NBER WORKING PAPER SERIES

TECHNOLOGY AND TAX CAPACITY: EVIDENCE FROM LOCAL GOVERNMENTS IN GHANA

James Dzansi
Anders Jensen
David Lagakos
Henry Telli

Working Paper 29923
http://www.nber.org/papers/w29923

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2022, Revised February 2025

For helpful comments on this project we thank Shan Aman-Rana, David Atkin, Augustin Bergeron, Anne Brockmeyer, Daniel Chachu, Ray Fisman, Andrew Foster, Rema Hanna, Gordon Hanson, Asim Khwaja, Justine Knebelmann, Gabriel Kreindler, Craig McIntosh, Karthik Muralidharan, Nicholas Li, Oyebola Okunogbe, Ben Olken, Tanner Regan, Dani Rodrik, Nii Sowa, Chris Udry, Silvia Vannutelli, Mazhar Waseem, Jonathan Weigel, Jaya Wen as well as seminar participants at Brown, BU, Dartmouth, George Washington, Harvard, Minneapolis Fed, NYU Abu Dhabi, Peking HSBC Business School, Syracuse, 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. The pre-analysis plan for this project was registered under AEA RCT Registry ID: AEARCTR-0007267. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Technology and Tax Capacity: Evidence from Local Governments in Ghana
James Dzansi, Anders Jensen, David Lagakos, and Henry Telli
NBER Working Paper No. 29923
April 2022, Revised February 2025
JEL No. H2, H71, O12, O33

ABSTRACT

This paper studies the role of technology in improving tax capacity in the developing world, focusing on local property taxation in Ghana. We randomize the use of a new technology designed to help revenue collectors locate property owners to deliver tax bills. We find that the technology increases bill deliveries by 27 percent and, surprisingly, increases tax collections by 103 percent. To reconcile these experimental findings, we build and estimate a dynamic time-use model in which revenue collectors respond to the new technology by shifting their allocation of time toward learning about households' propensity to pay and subsequently collecting from those with the highest payment propensity. The model's predictions are consistent with experimental evidence on collector time allocations, knowledge, and collection strategies. Our theory highlights how technology designed to solve one problem can help overcome other challenges once behavioral changes by users of the technology are taken into account.

James Dzansi
International Growth Centre
c/o Institute of Statistical, Social
& Economic Research
University of Ghana
Legon, Accra
Ghana
james.dzansi@theigc.org

Anders Jensen
Harvard Kennedy School
79 JFK Street
Cambridge, MA 02138
and NBER
anders_jensen@hks.harvard.edu

David Lagakos
Department of Economics
Boston University
270 Bay State Road
Boston, MA 02215
and NBER
lagakos@bu.edu

Henry Telli
International Growth Centre
c/o Institute of Statistical, Social
& Economic Research
University of Ghana
Legon, Accra
Ghana
henry.telli@theigc.org

1 Introduction

Developing countries generally struggle to collect taxes and provide 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 (see e.g. Gennaioli and Voth, 2015; 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 so as to finance productivity-enhancing public goods (Besley, Ilzetzki, and Persson, 2013; D'Arcy, Nistotskaya, and Olsson, 2024). Our paper studies the role of technology investments to improve tax capacity of local governments in Ghana. We focus on property taxes, a potentially significant but under-collected source of revenue in developing countries (Brockmeyer, Estefan, Arras, and Suárez Serrato, 2023).

We begin by conducting Ghana's first census of local governments, recording detailed measures of tax collection practices and outcomes in all 216 districts. The census highlights a key challenge at the beginning of the collection process, namely the limited ability to locate properties to deliver bills. We show that this challenge stems from incomplete addressing: in the average district, only 27 percent of properties are assigned a number on a named street. The location information available to local tax collectors is therefore imprecise, and most collectors around the country report struggling to find taxpayers. As a result, less than half of the printed tax bills in the average district end up being delivered. To help improve the delivery of property tax bills, a small number of local governments have invested in digital tax registries with exact geographic coordinates of properties. We document that investment in this geographic information system (GIS) technology is associated with more bills delivered and also more taxes collected per bill delivered. This raises the possibility that the GIS-technology may have helped improve collection in ways beyond locating properties to deliver bills.

Motivated by this association in the census, we collaborated with one local government to randomize the use of a new GIS technology and causally investigate how it affects the entire collection process. The experiment randomized the presence of an electronic tablet with GIS capabilities at the level of a tax collector. Both treatment and control collectors were given a stack of bills of similar average value in a randomly assigned work area and were tasked with collecting as much revenue as possible during the tax campaign. Each treatment collector was given a tablet with geo-spatial data that was supposed to increase delivery by improving navigation and taxpayer localization. All other aspects of the bill, collector, and work area were constant across groups.

We find that collectors using the technology delivered 27 percent more bills than control collectors by the end of the campaign. Treatment collectors reported substantially fewer challenges in navigating in their area and locating property owners and spent significantly less time per bill delivered, confirming technology's advantage to navigate and deliver bills in an environment with scant addressing. More surprisingly, treatment collectors collected 103 percent more taxes than control collectors, which implies a treatment effect on collections that is almost four times as large as on bill delivery. We frame our analysis of mechanisms as investigating how a technology that primarily facilitated bill delivery nonetheless generated disproportionately large improvements in collection.

We find no evidence that technology provided direct advantages for collection conditional on delivery. First, technology could increase the likelihood that a household pays by improving tax morale or increasing perceptions of enforcement capacity. However, we find that technology has no average or heterogeneous impacts on numerous dimensions of enforcement perceptions and tax morale, including views on government's efforts to improve the efficiency of collection. A second possibility is that the presence of technology discouraged control collectors by denying them access to a potentially valuable tool to improve an otherwise challenging work environment. However, control and treatment collectors report similar job satisfaction, hours of work, and work challenges in all areas apart from navigation. Control collectors that randomly differed in their exposure to technology were also comparable in these dimensions. Third, the presence of technology may have increased collectors' perception of monitoring or households' perceived bargaining position, leading treatment collectors to substitute away from collecting bribes toward collecting taxes. If anything, however, we find that bribes increase in treatment areas. Moreover, there are no differences between groups in how collectors report being monitored by supervisors or in the resistance they report facing by households.

Rather than providing direct advantages to collection conditional on delivery, our preferred mechanism is that technology led treatment collectors to re-optimize their behavior across multiple activities, focusing less on deliveries and more on collections and learning about property owners. In particular, we argue that collectors allocate more time to collecting overall, and to learning about the hard-to-observe household characteristics that determine payment propensity, such as income, liquidity, and awareness of the tax code. Collectors subsequently use this local knowledge to target the households with higher propensity to pay.

This mechanism, which is novel in the literature, is supported by results from collector panel surveys and endline household surveys. Collector surveys reveal that treatment collectors allocate a much larger share of their time to non-delivery activities than control collectors. Household surveys show that treatment collectors spend this non-delivery time on conducting more visits with property owners and on having longer interactions with owners during each visit. These interactions are targeted: treatment collectors conduct more visits than control collectors with households that have higher actual propensity to pay (e.g. with higher income, liquidity and taxpayer awareness). Propensity is hard to observe and none of the information on households' individual propensity was provided to collectors or visible in the GIS-tablet or on the tax bills. Instead, our interpretation is that the treatment collector learns about households' hard-to-observe payment propensity by having more and longer interactions with them. Indeed, panel collector surveys reveal an increase over time in treatment collectors' knowledge about which households have a higher propensity to pay. Moreover, additional results on the characteristics of visited households are inconsistent with alternative settings where propensity is either perfectly observable or perfectly unobservable to all collectors. Consistent with using the new knowledge in collection strategies, we find that the treatment group collects a larger share of their payments from high propensity households.

We formalize our theory in a dynamic 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. Collectors spend their time endowment each period trying to deliver bills, learn about taxpayers, and collect from households of high-, low-, or unknown types. The collection probability from each household type is the same in the treatment and control groups. The technology increases the return to time spent delivering bills; this delivery advantage reflects technology's navigational improvement. We assume that the same navigational improvement also provides treatment collectors with a learning advantage reflecting the improved ability to locate households for follow-up interactions. The learning and delivery advantages are assumed to be proportional, consistent with our mechanism results that suggest technology did not provide direct advantages for collection conditional on delivery.

We estimate the main model to match the 27 percent experimental treatment effect on bill delivery and several other moments representing average outcomes across both groups. The model's predicted treatment effect on collections, which is not targeted, is 77 percent, meaning the model explains about two-thirds of the experiment's difference between the treatment effects on collections and deliveries. The model also predicts, without targeting, the experimental result that a higher fraction of collections in the treatment group comes from the high-type households. We show that these predictions arise from endogenous differences in time allocations: Treatment collectors shift from bill delivery to learning and collections earlier in the campaign than control collectors, and spend a larger fraction of their time gaining knowledge and a smaller fraction delivering bills; control collectors, continuously hindered by limited navigation, engage in almost no learning and switch directly from delivery to collection, though later in the campaign.

To disentangle the direct effects of technology, stemming from the improved navigation that increases the returns to delivery and learning, from the indirect effects, arising from the collectors' time re-allocations, we simulate a counterfactual that endows the collector with technology but without allowing them to re-optimize time allocations. This counterfactual yields treatment effects of similar magnitude on deliveries (52 percent) and collections (46 percent), in stark contrast to the experimental results. The lesson is that collector re-optimization is a significant component of the overall experimental impact of technology. In particular, re-optimization amplifies the effect on collection by 67 percent relative to the 'pure' effect of technology without re-optimization (77/46 = 1.67). Thus, an envelope-theorem logic does not apply here, meaning that the total effects of technology are not well approximated by their direct effects, ignoring re-optimization.

We next simulate a counterfactual in which we allow collectors to re-optimize but where technology only provides a delivery advantage and no learning advantage. This counterfactual predicts treatment effects that are twice as large for collections as for deliveries. Thus, the combined effects of a delivery advantage and collector time-reallocation explain around half the gap between the treatment effects on deliveries and collections in our experiment. Moving to the main estimation adds the learning advantage, and results in a treatment effect that is three times larger for collection. Quantitatively, then, 60 percent of the model's predicted gap in treatment effects between collection and delivery stem from the delivery advantage, and 40 percent come from the learning advantage.

Finally, when combined, the counterfactual exercises reveal how different parts of the model interact for learning to emerge. Specifically, learning emerges in our setting in two sequential steps: first, technology's navigational enhancement causes a significantly large improvement to delivery that the collector re-optimizes time-allocations across all activities; second, by providing a (navigational) advantage to learn, technology ensures that a significant share of the re-allocated time is devoted to learning. In other words, time re-allocation and a learning advantage are each necessary but not sufficient for learning to emerge as an important activity in the field.

We conclude that a technology which provided a significant navigational advantage to deliver bills increased collections directly, by improving the challenging task of delivery, but also indirectly, as collectors devoted more time to other challenging tasks. In our setting, as in other developing countries, tax officials are challenged by limited enforcement capacity, which necessitates multiple follow-up visits with taxpayers after bill delivery, and by limited information about taxpayers, which makes it hard to know which households are worth targeting for collection efforts. By improving the collector's return to time spent delivering, the technology allowed collectors to allocate more of their scarce time to the challenging tasks of learning about taxpayers and attempting to collect from them, particularly those with the highest propensity to pay. Quantitatively, we find that these indirect effects of technology are large, and help explain the bulk of why a technology that provided a strong advantage in bill delivery ultimately had a disproportionately larger impact on collections.

Related literature Our paper contributes to the empirical literature on technology adoption, in particular worker responses to new technologies (Houeix, 2025; Atkin et al., 2017; Prescott and Parente, 1994). The tax collector in Ghana is the archetypal street-level bureaucrat (Lipsky, 2010), who faces limited oversight and exercises significant discretion in deciding how to achieve loosely defined objectives. A related literature studies technologies that monitor the bureaucrat's activities, including to evaluate their performance and potentially curb their discretion: recent studies include Callen, Gulzar, Hasanain, Khan, and Rezaee (2020); Dal Bó, Finan, Li, and Schechter (2021); Dodge, Neggers, Pande, and Moore (2021); Mattsson (2023); Muralidharan, Niehaus, Sukhtankar, and Weaver (2021). Our study instead focuses on how technology can enhance the bureaucrat by improving the return to time spent on challenging tasks. In so doing, we show how workers respond to a technology that helps improve one challenging task by productively re-allocating their time toward other challenging tasks.[1]

Learning valuable local information emerges as a key activity that officials re-allocate time toward.[2] Previous studies have shown how *pre-existing* information, from third-parties and local chiefs, can improve taxation and policy (Balan, Bergeron, Tourek, and Weigel, 2022; Kleven, Knudsen, Kreiner, Pedersen, and Saez, 2011; Manara and Regan, 2022; Naritomi, 2019; Pomeranz, 2015). We show how, in settings where such information-sources are practically non-existent, the state can still strengthen tax capacity by *directly building* locally relevant information.[3] The newly built information may also strengthen expenditure capacity, if local officials can use the information to improve

[1]Though it has received less attention in empirical economics, it is well known in public administration research that street-level bureaucrats face multiple challenges that interact with each other. Beyond the public sector context, Suri and Udry (2022) emphasize that no single challenge, but rather the combination of multiple challenges, explains the lack of widespread technology adoption in African agriculture.

[2]Allocation of time use has received limited empirical attention as a dimension of bureaucrat performance (Finan et al., 2017), with recent exceptions (Bandiera, Best, Khan, and Prat, 2021; Kalaj, Rogger, and Somani, 2022). For detailed survey measures of bureaucrats' activities, see also Rasul and Rogger (2018).

[3]Our results on how officials build local information over time complement previous studies which focus on measuring officials' local information at a point in time (Dal Bó et al., 2021; Duflo et al., 2018).

transfer targeting and reduce their reliance on non-state actors (Basurto et al., 2020).

Our paper studies the impacts on government performance of improving localization in a setting with incomplete addressing. Incomplete addressing is a well-known issue in Ghana, but is not unique to the country: the UN estimates that 4 billion people live in places without an address (link). Moreover, GIS-technologies to overcome challenges from incomplete addressing are being implemented by governments around the world (Knebelmann, 2022). Despite its ubiquity, there is little work in economics on the consequences of incomplete addressing. Our experimental results shed light on the value of an address for government's ability to carry out core activities in the field.

Finally, our work relates to studies on technology and tax capacity: Hjort and Tian (2024) and Okunogbe and Tourek (2024) provide recent reviews.[4] To our knowledge, our paper is the first to directly randomize the presence of technology in the field for tax officials. Moreover, by focusing on the localization benefits of a GIS-based registry, our paper complements recent work on enhancing tax registries. In the context of building a digital registry, Knebelmann et al. (2023) investigate the value of officials' discretion and Aman-Rana and Minaudier (2024) study the organizational impacts on officials' ability to collect taxes. Okunogbe (2021) studies the impacts on tax compliance of providing information to taxpayers based on an enhanced registry. Ferraz, Foremny, and Santini (2024), Gadenne (2017), and Martinez (2023) study the interactions between non-tax revenue flows and registry enhancements, and Casaburi and Troiano (2016) analyze the electoral impacts of an enhanced registry.[5]

2 Census of Tax Capacity in Local Governments

To understand the process of local taxation and the challenges of collection, we conducted Ghana's first census of all local governments in 2017. In each of the country's 216 local governments, we interviewed three sets of respondents: citizens, officials, and locally elected assembly members. Every local official involved in the tax collection process was surveyed, including: the chief executive (political head); the coordinating director (bureaucratic head); finance officers; physical planning officers; revenue supervisors; and, field collectors. The census contains responses from 5,375 citizens (approximately 25 per district) and 2,785 local government officials and assembly members (13 per district). We also digitized administrative records of tax collection across all 216 districts.

Tax collections are determined by the probability of bill delivery (the delivery margin)

[4]See also Brockmeyer and Somarriba (2022), Eissa and Zeitlin (2014), Fan, Liu, Qian, and Wen (2021), Mascagni, Mengistu, and Woldeyes (2021), Okunogbe and Pouliquen (2022) and Das et al. (2023).

[5]Our work also relates to the literature on experiments with tax collectors, including Bergeron, Bessone, Kabeya, Tourek, and Weigel (2022), Khan, Khwaja, and Olken (2015, 2019), and Weigel (2020).

Table 1: Characteristics of Local Tax Capacity in Ghana

	Mean	Median
Share of bills delivered (%)	43	43
Share of delivered bills that are paid (%)	30	29
Share of properties with address (%)	27	21
Common not to locate properties [0,1]	0.78	100
Take tax defaulters to court [0,1]	0.22	0
Citizen has tax awareness [0,1]	0.07	0
Citizen has public goods awareness [0,1]	0.35	29
Has electronic property database [0,1]	0.15	0
Number of local governments	216	216

Note: All variables are unweighted averages calculated at the local government level.See Section 2 for details on the variables.

and the amount paid conditional on delivery (the payment margin).[6] On the delivery margin, Table 1 shows that only 43 percent of property tax bills are delivered in the average local government (Table 1). The delivery margin is thus significantly challenged. Most studies abstract from bill delivery and focus instead on the payment margin. The payment margin is also limited in this setting: in the average district, the likelihood a property owner pays their property taxes after receiving a bill is 30 percent.

In the census data, absence of data on properties was the most frequently cited challenge for local tax collection by local officials and assembly members. For bill delivery, the absence of data on property owners and difficulties in locating them were cited as two of the three most important challenges.[7]

This lack of data information begins with the simple absence of street and property addressing. Table 1 shows that in the average district, only 27 percent of properties have an official address, meaning a property number on a named street. The property tax registry inherits this limited address information, and hence the location listed on most property tax bills is imprecise. Typically, the main reference is a nearby landmark. For example, Figure A1 provides an illustration of an actual tax bill where the location

[6]The denominator in the bill delivery measure is the set of properties that are registered; this delivery margin is different from the registration margin (the share of existing properties that are registered).

[7]The absence of precise information on taxpayers has been documented in national tax authorities around the world: of 61 assessments in lower income countries conducted by the World Bank, only 5% received a score of good or better for the accuracy of information in the taxpayer registry (Nyanga, 2021).

of the property is listed as "Opposite Presec School" (a secondary boarding school). In the absence of precise addressing, collectors typically navigate to the landmark and then attempt to locate the targeted property (Figure A1). These attempts often result in failure. Across all of Ghana's districts, 78 percent of local collectors reported that it is common not to be able to locate the intended property for delivery (Table 1).[8]

Incomplete addressing is a well-known issue in Ghana[9] but is not unique to the country. The United Nations estimates that 4 billion people live in places without an official address (link), while Farvacque et al. (2005) estimate that half of the global urban population does not have an official address. The existence of incomplete addressing has received little attention in economics. While the policy target may seem simple, a multi-disciplinary literature has identified several factors that limit broad addressing coverage (Marx, Stoker, and Suri, 2013).[10]

The low payment likelihood in part reflects limited enforcement. The most credible enforcement tool of local governments is to summon the delinquent taxpayer to court, yet only 22 percent of local governments resorted to this at all during the past year (Table 1). In high-capacity tax systems, enforcement is supported by third-party information coverage (Gordon and Li, 2009; Kleven et al., 2016; Pomeranz, 2015), but such sources of 'hard information' are virtually non-existent at the local government level in Ghana.

It is precisely in settings where hard information and enforcement are limited that 'soft' information on property owners' propensity to pay can be helpful (Luttmer and Singhal, 2014). Balan et al. (2022) show, for example, that traditional leaders in the DRC have relevant soft information on property owners that can be leveraged to collect taxes. The challenge is that propensity to pay varies significantly across households and is hard for officials to initially observe. For example, awareness of the tax code and local public good provisions are determinants of payment propensity. However, only 7 percent of citizens in the average district were aware of the "fee fixing resolution" that underlies the official local tax rates and regulations, and only 35 percent could name any project undertaken by their local government in the past two years (Table 1).

Local governments can invest in technology to alleviate challenges for tax capacity. One promising technology in this setting is a GIS-enhanced tax registry, which can help increase bill delivery by improving localization (World Bank, 2020). Constructing this

[8] Table A1 also shows a positive association across the country's districts between the share of properties with addressing and the share of bills that are delivered, suggesting that limited addressing ultimately creates significant challenges for local governments. These observations are consistent with the central government's message that local governments "have no accurate spatial reference" and that "the importance of street and property addressing cannot be overstated" (Government of Ghana, 2011).

[9] Some examples of discussions in the media include link #1, link #2 and link #3.

[10] The findings from this multi-disciplinary literature are reviewed in the online appendix (link).

registry requires digitally recording the geographic coordinates of all parcels using GPS coordinates and aerial data, and provides precise location information for governments in the absence of an official addressing system.[11]

Table 1 shows that only 15 percent of local governments have invested in a GIS-enhanced tax registry. Adoption of technology is at the discretion of each local government.[12] We find a robust correlation across districts between adoption of technology and improvements in both the delivery and the payment margins (Table A3). Indeed, districts with GIS-technologies both deliver more bills and collect more taxes per bill delivered; ultimately, GIS-adoption is associated with much higher tax revenue per capita.[13]

3 Experiment and Main Tax Results

The associations in the census suggest a potential role for GIS-technology to improve tax capacity, both by directly improving the delivery margin and by indirectly helping to address challenges on the payment margin. Motivated by these patterns across local governments throughout the country, we implement an experiment within one local government to causally establish the impacts of a GIS-based technology and to precisely investigate mechanism impacts on different challenges in the tax collection process.

3.1 Setting, Design and Data

Setting We conducted the experiment in 2021 in La Nkwantanang Madina Municipal Assembly (henceforth, Madina). Madina is more affluent and urban than the average district. We collaborated with the municipal government and a private Ghanaian firm that developed a new technology to help increase property taxes. The technology is based on a digital property tax registry with precise geo-coordinates of each structure that was constructed from high-resolution aerial photographs and in-person visits. The digital registry is accessed using a hand-held electronic tablet with GIS capabilities. The tablet provides the user's live location on a digital map as well as the location of a designated property (Figure A1) but does not automate navigation, such as by calculating the most efficient route or providing turn-by-turn directions.[14] The tablet aims to assist collectors by helping them navigate to specific properties in the field. As we detail below, what we vary across treatment and control groups is the presence of the tablet.[15]

[11]Non-technology initiatives to overcome limited addressing have faced challenges in Ghana (Abebrese, 2019) and elsewhere (Bigon and Njoh, 2012). The online appendix (link) summarizes these initiatives.

[12]Table A2 provides cross-district correlates of adoption choices. See also Knebelmann (2022).

[13]These associations are larger when the district's addressing coverage is smaller (Table A1), which may reflect the improved localization provided by GIS in the context of incomplete addressing.

[14]The tablet's features score only 2 out of 10 on the system automation scale of Parasuraman (2000).

[15]The GIS-registry was constructed prior to the experiment. The GIS-data was also used in a software to compile bills and issue enforcement notices, but these components are not randomized in the experiment.

During a fiscal year, the local government assigns collectors to designated geographical areas for six weeks at a time (a ’campaign’). The designated areas are called ’collection units’ and are defined with geographical boundaries that create a cluster of physically adjacent properties. During each six-week campaign, collectors are responsible for delivering bills and collecting payments from assigned property taxpayers in their unit. Each unit is only covered once during a fiscal year; after each campaign, the collector is assigned to a new unit. The relatively short duration of each campaign results from the large number of properties relative to the limited number of collectors. The duration is not specific to Madina: the local governance Act stipulates that all property owners in the country are legally required to pay within the six weeks that mark a campaign. Pay stations exist but virtually all payments are made to the collector, most often in cash.

Our experiment was embedded in the six-week campaign between March and April 2021. Before the campaign, collectors received training. 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, an 8-percent commission rate on taxes collected from assigned bills, was chosen by the local government and held constant across groups. Collectors also received a daily transportation allowance and a base salary.

Experimental design All of Madina’s collectors participated in the campaign. Out of the 56 collectors that were trained, 28 were randomly assigned to the treatment group and 28 to the control group. Of the 56 collectors, 39 had previously worked in Madina and 17 were hired shortly before the experiment. Of the 39 collectors with previous experience, 11 were rated as ‘high performing’ by the local government. Collectors worked individually in their assigned collection units and were assigned approximately 145 bills each. In the treatment group, the GIS-tablet contains precise localization details for all assigned properties but displays localization only for a single property at a time after the collector clicks on the property ID in the assigned list. The tablet provided to a collector only displays information for the collector’s specifically assigned properties.

All supervisors were randomly assigned to both treatment and control collectors for the duration of the campaign. Supervisors were in charge of monitoring the revenue collectors and assisting them with challenges in the field.[16]

[16] The implemented experiment differs in two ways from the protocol described in the pre-analysis plan. First, 56, rather than 60, collectors were included because 4 dropped out before assignment to treatment. Second, the pre-analysis plan indicates that the navigational tablet: records the payment status of all assigned properties; is monitored in real-time; and, issues a digital receipt upon payment. For logistical reasons, these features were not active or available in the tablet for the experiment.

At the beginning of the experiment, all collectors in the treatment group were given the tablet for use during the tax campaign. Other than the tablet, the treatment group was not provided with any other advantages. Both groups were provided with the printed bills for all properties in their respective collection units, which contain information on the property's location (though imprecise), the current tax liability (which includes arrears), and the previous year's amount of taxes paid and due (see Figure A1 for an example). Madina implements a presumptive tax schedule where the tax liability is based on observable characteristics rather than a directly assessed property value.[17] The characteristics are: number of floors; quality of material used to build the outer walls and roof; geographical zone.[18]

Both groups were provided with physical maps that provide limited, aggregated spatial information by delineating the collection unit's boundaries relative to a small set of main roads. Qualitative work from our pilot revealed that control collectors hardly made use of the maps due to their lack of detail on specific property locations, however (Figure A1). We therefore consider that the control group represents a reasonable approximation to the status quo where collectors do not rely on technology in their field work.[19]

Our randomization proceeded in two steps. First, we randomly assigned each collector to a collection unit.[20] Second, we randomly assigned the collector-collection unit pair to the treatment or control group. We stratified 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 specific to collectors: a dummy for previous work experience in Madina, and a dummy for high-performance rating. The other four variables were specific 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 A4 summarizes balance checks for characteristics at the tax bill level (Panel A), the collector-unit level (Panel B), and the household-level (Panel C). None of the individual characteristics is statistically sig-

[17]Our census reveals that less than 20% of local governments have property valuations. Presumptive schedules are more common in developing countries where comprehensive market-value information for valuations, including from banks and mortgage providers, is lacking (Franzsen and McCluskey, 2017).

[18]These characteristics were recorded by officials when the registry was built, prior to the experiment.

[19]The physical map was created based on the GIS-enhanced registry but, as noted, control collectors in practice did not make much use of this map. Moreover, apart from a potential improvement in the accuracy of the delineations, the aggregate spatial information conveyed in the map is similar before and after the GIS-enhanced registry. Finally, no other technological tools related to the GIS-registry were used during the experiment, including for supervisors.

[20]Random, rotating assignment of collectors to units is part of the local government's standard process.

nificantly different between groups at 10%. Moreover, we fail to reject the null that the difference in all characteristics are all zero at the tax bill-level ($F = 0.71, p = 0.66$), collector-unit level ($F = 0.16, p = 0.95$) and household-level ($F = 1.07, p = 0.38$).

Experimental data and estimation In this section, we describe the data sources used in our analyses. We use administrative data at the property level, covering 8,120 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.

The local government gathers daily data from each collector on the number of bills delivered and the amount of revenue collected. These data, gathered in a uniform manner from all collectors, allow us to study the activities of both groups at a high frequency. Collecting these data is part of the government's routine campaign process, which helps explain the very low attrition (uncorrelated with treatment).[21] To minimize idiosyncratic measurement error, we winsorize outcomes at the 95^{th} percentile by group and day.

In addition to the daily data, enumerators working for the research team conducted three rounds of detailed surveys with the 56 collectors. The first 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. There is 17% attrition in the collector surveys, but attrition is uncorrelated with assignment to the treatment group (coefficient of -0.017, standard error of 0.086). The main tables report treatment effects based on the unbalanced sample, though Table A5 shows that the results are similar in the balanced sample. This implies that any characteristic which predicts attrition is not a significant source of treatment effect heterogeneity. Notwithstanding, we use the balanced sample in all graphs that show levels of variables by group and survey round, to remove any influence from compositional effects between rounds.

Finally, the enumerators administered end-line surveys with 4,334 randomly selected households in May 2021. A random sample of equal size was drawn from each of the 56 collection units. Whenever an initially selected property could not be contacted, an adjacent property within the same collection unit would be randomly chosen. All variables created with these data are described when they first appear in the analyses.[22]

Given the random treatment assignment, we use OLS to estimate the causal impacts of technology. The econometric specification varies slightly depending on the unit of

[21] Each collector submits the daily information to the finance officer either in person or over the phone. Our research group provided support by randomly auditing the data on a collector-day level.

[22] We refer to the online appendix (link) for more details on all variables used in the paper.

observation. For outcomes that vary at the day and collector level, we estimate:

$$y_{cd} = \beta_d \cdot \mathbf{1}(Tech)_c + \theta_d + \Omega \cdot X_c + \epsilon_{cd}, \tag{1}$$

where y_{cd} is the outcome for collector-collection unit c on day d, θ_d are campaign-day fixed effects, and X_c contains time-invariant controls. In the main analysis, X_c only includes strata fixed 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 $\mathbf{1}(Tech)_c$ takes a value of 1 for all collector-units randomly assigned to treatment, and 0 for collector-units assigned to control. The treatment coefficient, β_d, is indexed by day because we estimate dynamic treatment effects by interacting the treatment dummy with the individual campaign-day fixed effects. In a robustness check, we leverage the panel-structure and include fixed effects for each collector-collection unit. 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, β_1 (the omitted category). Standard errors are clustered at the collector-unit level. We also use (1) to estimate impacts in the collector surveys, replacing day d by survey round s.

For outcomes at the household level, we estimate:

$$y_{hc} = \beta \cdot \mathbf{1}(Tech)_c + \Omega \cdot X_{hc} + \epsilon_{hc}, \tag{2}$$

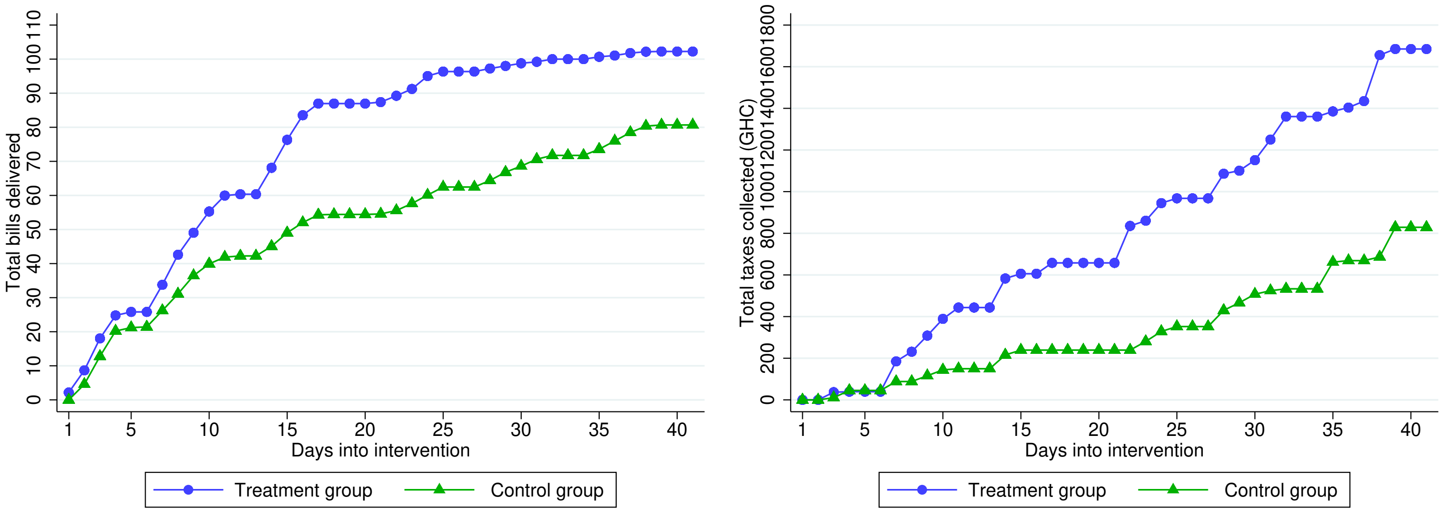
where h indexes households and c collector-units. Standard errors are clustered by collector-unit. X_{hc} always includes strata fixed effects. In robustness checks, we also include the controls at the collector-unit level from (1), as well as the household's property category and previous pay status (fully paid, partly paid, not paid).

3.2 Main Experimental Effects on Tax Outcomes

We begin by studying the impacts of technology on bill delivery and tax collection using the collector daily reports. Figure 1 shows the average total bills delivered (Panel A) and taxes collected (Panel B), by group and campaign-day. Figure A2 shows the corresponding daily treatment coefficients β_d (equation 1). Panel A shows that 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 deliver bills. The treatment coefficients are statistically significant at the 5 percent level in all campaign-days beyond the 10^{th} day (Figure A2). 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.

Figure 1: Impact of Technology on Bills Delivered and Taxes Collected

(a) Bills Delivered (b) Taxes Collected



Notes: This figure shows the impacts of technology on the total number of property tax bills delivered and taxes collected. Panel A shows the average total number of bills delivered by group and by day of the tax campaign. Panel B displays total taxes collected on average by group and day. Figure A2 shows the corresponding daily treatment coefficients (β_d in equation 1). See Section 3.1 for details.

Panel B of Figure 1 shows that technology causes a large increase in total taxes collected. There are no differences in collection performance during the first week, when 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 significant at the 5 percent level on all subsequent days and grows over time (Figure A2). In the end, 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.[23]

We can infer from these results that the treatment group collects more taxes per bill delivered. Figure A3 shows that this outcome grows over time; at the end of the campaign, the treatment group has collected substantially 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. This result mirrors the census cross-sectional regression, where we also found more taxes collected per bill delivered

[23]Total taxes net of costs increase by 96 percent. This calculation includes the running costs to fund the daily use of the GIS-tablet, but not the (primarily fixed) costs of building the underlying GIS-registry.

in local governments with GIS-technology than without (Table A3).

Robustness Figure A2 provides robustness checks. Estimates are similar when using non-winsorized outcomes.[24] Results are similar upon including the additional covariates contained in X_c (equation 1). Finally, the results are comparable when we include collector-unit fixed effects; in this case, β_d reflects the treatment effect based on changes within collector-unit over time (relative to the initial impact on day 1).[25]

An important concern is whether COVID-19 impacted the results. We conducted a pilot experiment in 2019 using the same location, technology and protocol (though with fewer collectors). We found similar qualitative and quantitative effects in the pilot as in the main experiment.[26] 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 Table A6 reports the treatment effects on tax outcomes based on equation (2) and using the independent household survey. Households in the treatment group are more likely to have received a tax bill and made a tax payment.[27] The magnitudes imply a higher payment amount conditional on bill delivery in treatment areas. The impact is statistically significant for tax payment, but not for bill delivery at conventional levels.[28] At the same time, we cannot reject the null that the effect-size on bill delivery is the same in the household survey and in the daily collector reports (p-value 0.30). We also fail to reject the null hypothesis that the effects on tax payment are similar across these data-sources (p-value 0.27).

Heterogeneity by collector We can leverage the random assignment of collectors to collection units to estimate the fixed effects for each collector-unit. Using the tax outcome from the household survey, Figure A4 shows that, while there is significant variation in performance between control collectors, the technology intervention appears to have increased performance at most parts of the control group distribution. For policy, this suggests that the effectiveness of technology does not seem to hinge on a particularly high or low initial level of collector capacity. For our investigation, this result motivates our focus on mechanism channels for the average collector.

[24]The effects are almost identical across all sub-samples which leave out one collector at a time (results available), alleviating concerns over undue influence by one outlier performing collector.

[25]The fixed effect technically captures variation within each collector-collection unit. However, we interpret it as reflecting 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.

[26]Qualitatively, the pilot and main experiment both produce an effect on bill delivery that is larger in the middle of the campaign than at the 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.

[27]Table A6 shows the robustness to removing all controls X_{hc} and including more extensive controls.

[28]The delivery effect increases with income, consistent with learning (see Section 5).

4 Potential Mechanisms For Experimental Tax Results

In this section and the following one, we frame our investigation of mechanisms as trying to explain how a technology that provided a substantial advantage in delivery nonetheless caused a treatment effect on collections that was 4 times larger than on delivery. In this section, we investigate three potential mechanisms that are motivated by the literature and through which technology could provide a direct advantage to collection conditional on delivery. In the following section, we focus on a mechanism that is new in the literature, which combines time use re-allocation and learning.

4.1 Tax Morale and Perceived Enforcement Capabilities

The first mechanism we consider is that technology improved households' tax morale or increased households' perceived enforcement capabilities of local government. Tax morale is broadly defined 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 more efficiently and equitably. Household perceptions of government enforcement may also change if seeing a 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 efficient ways; satisfaction with government services; government integrity and governance capacity. We also create an index of households' perceptions of government informational capacity and enforcement strength. Each index is based on several individual questions which are detailed in Table A7.

In Panel A of Table 2, we find null effects of the technology on all four indices of household tax morale and enforcement perceptions. In Table A7, we find null effects on 12 of the 13 individual underlying household survey questions used to build the four indices.[29] For example, there is no treatment impact on the households' perception that a non-complier will end up paying, or on the share of households agreeing that government will use tax revenues wisely. The only statistically significant impact is a decrease in the perception that everyone pays their fair share of taxes; if anything, this could suggest lower perceptions of equity in treatment areas, but it is not conclusive as all other equity questions have null effects. The average null effects may mask heterogeneity along the income distribution if, for example, morale is boosted only among the well off that previously paid taxes. Yet we find no significant heterogeneous effects on any of the indices by the income level of the household (Table A8).

[29] Based on the collector surveys, Tables A10 and A11 show that there are also no treatment effects on collectors' reports of resistance by households, in relation to acknowledging receipt of the bill delivery, complaining about the amount due, or mistrusting the collector to handle cash payments.

In the learning mechanism described below, we find that households with hard-to-observe propensity to pay (measured by income, liquidity and taxpayer awareness) are more likely to pay in treatment than in control areas. Importantly, Table A8 shows there are no heterogeneous effects on the four indices by propensity to pay – suggesting that the higher payment rate in treatment areas among those with higher propensity to pay is not driven by an increase in morale or enforcement specifically for those households.[30]

Table 2: Tax Morale, Enforcement Beliefs and Bribes

Panel A: Beliefs & Morale	Satisfaction w. gov't services (1)	Integrity of gov't (2)	Equity & efficiency (3)	Enforcement & information (4)
1(Technology)	-0.007	0.062	-0.014	-0.053
	(0.070)	(0.072)	(0.060)	(0.057)
Mean in CG	0.045	-0.039	-0.033	0.004
Observations	4334	4334	4334	4334
Panel B: Bribes	**1**(Any bribe) (5)	Total bribe (in %) (6)	Coercive bribe (in %) (7)	Collusive bribe (in GHC) (8)
1(Technology)	0.116***	0.025**	0.011*	6.160**
	(0.039)	(0.011)	(0.006)	(3.071)
Mean in CG	0.139	0.117	0.039	11.612
Observations	4334	4334	4334	4334

Notes: This table presents the impacts of technology on tax morale, enforcement beliefs and bribes based on (2). Panel A focuses on beliefs and morale: satisfaction with government's delivery of services; perceived integrity and competency of local government; perceived government efforts to collect taxes in an equitable and efficient manner; perceived enforcement capacity and informational knowledge of local government. Panel B focuses on bribe outcomes: a dummy which takes a value of 1 if the household estimates that the tax collector asks for an unofficial payment during visits to property owners (collusive bribe) or pockets some of the money collected from property owners (coercive bribe), and 0 otherwise; total bribe amount (in %), which is the average of the coercive bribe amount and the collusive bribe amount; coercive bribe amount (% of a hypothetical 1000 GHC); collusive bribe amount (in GHC). * p<0.10 ** p<0.05 *** p<0.01. See Table A6 and A7 for more details on the outcomes.

Finally, we leverage the fact that a subset of households was included for the first time in the tax registry for the experiment campaign. The average null effects may mask heterogeneity by extent of prior exposure to tax collectors; for example, positive impacts of technology on morale and enforcement may be concentrated amongst those that have no prior interaction with the taxation process. In Table A9, we find no significant treatment heterogeneity between newly and previously registered property owners. Finally,

[30] These results also hold with the 'high type' indicator of propensity to pay (Section 5.3).

the null results may be due to a lack of salience: the treatment group collectors were not instructed to show the tablet to property owners to convey changes to the tax collection process; moreover, the tablet mainly helps with navigation and the issuance of a receipt for payment still has to be done manually in both treatment and control areas.[31]

4.2 Bribes

The second mechanism we consider is that technology may have improved the payment margin, conditional on delivery, by reducing bribe activity. Bribes can take 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.

The effect of technology on these two types of bribe activities is ambiguous ex-ante. On one hand, technology can reduce these bribe activities through better monitoring by supervisors, or easier reporting of bribe taking by households. On the other hand, technology may increase households' perception of collectors' enforcement capacity and raise collectors' bargaining power, which could increase bribe taking. Technology could also free up time for the collectors, which they may use to attempt to collect bribes.

We use the household survey to measure bribes which, due to their illegal and culturally sensitive nature, come from indirect questions (e.g. we ask if it is likely that collectors *in the household's area* ask for bribes). In Panel B of Table 2, we find a positive treatment effect on bribes.[32] While technology causes an increase in the likelihood of coercive or collusive bribes (column 5), the treatment effects on bribe amounts are smaller (columns 6-8) – for example, the treatment effect is over 4 times smaller on collusive bribe amount in GHC (column 8) than on tax amount paid in GHC (Table A6).

These positive impacts suggest that the larger treatment effect on tax collection than on bill delivery does not operate through a substitution away from bribe activities.

[31] Our survey asked households in treatment areas if they had seen the tablet. In a household level OLS regression with collector fixed effects, having seen the tablet is not significantly associated with morale and enforcement perceptions (results available). Our null effects capture short run impacts; it is possible that tax morale or enforcement views may be shaped in the longer run, for example if repeated interactions with technology cause households to feel they are becoming more legible to the state (Okunogbe, 2021).

[32] One possible explanation for the positive bribe impact could be that the supervisors monitored the treatment group less than the control group. However, we find a null effect on monitoring as perceived by the collectors (Table A10). Motivated by the learning mechanism (Section 5), treatment collectors may have spent time discovering which households are more likely to pay a bribe: for example, households that have witnessed government's enforcement actions may more strongly perceive collectors' threat of retaliation and therefore be amenable to pay the bribe. We lack data on bribe incidence outside of our experiment, which limits our ability to construct predictors of bribes and pursue this hypothesis rigorously. With this caveat in mind, the online appendix (link) explores this hypothesis.

4.3 Collector Effort and Motivation

An important concern is that control collectors may have put in less effort, or felt less motivated, since they were not given access to the new technology.[33] This could explain the larger treatment effect on collection relative to delivery if control collectors' drop in activity occurred in the later parts of the campaign where more time was spent on collection than on delivery. Four results suggest that this is not a significant concern.

First, there are no treatment effects on collectors' hours worked (Table 3) or self-reported job satisfaction (Table A10). The job satisfaction variable is an index, and Table A11 shows that there are no treatment effects on the underlying individual questions used to create the index. Figure A5 shows that these null average effects do not mask significant heterogeneity over time. Importantly, the control group does not see a drop off in hours worked or job satisfaction over time. If anything, both hours worked and job satisfaction increase, though by small amounts, toward the end of the campaign in the control group (similar to the treatment group). These results support the interpretation that control collectors were not specifically discouraged over the course of the campaign.

Second, we leverage the fact that all collectors were randomly assigned to collection units to investigate if control collectors that worked in closer proximity to treatment collectors performed differently. For each control collector unit, we calculate the share of geographically adjacent collection units that are populated by treatment collectors. In the sample of control collectors, Table A12 shows that this variable does not cause any impact on a broad set of outcomes from the collector surveys (including job satisfaction, hours worked, strategies used, knowledge) and the daily administrative data (bills delivered and taxes collected). With the caveat that this variable is an imperfect measure for knowledge about the technology, these null results support the interpretation that the technology did not induce major artificial changes in control collectors' behavior.

Third, beyond navigational issues that the tablet was designed to alleviate, collectors face other challenges in the field – such as wrong information printed on the bills, complaints from households about the bill amount, household mistrust of the collector's handling of cash payments, or resistance from households to acknowledge receipt of the bill delivery. Tables A10 and A11 show that there are no significant differences between treatment and control collectors in any of these additional challenges.

Fourth, supervisors were randomly assigned to groups. Supervisors may have sought to disproportionately help treatment collectors or restrain control collectors to artificially create positive impacts of technology. While it is hard to fully rule this out, we reassur-

[33] In the training sessions, neither group was made aware of the other group's activities, nor was it suggested that the continued use of technology depended on the performance during the campaign.

ingly find no differences between groups in collectors' reports (based on confidential surveys with our independent enumerators) of how often supervisors were available to collectors, monitored them or checked for mistakes in their work (Table A10).

5 Collector Time Allocation and Learning

This section investigates the hypothesis that the disproportionately large effect on collection versus delivery arose because collectors respond to the new technology by shifting their allocation of time away from deliveries and toward learning about and collecting from households, particularly those with higher propensity to pay. As this mechanism is novel in the literature, we begin with background observations to motivate it. We then present evidence on time allocations, learning, and collector strategies.

5.1 Navigational Challenges and Household Types

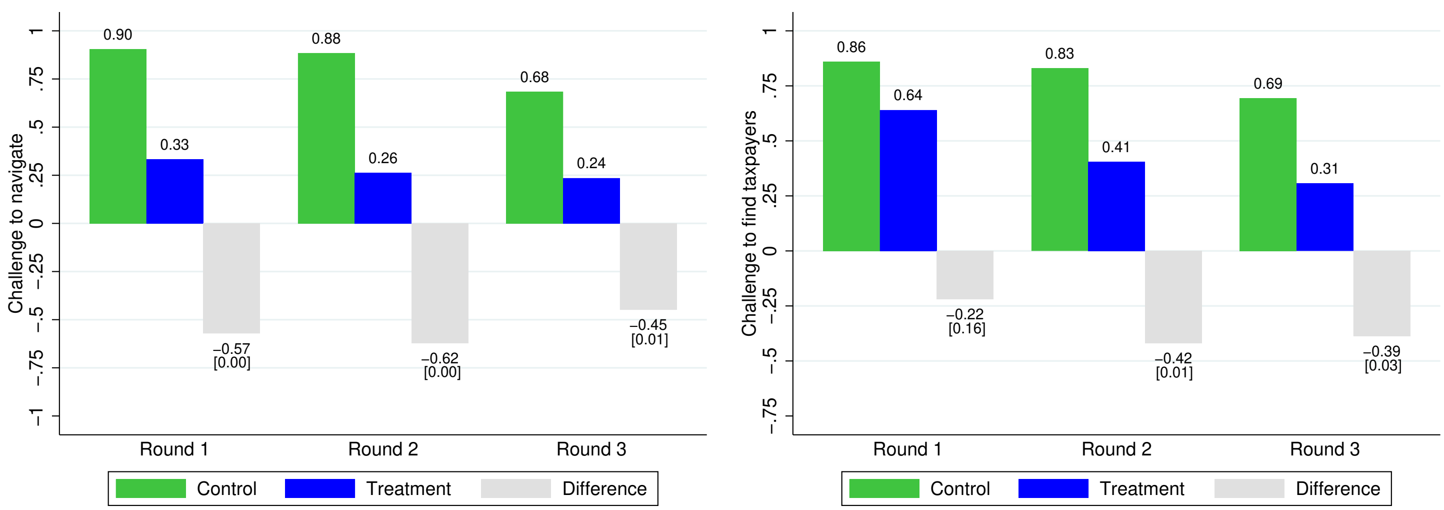
For this mechanism to help explain the results of Section 3, it has to be that collectors were initially challenged in their ability to locate property owners and in their knowledge of households' propensity. Several observations are consistent with such a setting.

Self-reported time-use data show that the average control collector would require 9 weeks to deliver all 145 assigned bills, let alone conduct follow up visits to collect payment, while the campaign lasts 6 weeks (based on Table 3 and Figure A5). Navigational challenges likely constrain delivery: at baseline, 90 percent of control collectors find it challenging to navigate in the field, and 86 percent find it hard to locate the assigned taxpayer (Figure 2). In the control group, the size of the collection unit, measured as the time it would take to travel to all assigned properties once for delivery, is negatively associated with the actual number of bills delivered (Figure A6). These observations suggest that eased navigation may improve the delivery margin.

A household's propensity to pay is an important determinant of tax payment when enforcement is limited. This observation was revealed in qualitative focus groups with local government officials and experienced tax collectors in our setting; it is also consistent with findings from the literature (Luttmer and Singhal, 2014; Balan et al., 2022). In our setting, propensity is determined primarily by a household's income, its liquidity and its awareness of the tax system. We use the household survey to measure propensity to pay by combining proxies for income, liquidity and awareness into an index at the household level.[34] This index strongly predicts actual tax payment in a sample outside

[34] The income proxy is the standardized household's total earnings in the past month. The liquidity proxy is the average of the standardized number of days the household finds itself short of cash for basic expenditures, and the standardized reported difficulty with which the household could find 300 GHC to pay an unexpected fee. The awareness proxy is the standardized sum of correct answers to six questions about the property tax in Madina. The average of the three standardized proxies is the propensity index.

Figure 2: Challenges to Navigate and Locate Taxpayers



Notes: This figure shows navigation and localization challenges, based on equation (1). In Panel A, the outcome equals 1 if a collector reports finding it challenging or very challenging to navigate in the field. In Panel B, the outcome equals 1 if a collector reports finding it challenging or very challenging to locate an assigned taxpayer. The randomization inference-based p-value of the difference between treatment and control is reported in brackets. For details on the outcomes, see Section 5 and Table A5.

of our experiment (Table A13).[35] We label propensity to pay the 'hard to observe' index because these household characteristics are hard to directly observe.

We construct an 'easy to observe' index, which combines the following characteristics: amount of taxes due; taxes paid in the past; easily observable assets (ownership of a car, motorbike or electric generator); distance to main roads, markets and the local government office. A collector can directly infer the current tax due and whether a property paid taxes in the past year based on the bill information (Figure A1); the last two characteristics can easily be inferred in the field.[36] The index is the average of the standardized values of these characteristics. This index is positively associated with compliance outside of our experiment, but less strongly so than the hard to observe index (Table A13) – suggesting that targeting properties with easily observable characteristics may be a relevant strategy for collectors, but with a lower potential yield than targeting

[35]These proxies are based on the end-line household survey, but they are not impacted by the treatment: see Table A4. Indeed, it is unlikely that the treatment effect on tax payment affects households' income choices within the six campaign weeks. The liquidity questions refer to a 'typical' month rather than the specific campaign months. Finally, no property owner from the experimental areas was summoned to court for non-payment during the campaign, which could otherwise have raised tax awareness.

[36]Results are unchanged if assets is instead considered a 'hard to observe' characteristic. This is because assets is not a dimension that is differentially targeted across groups: see Panel C of Figure A8.

hard to observe characteristics. Within a collection unit, the observable index accounts for only 4 percent of the variation across households in the hard to observe index.[37]

In summary, while it is ultimately a strong predictor of tax payment, propensity to pay is hard to observe, heterogeneous across households and weakly correlated with more easily observable characteristics. Reflecting these features, at baseline 88 percent of control collectors report not having a good understanding of which households in their assigned area have a higher propensity to pay (Figure 3).

5.2 Treatment Effects on Collector Time Allocations

From the outset of the campaign, treatment collectors report fewer challenges than control collectors to navigate and locate the assigned taxpayer (Figure 2). Both groups experience reductions in challenges over time, but technology's treatment effects are substantial throughout the campaign and, despite the small sample size, significant in almost all survey rounds. These improvements are not mechanical, as the tablet does not automate navigation (Section 3.1); the collector has to actively apply effort and switch between the visual information in the field and the tablet's map and localization features.

The positive effect of technology on bill delivery shows up in significant reductions in time spent per bill delivered. In particular, the treatment group spends 63 percent less time per bill delivered than the control group (Table 3). This reflects technology's intended impact, namely to enhance delivery in a setting with incomplete addressing.

Technology leads to significant changes in collectors' time allocations. In the average week, the treatment group spends two and a half times more hours on non-delivery activities than the control group, even though both groups work the same total number of hours (Table 3), causing them to allocate a larger share of time away from delivery to other activities. What do treatment collectors spend their extra non-delivery time on?

5.3 Treatment Effects on Learning and Collection Strategies

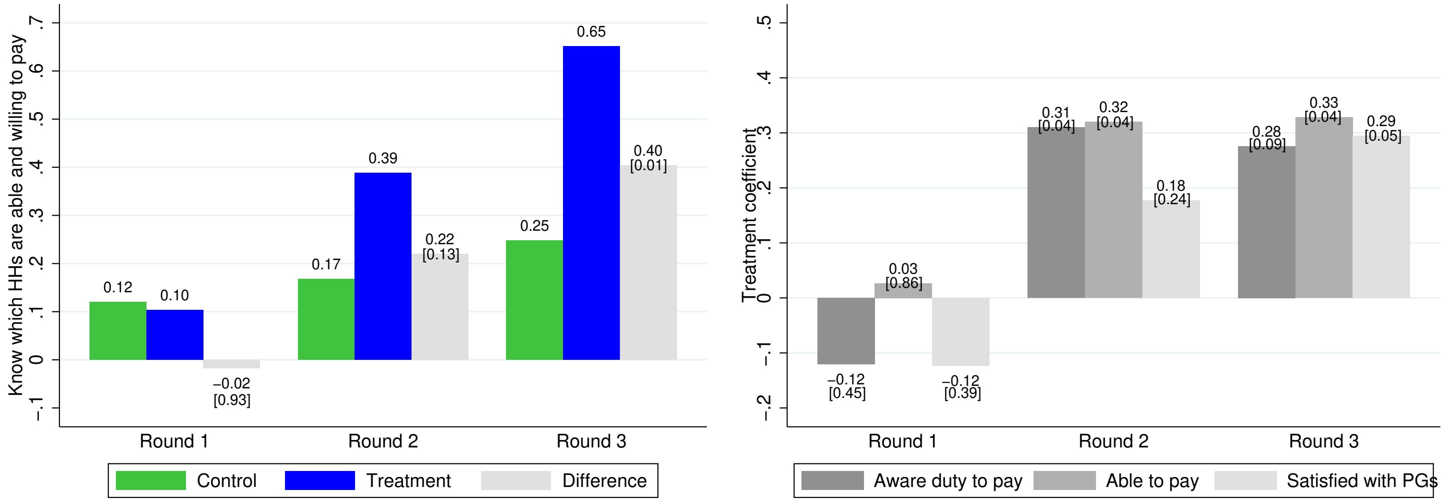
This sub-section provides evidence that treatment collectors spend their extra non-delivery time on learning about households' hard-to-observe propensity to pay, and using this new knowledge in their collection strategies. By learning, we mean that the household's propensity is initially not observed, but the collector can discover the type by paying the time cost to have longer and more frequent interactions with the household.

Frequency and duration of interactions Collectors can learn about propensity to pay in various ways based on interactions with the household. The collector can take time to ask questions that directly or indirectly relate to the property owner's liquidity, income

[37] The weak correlation arises in part because of the presumptive tax schedule (Section 3.1), which weakens the link between property tax liability and 'true' property value, let alone household income.

Figure 3: Collector Knowledge and Strategies

(a) Knowledge (b) Collection strategies



Notes: These panels show collector knowledge (Panel A) and strategies (Panel B), based on equation (1). 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. Panel B shows the estimated difference between groups in three collection strategies, focusing on: property owners that are more aware of their duty to pay; properties on specific days where the owners are more likely to be able to pay; property owners that are more satisfied with local public goods. Collection strategy takes a value of 1 if the collector uses the specific strategy all the time or often. The randomization inference-based p-value of the difference between treatment and control is reported in brackets. See Section 5 and Table A5 for details.

and taxpayer awareness. Moreover, the collector can infer propensity to pay based on the property owner's actions during repeated interactions (e.g. willingness to schedule a follow-up appointment may signal higher propensity). These channels are consistent with the literature on social learning, where an individual's discovery of another person's initially non-visible attributes is often modeled as a choice to pay the time-cost to have longer and repeated interactions with them (for example, Chakraborty et al., 2024). Finally, the collector can spend time 'surveying' the property and area to notice less obvious clues and talk to neighbors and community members. Although a household may understandably be reluctant to reveal its type directly to an official, collectors can learn in multiple ways based on interacting with the household and the environment.[38]

Panel B of Table 3 shows that treatment collectors conduct more visits and spend more time in total interacting with households. Panel B also shows that the duration

[38]This form of learning by interacting takes place prior to collecting and would be referred to as 'active learning' in the literature (Thompson, 2010). In principle, learning could also occur 'passively' if the collector, through repeated attempts to collect, uncovers the spatial clustering of types and forms beliefs about the household's type based on the immediate area's payment rate. However, we find that the spatial auto-correlation of types in our sample is very small (Moran's I is 0.025). See also online appendix (link).

Table 3: Collectors' Time Allocations and Interactions with Households

Panel A: Collector Surveys	Total hours worked (1)	Hours per bill delivered (2)	Hours on non-delivery activities (3)	Share of time on non-delivery (% total time) (4)
1(Technology)	-0.744	-0.772***	9.167**	0.501*
	(1.848)	(0.198)	(3.979)	(0.273)
	[0.632]	[0.001]	[0.062]	[0.091]
Mean in CG	19.057	1.230	3.601	0.188
Observations	141	141	141	141

Panel B: Household Survey	**1**(Any visit) (5)	Total # of visits (6)	Total time spent on visits (7)	Time spent per visit (8)
1(Technology)	0.088***	0.325**	0.622**	0.283*
	(0.028)	(0.144)	(0.298)	(0.155)
Mean in CG	0.549	0.876	0.357	0.121
Observations	4334	4334	4334	2570

Notes: This table shows the impacts of technology on collectors' time allocations and interactions with households, based on estimating (1) in the sample of collector surveys in Panel A and based on estimating (2) in the sample of household surveys in Panel B. Panel A provides treatment effects on: total weekly hours worked; hours per bill delivered; total weekly hours spent on non-delivery activities; and, the share of total weekly hours devoted to non-delivery activities. The randomization inference based p-value is reported in brackets. Panel B provides treatment effects for: a dummy for any visit from a collector; total number of visits from a collector; total time spent interacting with a collector; average time spent per interaction with a collector. The final outcome conditions the sample on having received any visit from a collector. For details on the outcomes, see Section 5 and Table A5. * p<0.10 ** p<0.05 *** p<0.01.

of each visit lasts longer in treatment areas, as there is a significant and positive treatment effect on average time spent per visit. This last result should be interpreted with caution, as it conditions on the endogenous variable of any visit; it is supported by the observation that the treatment effect on total time spent interacting with a household is disproportionately larger than the treatment effect on total visits.

Targeted interactions In addition to being of longer duration and more frequent, the interactions conducted by treatment collectors are also more targeted. To study targeting, we use the 'hard to observe' index which captures the fixed household characteristics of propensity to pay (income, liquidity, and tax awareness; see Section 5.1). Table 4

Table 4: Targeted Interactions

	Total # of visits		Total time spent on visits		Time spent per visit		**1**(Any payment)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
'Hard to observe' index	0.037	0.211***	0.052	0.191***	-0.01	-0.036	0.054**	0.155***
	(0.053)	(0.052)	(0.057)	(0.051)	(0.026)	(0.045)	(0.024)	(0.021)
Collector-unit FE	X	X	X	X	X	X	X	X
Sample	CG	TG	CG	TG	CG	TG	CG	TG
Observations	2163	2171	2163	2171	1187	1382	2163	2171
Clusters	28	28	28	28	28	28	28	28

Notes: This table shows associations between the hard to observe index and measures of interactions. Each column is based on regressing a proxy for interactions on the index, the household controls from (2) and collector-unit fixed effects. In odd (even) columns, the regression is estimated based on the household surveys in control (treatment) areas. Standard errors are clustered at the collector-unit level. See Section 5.1-5.3 for details. * p<0.10 ** p<0.05 *** p<0.01.

provides simple associations between measures of interactions and the hard to observe index in treatment and control areas. Treatment collectors are much more likely to conduct interactions and spend more time in total with households that have higher propensity to pay (columns 2 and 4); the associations are also positive in control areas but statistically insignificant and much smaller in magnitude (columns 1 and 3).

To precisely measure the treatment effect on targeted interactions, we estimate:

$$interaction_{hc} = \theta \cdot index_h + \beta \cdot [index_h \cdot \mathbf{1}(Tech)_c] + \Omega \cdot X_{hc} + \mu_c + \epsilon_{hc} \tag{3}$$

where $interaction_{hc}$ is the interaction between collector c and household h, and $index_h$ is the 'hard to observe' index of propensity to pay. The treatment coefficient β shows how the difference in propensity to pay between households that are visited more versus less frequently by collectors causally changes in treatment versus control areas; any non-zero β indicates that technology causes targeting in the characteristics of the households that the collector decides to interact more with. We include collector-unit fixed effects (μ_c) to focus on differences in characteristics between more versus less targeted households within each collection area. Standard errors are clustered by collector-unit.

Figure 4 plots β from (3). Treatment collectors are more likely than control collectors to have more interactions with high propensity households, and the differential targeting is significant. Panel B of Figure A8 shows that, qualitatively, there is positive differential targeting on each of the components of the propensity to pay index, though the magnitude is quantitatively larger for income and liquidity than for taxpayer awareness.

Figure 4 also shows that households with higher propensity to pay are more likely to make a payment in treatment than in control areas.[39] Importantly, it is not that households with higher propensity are unlikely to pay in control areas: Table 4 shows that if a control collector happens to interact with a high propensity household, the payment likelihood is significantly higher than for a low propensity household. Instead, the differential payment rate in Figure 4 reflects the fact that treatment collectors are more likely to interact with higher propensity households than control collectors.

Learning The differential targeting of interactions is consistent with learning, where the treatment collectors devote more time to learn about all households' propensity to pay, and subsequently allocate their follow-up interactions to those households with higher propensity to pay.[40] Additional results support the learning interpretation.

First, collector surveys show an increase over time in treatment collectors' self-reported knowledge about households' propensity to pay. Panel A of Figure 3 shows that, initially, there were no differences across groups in how well collectors knew which households had higher propensity to pay.[41] Over time, a positive knowledge gap opens up, which is statistically significant at 5% by the end of the campaign.[42] Table 5 provides regression results. In column 2, we leverage the panel nature of the collector surveys and include collector fixed effects, finding an even stronger treatment effect on knowledge. As the fixed effects isolate the treatment impact that varies within collector over time (relative to any initial impact at the beginning of the experiment), this result is consistent with treatment collectors discovering the hard to observe propensity by learning through interactions over the course of the campaign.

Second, a challenge with (3) is that any non-zero β reflects both a difference in the total number of interactions and in the composition of interactions between groups,

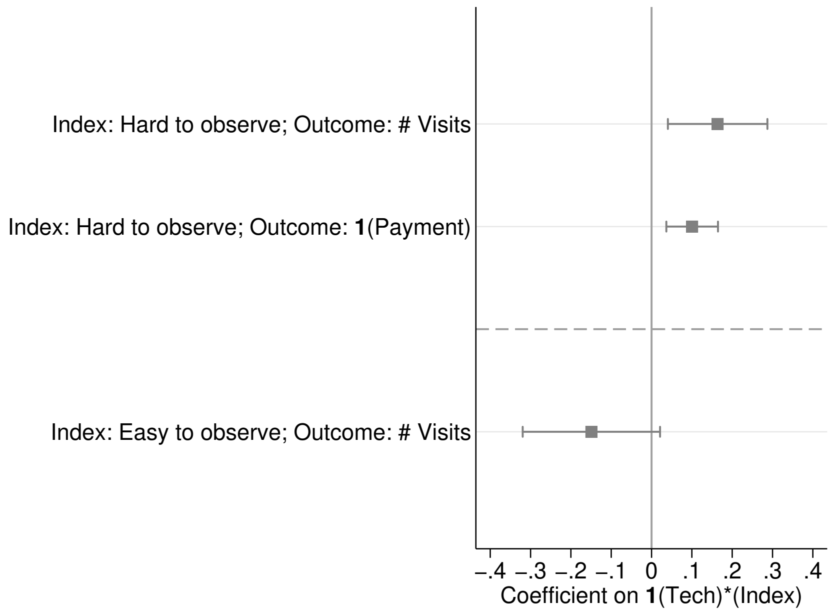
[39] This also holds with a dummy for full payment, payment amount (GHC), payment as % of tax due.

[40] The differential targeting holds with other measures of interactions (Panel A, Figure A8): total interaction time; bill delivered; any follow up visit, conditional on a first visit. The last two results suggest that learning occurs during both the first interaction and follow up interactions. Since it is costly time-wise to follow up on any delivered bill, collectors may pay attention to less visible physical clues of propensity or interact with neighbors when attempting to deliver. Conditional on a first visit for delivery, treatment collectors may further learn through longer interactions and multiple follow up visits.

[41] 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".

[42] The knowledge gap is driven by the six-fold, statistically significant increase in knowledge in the treatment group. The control group sees a positive but limited increase in knowledge; the small increase is, nonetheless, consistent with the mild positive targeting on propensity to pay in control areas (Table 4).

Figure 4: Differential Targeting of Visits and Payments



Notes: This figure plots the estimated β and its 95% confidence interval from (3). The first row plots β for total number of visits and the hard to observe index; the second row plots β for a dummy of any tax payment and the hard to observe index; the third row plots β for total number of visits and the easy to observe index. Standard errors are clustered at the collector-unit level. For details, see Section 5.1-5.3.

while Table 3 shows that the treatment group conducts more interactions in total.[43] We therefore provide a second targeting measure, which focuses on the composition of interactions. We create a 'high type' indicator, which equals 1 for values of the hard-to-observe index above the 95^{th} percentile and 0 otherwise. Data in the control group shows that the likelihood of tax payment and amount paid hardly vary with values of the hard-to-observe index until the top percentiles where they spike (Figure A10) – the returns to learning about hard to observe characteristics may be limited to the discovery of a small set of high-types.[44] The second targeting measure is, for each collector at the end of the campaign, the share of all visits that were conducted with high-types. Figure 5 shows that treatment collectors conduct a much larger share of their visits with high-types: the treatment effect (7.3 ppt) is statistically significant and represents a 110% increase over the share of visits made to high-types in the control group (6.6%).

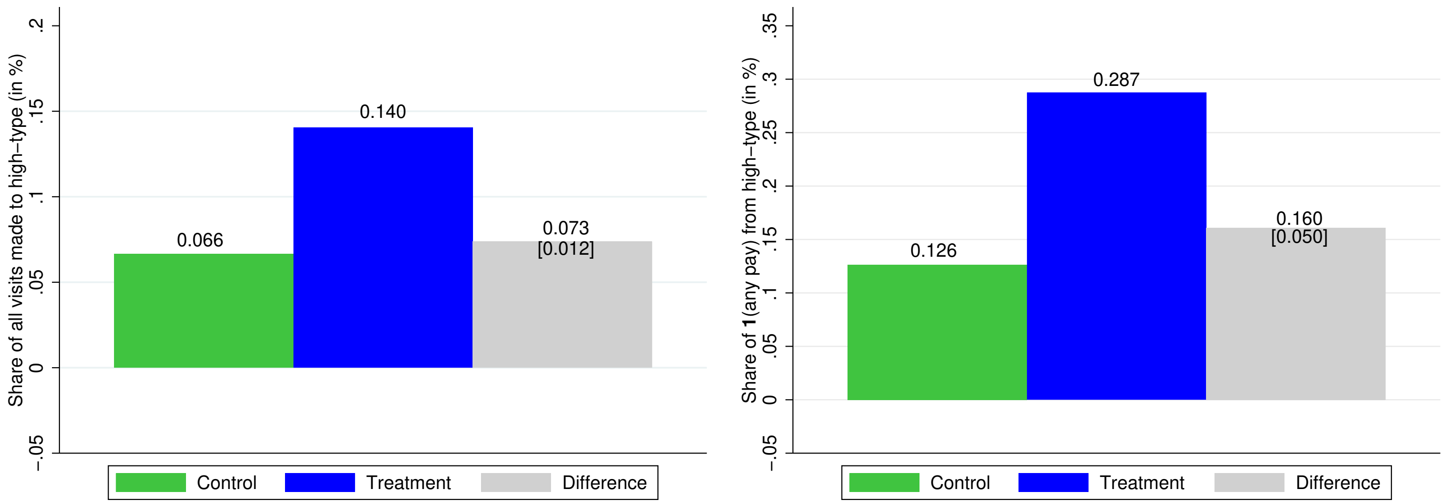
Third, additional results are inconsistent with settings where the hard to observe index of propensity is, rather than discoverable, in fact either perfectly observable or

[43]Even if 100% of control collectors' interactions were with high propensity households, β could be positive if treatment collectors conduct more visits in total with high propensity households.

[44]The differential targeting of interactions holds with this discrete formulation: see Figure A11.

Figure 5: Composition of Visits and Collected Payments

(a) Visits (b) Collections



Notes: These graphs show the impact of technology on the composition of collector visits and tax payments, based on equation (1) estimated in a single cross-section of collectors at the end of the campaign. In Panel A the outcome is the share of all visits in a collection unit conducted with households in the top 5% of the hard-to-observe index distribution. In Panel B, the outcome is the share of all payments in a unit collected from the top 5% of the index. The number in brackets is the randomization inference-based p-value on the difference between treatment and control. See Section 5.3 for details.

perfectly unobservable to all collectors throughout the campaign. We discuss this in detail in Appendix B, and provide a summary here. In the latter case, note that the positive differential targeting of interactions on propensity (Figure 4) is a priori inconsistent with propensity being unobservable. Even so, this result may reflect a correlation between propensity and the (observable) household characteristics that are targeted by collectors. However, Figure 4 shows there is *negative* differential targeting of interactions on the household's 'easy to observe' index.[45] Moreover, the differential targeting on propensity to pay is robust to controlling for the easy to observe index (Figure A9).[46]

In the former case, since treatment collectors have more available time for interactions, the positive differential targeting could arise if extracting a payment from a high

[45]See Panel C of Figure A8 for the components of the easy-to-observe index (tax liability; previous taxes paid; observable assets; distance). Panel D shows no differential targeting on additional observables: number of floors; house quality; registration status. Registration can be inferred from the tax bill.

[46]Figure A9 also shows that the negative differential targeting of the easily observable index is robust to controlling for the hard to observe index. This rules out a setting where treatment and control collectors apply the same targeting strategy (presumably of easily observable characteristics) up to the number of bills delivered in the control group, and treatment collectors reserve the search for households with hard-to-observe characteristics to the *marginal* bills delivered. Rather, this result suggests that treatment collectors apply a different strategy from control collectors to their full set of delivered bills.

propensity type requires a much larger total amount of interaction time than for a low type. For example, it may require a time-cost to travel to all the high-type households that are located far away from the collector's starting point; however, we find no differential targeting of interactions for the household's distance to main roads, markets and the local government office (Figure A8). Alternatively, trust, morale or enforcement perceptions may only be activated for high propensity households after a large amount of time spent interacting with them. However, there are no treatment effects on morale, trust and enforcement perceptions in general (Table 2) or by level of propensity to pay (Table A8). Table 4 also shows that neither treatment nor control collectors spend more time per visit with higher propensity households.[47] Finally, while treatment collectors conduct more visits than control collectors with households in the top 5% of the propensity index, they also conduct *fewer* total visits than control collectors with households in the bottom 5% (Figure A11). If propensity to pay was perfectly observable and control collectors could not afford the time-cost to visit the top 5%, it is not clear why they would devote more visits to the lowest 5% rather than the middle 90% with higher propensity.[48]

Collector strategies Technology has impacts on collectors' reported strategies. Panel B of Figure 3 shows that, over time and relative to the control group, treatment collectors report making increasingly use of strategies to visit households where they have identified higher propensity to pay, by: going to areas on specific days where they know property owners are more likely to be able to pay; going to properties where they know taxpayers are aware of their duty to pay; and, going to properties where owners are more satisfied with public services and willing to pay.[49] Table 5 shows regression results. The impacts hold with collector fixed effects, revealing a time-varying treatment effect on hard-to-observe strategies within collector over time. These strategic changes are likely driven by the accumulated knowledge on propensity to pay, which informs decisions on whom to target for collection. The treatment group also makes disproportionately more use of hard to observe strategies relative to easy to observe strategies.[50]

Due to the increased collection focus on high income households in treatment areas, Figure A7 shows that the tax system with technology becomes more progressive.

Motivated by these results, the model provides an explanation for why the navigational technology caused treatment collectors to change their time allocations and learn.

[47]This negative association is consistent with our set-up for learning – see Section 6 and Appendix B.

[48]This result is consistent with learning, where treatment collectors spend more time per visit to discover any household's type, and subsequently allocate less visits to the discovered low-type.

[49]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.

[50]The hard (easy) to observe collection strategies focus on the characteristics that make up the hard (easy) to observe index (Section 5.1). See also Table A6.

Table 5: Impacts of Technology on Collector Knowledge and Strategies

	Knowledge of hard-to-observe household characteristics		Focus on collections, hard-to-observe household characteristics		Focus on collections, easy-to-observe household characteristics		Difference in strategies: Hard versus easy to observe characteristics	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Average Effect								
1(Technology)	0.187**	0.610**	0.140*	0.667***	0.088	0.445**	0.051	0.223**
	(0.091)	(0.252)	(0.078)	(0.252)	(0.058)	(0.218)	(0.045)	(0.093)
	[0.030]	[0.003]	[0.070]	[0.000]	[0.068]	[0.049]	[0.238]	[0.001]
Panel B: Dynamic Effects								
1(Technology) × **1**(Round 1)	-0.001	–	-0.082	–	0.008	–	-0.091	–
	(0.131)	–	(0.104)	–	(0.069)	–	(0.079)	–
1(Technology) × **1**(Round 2)	0.211*	0.273*	0.264**	0.364**	0.123	0.137	0.140**	0.227**
	(0.121)	(0.156)	(0.126)	(0.153)	(0.093)	(0.101)	(0.066)	(0.095)
1(Technology) × **1**(Round 3)	0.369***	0.401**	0.253**	0.349**	0.138	0.130	0.114*	0.219**
	(0.135)	(0.167)	(0.119)	(0.151)	(0.089)	(0.106)	(0.064)	(0.093)
Collector-unit controls	X		X		X		X	
Survey round FE	X	X	X	X	X	X	X	X
Collector-unit FE		X		X		X		X
Mean in CG	0.195	0.195	0.280	0.280	0.239	0.239	0.041	0.041
Observations	141	141	141	141	141	141	141	141

Notes: This table presents the impacts of technology on collector knowledge and strategies, based on equation (1). In columns (1)-(2), 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 columns (3)-(4), 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 (taxpayer awareness, ability to pay, satisfaction with public goods). In columns (5)-(6), 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 (7)-(8), the outcome is the difference between the reliance on hard-to-observe versus easy-to-observe strategies. For details on the outcomes, see Section 5 and Table A5. Standard errors clustered at the collector-unit are reported in parentheses. In Panel A, the randomization inference based p-value is reported in brackets. * p<0.10 ** p<0.05 *** p<0.01.

6 Model

In this section, we formalize our theory of why a technology designed to assist in navigation for the main purpose of bill delivery ended up having a substantially larger effect on collections than on deliveries. We do so in a dynamic Beckerian model of time use by forward looking revenue collectors. We then estimate the model to match moments of the experimental data, evaluate its predictions for non-targeted moments, and illustrate our theory by simulating counterfactual scenarios.

6.1 Model Environment

Collectors are endowed with one unit of time each period, and the campaign lasts a total of T periods. Collectors spend their time delivering bills, learning about households to which they've delivered bills, and trying to collect payments. Each collector is endowed with a large number of bills in the initial period, and every bill has a face value of one local currency unit. Neither the exact initial number of bills nor variation in value across bills plays an important role in the experiment, so we abstract from these in the model.

There are treatment (T) and control (C) collectors. For each unit of time devoted to delivery, treatment collectors distribute θ_T bills, and control collectors distribute θ_C bills. The delivery advantage takes the form $\theta_T \geq \theta_C$, motivated by technology's reduction in navigational challenges (Figure 2) and time spent to deliver a bill (Table 3).

Households come in two types: "high" and "low," referring to their propensity to pay the tax. A fraction μ of households are high types, and μ is known to the collectors. Collectors do not know which type each household is at the outset of the campaign. To discover a household's type, we assume that collectors must spend additional time learning about them after their bill has been delivered. This learning time captures the follow-up visits needed to interact with the household and become informed about their propensity to pay (Section 5.1-5.3). Our discrete household type formulation is motivated by the finding that returns to learning about hard to observe characteristics appear to be limited to the discovery of a small set of high-types (Figure A10).

For each unit of time devoted to learning, treatment collectors have a probability η_T of discovering the household's true type, and control collectors have learning probability η_C. We assume that $\eta_T \geq \eta_C$. The learning advantage comes from the technology's navigational enhancement, which improves the collector's ability to locate households for the follow-up interactions that are required to learn. Consistent with this premise, in our benchmark we fix the learning advantage to be equal to the delivery advantage. This modeling choice is also motivated by the mechanism results which suggested that technology did not provide direct advantages for non-delivery activities, conditional on delivery (Section 4). Conditional on learning the household's type, the collector learns that the household is a high type with probability μ and a low type with probability $1-\mu$. Modeling learning as a time-cost choice is consistent with the social learning literature (Chakraborty et al., 2024); it allows for the possibility that treatment collectors spend more or less time learning about household types than the control group. It is not obvious ex ante whether technology increases or decreases time spent learning.

The collection technology in the model is exactly the same for treatment and control collectors. Each period, collectors can devote time to collecting from households to whom they have delivered a bill. Collectors can target their time toward households whose type is unknown, households known to be high types, and households known to be low types. For each unit of time spent trying to collect from high type households, the probability of collection is π_h. The collection probability per unit of time spent collecting from low type households is π_ℓ, where $\pi_h > \pi_\ell$. We let $\pi = \mu\pi_h + (1-\mu)(1-\pi_\ell)$ denote the probability of collection from unknown types. To be clear, when the collector faces a high type the likelihood of collecting a payment is π_h regardless of whether the collector knows the household's type; π reflects the ex ante probability of collecting from a household whose type is unknown to the collector.

Our assumption that the technology does not offer direct advantages in collection is based on two observations. The first is that the GIS-technology we randomized did not offer any direct or specific assistance in collecting from households once their bills had been delivered. The second is that the experiment showed a null treatment impact on household tax morale and enforcement perceptions, suggesting that households did not change their payment propensities directly by being visited by a collector with the tablet (Section 4). Still, the assumption of no collection advantages may be restrictive in the sense that the same navigational advantages offered by the technology may well increase the return to time spent trying to collect as well. Our approach is to be as conservative as possible, and to see whether the model can generate a much larger treatment effect on collection than on delivery without any direct advantage in collections.

A collector's state variables each period are: b, the number of bills delivered to households of unknown type; b_h, the number of bills delivered to known high-types; and b_ℓ, the number of bills delivered to known low-types. Their choice variables are time spent distributing bills, d, time spent learning about household types, x, and time spent trying to collect from unknown, high-type and low-type households: c, c_h, and c_ℓ.

The collector's dynamic problem is to maximize total tax collections:

$$V(b_h, b, b_\ell, t) = \max_{\{d,x,c_h,c,c_\ell\}} \{c\pi + c_h\pi_h + c_\ell\pi_\ell + V(b_h', b', b_\ell', t+1)\}$$

subject to:

$d + x + c + c_h + c_\ell \leq 1$ (the time constraint),

$c \leq b$, $c_h \leq b_h$, $c_\ell \leq b_\ell$ (the collection constraints),

$b' = b + \theta_j d - x\eta_j - c\pi$ (the law of motion for unknown-type bills delivered),

$b'_h = b_h + x\mu\eta_j - c_h\pi_h$ (the law of motion for high-type bills delivered), and

$b'_\ell = b_\ell + x(1-\mu)\eta_j - c_\ell\pi_\ell$ (the law of motion for low-type bills delivered).

The time constraint requires that the total time spent on deliveries, learning, and collection does not exceed the time endowment. The collection constraints ensure that collectors cannot collect from more than their current stock of bills. The law of motion for unknown-type bills govern how the stock evolves given inflows of new bills delivered and outflows of bills for which the collector learns the true type or collects upon. The laws of motion for high-type and low-type bills feature inflows from learning and outflows from collection. There is also a terminal period – omitted for brevity – in which collectors use all their time endowment trying to collect from their current stock of bills.

The dynamic tradeoffs for a collector can be summarized as follows. The benefit of time spent on deliveries is that it increases the stock of unknown-type delivered bills that can later be collected upon. The opportunity cost of additional delivery time is not collecting from the current stock of bills already delivered. Learning time also carries this opportunity cost. But it can help identify which bills are the high-type.

Conditional on time being allocated to collections, collectors have to decide how much time to devote to collecting from unknown-, high-, and low-types. It is easy to see that the optimal collection strategy involves first trying to collect from the high-type households, since they have the highest chances of paying. Any remaining time gets spent trying to collect from the unknown types, followed by the low types, which have the lowest propensity to pay. One can see then that learning time is valuable because it can help favorably shape the stock of bills, allowing collectors to focus more on households with higher propensities to pay.

6.2 Model Estimation

We estimate the model assuming that each period represents two days, meaning that $T = 21$. This choice allows the model to capture the rich dynamics of the collectors' problem while still allowing us to estimate the model in a reasonable amount of time. We set θ_C, the delivery parameter in the control group, to be 0.05. This choice amounts to a normalization on the units of bills in the model, and has no substantive impact on any of our results. We set the fraction of high types, μ, to be 0.05, consistent with the empirical observation that the higher likelihood of payment is concentrated amongst the

Table 6: Model Estimation

Panel A: Moments Targeted in the Estimation

Moments	Target	Model
Treatment Effect on Bills Delivered (%)	27.0	27.3
Ratio of Learning Time to Collection Time (Average)	15.0	15.8
Probability of Full Payment | Delivery (Average)	13.7	13.6
Fraction of Collections from High Type (Average)	21.9	22.1

Panel B: Estimated Parameters and Confidence Intervals

θ_T	η_C	π_ℓ	π_h
0.104	0.383	0.008	0.032
(0.077, 0.206)	(0.313, 0.501)	(0.007, 0.012)	(0.013, 0.046)

Note: Panel A reports the moments targeted in the estimation and their values in the data and in the model. Panel B reports the estimated parameter values and their bootstrapped 95-percent confidence intervals.

households with the highest 5 percent of the hard-to-observe index (Figure A10).

We then estimate the model to match four target moments, which we list in Table 6, Panel A. The first is the treatment effect on deliveries of 27 percent. This is the only experimental outcome we target directly. The remaining moments represent averages across all collectors in the experiment. These are: a ratio of learning time to collection time of 15 percent; the probability of a collector getting a full payment during the experiment conditional on bill delivery, which is only 13.7 percent; and, the fraction of collections coming from the high type, which is 21.9 percent. The learning time target is not directly observed, but consistent with a small but positive amount of learning on average (Figure 3). We have experimented with alternative targets but find that our results are not particularly sensitive to other values. The probability of getting a full payment is observed directly in our data, and implies low overall payment probabilities by the households. The fraction of collections coming from the high-type households is much larger than the fraction of households that are high types, which the model will attribute to collector learning plus a higher payment probability by the high type.

Our estimation strategy formally solves for the parameter vector $\{\theta_T, \eta_C, \pi_h, \pi_\ell\}$ that minimizes the sum of squared differences between these four moments and their counterparts in the model. We impose $\eta_T/\eta_C = \theta_T/\theta_C$, which implies that the treatment group's learning advantage is proportional to its delivery advantage (explained above).

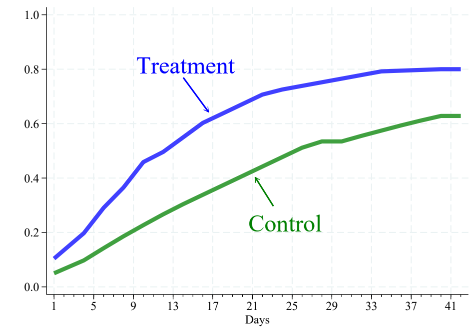
The estimated parameter values are reported in Table 6, Panel B, along with bootstrapped 95 percent confidence intervals. The value of θ_T is 0.104 implying that bill delivery probability per unit of time is just over twice as high with the technology as without it. This is substantially larger than what one might naively assume about technology's delivery advantage based on the experiment's 27 percent treatment effect on deliveries. The intuition is that collectors in the model respond to the technology by substantially reducing their time spent on deliveries, consistent with the empirical evidence in Section 5. At the same time, the estimate of θ_T is consistent with the experimental evidence that treatment collectors spend less than half as much time delivering each bill than control collectors (Table 3). Though the confidence interval for θ_T is fairly wide (stemming from the confidence interval around the delivery treatment effect), it does not include the value of return on time spent delivering in the control group $\theta_C = 0.05$.

The estimated value of η_C is 0.383, and implies that a control collector spending all of their time in a period trying to learn about the household types for a given set of delivered bills results in them learning about just under two in five of them. The resulting value of η_T is 0.801, meaning a treatment collector trying to learn about unknown type bills will do so for around four in five. The estimated values of π_ℓ and π_h are 0.008 and 0.032. Overall, these low values imply that efforts to collect from households for whom a bill has been delivered are quite unlikely to lead to a payment. This has to be the case in order for the model to match the observed overall low payment rate of 13.7 percent conditional on delivery. The confidence intervals for the two probabilities do not overlap, which indicates that the estimated model features two distinct household types. The probability of a payment conditional on a per unit of collection time is about four times as high for high-types as low-types. This highlights the potential returns to time spent learning in the model: if a collector knows the households' types, they can focus on the high-types and enjoy a substantially higher chance of collecting payment.

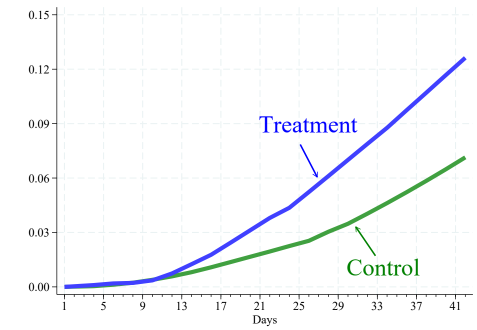
6.3 Quantitative Predictions of Model

The top two panels of Figure 6 plot the model's predicted deliveries and collections by group and by day. The treatment effect on deliveries of 27 percent, which we target in the estimation, is represented by the difference in deliveries between the two groups on the last day of the campaign. Before that, the model matches well the concave pattern of deliveries in the experiment and the peak effect on deliveries in the middle of the campaign (Figure 1). The model also matches well the convex pattern of collections in

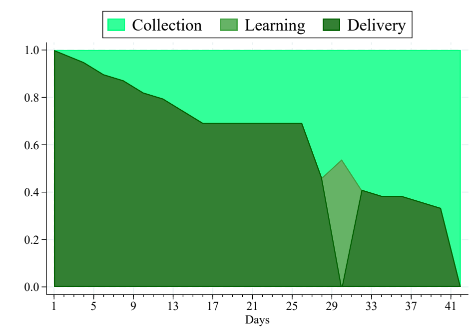
Figure 6: Predictions of Estimated Model



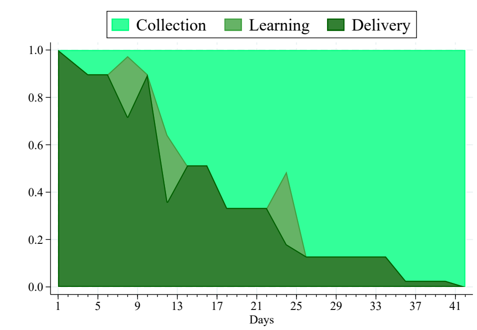
(a) Bills delivered



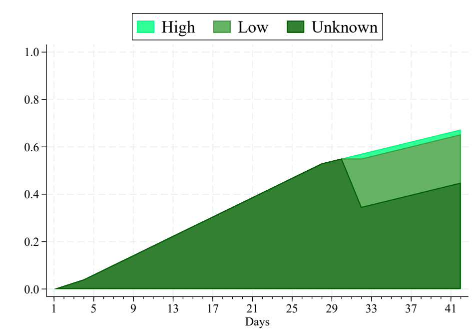
(b) Revenue Collections



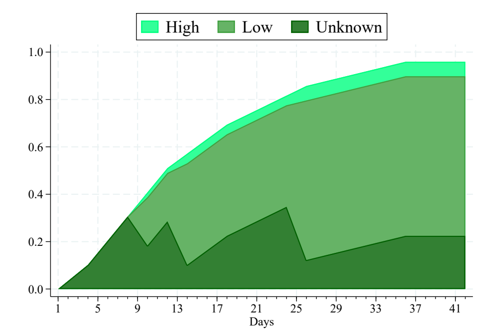
(c) Time Allocation (Control)



(d) Time Allocation (Treatment)



(e) Composition of Bills Delivered (Control)



(f) Composition of Bills Delivered (Treatment)

the experiment. By the end of the campaign, the model predicts – without targeting – a 77 percent increase in collections, which is around three times as high as the predicted effect on deliveries (compared to four times in the experiment). In differences, the model predicts treatment effects that are 50 percentage points higher for collections than for deliveries, compared to 76 percentage points in the experiment, meaning that the model can explain around two-thirds of the experimental difference (50/76).

The middle two panels of Figure 6 plot the allocations of time for the control and treatment collectors in the model, which are useful for understanding how the behavior of the two groups differs. In both groups, deliveries (the dark shaded area) starts out as the main activity and declines during the campaign. Yet delivery time declines more quickly in the treatment group, meaning an earlier and steeper drop in the dark shaded region. This highlights a key behavioral change resulting from technology. Because the treatment collectors can deliver more bills in the same amount of time, they choose to spend less time delivering. What do the treatment collectors do instead of bill deliveries? As the panels show, they spend more time collecting (lightest shaded area) and also more time learning (medium shade areas). The control group concentrates the little amount of learning it chooses to do late in the campaign, around day 30, whereas the treatment group learns earlier, starting around day 7, and more often throughout the campaign.

The bottom panels of Figure 6 plot the stocks of bills by type in each period in the control and treatment groups. In both groups, the stock of bills delivered grows over time and initially consists only of unknown types. In the control group, once the collectors engage in learning, the stock of unknown-type bills drops and are replaced with known low- and high-types. In the treatment group, the earlier and more frequent learning results in a much greater share of known types in the stock of bills delivered. At the end of the campaign, treatment collectors know the type of 61 percent of households in their unit, while control collectors have only learned the type for 19 percent.[51] This mimics our experimental results, where treatment collectors are far more knowledgeable about their assigned households' types by the end of the campaign (Figure 3).

The last two columns of Table 7 report the model's predictions for the fraction of collected payments from high-type households. This corresponds to the second measure of targeting from the experiment (Section 5). Panel B of of Figure 5 shows that, in the experiment, only 12.6 percent of collections in the control group came from the high-types, compared to 28.7 percent in the treatment group (a 127% increase). The first row of Table 7 reproduces this experimental result of positive differential targeting.

The estimation strategy targets the fraction of payments from high-types on average but not by group, and can thus serve as a validation exercise on our model. As in the experiment, our model generates positive differential targeting, with a larger fraction of collections coming from high-types in the treatment group (26 percent) than in the control group (19 percent). In the model, this positive differential targeting arises from the knowledge gained by treatment collectors over the course of the campaign.

[51] We calculate this by assuming that the measure of bills delivered in each group by the end of the model corresponds to the actual average number of bills delivered in each group by the end of the experiment (Figure 1). This is reasonable as the model targets the experimental effect on delivery.

Table 7: Model Predictions versus Experimental Data

	Treatment Effects (%)		Collections, High Type (%)	
	Deliveries	Collections	Control	Treatment
Experimental data	27	103	13	29
Estimated main model	27	77	19	26
No re-optimization	52	46	19	18
Easier collection (π_h, π_ℓ x 10)	58	76	15	23
No learning advantage ($\eta_T = \eta_C$)	33	64	19	19
No delivery advantage ($\theta_T = \theta_C$)	-2	8	19	24

Note: This table reports the treatment effects on deliveries and collections in the data and in the model, and the fraction of collections that comes from the high type in the control and treatment groups.

6.4 Insights Based on Counterfactual Simulations

Direct and indirect effects of technology One value of the model is to quantify the direct effects of technology, stemming from navigational improvements, while putting aside any indirect effects stemming from changes in collector behavior in response to technology. To do so, we conduct a counterfactual exercise in which the treatment group has the same advantages in delivery and learning as in the main estimation, but which keeps the treatment's groups *decision rules* the same as the control group.[52] In other words, this counterfactual simulates an "envelope theorem" type scenario where the technology is introduced but collector re-optimization is ignored. The third row of Table 7 shows that this counterfactual predicts treatment effects on deliveries and collections of similar magnitude (52 percent and 46 percent). Since the treatment effect on collection in the main estimation is 77 percent, we can conclude that collector re-optimization in response to the technology amplifies the direct effects of the technology by 67 percent (77/46). Similarly, collector re-optimization scales down the impact on deliveries roughly in half (27/52). Thus, an envelope theorem logic, where the technology simply scales up the performance of control collectors, does not give an accurate representation of the technology's actual experimental impacts. This model insight is consistent with additional experimental results which show that the treatment group's activities differ

[52]The dynamic predictions for all counterfactual exercises are reported in the online appendix (link).

from a ’scaled up version’ of the control collector’s path of activities (Figure A5).[53]

In this counterfactual, despite having a learning advantage, the treatment collector only learns the type for 22 percent of households - a muted knowledge impact compared to the control collector (19 percent), and a much smaller impact compared to the benchmark treatment collector (61 percent). This result reflects the fact that learning is not a fundamental activity in the field to begin with, given the control collector’s high opportunity cost of learning imposed by the challenging delivery – hence a direct, technology-induced enhancement of the return to time spent learning has little impact on actual learning. Table 7 shows that, in this counterfactual without a significant difference in knowledge between groups, there is also no significant differential targeting.

To help understand why the indirect effects of technology are so important, we simulate another counterfactual scenario in which we relax the collection constraint by making collections far easier than they are in reality. We do so by multiplying the collection probabilities, π_h and π_ℓ, by 10. We then introduce the technology as in the main model, and simulate its impact. This time, the treatment effect on deliveries is much higher, at 58 percent, and the treatment effect on collections is similar to the main model, at 76 percent (fourth row of Table 7). Collector time use is much more focused on deliveries than in the main model. The reason is that collections are now not so challenging as before, so the collectors respond to the better delivery technology by continuing to focus on deliveries. The lesson from this counterfactual is that it is not an artifact of the main estimated model, but a result, that collectors respond to the new technology by shifting their allocation of time toward other challenging activities.

A related insight is that, in this counterfactual with stronger enforcement, collectors choose to learn very little: treatment and control collectors end up knowing only about 18 percent and 9 percent of households’ types, respectively.[54] This result highlights how it is particularly in settings where broad and strong enforcement is limited, such as local taxation in Ghana (Section 2), that learning about households’ types is a relevant strategy, given the subsequent targeting that it permits for attempting to collect.

Relative importance of delivery and learning advantages To illustrate the individual importance of the delivery and learning advantages, we simulate two counterfactual scenarios in which we shut each of these advantages down one at a time. In the “no

[53] Panel C of Figure A5 measures the number of failed attempts per successful visit and shows that treatment collectors switch over time into more demanding tasks with a higher failure rate, while control collectors maintain a similar failure rate throughout. Panel D shows that the treatment group overhauls the organization of its field activities over time while the control group effectively makes no changes.

[54] The magnitude of targeting is amplified when π is scaled upwards, so the extent of differential targeting is not directly comparable between this and the other counterfactuals.

learning advantage" counterfactual we set $\eta_T = \eta_C$, and leave all other parameters as in the benchmark estimation. Note this leaves in place the treatment group's advantage in delivering bills, but gives them no additional advantage in learning the households' type. In the "no delivery advantage" counterfactual we set $\theta_T = \theta_C$ but keep other parameters the same, leaving the treatment group with the learning advantage. The bottom two rows of Table 7 summarize the model's predicted treatment effects for deliveries and collections in these counterfactual simulations.

With no learning advantage, the treatment group now delivers 33 percent more bills and collects 64 percent more revenues. Relative to the main model, the treatment collectors choose to spend more time both delivering and collecting, but less time learning. Since the treatment effects in this counterfactual are 31 percentage points larger for collections than for deliveries, we can say that about sixty percent (31/50) of the difference predicted by the estimated main model stems from the delivery advantage. The remaining forty percent then comes from the learning advantage. This counterfactual reveals that, with time that can be re-allocated to learning but no advantage in learning, treatment collectors effectively choose to not gain more knowledge than control collectors (learning 22 and 19 percent of households' types, respectively). Reflecting the absence of a strong effect on learning, there is also no differential targeting: treatment and control collectors, having approximately the same amount of knowledge, collect the same fraction of payments from the high-type.

With no delivery advantage, the model's predictions look markedly different from the benchmark model and actual experiment. Now, the treatment effects on deliveries and collection are -2 and 8 percent, respectively. The learning advantage leads the treatment collector to switch into non-delivery faster than the control collector, causing the negative treatment effect on delivery. The learning advantage also leads the treatment collector to spend a large share of non-delivery time on learning versus collecting, causing an (infinitely) large ratio of collection to delivery treatment effects.[55]

Discussion: Learning, Substitutes for Technology When combined, the counterfactual exercises reveal how the different parts of the model interact for learning to emerge as a quantitatively important activity – whereby the collector learns the household type for a significant share of properties in their area. Specifically, learning emerges in our setting in two sequential steps: by easing the delivery margin, technology first causes a re-optimization of time-allocations which can potentially be devoted to learning; by en-

[55]Due to the initial challenges in delivery, this treatment collector spends both more total time on delivery and less total time on learning than the benchmark treatment collector. Combined with the mechanically smaller pool of households that received a bill, this causes the treatment collector to learn about 31% of households' types, compared to 61% for the benchmark treatment collector.

hancing the return to time spent learning, technology in turn ensures that some of this re-allocated time is spent on learning. Technology's direct impact, through the learning advantage, and indirect impact, through time re-allocation, are each necessary but not sufficient for learning to emerge as a significant activity in the field.[56]

In the model, technology improves the return to time spent on otherwise challenging activities. By this logic, technology may have smaller impacts among collectors that initially face less severe challenges in their work setting. For example, the delivery (θ_T) and learning (η_T) advantages of technology may be smaller for a collector with some prior local experience, relating to knowledge of households' location and propensity to pay, than for a collector with no prior experience. Results from the experiment in Madina support this hypothesis. Panels A and B of Figure A12 show that technology has larger treatment effects on interactions and collections among collectors with no prior work experience in Madina; among collectors with prior experience, who also report less challenges at baseline, the positive impacts of technology remain significant, but smaller.[57] In fact, Panel C of Figure A12 shows that the positive, non-identified difference in performance between experienced and non-experienced collectors in the control group is similar in magnitude to the positive, identified difference in performance between treatment and control collectors in the inexperienced group. In other words, technology and collectors' pre-existing experience in the area, including local information on localization and propensity, appear to be substitutes in our setting.[58]

These results have policy implications in developing countries. Almost all governments require tax officials to rotate across areas (World Bank, 2019), in order to minimize familiarity and adverse behavior. However, doing so also limits officials' ability to learn

[56]In the model we assumed that the learning advantage is proportional to the delivery advantage, based on the idea that the navigational improvement helps to locate a property both for delivery and for follow-up visits. Additional potential channels exist through which technology may create a favorable environment that enhances the return to time spent trying to learn. In particular, motivated by findings in environmental psychology (Munzer et al., 2006), the specific features of the GIS-tablet, in contrast to the poor spatial information available to control collectors, may encourage treatment collectors to make the effortful decision to acquire survey knowledge – an understanding of the true spatial map, that increases the collector's ability to engage with the environment and pay attention to less visible details (Chuanxiuyue and Hegarty, 2020). Investigating the role of survey knowledge is an area for future research; we also refer to the online appendix (link) for more details.

[57]35% and 0% of collectors with and without prior experience, respectively, report at baseline having a good sense of which households have a high propensity to pay. Collectors with prior experience in Madina are also much less likely than those without to report challenges in locating property owners. Prior experience as a source of treatment effect heterogeneity was specified in the pre-analysis plan.

[58]Table A1 shows that GIS-technology and addressing coverage are substitutes at the level of local governments across the country's 216 districts. This result is consistent with the observation that property tax registries are not systematically associated with the use of GIS technologies in high capacity settings where broad addressing is sufficient to support tax collection (Knebelmann, 2022).

and build locally relevant knowledge for tax collection. Our results show that technology can counter some of the challenges that are compounded by rotation policies.[59]

7 Conclusion

This paper studies the role of technology in improving tax capacity, by focusing on the local property tax in Ghana. The experiment randomized the presence of an electronic GIS-tablet at the level of a tax collector. The technology was designed to help collectors locate property owners and deliver tax bills – an otherwise challenging task in an environment with incomplete addressing, as is common in many developing countries. Technology caused bill deliveries to increase significantly, as intended. Interestingly, however, technology caused tax collections to increase by a disproportionately larger amount. Experimental results and a dynamic model support our theory that technology, by improving the return to time spent delivering, led collectors to significantly re-optimize their time allocations across all activities, so as to focus more on other challenging activities. Specifically, treatment collectors allocated more of their scarce time to the challenging tasks of learning about taxpayers and attempting to collect from them, particularly those with the highest propensity to pay.

Our results show how the state can directly build locally relevant information. The information gathered by tax collectors may be useful for local officials that implement targeted transfer programs. Future research could explore how to effectively transfer this newly built information among state agents, which would enable states to revisit their reliance on non-state actors to target transfers (Basurto, Dupas, and Robinson, 2020).

Technology improved the collector's ability to locate taxpayers in a setting with scant addressing, but did not otherwise automate any tasks or directly provide other technical or organizational enhancements in the field. It is in this sense that we interpret our results as providing a first empirical step toward understanding what is the value of an address for government performance. More work is needed to establish the longer run impacts of improved localization, including the persistence of collectors' learning and how citizens adapt to increased legibility (Okunogbe, 2021; Scott, 1998).

Investments in GIS-technologies for taxation are limited but growing in Africa and other parts of the world (Knebelmann, 2022). In the context of this increasing investment rate, more work is needed to help ensure that the taxes collected with technology adopted at scale are ultimately used to fund beneficial local public goods.

[59] At the same time, technology also improves collectors' ability to extract bribes (Table 2). More work is needed to understand how technology can be re-designed to maintain the positive improvements in the collector's work environment while minimizing the risk of adverse, rent-seeking behavior.

References

Abebrese, K. (2019): "Implementing Street Addressing Systems in an Evolving Urban Center: A Case Study of the Kumasi Metropolitan Area in Ghana," Working Paper.

Aman-Rana, S. and C. Minaudier (2024): "Spillovers in State Capacity Building: Evidence from the Digitization of Land Records in Pakistan," *Working Paper*.

Atkin, D., A. Chaudhry, S. Chaudry, A. Khandelwal, and E. Verhoogen (2017): "Organizational Barriers to Technology Adoption: Evidence from Soccer-Ball Producers in Pakistan," *Quarterly Journal of Economics*, 132.

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.

Bandiera, O., M. Best, A. Khan, and A. Prat (2021): "The Allocation of Authority in Organizations," *Quarterly Journal of Economics*, 135.

Banerjee, A. V., S. Chassang, S. Montero, and E. Snowberg (2020): "A Theory of Experimenters," *American Economic Review*, 110.

Basurto, M., P. Dupas, and J. Robinson (2020): "Decentralization and Efficiency of Subsidy Targeting," *Journal of Public Economics*, 185.

Bergeron, A., P. Bessone, J. Kabeya, G. Tourek, and J. Weigel (2022): "Optimal Assignment of Bureaucrats: Evidence from Randomly Assigned Tax Collectors in the DRC," 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.

Bigon, L. and A. Njoh (2012): "The Toponymic Inscription Problematic in Urban Sub-Saharan Africa," *Journal of Asian and African Studies*, 50.

Brockmeyer, A., A. Estefan, K. Arras, and J. C. Suárez Serrato (2023): "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," *World Bank Policy Research Working Paper*.

Callen, M., S. Gulzar, S. A. Hasanain, M. Y. Khan, and A. Rezaee (2020): "Data and Policy Decisions," *Journal of Development Economics*, 146.

Casaburi, L. and U. Troiano (2016): "Ghost-House Busters: The Electoral Response to a Large Anti-Tax Evasion Program," *Quarterly Journal of Economics*, 131.

Chakraborty, A., A. Ghosh, M. Lowe, and G. Nellis (2024): "Learning About Outgroups: The Impact of Broad versus Deep Interactions," *Working Paper*.

Chuanxiuyue, H. and M. Hegarty (2020): "How Anxiety and Growth Mindset Are Linked to Navigation Ability: Impacts of Exploration and GPS Use," *Journal of Environmental Psychology*.

Dal Bó, E., F. Finan, N. Li, and L. Schechter (2021): "Information Technology and Government Decentralization," *Econometrica*, 89.

D'Arcy, M., M. Nistotskaya, and O. Olsson (2024): "Cadasters and Economic Growth: A Long-Run Cross-Country Panel," *Journal of Political Economy*, 132.

Das, S., L. Gadenne, T. Nandi, and R. Warwick (2023): "Does Going Cashless Make You Tax-Rich?" *Journal of Public Economics*, 224.

Dincecco, M. and G. Katz (2016): "State Capacity and Long-Run Economic Perfor-

mance," *The Economic Journal*, 126.
Dodge, E., Y. Neggers, R. Pande, and C. Moore (2021): "Updating the State: Information Acquisition Costs and Social Protection Delivery," *Working Paper*.
Duflo, E., M. Greenstone, R. Pande, and N. Ryan (2018): "The Value of Regulatory Discretion: Estimates from Environmental Inspections in India," *Econometrica*, 86.
Eissa, N. and A. Zeitlin (2014): "Using Mobile Technologies to Increase VAT Compliance in Rwanda," Working Paper.
Fan, H., Y. Liu, N. Qian, and J. Wen (2021): "Computerizing VAT Invoices in China," *NBER Working Paper Series*, 24414.
Farvacque, C., L. Godin, H. Leroux, F. Verdet, and R. Chavez (2005): "Street Addressing and the Management of Cities," *World Bank Directions in Development*, 32923.
Ferraz, C., D. Foremny, and J. F. Santini (2024): "Revenue Slumps and Fiscal Capacity: Evidence from Brazil," *NBER Working Paper*, 32440.
Finan, F., B. Olken, and R. Pande (2017): "The Personnel Economics of the Developing State," *Handbook of Economic Field Experiments*, 6.
Franzsen, R. and W. McCluskey (2017): "Property Tax in Africa," Lincoln Institute.
Gadenne, L. (2017): "Tax Me, But Spend Wisely? Sources of Public Finance and Government Accountability," *American Economic Journal: Applied Economics*, 9.
Gennaioli, N. and H.-J. Voth (2015): "State Capacity and Military Conflict," *Review of Economic Studies*, 82, 1409–1448.
Gordon, R. and W. Li (2009): "Tax Structures in Developing Countries: Many Puzzles and a Possible Explanation," *Journal of Public Economics*, 93.
Government of Ghana (2011): "National Street Naming and Property Addressing Policy," Ghana Ministry of Local Government and Rural Development.
Hjort, J. and L. Tian (2024): "The Economic Impact of Internet Connectivity in Developing Countries," *Annual Review of Economics*, Forthcoming.
Houeix, D. (2025): "Asymmetric Information and Digital Technology Adoption: Evidence from Senegal," *Working Paper*.
Kalaj, J., D. Rogger, and R. Somani (2022): "Bureaucrat Time-Use: Evidence From a Survey Experiment," *World Development*, 152.
Khan, A., A. Khwaja, and B. Olken (2015): "Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors," *Quarterly Journal of Economics*, 131.
——— (2019): "Making Moves Matter: Experimental Evidence on Incentivizing Bureaucrats through Performance-Based Postings," *American Economic Review*, 109.
Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. Pedersen, and E. Saez (2011): "Unwilling or Unable to Cheat?" *Econometrica*, 79.
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.
Knebelmann, J. (2022): "Digitalisation of Property Taxation in Developing Countries: Recent Advances and Remaining Challenges," *ODI Report Series*.
Knebelmann, J., V. Pouliquen, and B. Sarr (2023): "Discretion versus Algorithms: Bureaucrats and Tax Equity in Senegal," Working Paper.
Lipsky, M. (2010): *Street-Level Bureaucracy*, Russell Sage Foundation.
Luttmer, E. and M. Singhal (2014): "Tax Morale," *Journal of Economic Perspectives*, 28.
Manara, M. and T. Regan (2022): "Ask a Local: Improving the Public Pricing of Land

Titles in Urban Tanzania," *Review of Economics and Statistics*.
Martinez, L. (2023): "Natural Resource Rents, Local Taxes, and Government Performance: Evidence from Colombia," *Review of Economics and Statistics*.
Marx, B., T. Stoker, and T. Suri (2013): "The Economics of Slums in the Developing World," *Journal of Economic Perspectives*, 27.
Mascagni, G., A. Mengistu, and F. Woldeyes (2021): "Can ICTs Increase Tax Compliance," *Journal of Economic Behavior and Organization*, 189.
Mattsson, M. (2023): "Information Systems, Service Delivery, and Corruption: Evidence from the Bangladesh Civil Service," *Working Paper*.
Mayshar, J., O. Moav, and L. Pascali (2022): "The Origin of the State: Land Productivity or Appropriability?" *Journal of Political Economy*, 130.
Munzer, S., H. Zimmer, M. Schwalm, J. Baus, and I. Aslan (2006): "Computer-Assisted Navigation and the Acquisition of Route and Survey Knowledge," *Journal of Environment Psychology*, 26.
Muralidharan, K., P. Niehaus, S. Sukhtankar, and J. Weaver (2021): "Improving Last-Mile Service Delivery Using Phone-Based Monitoring," *American Economic Journal: Applied Economics*, 13.
Naritomi, J. (2019): "Consumers as Tax Auditors," *American Economic Review*, 109.
Nyanga, M. (2021): "Performance Outcome Area 1 Integrity of the Registered Taxpayer Base: Essentials of a High Integrity Taxpayer Register," *TADAT Insights*.
Okunogbe, O. (2021): "Becoming Legible to the State: The Role of Detection and Enforcement Capacity in Tax Compliance," *World Bank Policy Research Working Paper Series*.
Okunogbe, O. and V. Pouliquen (2022): "Technology, Taxation and Corruption," *American Economic Journal: Economic Policy*, 14.
Okunogbe, O. and G. Tourek (2024): "How Can Lower-Income Countries Collect More Taxes ? The Role of Technology, Tax Agents, and Politics," *Journal of Economic Perspectives*, 38.
Parasuraman, R. (2000): "Designing Automation For Human Use: Empirical Studies and Quantitative Models," *Ergonomics*, 43.
Pomeranz, D. (2015): "No Taxation without Information: Deterrence and Self-enforcement in the Value Added Tax," *American Economic Review*, 105, 2539–2569.
Prescott, E. C. and S. L. Parente (1994): "Barriers to Technology Adoption and Development," *Journal of Political Economy*, 102.
Rasul, I. and D. Rogger (2018): "Management of Bureaucrats and Public Service Delivery," *Economic Journal*, 128.
Scott, J. (1998): *Seeing Like a State*, Yale University Press.
Suri, T. and C. Udry (2022): "Agricultural Technology in Africa," *Journal of Economic Perspectives*, 36.
Thompson, P. (2010): "Learning by Doing," in *Handbook of the Economics of Innovation*, Oxford University Press.
Weigel, J. L. (2020): "The Participation Dividend of Taxation," *Quarterly Journal of Economics*, 135.
World Bank (2019): "Codes of Ethics and Conduct in Revenue Administratons: What Does International Practice Tell Us," World Bank Group.
——— (2020): "Property Tax Diagnostic Manual," World Bank Group.

Supplementary Appendix

A Additional Figures and Tables

Table A1: Address Coverage, Technology and Bill Delivery

Outcome: Share of bills delivered (%)	(1)	(2)	(3)	(4)	(5)
Panel A					
Share of properties with address (%)	0.305**	0.170*	0.250**	0.236**	0.193**
	(0.120)	(0.089)	(0.107)	(0.082)	(0.061)
Panel B					
Share of properties with address (%)	0.270*	0.210*	0.263*	0.251**	0.246**
	(0.127)	(0.110)	(0.120)	(0.088)	(0.084)
1(Technology)	0.358***	0.322***	0.296**	0.195**	0.199***
	(0.095)	(0.078)	(0.107)	(0.077)	(0.041)
(Share w address) $\times$ **1**(Technology)	-0.353	-0.481**	-0.385	-0.272	-0.402**
	(0.203)	(0.163)	(0.201)	(0.180)	(0.145)
Technology impact at level of share with address					
1^{st} percentile (0%)	0.358***	0.322***	0.296**	0.195**	0.199***
	(0.095)	(0.078)	(0.107)	(0.077)	(0.041)
50^{th} percentile (20%)	0.287***	0.226***	0.219**	0.140**	0.118***
	(0.064)	(0.056)	(0.070)	(0.043)	(0.022)
99^{th} percentile (94.5%)	0.024	-0.132	-0.067	-0.062	-0.181
	(0.122)	(0.105)	(0.095)	(0.096)	(0.104)
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: This is a cross-sectional regression of all 216 districts. The outcome is the share of bills delivered, which in Panel A is regressed on the share of properties with addresses. Panel B adds as regressors **1**(Technology), a dummy variable taking a value of 1 if the local government has a GIS-tax registry and 0 otherwise, and its interaction with the address-share variable. The bottom of Panel B shows the estimated impact of technology at percentiles of the address variable: 1^{st} percentile (districts where 0% of properties have addresses), 50^{th} percentile (20%); and 99^{th} percentile (94.5%). The same controls are used as in Table A3. Standard errors, clustered at the region level, are shown in parentheses. * p<0.10 ** p<0.05 *** p<0.01.

Table A2: Associations with Technology Adoption

	1(Technology exists) (1)	(2)
Total population	0.103***	0.065**
	(0.026)	(0.024)
Income per capita	0.083**	0.015
	(0.031)	(0.020)
Urban share of population	0.118***	0.073*
	(0.025)	(0.032)
Share of properties with address	0.125**	0.094**
	(0.054)	(0.040)
Share of properties with valuation	0.177***	0.134***
	(0.028)	(0.025)
1(Legal capacity to enforce taxes)	0.083*	0.049*
	(0.041)	(0.022)
1(Tax-delinquents taken to court)	0.017	0.012
	(0.031)	(0.021)
Officials' years of work experience	0.046*	0.047
	(0.024)	(0.033)
Officials' years of education	0.029	0.017
	(0.030)	(0.025)
Trust in officials	-0.002	-0.001
	(0.014)	(0.015)
Citizen tax awareness	-0.010	-0.011
	(0.012)	(0.019)
Citizen compliance attitude	-0.009	0.016
	(0.019)	(0.010)
Region FE		X
Observations	216	216
Clusters	10	10

Notes: Each cell represents the β coefficient from a separate cross-district regression, based on the model $\mathbf{1}(Technology)_{dr} = \beta \cdot X_d + \mu_r + \epsilon_{dr}$, where $1(Technology)_{dr}$ is a dummy equal to 1 if the local government in district d in region r has an electronic tax registry of properties (see Section 2). X_d is the district characteristic which varies between rows; in column (2), region fixed effects (μ_r) are included. All district characteristics are standardized, for ease of comparison across rows. Standard errors are clustered at the regional level. * p<0.10 ** p<0.05 *** p<0.01.

Table A3: Tax Outcomes and Technology Across Local Governments

	(1)	(2)	(3)	(4)	(5)
Panel A: Taxes Collected per Capita					
1(Technology)	5.525**	2.482***	3.116***	2.919***	2.667*
	(1.742)	(0.768)	(0.796)	(0.835)	(1.372)
Mean outcome variable	4.153	4.153	4.153	4.153	4.153
Panel B: Share of Bills Delivered (%)					
1(Technology)	0.239***	0.097*	0.177***	0.134***	0.079**
	(0.052)	(0.051)	(0.034)	(0.015)	(0.035)
Mean outcome variable	0.430	0.430	0.430	0.430	0.430
Panel C: Taxes per Bill Delivered					
1(Technology)	6.868*	6.685**	4.629**	3.593***	2.911*
	(3.617)	(2.475)	(1.952)	(0.809)	(1.501)
Mean outcome variable	11.453	11.453	11.453	11.453	11.453
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 has an electronic tax registry of properties. 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). Across columns, the specifications 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 with official addresses, legal capacity, officials' years of work experience) in column (2); the share of each district's geographically adjacent neighbor governments with technology in column (3); region fixed effects in column (4); all three sets of controls in column (5). These results also hold when we control for whether the district made additional investments in technologies that automate the creation of tax bills and that assist with payment recording and enforcement (results available). Standard errors are clustered at the region level. * $p<0.10$ ** $p<0.05$ *** $p<0.01$.

Table A4: Randomization Balance

	N	Control mean	Treatment coefficient
	(1)	(2)	(3)
Panel A: Tax Bill Characteristics			
Current tax amount	8120	322.8	-9.0
			(16.4)
Total tax amount	8120	692.5	-5.5
			(29.1)
Previous pay status	8120	1.2	0.0
			(0.0)
Previous tax payment	8120	59.7	-6.6
			(9.4)
Residential	8120	0.5	0.0
			(0.0)
Property quality	8120	0.5	0.0
			(0.1)
F-test joint significance [F, p]			[0.71,0.66]
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 significance [F, p]			[0.16,0.95]
Panel C: Household Characteristics			
Income index	4334	-0.014	0.003
			(0.106)
Liquidity index	4334	0.051	-0.177
			(0.119)
Taxpayer awareness index	4334	0.011	-0.01
			(0.039)
F-test joint significance [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). The treatment coefficient in column (3) is the coefficient on technology in a cross-sectional regression with strata fixed effects. Standard errors are clustered at the collector-unit level. At the bottom of each panel, the F-test on the joint significance of all characteristics is reported along with the p-value. For more information on the characteristics, see Section 3 and Section 5. * p<0.10 ** p<0.05 *** p<0.01.

Table A5: Change in Sample Balance for Collector Survey Outcomes

Outcome	Chall. navigate	Chall. locate	Know HH-type	Focus aware	Focus able	Focus PGs	Focus hard obs.	Focus easy obs.	Focus hard-easy	# failed/ per success	Hours worked	Hours per bill	Hours non-deliv	Fieldwork prepare	Content in job	Wrong info	Resistance from HH	Superv. monitor	Superv. unavai.	Superv. mistakes
Panel A: Unbalanced sample	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
1(Technology)	-0.55***	-0.31***	0.18**	0.13	0.19**	0.08	0.14*	0.08	0.05	-0.93	-0.74	-0.77***	9.17**	0.04	0.10	-0.26	-0.06	-0.14	-0.21	-0.07
	(0.07)	(0.06)	(0.09)	(0.09)	(0.09)	(0.09)	(0.07)	(0.05)	(0.04)	(1.34)	(1.84)	(0.19)	(3.97)	(0.07)	(0.15)	(0.18)	(0.16)	(0.21)	(0.21)	(0.21)
Observations	141	141	141	141	141	141	141	141	141	141	141	111	141	141	141	141	141	141	141	141
Panel B: Balanced sample	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)	(31)	(32)	(33)	(34)	(35)	(36)	(37)	(38)	(39)	(40)
1(Technology)	-0.54***	-0.30***	0.20*	0.15	0.22**	0.11	0.16*	0.08	0.07	-1.57	-1.09	-0.80***	8.64*	0.08	0.15	-0.22	-0.10	-0.08	-0.25	-0.03
	(0.07)	(0.07)	(0.10)	(0.09)	(0.10)	(0.10)	(0.08)	(0.06)	(0.04)	(2.18)	(2.00)	(0.20)	(4.69)	(0.08)	(0.17)	(0.19)	(0.16)	(0.22)	(0.23)	(0.23)
Observations	123	123	123	123	123	123	123	123	123	123	123	92	123	123	123	123	123	123	123	123
Collector-unit controls	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Survey round FE	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Notes: This table estimates equation (1), in the unbalanced sample (Panel A) and the balanced sample (Panel B). In column (1), the outcome is a dummy variable which takes a value of 1 if the respondent agrees with the statement "Finding my way around my collection unit was a challenge for me this week", and 0 otherwise. Column (2) is coded similarly, for the statement "Locating bill recipients was challenging for me this week". In column (3), the outcome is a dummy variable which takes a value of 1 if the respondents chooses statement A "I think I have a good understanding of which properties are more able and willing to pay and am able to focus my efforts on them" rather than statement B "I put a lot of effort to get my job done, but it remains unclear to me which exact properties are more likely or willing to pay their property rates". The variable takes a value of 0 if the respondent picks statement B. In column (4), the outcome is a dummy variable which takes a value of 1 if the respondent uses 'all the time' or 'often' the collection strategy "Go to areas where I know most taxpayers are aware of their duty to pay property rates"; the variable takes a value of 0 if the respondents uses this strategy 'only from time to time', 'not much' or 'never'. Column (5) is coded similarly, for the collection strategy "Go to areas on specific days where I know property owners are more likely to be able to pay. Column (6) is coded similarly for collection strategy "Go to areas where I know owners are more satisfied with the delivery of public services and are more likely to pay". In column (7), the outcome is the average of the outcomes in columns (4), (5) and (6). In column (8), the outcome is the average over several strategy questions. Each strategy question is coded similarly to columns (4)-(6). The six strategies are to go to: "areas where I know most taxpayers have paid property rates in the past year"; "go to areas where I know there are many properties with high property rates"; "go to areas where I know there are many property rate payers that have not yet paid this year's rates"; "go to areas which are close to the main road/center of activity"; "go to areas which are close to my home"; "go to areas which are closer to the Madina headquarters". In column (9), the outcome is the difference between the outcomes in column (7) and (8). In column (10), the outcome is the self-reported number of unsuccessful visits for every successful visit. In column (11), the outcome is the total number of weekly hours worked. In column (12), the outcome is the average number of hours spent per bill delivered. In column (13), the outcome is the total number of weekly hours spent on non-delivery activities. For column (14), see Table A10. In column (15), the outcome is the average (standardized) agreement on a 5-point scale with three statements: "Overall, this was a productive week for me"; "Overall, I was content while working this week"; "Overall, I am satisfied with my job". Each statement can be answered on a 5-point scale, from 'strongly disagree', to 'strongly agree'. In column (16), the outcome measures the collectors' extent of agreement, on a 5-point scale, with the two statements: "Some of the bills I tried to deliver this week had the wrong addresses"; "Some of the bills I tried to deliver this week had the wrong amounts". In column (17), the outcome measures the collectors' extent of agreement, on a 5-point scale, with three statements: "Collection was challenging this week because bill recipients preferred not to pay in cash"; "Collection was challenging this week because bill recipients preferred mobile payments, but I was not able to accept mobile payments"; "Collection was challenging this week because bill recipients said that they did not trust me to collect their payment". In columns (18), (19) and (20), the outcome is the collector's extent of agreement with the following statements: "My supervisors spent a lot of time monitoring my work this week"; "My supervisors were available to help me this week when I needed them"; "My supervisors checked on me regularly this week to make sure I was not making mistakes". * p<0.10 ** p<0.05 *** p<0.01.

Table A6: Robustness Checks for Technology Impacts in Household Survey

	Bill delivered	Any positive tax payment	Full tax payment	Total payment amount (in GHC)	Any bribe (coercive or collusive)	Total bribe amount (in %)	Coercive bribe amount (in %)	Collusive bribe amount (in %)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Benchmark								
1(Technology)	0.054	0.043**	0.023**	25.910**	0.116***	0.025**	0.011*	0.040*
	(0.036)	(0.021)	(0.010)	(10.901)	(0.039)	(0.011)	(0.006)	(0.021)
Panel B: No Controls								
1(Technology)	0.049	0.039**	0.021*	24.931**	0.120***	0.027**	0.012**	0.042*
	(0.036)	(0.017)	(0.012)	(10.891)	(0.038)	(0.012)	(0.005)	(0.023)
Panel C: Extensive Controls								
1(Technology)	0.055	0.047**	0.027**	27.211**	0.113***	0.025**	0.014**	0.036*
	(0.034)	(0.020)	(0.012)	(11.181)	(0.037)	(0.010)	(0.005)	(0.019)
Strata FE	X	X	X	X	X	X	X	X
Mean in CG	0.506	0.162	0.061	40.951	0.139	0.107	0.022	0.190
Observations	4334	4334	4334	4334	4334	4334	4334	4334
Clusters	56	56	56	56	56	56	56	56

Notes: This table presents technology impacts on the main set of tax and bribe outcomes, based on the household survey. Panel A uses the benchmark specification in equation (2). In Panel B, the estimation model is the same except that all household and collector controls are removed. Panel C augments (2) with additional controls, which are the set of fixed, hard-to-observe characteristics used in the targeting analysis, specifically income, liquidity and taxpayer awareness from Table A4. See also Section 5 and Figure 4. Standard errors clustered at the collector-unit are reported in parentheses. In column (1), the outcome takes a value of 1 if the household reports that they received a property tax bill from the tax collector in the past 6 weeks, and 0 otherwise. In column (2), the outcome takes a value of 1 if the household reports that they made a positive payment for the property tax liability that was due in the past 6 weeks, and 0 otherwise. In column (3), the outcome takes a value of 1 if the household reports that they made a positive payment for the property tax which corresponds to the full amount due (based on administrative data), and 0 otherwise. In column (4), the outcome is the total amount in Ghanaian Cedi that the household reports having paid in property taxes during the past 6 week. In column (5), the variable is based on two dummy variables. The first dummy variable takes a value of 1 if the household estimates that tax collectors will ask for any strictly positive unofficial payments when they are working in the field, and zero otherwise. The second dummy variable takes a value of 1 if the household reports that the tax collector will pocket any positive amount out of a hypothetical 1000 Ghanaian Cedi collected from households (coercive bribe). The variable in column (5) takes a value of 1 if either of the two dummy variables is equal to 1, and takes a value of 0 otherwise. In column (6), the outcome is the average of the outcomes in column (7) and (8). In column (7), the outcome is the share that the household estimates will be pocketed by the tax collector out of a hypothetical 1000 Ghanaian Cedi that the official has collected as payments from households while working in the field. In column (8), the household estimates how much will be asked by the tax collector as unofficial payment while conducting visits to the household. In turn, the outcome is this amount expressed as a percent of the households own, actual property tax liability. * p<0.10 ** p<0.05 *** p<0.01.

Table A7: Beliefs about Enforcement and Tax Morale

	Technology coefficient ($\hat{\beta}$) (1)	Mean in CG (2)	*N* (3)
Panel A: Enforcement & Information Capacity			
Share of HHs that comply with taxes	0.80	60.32	4334
	(2.38)		
Likelihood non-complier will end up paying	-0.07	3.08	4334
	(0.07)		
Likelihood Gov't has info about my tax status	-0.13	2.95	4334
	(0.13)		
Likelihood Gov't has info about my job	0.03	2.52	4334
	(0.09)		
Panel B: Equity & Efficiency of Tax Collection			
Agree efforts to collect taxes efficiently	0.01	3.58	4334
	(0.07)		
Agree efforts to ensure fair share paid	-0.18***	3.42	4334
	(0.07)		
Agree efforts to collect for useful purposes	0.08	3.04	4334
	(0.11)		
Panel C: Government Integrity and Competency			
Share of taxes wastefully spent	-3.48	55.81	4330
	(4.64)		
Agree Gov't has capacity to improve roads	0.04	3.94	4334
	(0.11)		
Overall Gov't competency rating	0.07	2.41	4334
	(0.07)		
Panel D: Satisfaction with Government Services			
Quality of tax collector services	-0.003	2.31	4334
	(0.05)		
Quality of tax authority services	-0.02	2.31	4334
	(0.05)		
Quality of overall Gov't services	-0.01	2.20	4334
	(0.05)		

Notes: Each row presents the technology treatment coefficient (in column 1) from estimating equation (2) on different outcomes (which are described to the left). Standard errors are clustered at the collector-unit. 'Likelihood' questions are on a 4-point scale, from 'very unlikely' to 'very likely'. 'Agree' questions are on a 5-point scale, from 'strongly disagree' to 'strongly agree'. 'Quality' questions are on a 5-point scale, from 'very unsatisfied' to 'very satisfied'. In Panel C, the 'rating' question is on a 4-point scale, from 'not competent at all' to 'very competent'. * p<0.10 ** p<0.05 *** p<0.01.

Table A8: Heterogeneity in Beliefs about Enforcement and Tax Morale

	Technology coefficient ($\hat{\beta}$)	Heterogeneity coefficient ($\beta \hat{\times} H$)	N
Outcome: Enforcement and Information Capacity			
Heterogeneity H: Liquidity index	-0.050	-0.016	4334
	(0.056)	(0.053)	
Heterogeneity H: Income index	-0.051	0.002	4334
	(0.057)	(0.042)	
Heterogeneity H: Taxpayer awareness index	-0.050	-0.021	4334
	(0.056)	(0.057)	
F-test joint significance of interaction terms [F, p]		[0.09, 0.96]	
Outcome: Equity & Efficiency of Tax Collection			
Heterogeneity H: Liquidity index	-0.016	0.048	4334
	(0.059)	(0.055)	
Heterogeneity H: Income index	-0.010	0.059	4334
	(0.060)	(0.039)	
Heterogeneity H: Taxpayer awareness index	-0.012	0.068	4334
	(0.061)	(0.069)	
F-test joint significance of interaction terms [F, p]		[1.27, 0.29]	
Outcome: Government Integrity and Competency			
Heterogeneity H: Liquidity index	0.048	0.039	4334
	(0.070)	(0.063)	
Heterogeneity H: Income index	0.060	0.012	4334
	(0.073)	(0.040)	
Heterogeneity H: Taxpayer awareness index	0.064	-0.036	4334
	(0.072)	(0.048)	
F-test joint significance of interaction terms [F, p]	[0.32, 0.81]		
Outcome: Satisfaction with Gov't Services			
Heterogeneity H: Liquidity index	-0.018	0.042	4334
	(0.069)	(0.059)	
Heterogeneity H: Income index	-0.009	0.011	4334
	(0.069)	(0.032)	
Heterogeneity H: Taxpayer awareness index	-0.017	0.041	4334
	(0.070)	(0.064)	
F-test joint significance of interaction terms [F, p]		[0.45,0.72]	

Notes: This table investigates heterogeneous technology impacts on beliefs and tax morale. Each row presents the technology treatment coefficient and the interaction coefficient, 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 significance of the three interaction coefficients for a given outcome. The outcomes and heterogeneity dimensions are described in Section 4-5. Standard errors clustered at the collector-unit are reported in parentheses. * p<0.10 ** p<0.05 *** p<0.01.

Table A9: Heterogeneity in Tax Outcomes, Beliefs and Morale by Registration Status

	Technology coefficient $\hat{\beta}$	Heterogeneity coefficient $\beta \times \hat{1(\text{New register})}$	N
Panel A: Tax Payment Outcomes			
Outcome: 1(Any payment)	0.043**	-0.003	4334
	(0.020)	(0.040)	
Outcome: 1(Full payment)	0.024*	0.012	4334
	(0.012)	(0.026)	
Outcome: Amount paid (in GHC)	25.796**	0.901	4334
	(11.608)	(22.241)	
Panel B: Beliefs and Tax Morale Outcomes			
Outcome: Satisfaction with gov't services	-0.030	0.094	4334
	(0.066)	(0.140)	
Outcome: Integrity of gov't	0.063	-0.044	4334
	(0.074)	(0.134)	
Outcome: Equity & efficiency	-0.021	0.016	4334
	(0.059)	(0.124)	
Outcome: Enforcement & information	-0.056	0.019	4334
	(0.055)	(0.131)	

Notes: This table investigates heterogeneous technology impacts by registration status. A subset of the properties were registered in the tax registry for the first time during the experiment's campaign; 1(New register) takes a value of 1 if the property is newly registered, and 0 otherwise. Each row corresponds to a different outcome. In each row, we present the technology treatment coefficient and the interaction coefficient, from estimating equation 2 augmented with the interaction between technology and 1(New register). Panel A focuses on tax outcomes: a dummy for any tax payment made; a dummy for a full tax payment made; and, the amount of taxes paid (in GHC). Panel B focuses on household beliefs and tax morale outcomes: an index for satisfaction with government services; an index for the perceived integrity of local government; an index for the perceived equity and efficiency of the tax collection process; and, an index for the perceived enforcement and information capacity of the local government. The outcomes are the same as in Table 2 and Table A6. Standard errors clustered at the collector-unit are reported in parentheses. * p<0.10 ** p<0.05 *** p<0.01.

Table A10: Collector Performance and Challenges Reported in the Field

Panel A: Challenges	Wrong information printed on bills (1)	Resistance from property to accept bill (2)	Supervisors do not monitor activities in the field (3)	Supervisors unavailable for support if needed (4)	Supervisors check for mistakes in the field (5)
1(Technology)	-0.265	-0.060	-0.141	-0.213	-0.078
	(0.186)	(0.163)	(0.214)	(0.213)	(0.216)
	[0.084]	[0.628]	[0.420]	[0.278]	[0.668]
Mean in CG	0.131	0.04	0.076	0.158	0.032
Observations	141	141	141	141	141

Panel B: Performance	# unsuccessful visits per successful visit (6)	Fieldwork is prepared and narrowly focused (7)	Overall satisfaction in job (8)
1(Technology)	-0.933	0.041	0.103
	(1.345)	(0.077)	(0.158)
	[0.420]	[0.482]	[0.524]
Mean in CG	8.028	0.536	-0.065
Observations	141	141	141

Notes: This table provides estimates based on equation (1). All regressions include collector-unit controls and survey round fixed effects (Section 3). Panel A measures the extent to which the collector agrees that a certain challenge characterizes their field work: wrong information on the bills; resistance from property owners to accept the bill; and, supervisors do not monitor activities, are unavailable for support, and check for mistakes. Panel B focuses on performance measures: # of unsuccessful visits per successful one; fieldwork organization; satisfaction in the job. In column (7), the outcome is the average of two dummy variables. The first dummy takes a value of 1 if the collector reports that their field-work this week is best characterized by a focus on specific properties in their unit or 0 if their work is best characterized by working in broad areas throughout their unit. The second dummy takes a value of 1 if the collector reports that their field-work this week is best characterized by organizing which properties to visit before going into the field each day or 0 if their work is best characterized by going into the field immediately and making the most of it. Standard errors clustered at the collector-unit are reported in parentheses. The randomization inference based p-value is reported in brackets. Results are based on the full sample of collector surveys (results from the balanced sample are in Table A5). For details on the outcomes, see Section 5 and Table A5 * p<0.10 ** p<0.05 *** p<0.01.

Table A11: Collector Survey Outcomes

	Technology coefficient ($\hat{\beta}$) (1)	Mean in CG (2)	N (3)
Panel A: Contentment in Job			
This was a productive week for me	0.239	3.391	141
	(0.184)		
I was content while working this week	0.024	3.260	141
	(0.175)		
I am satisfied with my job	0.054	3.391	141
	(0.194)		
Panel B: Collection Challenges Faced This Week			
Some of my bills had the wrong address	-0.304	3.449	141
	(0.279)		
Some of my bills had the wrong amount to be paid	0.004	3.594	141
	(0.300)		
Bill recipients preferred not to pay in cash	0.015	3.115	141
	(0.233)		
Bill recipients did not trust me to collect their payment	-0.200	3.507	141
	(0.322)		
Bill recipients felt the amount to pay was too high	-0.052	3.782	141
	(0.335)		
Panel C: Supervisors "..."			
Spent a lot of time monitoring my work	0.163	3.362	141
	(0.247)		
Were available when I needed them	0.335	3.347	141
	(0.228)		
Checked on me regularly to prevent mistakes	0.089	3.405	141
	(0.248)		

Notes: This table shows the average impacts of technology on collectors' job satisfaction (panel A), work challenges (panel B), and perceptions of supervisors (panel C). Each row presents the technology treatment coefficient (in column 1) from estimating equation (1) on different outcomes (which are described to the left). All outcomes are questions which take on a value from 1 (strongly disagree) to 5 (strongly agree). Column (2) presents the mean of the outcome variable in control areas, while column (3) shows the sample size. For the outcomes in Panel C, note that they are reverse coded in the other tables and figures of the paper (e.g. Table A10) to indicate challenges. Standard errors clustered at the collector-unit are reported in parentheses. For details on the outcomes, see Section 5 and Table A5. All regressions include collector-unit controls (Section 3). * p<0.10 ** p<0.05 *** p<0.01.

Table A12: Exposure of Control Group Collectors to Technology

	Data: Collector Surveys							Data: Collector Daily Information	
	Challenge locate (1)	Knowledge of HH type (2)	Coll. focus hard vs easy (3)	Content in job (4)	Total hours worked (5)	# failed per successful visit (6)	# hours per bill delivered (7)	Total bills delivered (8)	Total taxes collected (GHC) (9)
TG-share in adjacent units (%)	0.125	0.039	0.009	-0.018	-0.252	0.202	0.413	-1.744	-27.577
	(0.108)	(0.174)	(0.080)	(0.221)	(4.390)	(1.515)	(0.522)	(13.130)	(191.255)
Mean in CG	0.794	0.197	0.041	-0.065	19.057	8.028	1.662	52.515	329.206
Observations	71	71	71	71	71	71	71	1164	1164
Collector-unit controls	X	X	X	X	X	X	X	X	X
Time FE	X	X	X	X	X	X	X	X	X

Notes: The analysis in this table is restricted to the control collectors. The estimation is based on equation (1) and is estimated at the collector-survey round level in columns (1) to (7), and at the collector-day level in columns (8)-(9). The estimating equation replaces the treatment assignment, $\mathbf{1}(Tech)_c$, with the variable $(TG - share)_c$ which measures for each collector c the share of geographically adjacent collection units which are assigned to the treatment group. In columns (1) to (7), the outcome variables are from the collector surveys: a dummy equal to 1 if the collector reports finding it challenging or very challenging to locate assigned taxpayers (and 0 otherwise); a dummy equal to 1 if the collector reports having a good understanding of which properties are more willing and able to pay (and 0 otherwise); the difference in strategy focus between hard to observe characteristics and easy to observe characteristics; contentment in job; hours worked; # of failed visits per succesfull visit; and # of hours worked per bill delivered. In columns (8)-(9), the outcome variables are from the collector daily administrative data (Section 3): total bills delivered; total taxes collected (in GHC). All regressions include the collector-unit controls (Section 3). All regressions include time fixed effects, which are survey round fixed effects in columns (1)-(7) and campaign-day fixed effects in columns (8)-(9). Standard errors clustered at the collector-unit are reported in parentheses. For details on the variables from the collector surveys, see Section 5 and Table A5. For details on the variables from the collector daily data, see Section 3. * p<0.10 ** p<0.05 *** p<0.01.

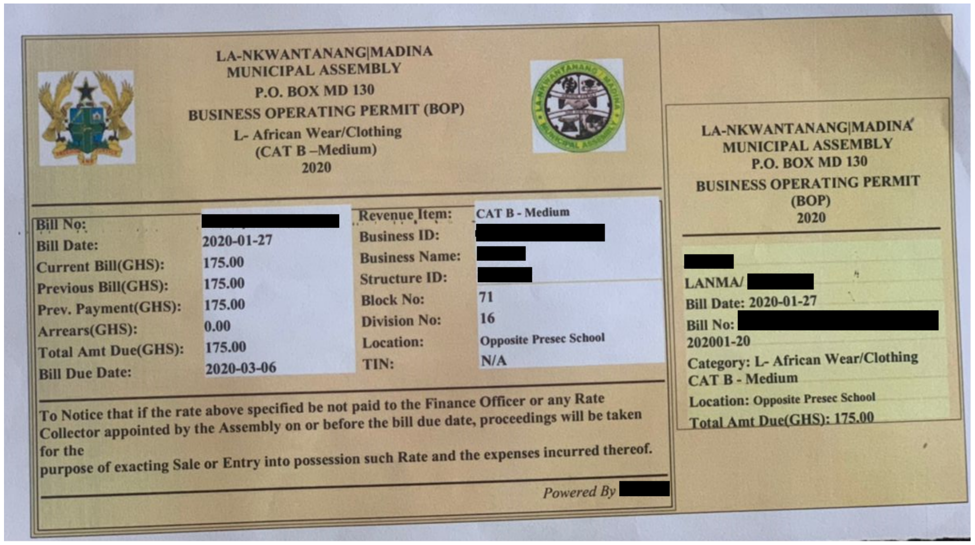
Table A13: Hard to Observe and Easy to Observe Indices as Predictors of Tax Payment

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Outcome is 1(Any payment)						
Income index	0.016**					
	(0.008)					
Liquidity index		0.017**				
		(0.007)				
Awareness index			0.011			
			(0.0139)			
Hard to observe index				0.035**		0.029**
				(0.014)		(0.014)
Easy to observe index					0.021***	0.017**
					(0.006)	(0.008)
Panel B: Outcome is 1(Full payment)						
Income index	0.012**					
	(0.005)					
Liquidity index		0.012**				
		(0.005)				
Awareness index			0.005			
			(0.009)			
Hard to observe index				0.024***		0.025***
				(0.009)		(0.009)
Easy to observe index					-0.003	-0.007
					(0.009)	(0.010)
Block FEs	X	X	X	X	X	X
Observations	4334	4334	4334	4334	4334	4334
Clusters	56	56	56	56	56	56

Notes: This table estimates how the hard to observe index, its components, and the easy to observe index are predictors of tax payment outside of the experimental sample. Specifically, the tax outcomes are measured at the property level in the previous fiscal year's tax campaign. In Panel A, the outcome is a dummy which takes a value of 1 if the household makes any tax payment (and 0 otherwise). In Panel B, the outcome is a dummy which takes a value of 1 if the household makes a tax payment equal to the entire tax liability due. Across columns, the outcome is regressed on different explanatory variables: the income index; the liquidity index; the taxpayer awareness index; the hard to observe index (which combines income, liquidity and awareness); the easy to observe index; and, the hard and easy to observe indices. For a description of the indices, see Section 5. All regressions include block fixed effects (7-8 properties per block, on average). The outcome is not defined for the subset of property owners that were newly registered for the experiment's campaign (Section 4.1). We assign an arbitrary value of the outcome for those property owners, and include a dummy for new registration. Standard errors are clustered at the collector-unit level. * p<0.10 ** p<0.05 *** p<0.01

Figure A1: Illustrations of Tax Bill and Navigation in Treatment and Control Groups

(a) Typical Tax Bill in Experimental Sample



LA-NKWANTANANG|MADINA
MUNICIPAL ASSEMBLY
P.O. BOX MD 130
BUSINESS OPERATING PERMIT (BOP)
L- African Wear/Clothing
(CAT B –Medium)
2020

Bill No:	[redacted]	Revenue Item:	CAT B - Medium
Bill Date:	2020-01-27	Business ID:	[redacted]
Current Bill(GHS):	175.00	Business Name:	[redacted]
Previous Bill(GHS):	175.00	Structure ID:	[redacted]
Prev. Payment(GHS):	175.00	Block No:	71
Arrears(GHS):	0.00	Division No:	16
Total Amt Due(GHS):	175.00	Location:	Opposite Presec School
Bill Due Date:	2020-03-06	TIN:	N/A

To Notice that if the rate above specified be not paid to the Finance Officer or any Rate Collector appointed by the Assembly on or before the bill due date, proceedings will be taken for the
purpose of exacting Sale or Entry into possession such Rate and the expenses incurred thereof.

Powered By [redacted]

LA-NKWANTANANG|MADINA
MUNICIPAL ASSEMBLY
P.O. BOX MD 130
BUSINESS OPERATING PERMIT
(BOP)
2020

[redacted]
LANMA/ [redacted]
Bill Date: 2020-01-27
Bill No: [redacted]
202001-20
Category: L- African Wear/Clothing
CAT B - Medium
Location: Opposite Presec School
Total Amt Due(GHS): 175.00

(b) Navigation with Tablet in Treatment Group

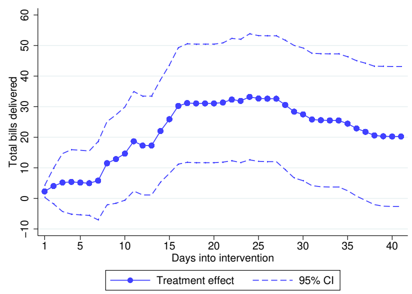
(c) Navigation Without Tablet in Control Group

Notes: Panel A illustrates a typical property tax bill in Madina. Panel B illustrates the navigational assistance provided by the GIS-tablet. Panel C illustrates navigation in the control group, where collectors sometimes ask local residents for assistance to navigate.

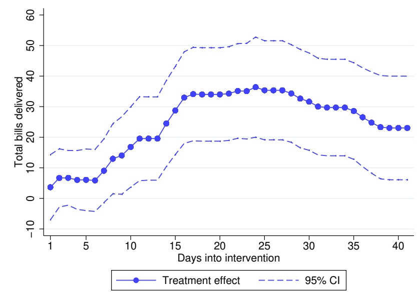
Figure A2: Robustness of Treatment Effects for Bills Delivered and Taxes Collected

(a) Bills: Benchmark

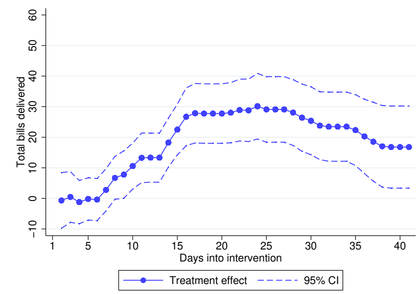
(b) Bills: Winsorize



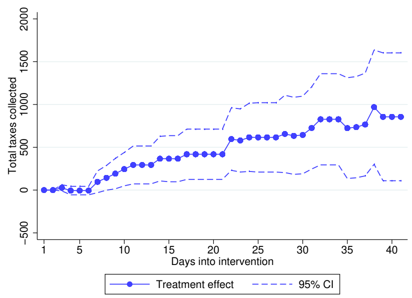
(c) Bills: Covariates



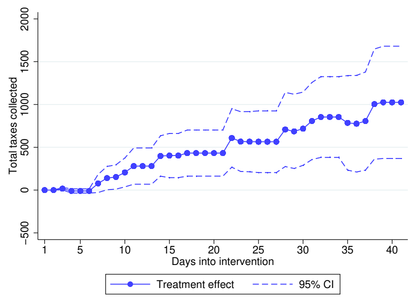
(d) Bills: Collector FEs



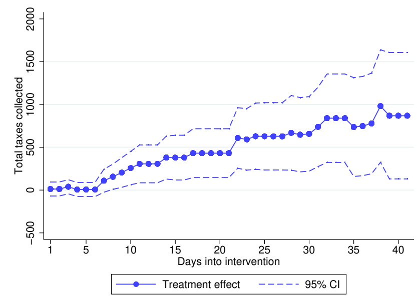
(e) Taxes: Benchmark



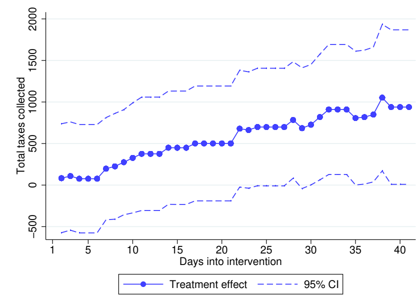
(f) Taxes: Winsorize



(g) Taxes: Covariates



(h) Taxes: Collector FEs

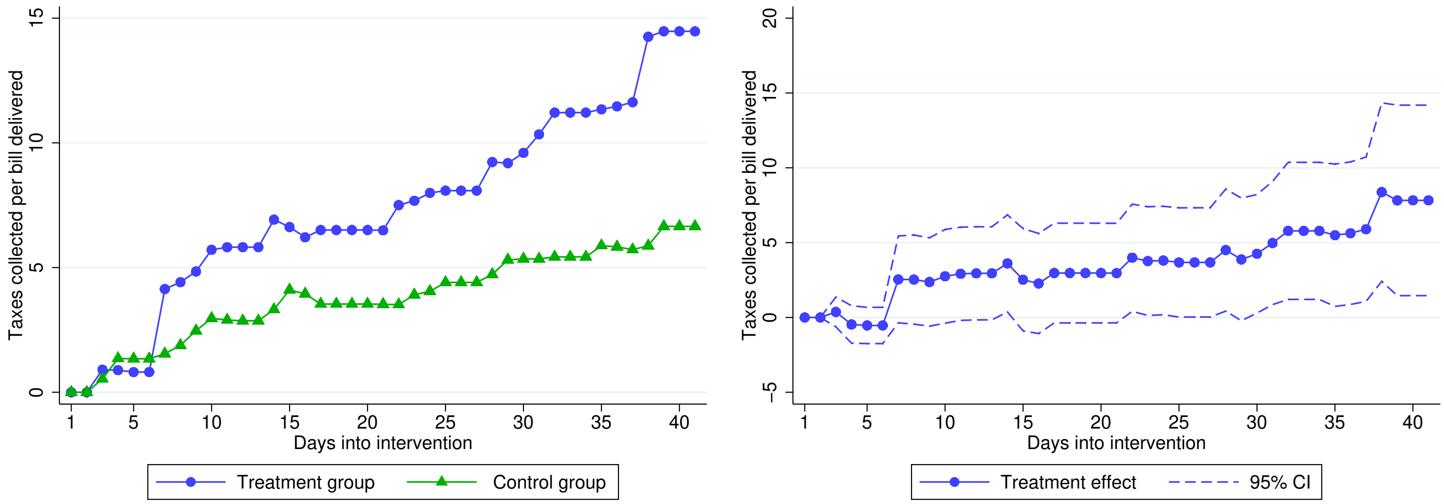


Notes: These panels show robustness for the estimated treatment effects of technology on the number of property tax bills delivered (panels in top row) and the total amount of taxes collected (panels in bottom row). All regressions are based on estimating equation (1). In both rows, the first panel to the left shows the treatment effect from the benchmark specification. The second panel from the left changes the benchmark by using the non-winsorized outcome. The third panel from the left changes the benchmark by including control variables: a dummy for whether the collector has previously worked in Madina; a dummy for whether the collector is assessed to be high performing; the total number of bills assigned to the collector; and, the average tax bill value per bill assigned. The fourth panel from the left changes the benchmark by including collector-unit fixed effects – in this case we omit β_1, the treatment category in day 1 (see equation 1). Coefficients together with the 95% confidence intervals are displayed. Standard errors are clustered at the collector-unit level. The analysis is based on the daily collector data, described in Section 3.1.

Figure A3: Impacts of Technology on Taxes Collected per Bill Delivered

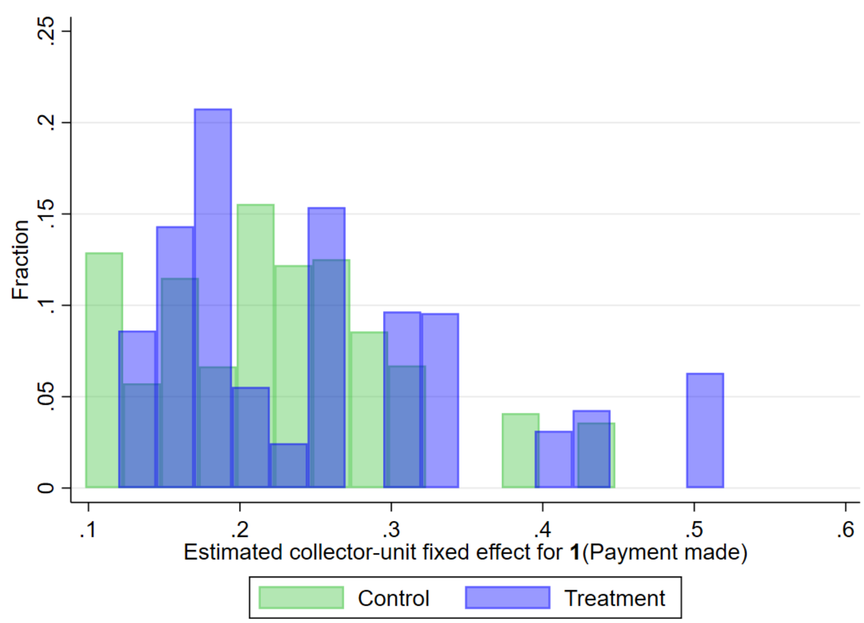
(a) Taxes Collected per Bill Delivered by Group

(b) Treatment Effect



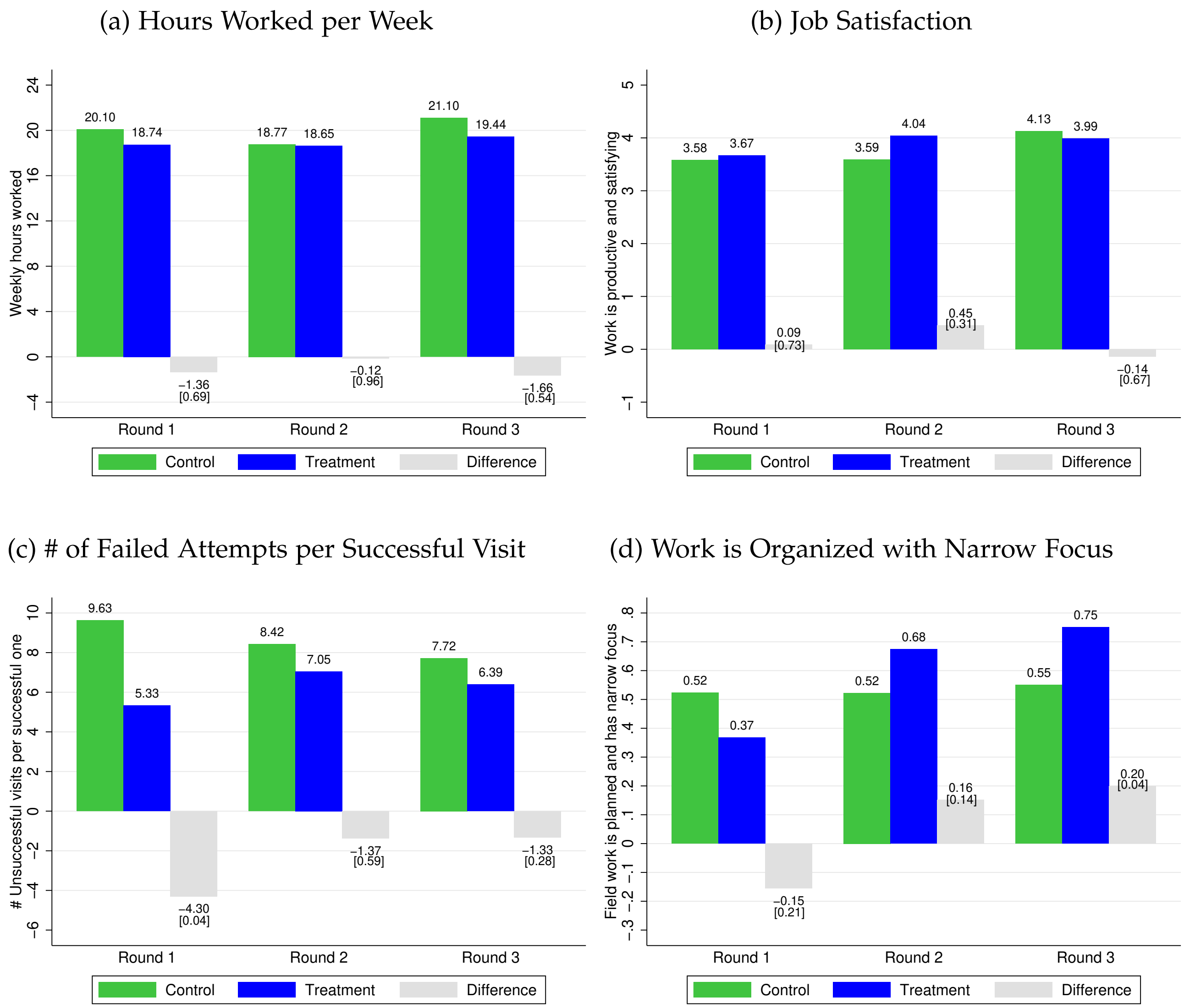
Notes: These panels show the impact of technology on the amount of taxes collected per bill delivered. Panel A shows the average amount of taxes collected per bill delivered by group (treatment, control) and by day of the intervention. Panel B displays the treatment effect coefficients on technology, separately by day, based on estimating equation (1).

Figure A4: Estimated Collector Fixed Effects



Notes: This figure is based on the household survey sample. We regress a dummy for any tax payment on the set of collector-unit fixed effects, and plot the density distribution of these estimated fixed effects for treatment and control collector-units. Collectors were randomly assigned to collection units, and each collector-unit pair was subsequently randomly assigned to the treatment or control group.

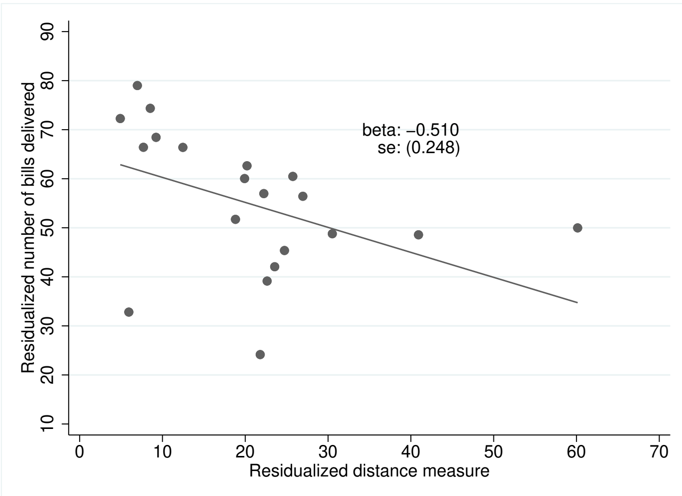
Figure A5: Collector Outcomes Across Rounds of the Tax Campaign



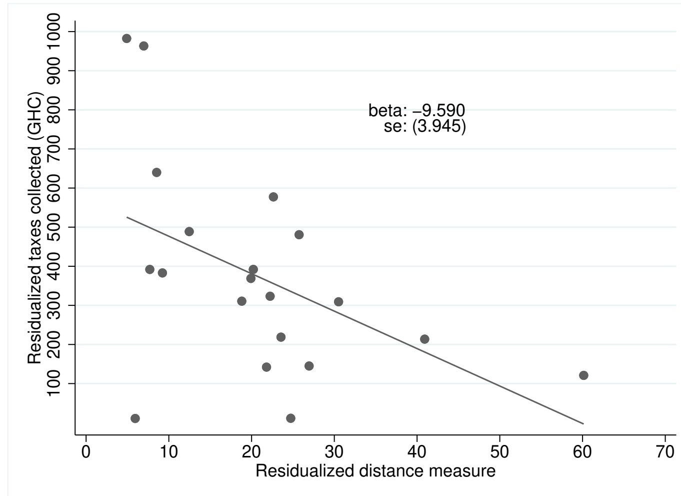
Notes: These panels show results on collector outcomes, based on equation (1). In all panels, Round 1, 2 and 3 correspond to the baseline, mid-line and endline collector survey rounds, respectively. In all panels, the gray bar measures the difference in outcome between the treatment and control groups; the number in brackets is the randomization inference-based p-value on the statistical significance of the difference. In panel A, the outcome is the collector's self-reported number of hours worked per week. In panel B, the outcome is the average job satisfaction, which combines measures of how much the collector, during the past week, agrees that: their work was productive; they were content while working; and, they were satisfied with their job. The analysis is based on the balanced sample of collector surveys (Section 3.1). In panel C, the outcome is the collector's self-reported number of failed attempts per successful visit. In panel D, the outcome is the average of two dummy variables. The first dummy takes a value of 1 if the collector reports that they focus on specific properties in their unit or 0 if they instead report working in broad areas throughout their unit. The second dummy takes a value of 1 if the collector reports that they organize properties to visit before going into the field each day or 0 if they instead report going into the field directly. The analysis is based on the balanced sample of collector surveys (Section 3.1); results based on the unbalanced sample are in Table A5. For details on the outcomes, see Section 5 and Table A5.

Figure A6: Size of Collection Unit and Tax Outcomes in the Control Group

(a) Bills Delivered

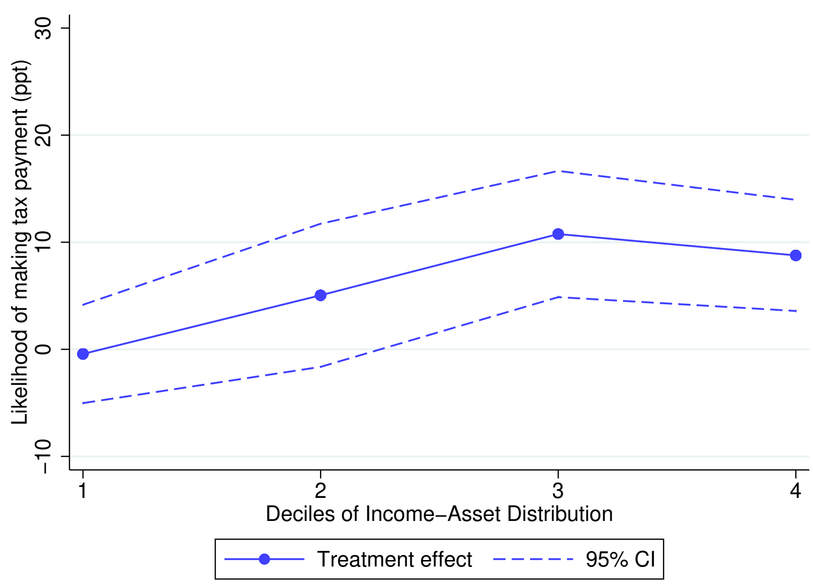


(b) Taxes Collected



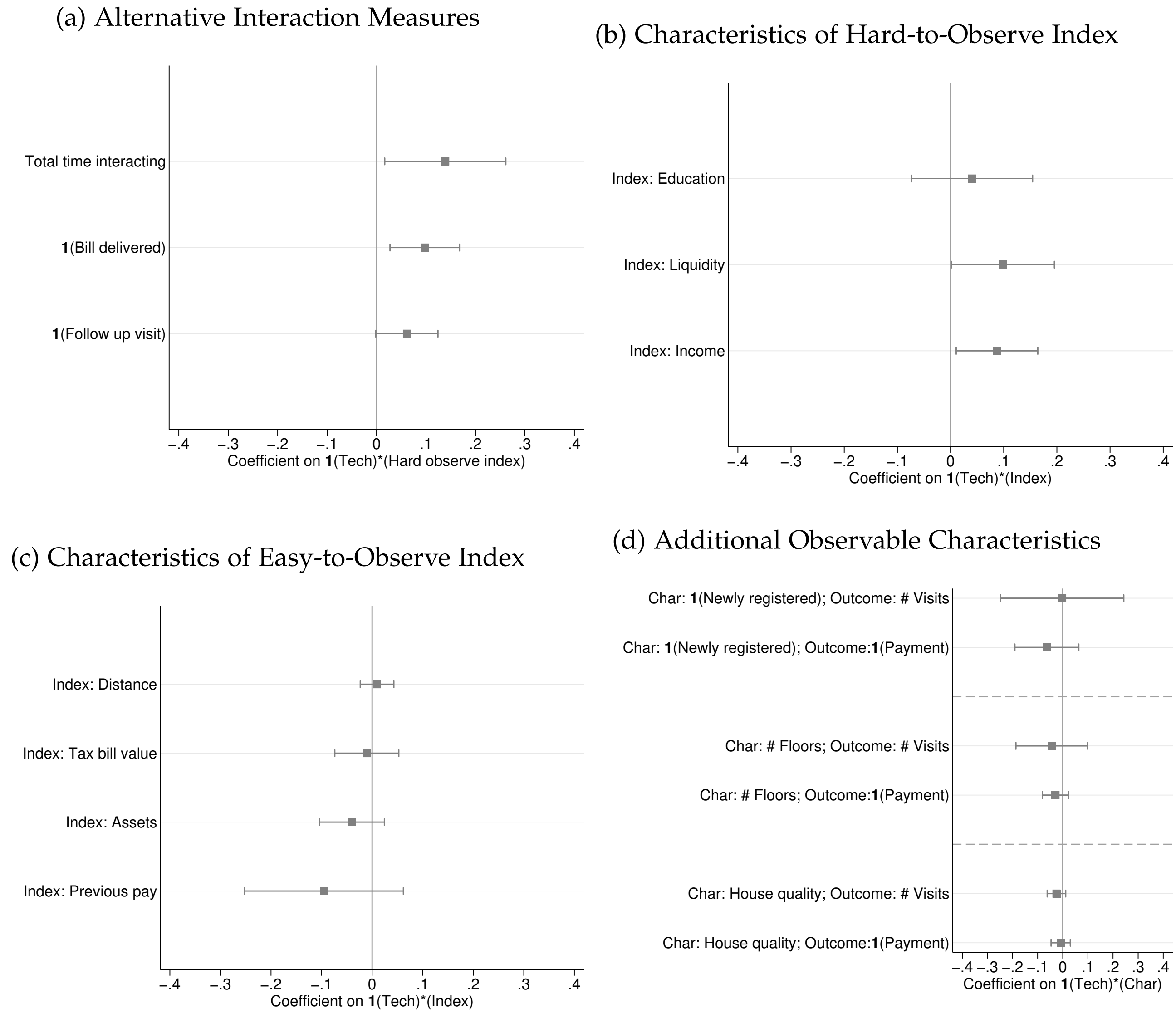
Notes: In each collection unit, we measure the shortest total distance that is required to visit every property once that the collector is assigned to. The starting point is the location of the local government office. Panel A and B show the conditional association between the distance measure and bills delivered and taxes collected, respectively. Bills delivered and taxes collected are measured in the daily collector data. All variables are first regressed on the total number of bills assigned to the collection unit and campaign-day fixed effects. In turn, twenty equal sized bins of the residualized distance measure are created and the dots represent the average residualized outcome by bin.

Figure A7: Distributional Effects of Technology on Taxes



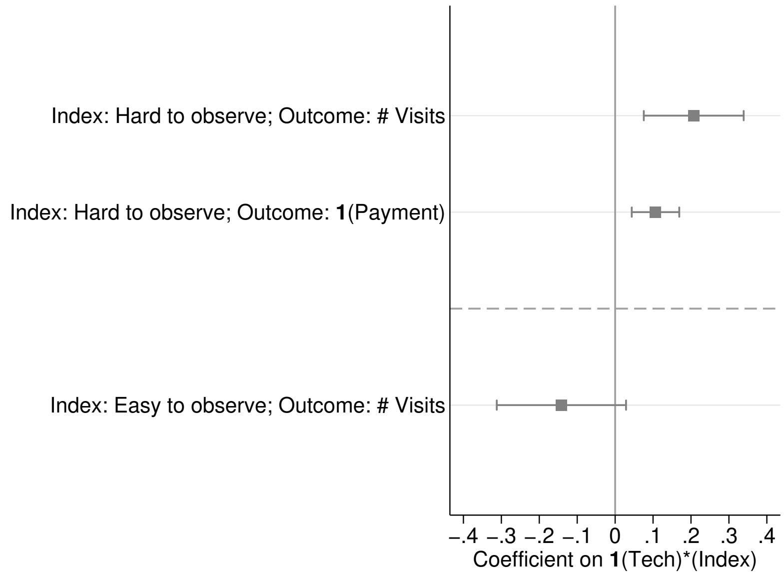
Notes: This figure displays the treatment effect coefficient on technology, separately by quartile of the income-asset distribution, based on estimating equation (2) augmented with a set of interactions between the treatment assignment $\mathbf{1}(Tech)_c$ and dummies for quartiles of the household income-asset distribution.

Figure A8: Differential Targeting: Robustness and Additional Results



Notes: Panel A reports estimates of the differential targeting coefficient β from estimating (3). Across rows, the outcome variable changes: total number of hours spent interacting between the collector and household; a dummy variable for whether a bill was delivered; a dummy variable for any follow up visit after an initial visit. In the third row, the sample conditions on an initial visit. In Panel A, the characteristic is the hard-to-observe index. In Panels B and C, each row reports the coefficient β from a separate regression that estimates (3). In Panels B and C, the outcome is always the total number of visits. Across rows, the household characteristic, in a standardized index format, varies. Panel B focuses on the characteristics of the hard-to-observe index (household tax awareness/education; household liquidity; household income). Panel C focuses on the four characteristics of the easy to observe index: property's distance to main roads and main commercial areas; tax bill value; observable assets; previous payment of property taxes. Panel D is estimated similarly to Panel C; across rows, the indexed characteristic and the outcome both vary. In the top rows, the characteristic is a dummy for whether the taxpayer is included in the tax registry for the first time in the experimental campaign; in the middle rows, the characteristic is the number of floors of the property; in the bottom row, the characteristic is the house quality (material used to build outer walls and roof). All characteristics are standardized. For more details on the variables, outcomes and specification, see Section 4 and 5.

Figure A9: Differential Targeting of Visits and Payments: Robustness



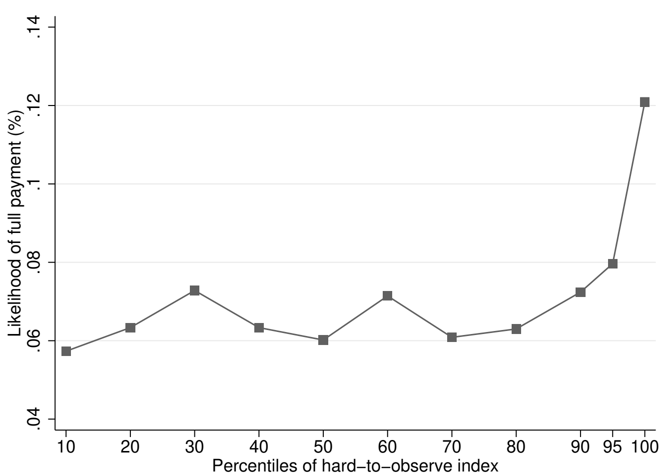
Notes: This figure shows robustness of the results in Figure 4 to controlling for the other index of household characteristics. Specifically, the coefficient in the top row reports the β coefficient from estimating:

$$y_{hc} = \theta \cdot HO - index_h + \beta \cdot [HO - index_h \cdot (Tech)_c] + \psi \cdot EO - index_h + \pi \cdot [EO - index_h \cdot (Tech)_c] + \Omega \cdot X_{hc} + \mu_c + \epsilon_{hc}$$

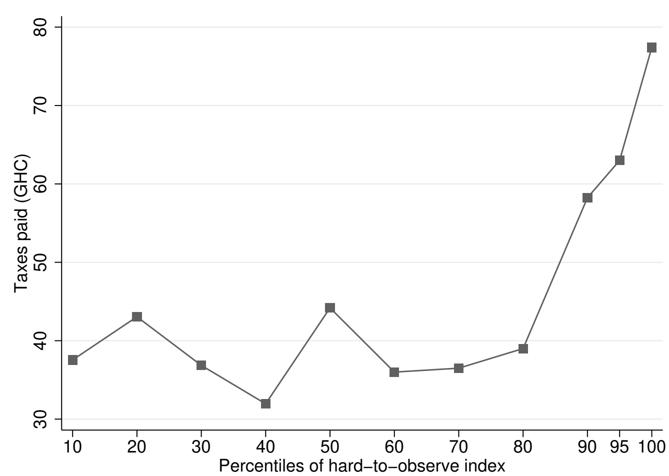
where y_{hc} is total number of visits, $HO - index_h$ and $EO - index_h$ are the hard to observe and easy to observe indices, respectively. The second row reports the estimate of β when the outcome is a dummy for any tax payment. The third row reports the estimate of π, when the outcome is total number of visits.

Figure A10: Tax Outcomes by Values of the Hard-to-Observe Index

(a) Likelihood of Tax Payment

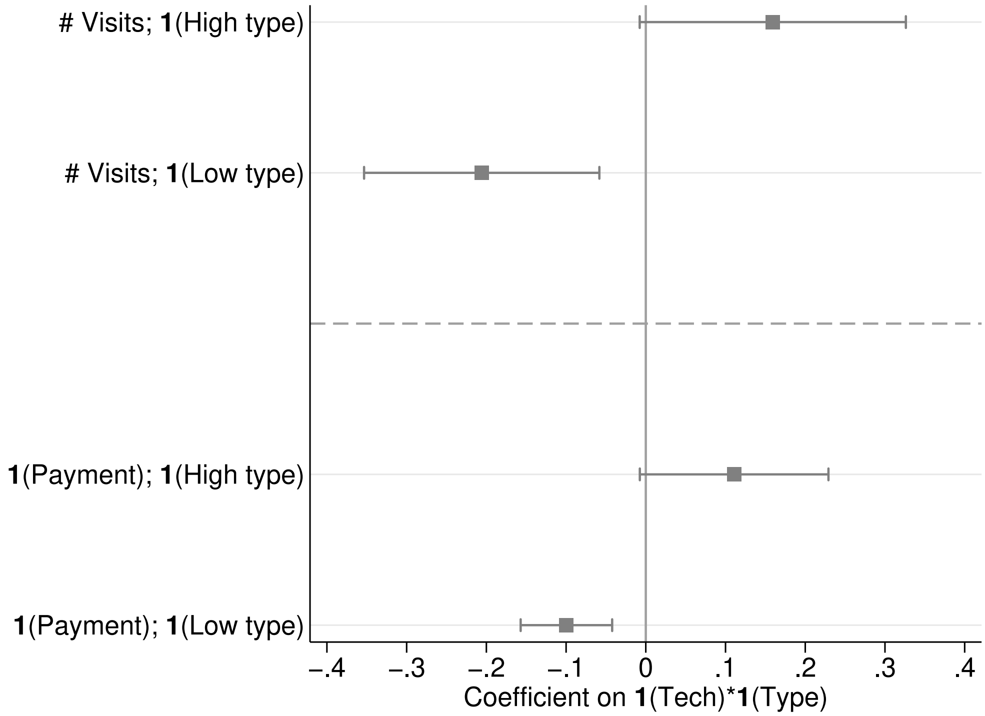


(b) Amount of Taxes Paid (GHC)



Notes: These panels are constructed using the household survey sample in control areas. The hard-to-observe index is partitioned into 100 percentiles; in turn, the figure shows the average value of the outcome in each decile, as well as the in $90 - 95^{th}$ percentile range and the $95 - 100^{th}$ percentile range. In Panel A (B), the outcome is a dummy which takes a value of 1 if the household made any tax payment (full tax payment), and 0 otherwise.

Figure A11: Targeting Based on Binary Measures of Hard to Observe Index



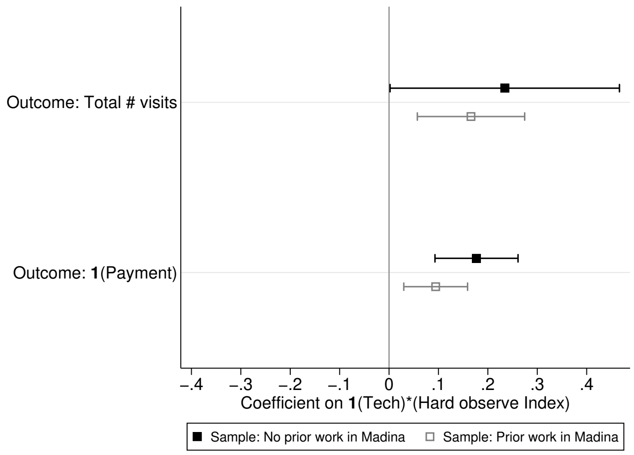
Notes: This figure estimates differential targeting, based on an augmented version of (3). The top two rows display the coefficients β and π from the following regression:

$$y_{hc} = \theta \cdot 1(Hightype)_h + \beta \cdot [1(Hightype)_h \cdot 1(Tech)_c] + \psi \cdot 1(Lowtype)_h + \pi \cdot [1(Lowtype)_h \cdot 1(Tech)_c] + \Omega \cdot X_{hc} + \mu_c + \epsilon_{hc}$$

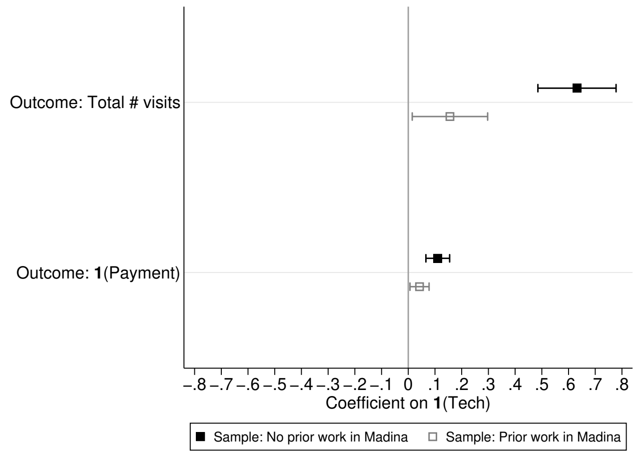
where y_{hc} is total number of visits, $1(Hightype)_h$ is a dummy taking a value of 1 if the household's value of the hard to observe index is in the top 5% of the index distribution, and $1(Lowtype)_h$ is a dummy taking a value of 1 if the household's value of the hard to observe index is in the bottom 5% of the index distribution. The 95% confidence intervals are reported, based on standard errors clustered at the collector-unit level. The bottom two rows report the coefficients from the same regression, but where the outcome y_{hc} is a dummy for any tax payment made. For more details on the high and low type dummies, see Section 5.3.

Figure A12: Heterogeneity by Prior Work Experience

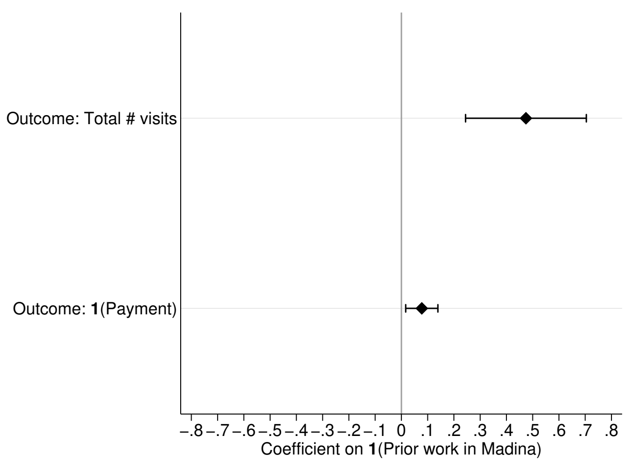
(a) Targeted Interactions



(b) Volume of Interactions and Payment



(c) Prior Experience (Control Group)



Notes: In Panel A, each coefficient reports β from a separate regression that estimates (3); all regressions use the hard to observe index. The filled square coefficients are based on estimating the equation in the sample of collectors with no prior work experience in Madina; the hollow square coefficients are based on estimating the equation in the sample of collectors with prior work experience in Madina. Outcomes differ by row and are specified in the y-axis. In Panel B, each coefficient reports β from a separate regression that estimates (2). The hollow (filled) square coefficients are based on estimating the equation in the sample of collectors with (no) prior work experience in Madina. Outcomes differ by row and are specified in the y-axis. Panel C shows the association between prior work experience in Madina and outcomes of interest, based on the sample of control areas only. The outcome varies by row and is specified in the y-axis. The filled diamond coefficient reports the estimated β from the following regression:

$$y_{hc} = \beta \cdot \mathbf{1}(\text{Prior work})_c + \Omega \cdot X_{hc} + \epsilon_{hc},$$

where $\mathbf{1}(\text{Prior work})_c$ is a dummy which takes a value of 1 if collector c has prior experience working as a collector in Madina. In all panels, 95% confidence intervals are reported, based on clustering of standard errors at the collector-unit level. For details on the prior work experience variable, see Section 3.1; for details on the targeting specification and the index, see Section 5.1-5.3.

B Observability of Household Types: Discussion

In the main text, we have focused on a setting where types are hard to observe – in the sense that the household's propensity to pay is initially not observed but the collector can discover it through a time-costly process of engaging with the household. This setting is consistent with the results from the collector surveys: treatment collectors initially report a limited understanding of which households have higher propensity, but report significantly better knowledge of households' types over time (Figure 3). However, those results are based on collectors' self reports. In this sub-section, we will instead focus on the independent results based on the household reports. We will investigate whether the mechanism results based on the household survey (Section 5) can be made consistent with alternative settings. We discuss in turn the settings where the household's propensity is either perfectly observable or perfectly unobservable to all collectors at all points during the campaign.

Before proceeding, let us summarize the household survey results that are consistent with the setting where propensity is initially hard to observe, but can be discovered. Relative to control collectors, treatment collectors: have more and longer interactions with households in general (Table 3); have more interactions with households that have higher propensity to pay (Figure 4), and fewer interactions with households that have very low propensity to pay (Figure A11); if anything, spend slightly less time per interaction with higher propensity households (Table 4); conduct a larger share of their interactions with high propensity types (Figure 5); and, have fewer interactions with households that have high values of easily observable characteristics (Figure 4), even after controlling for the hard to observe characteristics (Figure A9). Finally, hard-to-observe characteristics are stronger predictors of tax payment than the easy-to-observe characteristics (Table A13).

These results are consistent with learning, where longer interactions are required to learn any household's type, and where the discovery of the high (low) types lead treatment collectors to focus more (less) of their subsequent interactions on them. Treatment collectors may spend less time per interaction with high propensity households if the time cost to learn the type is only required once, and the follow up interactions that focus on collection, which are targeted to high types, are shorter in duration. The negative differential targeting on easily observable characteristics suggests that, as treatment collectors pay the time-cost and learn more households' types, they switch out of the control collectors' strategy, which is less time-costly but also offers less returns, to target households based on observable predictors of payment. The fact that the negative differential targeting of easily observable characteristics holds when controlling for the hard to observe characteristics goes against a setting where treatment and control collectors apply the same targeting strategy of focusing on easily observable characteristics up to the number of bills delivered in the control group, and treatment collectors only seek to learn about propensity to pay for the *marginal* bills delivered in excess of the control group. Instead, and consistent with our model (Section 6), this result suggests that treatment collectors apply the learning strategy for their full set of delivered bills.

Propensity to pay is fully unobservable We begin with the setting where the hard to observe index of propensity to pay is perfectly unobservable, to all collectors at all points in time. In this setting, the allocation of interactions by collectors should be unrelated to the unobservable characteristics of households. This is a priori inconsistent with the differential targeting results: treatment collectors have more interactions than control collectors with households that have higher propensity to pay, both in terms of total number of visits and in terms of total time spent (Table 4 and Figure 4). Treatment collectors also conduct a larger share of their visits with households that are in the top 5% of the propensity to pay index (Figure A11).

Nonetheless, the differential targeting results could emerge if households' propensity to pay was truly unobservable but positively correlated with the observable household characteristics that treatment collectors do (differentially) target. These would likely be the observable characteristics that predict tax compliance – tax bill value, previous tax payment, observable assets, distance to main roads and commercial centers (Section 5.1 and Table A13). However, treatment collectors are less likely than control collectors to target these observable characteristics in their interactions, resulting in a negative selection on the 'easy to observe' index (Figure 4). Moreover, the positive differential targeting result on the hard to observe index holds when we control for the household's 'easy to observe' index (Figure A9).

In summary, there is little evidence to suggest that the household's propensity to pay is perfectly unobservable to all collectors throughout the tax campaign.

Propensity to pay is fully observable Let us now consider the alternative setting where a household's propensity to pay is perfectly observable to all collectors throughout the tax campaign. We will focus on the second targeting measure introduced in the main text (Figure 5), which shows positive differential targeting: treatment collectors conduct a larger share of their visits with the high propensity type households.

If all households are immediately available for interaction, both treatment and control collectors would start 'at the top' and initially focus their interactions on the households with the highest propensity to pay. Treatment collectors have more visits to conduct. There would be no differential targeting if, despite conducting more interactions, both treatment and control collectors allocate all of their interactions with the highest propensity households. There would be negative differential targeting if, by conducting more interactions, the treatment collector 'exhausts' the highest propensity type and conducts some visits with households that have lower propensity to pay. These scenarios are not consistent with the experiment's results, where treatment collectors conduct a larger share of their visits with the highest type.

To achieve positive differential targeting when propensity is perfectly observable requires a set-up where only past a certain threshold of total interaction time does the return per unit of time interacting with a high propensity household dominate the return per unit of time spent interacting with a low propensity household. In this case, because they have a larger total time available, treatment collectors may afford more interactions with the high-type than control collectors. Note that the positive differential targeting is not theoretically guaranteed as a general case in this set-up. That is, there are combinations of the parameter values of the threshold and the difference in total time available between groups which would generate negative differential targeting. For example, the parameters may be such that control collectors manage to allocate all their visits to the high type while the treatment collectors have some time 'left over' that falls below the threshold and which is therefore allocated to low type visits. In this case, treatment collectors would conduct a lower share of their interactions with the high-type than control collectors (even though treatment collectors can afford to conduct more visits in total with high type households).

Two sets of results reduce the plausibility of this alternative setting that generates positive differential targeting. First, it not clear what factors would generate these differential returns per unit of time, given our findings. The differential return per unit of time could capture the fact that trust, morale or enforcement perceptions may only be activated among high propensity households after a collector spends a long amount of time interacting with them during each visit. However, neither treatment nor control collectors spend longer time per visit with households that have higher propensity to pay - in fact, the associations in Table 3 suggest that collectors in both groups spend slightly less time per interaction with households that have higher propensity. We also find no treatment effects on morale, trust and enforcement perceptions – neither in general nor by level of the propensity to pay index (Table 2 and Table A8). The differential return per unit of time could alternatively reflect that the time cost paid to reach a household, before any interaction can take place, is larger for a high type household than for a low type household. For example, high type households may live further away from the collector's starting point than low type households and it may require a significant fixed time cost for the collector to travel to the high type households' location for interactions. However, we find that treatment collectors allocate no fewer interactions than control collectors to households that live further away from main roads, commercial centers and the local government office (which is the official daily starting point for collectors) – see Panel C of Figure A8.

Second, while treatment collectors conduct more visits than control collectors with high-type households that are in the top 5% of the propensity index, they also conduct *fewer* visits in total than control collectors with households that are in the bottom 5% of the propensity index (Figure A11). If propensity to pay was perfectly observable and control collectors could not afford to visit the highest 5% of types, it it not clear why they would devote more visits to the lowest 5% of types versus with any of the remaining middle 90% of households that have higher propensity.

In summary, the observed targeting of interactions could arise in a setting with perfectly observable propensity, but such a setting must accommodate the following: treatment collectors spend slightly less time per interaction with high propensity types; the factors that determine the difference in return to visiting high versus low propensity types cannot be captured by any of the observable characteristics or measures of morale, trust and beliefs; and, control collectors would have to be both less likely to visit the top 5% with highest propensity but also more likely to visit the bottom 5% with lowest propensity (despite all propensity values being fully observable).

Beyond the household survey results, the perfectly observable setting is also at odds with collectors' self-reported limited knowledge of households' location and their propensity to pay (Figure 2 and Figure 3).