Mobilization," World Bank Policy Research Working Paper Series.

Peiffer, C. and R. Rose (2018): "Why Are the Poor More Vulnerable to Bribery in Africa? The Institutional Effects of Services," Journal of Development Studies, 54, 18–29.

Pomeranz, D. (2015): "No Taxation without Information: Deterrence and Self-enforcement in the Value Added Tax," American Economic Review, 105, 2539–2569.

Scott, J. (1998): Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed, Yale University Press.

Vannutelli, S. (2022): "From Lapdogs to Watchdogs: Random Auditor Assignment and Municipal Fiscal Performance," Unpublished Working Paper.

Waseem, M. (2022): "Overclaimed Refunds, Undeclared Sales, and Invoice Mills: Nature and Extent of Noncompliance in a Value-Added Tax," Unpublished Working Paper.

Figure 1: Constraints on Tax Collection and Bill Delivery

(a) Perceived Importance of Different Constraints on Tax Collection

(b) Most Important Perceived Constraint on Bill Delivery

Notes These panels show the perceived constraints on tax collection and bill delivery as reported by local government of cials and politicians. In Panel A, the bars show the percent of all respondents that consider a particular constraint to be 'most important', on a ve-choice scale from 'least important' to 'most important'. In Panel B, the bars show the percent of all respondents who consider a particular constraint to be the most important constraint (mutually exclusive choices). Responses are pooled across all respondents in all 216 local governments.

Figure 2: Impact of Technology on Bills Delivered

(a) Bills Delivered per Collector By Group

(b) Treatment Effect

Notes These panels show the impact of technology on the number of property tax bills delivered. Panel A shows the average number of bills delivered by group and by day of the intervention. Panel B displays the treatment effect coef cients on technology, separately by day, based on estimating equation (1). The analysis is based on the daily collector data, described in Section 3.3.

Figure 3: Impact of Technology on Taxes Collected

(a) Taxes Collected per Collector by Group

(b) Treatment Effect

Notes These panels show the impact of technology on the amount of property taxes collected. Panel A shows the average total amount of taxes collected by group (treatment, control) and by day of the intervention. Panel B displays the treatment effect coef cients on technology, separately by day, based on estimating equation (1). The analysis is based on the daily collector data, described in Section 3.3.

Figure 4: Impact of Technology on Search Challenges in the Field

(a) Challenging to Navigate in the Field?

(b) Challenging to Locate Taxpayers?

Notes In Panel A, the outcome is a dummy variable which takes a value of 1 if the collector nds it challenging or very challenging to navigate in the eld (and 0 otherwise). In Panel B, the outcome is a dummy variable which takes a value of 1 if the collector nds it challenging or very challenging to locate assigned taxpayers (and 0 otherwise). The grey bar measures the difference in outcome between treatment and control; the number in parentheses is the randomization inference-based p-value on the statistical signi cance of the difference. For a description of the challenge measures, see Section 4.3 and Data Appendix B.5. The analysis is based on the collector surveys, described in Section 3.3.

Figure 5: Collector Knowledge of and Focus on Households that are Able to Pay

(a) Knowledgeable about Which Households are Able and Willing to Pay?

(b) Focus on Households that are Able to Pay?

Notes In Panel A, the outcome takes a value of 1 if the collector reports having a good understanding of which properties are more able and willing to pay (and 0 otherwise). In Panel B, the outcome is a dummy variable which takes a value of 1 if the collector uses all the time or often the collection strategy to focus on properties on speci c days where property owners are more likely to be able to pay (and 0 otherwise). The grey bar measures the difference in outcome between the treatment and control groups; the number in parentheses is the randomization inference-based p-value on the statistical signi cance of the difference. For a detailed description of the knowledge and strategy measures, see Data Appendix B.5. The analysis is based on the collector surveys, described in Section 3.3.

Figure 6: Collector Focus on Those Aware of Tax Duties and Satis ed with Public Goods

(a) Focus on Households that are Aware of Tax Payment Duty?

(b) Focus on Households that are Satis ed with Public Goods?

Notes In Panel A, the outcome takes a value of 1 if the collector uses all the time or often the collection strategy to focus on properties where the collector knows the owners are aware of their duty to pay property taxes (and 0 otherwise). In Panel B, the outcome takes a value of 1 if the collector uses all the time or often the collection strategy to focus on properties where the collector knows the owners are more likely to pay because of satisfaction with delivery of public services (and 0 otherwise). The grey bar measures the difference in reliance on these strategies between the treatment and control groups; the number in parentheses is the randomization inference-based p-value on the statistical signi cance of the difference. For a detailed description of the strategy measures, see Data Appendix B.5. The analysis is based on the collector surveys, described in Section 3.3.

Figure 7: Characteristics of Households that Made a Tax Payment by Treatment Status

(a) Hard-to-Observe Characteristics

(b) Easy-to-Observe Characteristics

Notes These gures show targeting of property owner characteristics for tax payment, based on estimating equation (3). Characteristics vary by row and the bottom row of each graph is an index which is the unweighted average of the characteristics. The characteristics in Panel A are harder to observe, while the characteristics in Panel B are easier to observe (see Section 4.3 for further explanation). Data Appendix B.4 provides details on the characteristics and the indices.

Figure 8: Predictions of Benchmark Model

Notes: This gure shows the predictions of the benchmark model for the control and treatment groups. The upper panels plot the model's predictions for bills delivered and revenues collected over the course of the experiment. The bottom panels plot the stock of high-type bills that have been delivered (but not collected from) and the fraction of time spent on collections (rather than bill delivery).

Figure 9: Predictions of Model with No Learning Advantage from Technology

Notes: This gure shows the predictions of the model when the learning probability is set to be the same in the treatment and control groups. The upper panels plot the model's predictions for bills delivered and revenues collected over the course of the experiment. The bottom panels plot the stock of high-type bills delivered (but not collected from) and the fraction of time spent on collections (rather than bill delivery).

Figure 10: Distributional Effects of Technology on Taxes and Bribes

(a) Treatment Effect on Tax Payment

(b) Treatment Effect on Bribe

Notes These panels show the impact of technology on the likelihood of making any positive tax payment (Panel A) and on the likelihood of bribes (Panel B). The bribe dummy takes a value of 1 if the household estimates that the tax collector will ask for any positive amount of unof cial payment during interactions with property owners (collusive bribe); or, if the household estimates that the collector will keep for themselves any positive amount of money collected from property owners (coercive bribe). The bribe variable takes a value of 0 only if the estimated amounts of collusive bribe and coercive bribe are both equal be 0. Both panels display the treatment effect coef cient on technology, separately by quartile of the income-asset distribution, based on estimating equation (5). The income-asset distribution is calculated as the unweighted average, by household, of the income index and the assets index. For more detail on the index measures and the bribe and tax payment measures, see Data Appendix B.3-B.4.

Table 1: Characteristics of Local Tax Capacity in Ghana

	Mean	Median
Panel A: Tax outcomes		
Taxes collected per capita (GHC)	4.2	2.6
Share of bills delivered (%)	43.0	43.3
Taxes collected per bill delivered (GHC)	11.5	6.7
Share of bills that are paid (%)	30.2	29.3
Panel B: Information and technology		
Share of properties with address (%)	26.7	20.0
1(Common not to locate property/owner)	0.74	1
1(Technology: database or software)	0.17	0
Panel C: Other capacity dimensions		
Share of properties with valuation (%)	17.1	0
Share of tax payments made in cash (%)	72.1	76.6
Cost of collection (% taxes collected)	64.1	47.5
1(Take tax defaulters to court)	0.22	0
Number of local governments	216	216

Notes All variables are calculated at the district level (N=216), using unweighted averages. In Panel A, the variables measure tax outcomes: total local taxes collected per capita; the share of bills that are actually delivered to property owners; total taxes collected per bill delivered; the share of bills delivered which are paid. In Panel B, the variables relate to information constraints and use of technology: the share of properties which have a physical address; a dummy variable which equals 1 if it is common for a collector not to be able to locate an assigned property or their owner (and 0 otherwise); a dummy equal to 1 if the local government has an electronic database of properties or a tax revenue management software (and 0 otherwise). In Panel C, the variables relate to additional constraints on tax capacity: the share of properties with of cial valuation; the share of tax payments that are made in cash to the collector; the cost of collecting taxes, as a share of total taxes collected; the likelihood that a local government takes tax defaulters to court. For a detailed description of all variables, see Data Appendix B.1.

Table 2: Associations with Technology at the Local Government Level

	(1)	(2)	(3)	(4)	(5)
Panel A: Taxes collected per Capita					
1(Technology)	6.32***	3.71***	4.06***	3.08**	3.24***
	(1.62)	(0.88)	(0.60)	(1.16)	(0.94)
Mean outcome variable	4.15	4.15	4.15	4.15	4.15
Panel B: Share of Bills Delivered (%)					
1(Technology)	0.26***	0.08**	0.18***	0.14***	0.09**
	(0.06)	(0.03)	(0.04)	(0.03)	(0.03)
Mean outcome variable	0.43	0.43	0.43	0.43	0.43
Panel C: Taxes per Bill Delivered					
1(Technology)	6.9*	5.1**	4.8**	3.1***	4.2***
	(3.6)	(2.0)	(1.8)	(0.9)	(1.1)
Mean outcome variable	11.5	11.5	11.5	11.5	11.5
Panel D: Cost of collection (% taxes collected)					
1(Technology)	-12.09**	-11.14**	-8.74*	-8.83*	-8.54*
	(4.67)	(4.22)	(4.22)	(4.42)	(4.35)
Mean outcome variable	64.1	64.1	64.1	64.1	64.1
District controls		x			x
Share neighbors with tech			x		x
Region FE				x	x
Observations	216	216	216	216	216
Clusters	10	10	10	10	10

Notes The regression model is a cross-sectional regression of all 216 districts in Ghana, with one local government per district. The variable 1(Technology) is a dummy variable taking a value of 1 if the local government either has an electronic database of properties or a revenue management software that assists with bill printing, payment recording, and follow-up enforcement. Across panels, the outcome is: local taxes collected per capita (Panel A); the share of bills that are delivered (Panel B); local taxes collected per bill delivered (Panel C); the cost of collecting taxes as a share of total taxes collected (Panel D). Across columns, the speci cations are: no controls in column (1); district controls (log per capita income, log population, urban share of population, share of properties with valuations, share of properties on of cial addresses, legal capacity, of cials' years of work experience) in column (2); the share of each district's geographically adjacent neighbor governments with technology in column (3); region xed effects in column (4); all three sets of controls in column (5). Standard errors are clustered at the region level. See Data Appendix B.1 for more detail on the variables.

Table 3: Impacts of Technology on Visits, Compliance and Revenues

	Any visit by tax collector (1)	Total visits (in %) (2)	Bill delivered by tax collector (3)	Any positive tax payment (4)	Total payment amount (in GHC) (5)	Payment amount per bill delivered (6)
Technology	0.087	0.094	0.054	0.043	25.9	47.3
	(0.033)	(0.050)	(0.036)	(0.021)	(10.9)	(19.6)
Household controls	X	X	X	X	X	X
Collector-unit controls	X	X	X	X	X	X
Strata FE	X	X	X	X	X	X
Mean in CG	0.55	0.67	0.51	0.16	41.0	80.9
Observations	4334	4334	4334	4334	4334	2276
Clusters	56	56	56	56	56	56

Notes This table presents the impacts of technology on main tax outcomes of interest. All coef cients are based on estimating equation (2), and using the household sample (Section 3.3). Across columns, the outcome is: a dummy for any visit received by a tax collector; the total number of visits (expressed in %, using the inverse hyperbolic sine transformation); a dummy for bill delivered; a dummy for any tax payment made; total tax amount paid (in GHC); total tax paid, restricted to households that received a bill. For a description of household controls and collector-unit controls, please refer to Section 3. The robustness of these results to the removal of control variables, or the inclusion of more extensive controls, is presented in Appendix Table A4.

Table 4: Impacts of Technology on Citizen Beliefs and Tax Morale

	Satisfaction with government services (1)	Integrity of government (2)	Tax equity & ef ciency efforts by government (3)	Enforcement information capacity of government (4)
Technology	-0.00771	0.0629	-0.0143	-0.0536
	(0.0701)	(0.0728)	(0.0604)	(0.0572)
Household controls	X	X	X	X
Collector-unit controls	X	X	X	X
Strata FE	X	X	X	X
Mean in CG	0.05	-0.04	-0.03	0.00
Observations	4324	4326	4326	4326
Clusters	56	56	56	56

Notes This table presents the impacts of technology on beliefs and tax morale. All coef cients are based on estimating equation (2), and using the household sample (Section 3.3). The outcome in each column is a (standardized) index variable, which averages over multiple (standardized) household survey questions. For the description of each underlying question that is used in each index, please see Data Appendix B.2. Across columns, the outcome index measures: the extent of satisfaction with government's delivery of services (column 1); the perceived integrity and competency of the local government (column 2); the local government's efforts to collect taxes in an equitable and ef cient manner (column 3); the perceived enforcement capacity of the government and the informational knowledge that the government possesses about its citizens (column 4). For a description of household controls and collector-unit controls, please refer to Section 3.

Table 5: Impacts of Technology on Bribe Activity

	Any bribe (coercive or collusive) (1)	Total bribe amount (in %) (2)	Collusive bribe amount (% of tax due) (3)	Coercive bribe amount (% of payment) (4)	Collusive bribe amount (in GHC) (5)
Technology	0.12	0.03	0.01	0.04	6.16**
	(0.04)	(0.01)	(0.00)	(0.02)	(3.07)
Household controls	X	X	X	X	X
Collector-unit controls	X	X	X	X	X
Strata FE	X	X	X	X	X
Mean in CG	0.14	0.11	0.02	0.19	11.6
Observations	4334	4334	4334	4331	4334
Clusters	56	56	56	56	56

Notes This table presents the impacts of technology on measures of bribe activity by collectors. All coef cients are based on estimating equation (2), and using the household survey (Section 3.3). In column 1, the outcome is a dummy which takes a value of 1 if the household estimates that the tax collector will ask for any positive amount of unof cial payment during visits to property owners (collusive bribe); or, if the household estimates that the collector will keep for themselves any positive amount of money collected from property owners (coercive bribe). The bribe dummy takes a value of 0 only if the estimated amounts of collusive bribe and coercive bribe are both equal be 0. In column 2, the outcome is the total bribe amount, in percent. At the household level, it is the average of the coercive bribe amount, expressed as a percent of a hypothetical 1000 GHC collected by the tax collector, and the collusive bribe amount, expressed as a percent of the household's true tax liability. In column 3, the outcome is the collusive bribe amount, expressed as a percent of the household's true tax liability. In column 4, the outcome is the coercive bribe amount, expressed as a percent of a hypothetical 1000 GHC collected by the tax collector. In column 5, the outcome is the collusive bribe amount in GHC that the household estimates will be asked by the tax of cial as unof cial payment during visits to households. The bribe variables are described in detail in Data Appendix B.3. For a description of household controls and collector-unit controls, please refer to Section 3. The robustness of these results to the removal of control variables, or the inclusion of more extensive controls, is presented in Appendix Table A4.

Table 6: Impact of Technology on Collector Strategies in the Field

		Focus on collections, hard-to-observe household characteristics		Focus on collections, easy-to-observe household characteristics		Difference in strategies: Hard versus easy to observe	
		(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Average effect							
Technology		0.141**	0.357**	0.087*	0.133	0.053	0.223**
		(0.066)	(0.136)	(0.046)	(0.092)	(0.039)	(0.102)
		[0.032]	[0.000]	[0.076]	[0.089]	[0.190]	[0.001]
Panel B: Dynamic effect							
Technology	Round 1	-0.082	–	0.007	–	-0.089	–
		(0.103)	–	(0.068)	–	(0.077)	–
Technology	Round 2	0.264**	0.364**	0.123	0.137	0.143**	0.227**
		(0.125)	(0.153)	(0.091)	(0.102)	(0.058)	(0.095)
Technology	Round 3	0.253**	0.349**	0.138	0.130	0.117*	0.219**
		(0.117)	(0.151)	(0.088)	(0.106)	(0.061)	(0.093)
Collector-unit controls		X		X		X	
Survey round FE		X	X	X	X	X	X
Collector-unit FE			X		X		X
Mean in CG		0.280	0.280	0.239	0.239	0.041	0.041
Observations		141	141	141	141	141	141

Notes This table presents the impacts of technology on collector strategies. In columns (1)-(2), the outcome is the likelihood that a collector makes uses all the time or often of collection strategies which focus on hard-to-observe household characteristics (liquidity, income, taxpayer awareness, satisfaction with public goods). In columns (3)-(4), the outcome is the likelihood that a collector makes use often or all the time of collection strategies which focus on more easily observable household characteristics (value of tax bill, past tax payment, geographical location). In column (5)-(6), the outcome is the difference between the reliance on hard-to-observe versus easy-to-observe strategies. For a detailed description of these collector strategies, see Data Appendix B.5. All regressions use the panel of three collector survey rounds (at the beginning, middle and end of the campaign), and include survey round xed effects. Odd columns include collector-unit controls. Even columns include collector-unit xed effects, in which case the omitted treatment category is round 1. Panel A reports the average effect of technology, while Panel B reports the round-by-round treatment effect (based on interacting round xed effects with the technology variable). Standard errors clustered at the collector-unit are reported in parentheses. In Panel A, the randomization inference based p-value is reported in brackets.

Appendix
Technology and Tax Capacity: Evidence from Ghana

A Additional Figures and Tables

Table A1: Associations with Technology Adoption

	1(Technology exists)	
	(1)	(2)
Total population	0.114***	0.084**
	(0.025)	(0.027)
Income per capita	0.073**	0.009
	(0.027)	(0.021)
Urban share of population	0.096***	0.048**
	(0.024)	(0.020)
Share of properties with address	0.112*	0.088*
	(0.053)	(0.044)
Share of properties with valuation	0.174***	0.130***
	(0.029)	(0.026)
Legal capacity to enforce taxes	0.084*	0.053**
	(0.042)	(0.022)
Tax-delinquents taken to court	-.001	-.001
	(0.016)	(0.011)
Of cials' years of work experience	0.058**	0.058*
	(0.021)	(0.030)
Of cials' years of education	0.014	-0.007
	(0.025)	(0.016)
Trust in of cials	0.004	-0.002
	(0.013)	(0.013)
Citizen tax awareness	-0.030	-0.026
	(0.015)	(0.026)
Citizen compliance attitude	-0.006	0.013
	(0.016)	(0.010)
Region FE		X
Observations	216	216
Clusters	10	10

Notes Each cell represents the b coef cient from a separate cross-district regression, based on the model $1(\text{Technology})_{dr} = b\ X_d + m_r + e_{dr}$ where $1(\text{Technology})_{dr}$ is a dummy equal to 1 if the local government in district d in region r uses technology for tax collection (see Section 2). X_d is the district characteristic which varies between rows; across columns, region xed effects (m_r) are included. All district characteristics are standardized, for ease of comparison across rows. Standard errors are clustered at the regional level. For a description of all the district characteristics, see Data Appendix B.1.

Table A2: Challenges Reported in the Field by Collectors

	Unable to locate properties owners (1)	Wrong information printed on bills (2)	Resistance from property to accept bill (3)	Supervisors don't monitor activities in the eld (4)	Supervisors unavailable for support if needed (5)
Technology	-1.040	-0.266	-0.100	-0.154	0.181
	(0.130)	(0.145)	(0.124)	(0.164)	(0.153)
	[0.00]	[0.07]	[0.43]	[0.35]	[0.26]
Collector-unit controls	X	X	X	X	X
Survey round FE	X	X	X	X	X
Mean in CG	0.51	0.13	0.04	0.07	-0.09
Observations	136	135	135	139	140

Notes This table presents the impacts of technology on the extent of challenges encountered by collectors in the eld. All regressions pool the collector survey responses across the survey rounds, and include survey round xed effects. All regressions also include the collector-unit controls described in Section 3. The outcomes measure the extent to which collectors agree (on a scale from 1 to 5) that a particular challenge characterizes their weekly work in the eld: inability to locate property owners; wrong information (address or tax amount due) printed on bills; resistance from property owners (mistrust of collector or refusal to pay); absence of monitoring by supervisors; unavailability of supervisors to support in the eld. The outcomes are standardized based on the set of underlying questions that characterize each challenge. Robust standard errors are reported in parentheses; the randomization inference-based p-value is reported in brackets. For a detailed description of the outcomes, see Data Appendix B.5

Table A3: Randomization Balance

	N	Control mean	Treatment coef cient
	(1)	(2)	(3)
Panel A: Tax bill characteristics			
Current tax amount	7560	322.8	-9.0
			(16.4)
Total tax amount	7560	692.5	-5.5
			(29.1)
Previous pay status	7560	1.2	0.0
			(0.0)
Previous tax payment	7560	59.7	-6.6
			(9.4)
Residential	7560	0.5	0.0
			(0.0)
Property quality	7560	0.5	0.0
			(0.1)
F-test joint signi cance [F, p]			[0.7,0.66]
Panel B: Collector-unit characteristics			
Experience in Madina	56	0.7	-0.1
			(0.1)
Performance rating	56	0.2	-0.1
			(0.1)
Total bills to deliver	56	135.2	1.7
			(4.7)
Average amount per tax bill	56	322.6	-7.4
			(16.5)
F-test joint signi cance [F, p]			[0.2,0.95]
Panel C: Household characteristics			
Income index	4353	-0.014	0.003
			(0.106)
Liquidity index	4353	0.051	-0.177
			(0.119)
Taxpayer awareness index	4353	0.011	-0.01
			(0.039)
F-test joint signi cance [F, p]			[1.07,0.38]

Notes This table presents balance checks of the randomization assignment for characteristics at the bill level (Panel A), the collector-unit level (Panel B), and the household level (Panel C). Details on the variables used in this table are provided in Section 3.3. The treatment coef cient in column (3) is the coef cient on technology in a cross-sectional regression with strata xed effects. Standard errors are clustered at the collection unit level. At the bottom of each panel, the F-test on the joint signi cance of all characteristics is reported along with the p-value.

Table A4: Robustness Checks for Technology Impacts on Tax and Bribe outcomes

	Any visit by tax collector (1)	Total visits (%) (2)	Bill delivered (3)	Any positive tax payment (4)	Total payment amount (in GHC) (5)	Any bribe (coercive or collusive) (6)	Total bribe amount (in %) (7)	Coercive bribe amount (in %) (8)	Collusive bribe amount (in %) (9)
Panel A: No Controls									
Technology	0.082	0.085	0.049	0.039	24.93	0.120	0.027	0.012	0.042
	(0.034)	(0.04)	(0.036)	(0.017)	(10.89)	(0.038)	(0.012)	(0.005)	(0.023)
Panel B: Extensive Controls									
Technology	0.086	0.087	0.055	0.047	27.21	0.113	0.025	0.014	0.036
	(0.032)	(0.049)	(0.034)	(0.020)	(11.18)	(0.037)	(0.010)	(0.005)	(0.019)
Strata FE	X	X	X	X	X	X	X	X	X
Mean in CG	0.55	0.67	0.51	0.16	40.95	0.14	0.11	0.02	0.19
Observations	4353	4353	4353	4353	4353	4353	4353	4353	4350
Clusters	56	56	56	56	56	56	56	56	56

Notes This table presents technology impacts on the same set of outcomes as in Table 3 and Table 5. The estimation model is the same, except that: in Panel A, all household and collector controls are removed; in Panel B, additional controls are added. The additional controls in Panel B are the set of 6 xed characteristics used in the targeting analysis – see Section 4.3 and Figure 7. For a description of the bribe variables, see Data Appendix B.3.

Table A5: Beliefs about Enforcement and Tax Morale

	Technology coef cient ($\hat{b}$) (1)	Mean in CG (2)	N (3)
Panel A: Enforcement and Information			
Share of HHs that comply with taxes	0.80	60.32	4330
	(2.38)		
Likelihood non-complier will end up paying	-0.07	3.08	4330
	(0.07)		
Likelihood Gov't has info about my tax status	-0.13	2.95	4330
	(0.13)		
Likelihood Gov't has info about my job	0.03	2.52	4330
	(0.09)		
Panel B: Gov't Efforts to Improve Tax Collection			
Agree efforts to collect taxes ef ciently	0.01	3.58	4330
	(0.07)		
Agree efforts to ensure fair share paid	-0.18***	3.42	4330
	(0.07)		
Agree efforts to collect for useful purposes	0.08	3.04	4330
	(0.11)		
Panel C: Government Capacity and Competency			
% of taxes wastefully spent	-3.48	55.81	4330
	(4.64)		
Agree Gov't has capacity to improve roads	0.04	3.94	4330
	(0.11)		
Overall Gov't competency rating	0.07	2.41	4330
	(0.07)		
Panel D: Satisfaction with Gov't Services			
Quality of tax collector services	-0.003	2.31	4330
	(0.05)		
Quality of tax authority services	-0.02	2.31	4330
	(0.05)		
Quality of overall Gov't services	-0.01	2.20	4330
	(0.05)		

Notes This table presents technology impacts on beliefs and tax morale. Each row presents the technology treatment coef cient (in column 1) from estimating equation (2) on different outcomes (which are described to the left). Standard errors are clustered at the collection-unit. Column (2) presents the mean of the outcome variable in control areas, while column (3) shows the sample size. For a description of all the outcomes, see Data Appendix B.2. All regressions include household and collector controls (Section 3).

Table A6: Proxy for Propensity to Pay is a Predictor of Tax Payment

	(1)	(2)	(3)	(4)
Panel A: Outcome is tax payment status				
Propensity to pay index	0.056**			
	(0.022)			
Income index		0.026**		
		(0.012)		
Taxpayer awareness index			0.016	
			(0.020)	
Liquidity index				0.029**
				(0.012)
Panel B: Outcome is total taxes paid				
Propensity to pay index	20.64***			
	(7.000)			
Income index		14.73***		
		(4.774)		
Taxpayer awareness index			1.980	
			(5.849)	
Liquidity index				5.360
				(3.718)
Outcome mean: Panel A	1.255	1.255	1.255	1.255
Outcome mean: Panel B	60.99	60.99	60.99	60.99
Control tax liability	X	X	X	X
Control block FEs	X	X	X	X
Observations	4353	4353	4353	4353
Clusters	56	56	56	56

Notes This table shows that the characteristics of propensity to pay are robust predictors of tax payment outside of the experimental setting. In Panel A, the outcome is the tax payment status in the year prior to the experiment, which can take a value of 1 (=no payment), 2 (=partial payment), 3 (=full payment). In Panel B, the outcome is the total amount of taxes paid in GHC, in the year prior to the experiment. Across columns, the outcome is regressed on different explanatory variables: the propensity index in column 1; the income index in column 2; the taxpayer awareness index in column 3; the liquidity index in column 4. The propensity index is the household-level unweighted average of the income, liquidity and awareness indices. For a discussion of the propensity to pay index, please see Section 4.3. For a detailed description of all the indices, please see Data Appendix B.4. All regressions include the level of property tax liability due as a control as well as block xed effects (approximately 7-8 properties per block). Standard errors are clustered at the collection-unit level.

Table A7: Impact of Technology on Collector Performance Measures

	# of unsuccessful visits per successful visit	Total hours worked per week	Average # of hours spent to deliver one bill	Satisfaction & happiness on the job
	(1)	(2)	(3)	(4)
Technology	-1.222	-2.382	-0.798***	0.117
	(1.024)	(1.412)	(0.207)	(0.150)
	[0.23]	[0.11]	[0.00]	[0.42]
Collector-unit controls	X	X	X	X
Survey round FE	X	X	X	X
Mean in CG	7.67	18.84	1.66	-0.07
Observations	141	141	111	139

Notes This table presents impacts of technology on collector performance measures. All regressions pool the collector survey responses across the survey rounds, and include survey round xed effects. All regressions also include the collector-unit controls described in Section 3. Across columns, the outcome is: number of unsuccessful visits per property for every successful visit; hours worked per week; hours worked per bill delivered; and, satisfaction and happiness in the job. The last outcome is a standardized variable, based on responses to three underlying questions: how productive the week was for the collector; how content the collector is while working; and, how satis ed the collector is with their job. Robust standard errors are reported in parenthesis; the randomization inference-based p-value is reported in brackets. For a detailed description of the outcomes, see Data Appendix B.5.

Table A8: Heterogeneity in Beliefs about Enforcement and Tax Morale

	Technology coef cient (b)	Heterogeneity coef cient (b H)
Outcome: Enforcement and Information Capacity Index		
Heterogeneity H: Liquidity index	-0.050	-0.016
	(0.056)	(0.053)
Heterogeneity H: Income index	-0.051	0.002
	(0.057)	(0.042)
Heterogeneity H: Taxpayer awareness index	-0.050	-0.021
	(0.056)	(0.057)
F-test joint signi cance of interaction terms [F, p]	[0.09, 0.96]	
Outcome: Efforts to Improve Tax Collection Index		
Heterogeneity H: Liquidity index	-0.016	0.048
	(0.059)	(0.055)
Heterogeneity H: Income index	-0.010	0.059
	(0.060)	(0.039)
Heterogeneity H: Taxpayer awareness index	-0.012	0.068
	(0.061)	(0.069)
F-test joint signi cance of interaction terms [F, p]	[1.27, 0.29]	
Outcome: Satisfaction with Gov't Services Index		
Heterogeneity H: Liquidity index	-0.018	-0.042
	(0.069)	(0.059)
Heterogeneity H: Income index	-0.009	0.011
	(0.069)	(0.032)
Heterogeneity H: Taxpayer awareness index	-0.007	0.041
	(0.070)	(0.064)
F-test joint signi cance of interaction terms [F, p]	[0.45,0.72]	

Notes This table presents heterogeneous technology impacts on beliefs and tax morale. Each row presents the technology treatment coef cient and the interaction coef cient, from estimating equation 2 augmented with the interaction between technology and the heterogeneity dimension H. Rows differ in the interaction (liquidity, income or taxpayer awareness), and panels differ in the outcome. The F-test at the bottom of each panel tests the joint signi cance of the three interaction coef cients for a given outcome. The outcomes are the same as in Table 4 (also described in Data Appendix B.2). The heterogeneity dimensions are the indices used in Panel A of Figure 7 (also described in Data Appendix B.4).

Table A9: Model Sensitivity Analysis

	Treatment Effect on Collections		
	Calibrated Model	No Learning	Ratio
Benchmark calibration	51	26	1.9
Higher base delivery ef ciency (q_C=1.5)	50	26	1.9
Higher frequency of high type (h_C=0.2)	52	27	1.9
Lower curvature in collections (m=0.4)	52	25	2.0
Lower ef ciency in collections (l =0.4)	51	26	1.9

Notes: This table reports the treatment effects on collections in the calibrated model and in the counterfactual equilibrium of the model when there is no collector learning ($h_T = h_C$). In the top row, the benchmark calibration sets $q_C = 1$, $h_C = 0.1$, $m = 0.5$, $l = 0.5$, sets $q_T = 1.38$ in order to match the 27 percent treatment effect on bills delivered (Section 3.4), and sets $h_T = 0.13$ to match the 22 percent increase in the disproportionate reliance on collector strategies targeting hard-to-observe versus easy-to-observe household characteristics (Section 4.3). In each subsequent row, the calibrated model matches the treatment effect on bill deliveries of 27 percent, and keeps all other parameters other than the change indicated as in the benchmark calibration.

Table A10: Cost-Bene t Analysis per Collector

			Control	Treatment
Cost items	Unit price	Total units		
Daily allowance	10	42	420	420
Commission rate	8%	42	66.3	134.8
Tablet rental	10	42	0	420
Network connection	40	1	0	40
Total cost			486.3	1014.8
Totals				
Total taxes collected			829	1685
Total taxes net of cost			342.7	670.2

Notes This table presents a cost-bene t analysis for the running costs of the average collector in the treatment and control groups during the 42 days of the tax experiment campaign. Some cost items are common to all collectors. Each collector receives 10 GHC in daily allowance. Moreover, each collector in both groups receives an 8% commission for taxes collected – which corresponds to 66.3 GHC (0.08 828 = 66.3) for the average collector in the control group and 134.8 GHC (0.08 1685 = 134.8) for the average collector in the treatment group (based on Figure 3). Some cost items are speci c to the treatment group. In particular, the private rm pays a 10 GHC daily rental price to the tablet provider; moreover, the tablet requires network connection. The top panel reports the total costs for the average collector over the 6-week campaign. In the bottom panel are reported the average taxes collected at the end of the campaign, as well as the taxes collected net of total cost. It is important to note that the cost items in the treatment group refer to the running cost of using the tablet on a daily basis – they do not account for any cost of building the geo-spatial database which serves as the input to the tablet (see Section 3).