Government Decentralization Under Changing State Capacity: Experimental Evidence from Paraguay*

Ernesto Dal Bó
UC Berkeley

Frederico Finan
UC-Berkeley

Nicholas Y. Li
UC Berkeley

Laura Schechter
UW-Madison

March 12, 2018

Abstract

Standard models of delegation assume that agents are better informed than principals about how to implement a particular task. We estimate the value of the informational advantage held by supervisors (the agents) when ministerial leadership (the principal) introduced a new monitoring technology aimed at improving the performance of agricultural extension agents (AEAs) in rural Paraguay. Our approach employs a novel experimental design in which, before randomization of treatment, we first elicited from supervisors which AEAs they believed should be prioritized for treatment. We semi-parametrically estimate marginal treatment effects (MTEs) and perform counterfactual exercises varying the principal's allocation rule and access to information. We find that supervisors did have valuable information—they prioritized AEAs who would be more responsive to the monitoring treatment. The AEAs' responsiveness is not easily observable to principals or analysts. We show both theoretically and empirically that the value of information and the benefits to decentralizing depend crucially on the sophistication of the principal and on the scale of rollout (i.e. the share of AEAs to receive treatment). When the principal is uninformed, decentralization usually dominates. A partially informed principal with data on basic observable AEA characteristics can outdo supervisors. The principal's advantage is largest if he can conduct a pilot RCT and subsequently expand roll-out based on predicted response to treatment. These results highlight the potential for evolving state capabilities for data analysis to alter government structure.

Keywords: Decentralization, Delegation, Bureaucracy, Monitoring, Marginal Treatment Effects

*We would like to thank José Molinas, the Minister of Secretaria Técnica de Planificación del Desarrollo Económico y Social (STP), whose support and guidance made this project possible. Maureen Stickel provided excellent research assistance. We also gratefully knowledge the IGC and JPAL-Governance Initiative for their generous financial support.

1 Introduction

Standard models of delegation assume that agents possess superior information to that of principals and conclude that devolution of decision-making powers to agents is a way to take advantage of that superior information (Aghion and Tirole (1997), Dessein (2002), Mookherjee (2006)). Decentralization – a key way in which governments delegate power – comes with informational gains but also costs as it requires administrative capabilities at lower levels. Moreover, the value of the information available under decentralization may depend on the planned scale of operation, which may or may not justify the costs. Suppose for example that an organization plans to provide assistance to low-income families. If resources are sufficient to cover all households, then it does not pay to decentralize the selection decision: the program can be rolled out from the center with universal coverage. If resources are insufficient to cover all households, then it might be important to decentralize the program by creating local branches to screen households and prioritize assignment according to need. But if resources are so meager that very few families can be covered in each district, it may again be inconvenient to pay the cost of developing the local branches and preferable for the center to pick recipients based on the knowledge at hand.

How the state rolls out a new monitoring technology among its front-line providers—the subject of this paper—raises similar considerations. In 2014, the government of Paraguay decided to roll out a new monitoring technology (a GPS-enabled cell phone) for supervisors to track their agricultural extension agents (AEAs). AEAs are tasked with visiting farmers scattered over large tracts of land and give them access to various support services including timely information about prices and best farming practices. In the eyes of the central government, AEAs were likely shirking due to the monitoring difficulties afflicting their supervisors, and GPS phones could help mitigate the problem. Because the government did not have the resources to provide phones to all the AEAs at once, they faced two questions: 1) what should the extent of the roll-out be? and 2) should the supervisors, who presumably had some understanding of which AEAs would best respond to the new technology, decide who should receive the phones or should the central government allocate them? In addition to administrative costs of devolving this decision to supervisors, the answers to these questions hinge not only on the amount of the information supervisors possess but also on the fact that the value of that information depends on the scale of planned roll-out.

In this paper, we examine the impact that the new monitoring technology had on the performance of AEAs as measured by the share of their assigned farmers that they visited in a given week. Based on a novel experimental design, we develop an approach that allows us to measure not only the value of supervisors' information, but also how the value of information varies at different levels of coverage. Specifically, we first elicited the preferences of supervisors regarding which half of their AEAs should be prioritized to receive the phone. We then randomly assigned phones to AEAs. AEAs thus fell into one of four cells of a 2-by-2 treatment-by-selection matrix. This simple design allows us to measure both the average impact of the monitoring technology as well as the differential impact on the AEAs that the supervisors thought the treatment would impact most. This latter estimate quantifies the advantage supervisors possess in targeting the cell phones over random assignment.

We find that the cell phones had a sizeable effect on AEA performance, increasing the share of farmers visited in the last week by an average of 6 percentage points. This represents a 22 percent increase over the AEAs in the control group. Because we do not find any impact of cell phones on AEAs who do not have supervisors, we interpret this effect to be a result of increased monitoring as opposed to the cell phones directly improving productivity. Also consistent with our interpretation, we find that AEAs under new monitoring are 21 percent more likely to agree with the statement that their supervisor knows their whereabouts. We find no evidence that treated AEAs increased the number of visits at the cost of conducting shorter ones.

Importantly, supervisor-chosen AEAs respond more to increased monitoring, entirely driving the average increase of 6 percentage points. Among these AEAs, treatment increased the likelihood that a farmer was visited in the past week by 12 percentage points compared to a statistically insignificant 2 percentage points decrease among those who were not selected. This finding corroborates the notion that going down the hierarchy from the top program officers to local supervisors on the ground could allow the organization to leverage superior, dispersed knowledge about how best to allocate treatment.

Supervisors have superior information regarding AEA characteristics, only some of which are observable to the principal or an econometrician. Having collected a rich dataset on the AEAs, including information on both cognitive and non-cognitive traits, we develop a two-step estimation procedure in the spirit of a sample selection model to decompose the value of information into observable and unobservable traits of an AEA. We use this to compute a series of marginal treatment effects under various selection rules and coverage rates. These marginal treatment effects are critical inputs to decide whether to decentralize the treatment assignment decision. In addition, the approach we develop would allow program leadership to optimize the program's roll-out scale.

We find that the information which supervisors use to most efficiently target the monitoring technology is neither explained by commonly observed demographic traits of the AEAs (e.g., age or gender) nor by harder-to-measure characteristics such as their IQ or personality type. When we allow the treatment effects to vary by a rich set of observable characteristics, it is the unobservable component of the supervisors' choices that most robustly predicts the responsiveness of an AEA to the additional monitoring.

While our findings suggest that supervisors have valuable information, the decision of whether to decentralize depends on the information held by the principal, the feasible allocation rules she could adopt, and the extent of available resources. In order to explore the potential for centralized vs decentralized assignment, we construct a number of counterfactuals corresponding to different degrees of sophistication of the central authority. Given our model estimates, we compute marginal treatment effects for every AEA in our sample to measure the program's impact under four different allocation rules at varying coverage rates. In particular, we consider 1) a totally uninformed principal who allocates randomly; 2) a minimally informed principal who targets AEAs who have to travel longer distances; 3) a more sophisticated principal who collects and analyzes baseline data on AEAs and targets predictably low productivity AEAs; and 4) the most sophisticated principal who pilots an experiment and thereafter targets AEAs in descending order of predicted responsiveness to treatment.

We find that the value of supervisor information is substantial relative to a regime in which the principal simply allocates phones at random and that this difference in program impact is maximized at 54 percent coverage. At this coverage, the supervisor allocation increases the share of farmers visited by 7.9 percentage points versus only a 3.6 percentage point increase under random assignment. A slightly more effective approach would be to simply allocate the phones to the AEAs who have to travel the longest distance to attend to their farmers. This method generally outperforms random assignment (a 2.4 percentage point advantage at 50 percent coverage), but the supervisor still outperforms this simple assignment mechanism.

A more effective centralized policy identifies and prioritizes the workers who are expected to be the least productive. This strategy does not rely on reports from the supervisors. We operationalize this policy by estimating among the control AEAs the relationship between AEA productivity and observable characteristics, and then predicting productivity without a GPS-phone for all AEAs. Assuming that the government has the information and capacity to approximate this procedure, we find that the government can perform at least as well as, and in many cases better than, the supervisors. The reason is that while this minimally informed principal does not have as much information as the supervisor, it is possible to make better use of it than supervisors appear to do. This in turn suggests that imperfect processing of information, or bias, prevents supervisors from being as effective as they could.

The most effective but most information-demanding centralized policy we consider uses AEA observables to predict response to treatment rather than to predict baseline productivity. This would require the principal to first conduct a pilot experiment, the results of which would be used to predict responsiveness among the remaining untreated AEAs. Treating AEAs in descending order of predicted responsiveness, even when lacking information on unobservables, vastly outperforms decentralized assignment by the supervisors. The high performance of these last two methods highlight that innovations in information and communication technologies, as well as the introduction of experimental methods to inform policy, can play a role in reducing information frictions and alter optimal organizational structure.

Our study speaks to several literatures. First and foremost, our study contributes to a large but mostly theoretical literature on why organizations decentralize decision-making authority.[1] Recently, some empirical progress has been made in understanding why private-sector firms decentralize. For instance, based on the insight by Aghion and Tirole (1997) that organizations are more likely to decentralize if the principal and agent have congruent preferences, Bloom et al. (2012) show that plants are more decentralized in regions where people are more trusting, even when measured bilaterally (i.e. trust that people in the headquarters region have in people in the plant location).

Given the standard assumption that agents are better informed than the principal, access to costly information can also determine a firm's decision to decentralize. For example, Acemoglu et al. (2007) show using data on French and British firms in the 1990s that firms closer to the technological frontier, firms in more heterogeneous environments, and younger firms are more likely to choose decentralization—settings that presumably proxy for environments where learning is more difficult. Despite the progress that these and other studies have made, direct empirical evidence on the existence of superior information by agents is still lacking.

One notable exception is Duflo et al. (2016) who conducted a field experiment that increased the frequency of inspections of industrial plants in Gujarat, India. In the control group, plants were audited as usual at the discretion of the inspectors, whereas in the treatment group, the audits were conducted more frequently but at random. They found that despite the increased regulatory scrutiny, the treatment plants did not significantly reduce pollution emissions. This is because the discretionary inspections targeted the plants with higher pollution signals. Because the largest penalties are reserved for extreme pollution violations, this is the population whose behavior is most likely to be impacted by audits.

[1] Mookherjee (2006) provides an excellent review of the theory on decentralization.

We complement this study in two important ways: 1) we show theoretically as well as empirically that the decision to decentralize depends critically on the scale of roll-out, while introducing an approach to quantify the value of information at different levels of roll-out – to this effect, we develop a method to rank potential recipients of treatment by estimating heterogenous marginal treatment effects; 2) while a random allocation mechanism is a useful benchmark, it is unlikely that the principal is completely uniformed. In our study, we quantify the value of decentralized information against not only a random allocation rule, but other selection rules that proxy for centralized principals with different levels of information.

Similar to the public sector, private sector employers also need to monitor their employees. de Rochambeau (2017) discusses the roll-out of GPS tracking devices in a trucking company. She finds that managers allocate the tracking device to drivers who perform less well at baseline, and that these truckers benefit most from the device. In fact, de Rochambeau finds that monitoring high-performing drivers can be counter-productive as their intrinsic motivation decreases.

The problem of how best to deploy monitoring technology is similar to the issue of how best to target social programs. In this regard, our paper is most related to two studies. Alderman (2002) examines the targeting of an Albanian social assistance program. He shows that even after controlling for the assets that were used in the targeting of the program, household consumption was predictive of the targeting. The author interprets this as evidence that local officials responsible for targeting the program relied on their local information. Alatas et al. (2012) conducted a field experiment in 640 villages in Indonesia to compare proxy-mean testing against community-based targeting of a social program. They find that proxy-mean testing targets consumption better than community-based targeting. They argue that this difference is not due to elite capture or local information, but rather a difference in how local communities define poverty. Similar to difficulties in targeting social programs, banks could benefit from community knowledge to help them lend to the most entrepreneurial people. Hussam et al. (2017) find that community members have useful information on marginal returns and this information is useful above and beyond what a machine learning algorithm would predict from observables.

Our study also has clear parallels to the literature on marginal treatment effects (MTE) (Heckman and Vytlacil (2005)). As in the MTE literature, we express our evaluation problem as a joint model of potential outcomes and selection as determined by a latent index crossing a threshold. Different from the standard MTE setup, our selection equation does not model an AEA's self-selection into treatment but rather the selection by a supervisor (in our case, the agent). Crucially, treatment was given to only those AEAs who were chosen by our design. Thus, when we compute the MTEs we do not have to extrapolate across subgroups of “always-takers" and “never-takers" because we only have compliers by design.

Finally, our study adds to a growing body of experimental evidence on the impact of new monitoring technologies for reducing shirking in the public sector. Similar to our setting, some of these studies involve weak or no explicit financial incentives. For example, Aker and Ksoll (2017) monitored teachers of adult education in Niger by calling both the teacher and the students to ask whether the class was held and how many students attended. They found that the calls led to fewer canceled classes and better student test scores. Callen et al. (2015) used a similar cell phone technology to monitor health facility inspectors and found that this increased the frequency of inspections, especially for those with ‘better’ personality traits.

Other studies have introduced new technologies for monitoring but have also overlaid financial incentives. For instance, Duflo et al. (2012) required teachers to take a picture of themselves with their students at the beginning and end of each school day using a camera with tamper-proof date and time functions, whereas Banerjee et al. (2008) asked nurses to time-stamp a register at the beginning, middle, and end of the day. Both studies found these treatments increased teacher and nurse attendance, but in both cases, the impact was found to be mostly due to concomitant financial incentives. Dhaliwal and Hanna (2017) found that fingerprint readers in health centers decreased absence even though financial incentives provided by the monitoring technology were rather weak. Banerjee et al. (2015) and Khan et al. (2016) look at on-the-job performance rather than attendance (among police and tax collectors respectively) and employ both monitoring and incentives. These papers do not give a definitive answer regarding whether most of the improvement in performance is due to the monitoring or the incentives. The first paper suggests a significant impact of monitoring alone, while the second suggests an insignificant impact.

We contribute to this literature by showing that a cell phone technology can be effective in reducing shirking for individuals such as agricultural extension agents whose job requires them to visit farmers who live out in rural areas, often quite far from the local agricultural ministry offices in town.

2 Background

Agricultural extension services in Paraguay are centered around the Ministry of Agriculture based in Asunción. Below the central ministry are 19 Centros de Desarrollo Agropecuario (CDAs, which exist at the department level, similar to a state in the United States) and below the CDA level there are 182 Agencias Locales de Asistencia Técnica (ALATs, which are at the municipality level, similar to a county in the United States). The Paraguayan Ministry of Agriculture has close to 1000 agricultural extension agents working within ALATs spread across four main agencies. We work with the biggest of these agencies, Dirección de Extensión Agraria (DEAg).

The main job of extension agents is to help farmers access institutional services that will help them improve their production. The goal is to increase farmers' output directed both for own consumption as well as the market. Another goal is to increase farmers' connection to, and participation in, markets. The official thematic areas are soil improvement, food security, product diversification, marketing, improvement of life quality, and institutional strengthening. Much of what extension agents do resembles the role of middlemen, connecting farmers with cooperatives, private enterprises, and specialists. Extension occurs both one-on-one and in groups. Extension agents conduct group meetings to talk about technical topics, conduct farm visits, and lead product or process demonstrations. This includes working on demonstration plots and organizing farmer field trips. Each extension agent is assigned to work with approximately 80 producers. Extension agents do not usually offer free goods or services to farmers. Although the headquarters for extension agents are in towns, most of their daily work involves driving out to visit farmers in the rural areas where these farmers live and work. Extension agents come from a variety of backgrounds including agricultural sciences, veterinary sciences, nutrition, law, and teaching.

Within every ALAT there is a supervisor who, in addition to working with his own farmers, must also monitor the other extension agents working in the ALAT. We will refer to individuals who work purely as agricultural extension agents, as 'AEAs.' By this definition, DEAg has over 200 AEAs working within the organization at any time.

In June 2014, the Ministry of Planning, in association with the Ministry of Agriculture, decided to provide AEAs with GPS-enabled cell phones. This initiative had several objectives. One was to improve coordination and communication between the AEAs and their supervisors. For example, it would give the AEA a mechanism to take a picture of a farmers' crop which was suffering from some pest, circulate it, and get a response for the farmer of how to deal with that pest. But crucially, it would allow the supervisors to see where AEAs were at all times, how long they spent in each place, and what they did there (since the AEA is supposed to document every meeting he participates in). AEAs can submit reports and review reports they have already submitted through the phone. Supervisors, in turn, as well as CDA-level managers, can view reports submitted by all the AEAs they oversee.

In the terms of hierarchical agency, we view the ministerial leadership introducing the new technology as the principal, the ALAT-level supervisors as the agents, and the AEAs as the lowest-tier.

3 Model

Consider a hierarchy composed by a principal, a supervisor, and a very large number n of agents. The supervisor is needed to routinely monitor the agents, and as will be discussed later, may also be relied upon to deploy a new monitoring technology.

Agents and monitoring The n agents each cater to a mass 1 of farmers. A visit by an agent i yields a constant benefit B to each farmer. Agents receive a wage w and chooses the share $s_i \in [0,1]$ of farmers to visit. The agent faces a cost $a_i s + b\frac{s_i^2}{2}$ of expanding visits, where a_i can potentially be negative for each $i = 1,...,n$ since it captures both linear costs and a linear intrinsic motivation term. The supervisor operates a monitoring technology that allows him to obtain a signal of the agent's effort. With probability q_i the supervisor learns s_i and reprimands the agent in proportion to the extent by which his effort fell short, namely by an amount $(1-s_i)$. The agent gets a disutility from this reprimand equal to $(1-s_i)\,r_i$, where r_i is an agent-specific type independently drawn from a distribution $F(r_i)$ over a support $[r_l, r_h]$. The highest type suffers the most from being reprimanded such that $r_h > 0$ while r_l could potentially be negative if the lowest type enjoys being reprimanded. We assume $r_l \geq \frac{a_l}{q_l}$ to guarantee non-negative effort by agents.[2] q_i tracks the state of monitoring technology over agent i. It is possible to extend the model to make q_i a function of an endogenous choice of supervision effort by the supervisor as well as technology, but here we stick with the simplest formulation.[3]

The timing is as follows: each agent learns q_i as given by the monitoring technology applied to him, and he chooses the share s_i of farmers to visit to maximize utility

$$u_i(s_i) = w - a_i s_i - b\frac{s_i^2}{2} - q_i\,(1-s_i)\,r_i.$$

Thus, the choice of s_i by the agent is both a measure of his effort and his output in terms of a level

[2] An alternative interpretation is that with probability q_i the supervisor can take an effective draw and inspect whether a farmer under the watch of the agent has been visited. In case the farmer has not been visited (which happens with probability $(1-s_i)$), the supervisor reprimands the agent.

[3] Such extension can also add an agency problem in the supervisor's choice of monitoring effort without affecting the essence of our results. Here we maintain our sights on the potential tension between principal and supervisor when it comes to the deployment of monitoring technology.

of service to farmers. Optimization by the agent immediately yields optimal agent effort to be,

$$s_i^*(q_i) = \frac{q_i r_i - a_i}{b}. \tag{1}$$

While a_i and r_i both affect the level of effort, only r_i affects the response of effort to a change in monitoring technology. Since b only affects effort through ratios involving a_i and r_i, parameters that we can scale arbitrarily, we henceforth adopt the normalization b=1.

New technology and treatment effects Assume that in an initial state of affairs, monitoring technology is relatively bad, so $q_i = q_l$ for all agents i. Suddenly, a better monitoring technology becomes available that raises q_i to $q_h > q_l$ if agent i is placed under the new technology. That means an agent of type r_i will increase his effort/output by $r_i(q_h - q_l) \equiv r_i \Delta q \equiv T(r_i)$, which captures the treatment impact of the new technology over agent i. Note that since r_l can be negative, $T(r_i)$ can be negative for some types. Deploying the new monitoring technology on any given agent costs an amount c per agent. If the new technology is deployed over all agents, then all agents are "treated," and given a very large number of agents, we can approximate the total (and average) treatment impact with $\int_{r_l}^{r_h} T(r) f(r) dr$, achieved at a total (and average) cost c. If the new technology is deployed over all agents above a cutoff k, the total treatment impact is $\int_k^{r_h} T(r) f(r) dr$, achieved at a total cost $c(1 - F(k))$.

Inspection of the expression for the average treatment impact $\int_{r_l}^{r_h} T(r) f(r) dr$ and for $T(r_i)$ immediately yields the following:

Remark 1. *If $r_l \geq 0$, the average treatment impact is guaranteed to be positive. If $r_l < 0$, the average treatment impact is positive if and only if the density $f(\cdot)$ places enough weight on positive types and the difference Δq is large enough.*

This remark is obvious yet highlights conditions for an impact evaluation of the new technology rolled out at 100% to yield positive results.

Remark 2. *Equilibrium agent effort s_i^* increases in monitoring technology q_i, and an improvement in monitoring technology (an increase in q_i) has a stronger effect on the effort of agents with a higher type r_i.*

The value of information and optimal decentralization We now isolate a central advantage of decentralization stemming from information possessed by the supervisor under strong assumptions that will be relaxed in our empirical implementation. We assume the principal knows the distribution of types $F(r)$, but does not know which agents have what type. The supervisor knows $F(r)$ and can also costlessly learn the type of each agent. Both principal and supervisor know the rest of the parameters in the model.

Centralization Since the principal has no information to recommend changing monitoring over any one agent versus another, her only decision is whether to adopt the new technology over all agents or not at all. She will adopt fully whenever $\int_{r_l}^{r_h} T(r) f(r) dr \geq c$ (breaking indifference in favor of adoption), i.e., whenever the average treatment effect of the new technology is larger than its marginal cost. If this condition is met, a roll-out at 100% would produce a total treatment impact equal to the average treatment effect. If roll-out is done at any lower scale m while selecting the recipients of treatment at random, the total treatment impact will be $m \int_{r_l}^{r_h} T(r) f(r) dr$ which increases linearly in m as illustrated by the strictly linear curve in Figure 1.

Decentralization Consider the case where the principal can pay a cost $d \geq 0$ to delegate to the supervisor the decision over what agents to place under the new monitoring technology.[4] For simplicity, we focus now on a well-meaning supervisor who deploys the new technology to maximize its effect over agent output and deal later with potential supervisor bias. If the marginal cost of the new technology is lower than its treatment effect over the highest type r_h, a benevolent supervisor will place agents under the new monitoring system starting from the highest type r_h and work downwards. How far down he goes depends on the scope of roll-out for the new technology. If the supervisor has discretion over what agents to treat but not over the scale of roll-out, he will have to treat some mandated share of agents and will choose the highest types to fill that quota. Thus, if the supervisor is told to place a measure m of agents under the new technology, he will treat every agent with type $r \in \left[r_m = F^{-1}(1-m), r_h\right]$. If in addition, the supervisor also has control over the scale of technology adoption, he will choose the lowest treated type k to maximize,

$$\int_{k}^{r_h} T(r) f(r) dr - c(1 - F(k)),$$

which immediately yields $T(k^*) = c$. In words, the supervisor will choose to treat every agent down to a type k^* whose marginal treatment effect from the new technology equals the marginal cost.

Optimal decentralization

It is helpful to consider the case when the supervisor, under decentralization, has authority over the selection of agents to be treated but not over the scale of technology adoption. A realistic example of this is when a new technology is acquired by government and is then made available

[4]The delegation cost can arise due to the need to transfer certain administration means to the supervisor or from establishing additional communication and administration channels to track not only the supervisor's routine oversight over agents, but also the supervisor's recommendations and/or technology deployment decisions.

to an agency in a fixed amount. Given a scale of adoption that allows a measure m of agents to be treated, the principal will choose to decentralize if and only if $\int_{r_m}^{r_h} T(r) f(r) dr - cm - d \geq \left(\int_{r_l}^{r_h} T(r) f(r) dr - c\right) m$, or equivalently, iff,

$$\iota(m) \equiv \int_{F^{-1}(1-m)}^{r_h} T(r) f(r) dr - m \int_{r_l}^{r_h} T(r) f(r) dr \geq d, \tag{2}$$

where $\iota(m)$, graphed in the bottom panel of Figure 1, captures the informational gain from decentralization. This gain is the difference between the total treatment effect that can be attained through the centralized and decentralized approaches, graphed in the top panel of Figure 1. Note that when $m = 0$ and $r_m = r_h$, the marginal gain from expanding roll-out under the decentralized scheme is at a maximum since the supervisor would treat the most responsive agent first, but it applies to almost no agents so the value of the informational gain from decentralization is zero and does not justify paying a positive cost d. On the other extreme, where $m = 1$ and $r_m = r_l$, the difference in value again goes to zero because the advantage of treating the more responsive agents first is completely diluted. For every value m strictly between 0 and 1, the total treatment effect attained by a supervisor who treats the most responsive first is larger than that which can be attained by assigning treatment at random. As illustrated in Figure 1, it is easy to show that the value of information $\iota(m)$ is increasing in r_m near r_l (or decreasing in m near m=1) up to a type $\bar{r}$ whose marginal treatment effect $T(\bar{r})$ is equal to the average treatment effect $\bar{T} \equiv \int_{r_l}^{r_h} T(r) f(r) dr$ and then decreasing as r_m approaches r_h (or increasing as m approaches 0).[5] This implies

Proposition 1. *(i) If $0 < \iota(\bar{r}) < d$, then decentralization is never optimal and if $d = 0$, decentralization is always optimal. (ii) If $0 < d < \iota(\bar{r})$, there exist two values $m' < m''$ in $[0,1]$ such that for any extent of roll-out of the new technology $m \in [m', m'']$ the principal prefers decentralization to centralization; for $m \notin [m', m'']$, centralization is preferred.*

This proposition establishes that the case for decentralization rests on the value of its informational gain relative to its cost, but the value of the informational gain depends crucially on the extent to which the new technology is to be adopted.

The model makes clear that there are interventions which can never yield positive value if implemented centrally and/or fully, but could deliver value if implemented in a decentralized manner with a limited roll-out. Such an intervention satisfies $\int_{r_l}^{r_h} T(r) f(r) dr < c$, yielding an average treatment impact below marginal and average cost but also satisfying $T(r_h) > c$, so that there are treated types with treatment effects larger than the marginal cost. In this sense, there is no such thing as a

[5]Differentiating $\iota(m)$ we get $(T(r_m) - \bar{T}) f(r_m)$, where $f(\cdot) > 0$.

pure impact evaluation that abstracts from the extent of roll-out and its implementation mode. It is important to be able to determine the likely total treatment impact under different roll-out extents under both the centralized and decentralized approaches.

Our empirical study will investigate three claims stemming from the two remarks and proposition derived in this section. First, does the intervention at hand deliver a positive treatment impact on average? Second, do supervisors have valuable knowledge about which agents ought to be treated given partial roll-out? Third, does roll-out extent could alter the relative convenience of decentralization vs centralization? This will require developing a method for ascertaining the marginal treatment impact on different types of agents while contemplating centralization approaches with varying levels of information.

4 Research Design

Our experiment was conducted on 180 local technical assistance agencies (ALATs, Agencia Local de Asistencia Técnica). On average, each ALAT consists of a supervisor and three agricultural extension agents. Some ALATs have a single extension agent, but 48 ALATS have at least 2 agents. We asked the supervisors of the latter group to indicate which half of his agents should receive the phones first given the program's objective to increase agent performance. We refer to these agents as the "selected agents." These 48 ALATS were then randomly assigned into three groups according to how and when the agents would receive their phones.

The main group of ALATs is in cells A, B, D and E in Figure 2. The ALATs in cells B and E (25 percent of the total) had both selected and non-selected agents receiving the phone. The ALATs in cells A and D, which contains half of the ALATs, serve as our control group. The difference in performance between agents in cells B and E and agents in cells A and D estimates the average impact of treatment. And, the difference-in-differences computed as the performance by agents in cells $B-A$ against cells $E-D$ estimates whether selected agents have a larger impact of treatment than the non-selected agents, which allows us to determine whether supervisors had valuable information on how to direct treatment. A third group of ALATs (in cells C, F) had only those agents who had been selected being treated immediately (in cell C). This helped make the elicitation of supervisors' preferences credible and relevant, and it in principal allowed us to detect spillover effects.[6] Eight months after the delivery of these phones, a second wave of phones were delivered to the ALATs in group F. All our main analysis restricts attention to the shaded cells (A, B, D, E).

[6]We do find some evidence of spillovers; these results are available upon request.

4.1 Taking the Theory to the Data

Recall that the performance of AEAs is given by their effort expression:

$$s_i^* = r_i q_i - a_i.$$

To operationalize this equation, we assume that the level of monitoring for each AEA, q_i, is a function of the monitoring technology t_i, according to the expression $q_i = q_l + q_h t_i$ where t_i takes value 0 when AEAs do not get a cell phone and 1 when they do. Because our objective is to see agents respond to exogenous changes in q_i, we normalize $q_l = 0$ and can rewrite the expected disutility of being reprimanded to be $r_i q_i = \beta_i t_i$, where $\beta_i = q_h r_i$. The empirical analogue of a_i is denoted $-\alpha_i$.

The central aspect of our approach is modeling various selection criteria and estimating the marginal treatment effects created under a given criterion for varying levels of roll-out. A key element in this approach will be to consider different degrees of observability of the individual parameters (α_i, β_i) in an individual agent's effort function in equation $s_i^* = \alpha_i + \beta_i t_i$. In particular, we will map these parameters into the vector of fixed characteristics (X_i) and two independently random characteristics (ε_i, η_i), to write: $\alpha_i(X_i, \varepsilon_i)$ and $\beta_i(X_i, \eta_i)$. While the vector X_i is potentially observable by the principal, the elements (ε_i, η_i) are only potentially partially observable by the supervisor.

Average treatment effect We can estimate the average treatment impact of the cell phone on effort by imposing some familiar (but mild) structure on individual parametric heterogeneity as follows: $\alpha_i = \alpha' X_i + \varepsilon_i$ and $\beta_i = \beta_0$. An individual agent's effort function becomes

$$s_i^* = \alpha' X_i + \beta_0 t_i + \varepsilon_i, \tag{3}$$

where s_i^* measures the share of farmers AEA i visited in the past week. The coefficient β_0 provides a causal estimate of the difference in performance between agents in both treated cells B and E relative to agents in the control cells A and D. Thus, the first theoretical claim that the intervention yields positive value is captured by contrasting the null hypothesis $\beta_0 = 0$ against the alternative $\beta_0 > 0$.

Given our research design, we cluster the standard errors at the ALAT level and also report p-values based on a wild-bootstrap procedure to account for the fact that we have relatively few clusters. In estimating Equation 3, we can also include the single agent ALATs, which were also assigned phones at random. In this randomization, one-third of agents initially received a phone, with two-thirds serving as a control. When including these ALATs, the vector X_i contains an indicator for whether or not the ALAT has a single agent.

Average treatment effect by supervisor's choice To test whether supervisors are able to select those agents who would most benefit from being monitored, we can simply re-parameterize $\beta_i = \beta_0 + \beta_1 D_i^S$, where D_i^S is an indicator for whether AEA i was selected to receive a phone. Equation 3 then becomes

$$s_i^* = \alpha' X_i + \beta_0 t_i + \beta_1 (D_i^S \times t_i) + \varepsilon_i, \tag{4}$$

where included within the vector X_i is the indicator D_i^S. With this specification, we can compare the difference in performance between selected agents and non-selected agents in the treatment group (cells $B-E$) net of the difference in selected versus non-selected in the control group (cells $A-D$). Thus, the second theoretical claim that supervisors have valuable information about which AEAs should be targeted is captured by contrasting the null hypothesis $\beta_1 = 0$ against the alternative $\beta_1 > 0$. We directly observe s_i^* and can thus estimate α' and $\beta = (\beta_0, \beta_1)$ directly via ordinary least squares since t_i is randomly assigned by design, D_i^S is elicited in a way that does not affect treatment assignment.

4.2 Estimating the Marginal Treatment Effects of the Program

Since supervisors were asked to select 50 percent of their AEAs, a value $\beta_1 > 0$ when estimating equation 4 is a necessary condition for a decentralized approach to be preferred, but it is not a sufficient condition. Two other considerations are pertinent. First, is the value β_1 large enough to justify paying the cost d of decentralization? Second, what would be the average treatment effect at scales other than 50 percent? The 50 percent pilot implementation does not directly tell us what β_1 would be at a revised scale. In this section we develop a method for tracing out the value of impact for all possible roll-out scales under different implementation regimes that vary the degree of informational advantage associated with decentralization.

Marginal treatment effects under different selection models

In order to lay out the main intuitions surrounding the value of decentralization, our theory considered the stark contrast between a totally uninformed principal and a fully informed, benevolent supervisor. We will allow for intermediate cases in our empirical approach – the econometric operationalization of the theory will in fact extend it in two directions. First, we allow for supervisors to be less than fully benevolent. Second, we allow them to be less than perfectly informed about the responsiveness of agents to treatment. In addition, this framework will allow us to consider versions of a principal who is partially informed.

Each organizational situation – decentralization or centralization under different informational capabilities of the principal – will be modeled as leading to the selection of agents according to a suitably defined latent index model.

In the case of decentralization, supervisors select agents according to

$$v_i = \beta_i(X_i, \eta_i) + \psi_i(X_i, \zeta_i),$$

where v_i is agent i's desirability for selection as seen by the supervisor, $\beta_i(X_i, \eta_i)$ represents the heterogeneous effect of receiving the cell phone and ψ_i is a preference for giving AEA i monitoring technology that depends on X_i and an independent, idiosyncratic preference term ζ_i. A benevolent supervisor would only select agents based on an index $v_i = \beta_i(\cdot)$. Thus, the additional term ψ_i captures the potential non-benevolence of the supervisor. In addition, supervisors do not observe η_i perfectly but, instead, observe a signal $\theta_i = \eta_i + \xi_i$, where $\xi_i \sim F_\xi(\cdot)$ is a white noise (hence mean zero) term; as the variance of ξ_i goes to zero, the supervisor gets closer to being perfectly informed. Given the random elements (ζ, ξ), the supervisor faces uncertainty, and will assign monitoring technology to individuals depending on the expected value $\mathbb{E}\{v_i|X_i, \theta_i, \zeta_i\}$.

Given a selection criterion (such as v_i), and a well defined measure of diversity across agents as given by a joint distribution over (X_i, θ_i, ζ_i), it is possible for the supervisor to rank order all agents according to the value $\mathbb{E}\{v_i|X_i, \theta_i, \zeta_i\}$, with minimum element $\underline{Ev}$ and maximum element $\overline{Ev}$ to such order. We assume there is enough variation that the rank order is strictly monotonic. Therefore, any roll-out of extent m under a selection criterion based on v_i implies treating all agents who satisfy $\mathbb{E}\{v_i|X_i, \theta_i, \zeta_i\} \geq c_p(m)$, where $c_p(m)$ is a putative cost (hence the subscript). This cost is putative in the sense that it is the cost of treatment that the supervisor would have to perceive in order to decide to treat a share m of agents. Thus, $c_p(m)$ satisfies $\frac{dc_p}{dm} < 0$, $\lim_{m \to 0} c_p(m) = \overline{Ev}$, and $\lim_{m \to 1} c_p(m) = -\underline{Ev}$. These conditions say that if the supervisor wants to treat more agents, the cost of treatment must be lower; if the supervisor wants to treat no agents, the marginal cost of treating a single agent exceeds the benefit of treating the most valuable agent; and that if the supervisor wants to treat all agents, the expected desirability of treating the least valuable agent exceeds the cost. When $\mathbb{E}\{v_i|X_i, \theta_i, \zeta_i\} \geq c_p(m)$ is true, the selection indicator denoted by

$D_i^M(X_i, \theta_i, \zeta_i, c_p)$ takes the value 1, and 0 otherwise.

The fundamental difference between X_i and η_i is that elements in the vector X_i are potentially observable by a sophisticated principal who can gather and analyze data. Elements in X_i could contain agent-related demographic and psychometric data. The term η_i is fully unobservable to the principal, and can potentially be known only to a supervisor who establishes a more personal connection with the agent. Thus, decentralization has two potential informational advantages: supervisors may (or may not) know and use data on X_i better than the principal, and they are the only ones who can potentially know something about η_i. To the extent that η_i enters the function $\beta_i(.)$ the supervisor will have an unassailable informational advantage over the principal.

To make further progress, we need to parameterize the dependence of $\alpha_i(\cdot)$, $\beta_i(\cdot)$, and $\psi_i(\cdot)$ on X_i. We parameterize each of these linearly. Slightly abusing notation we can re-write equation 3 as,

$$\begin{aligned} s_i^* &= \underbrace{(\alpha' X_i + \varepsilon_i)}_{-\alpha_i(\cdot)} + \underbrace{(\beta' X_i + \eta_i)}_{\beta_i(\cdot)} t_i \\ &= \alpha' X_i + (\beta' X_i) \times t_i + \varepsilon_i + \eta_i \times t_i. \end{aligned} \tag{5}$$

Marginal treatment impact under an uninformed principal An uninformed principal knows nothing about individual values of β_i, so she can only select which agents should be placed under the new technology at random. Given a scale of roll-out m (for the measure of agents to be treated), and conditional on X_i, the total difference in expected performance of those selected by the principal against the non-selected is

$$\int_{X_i} \left(\left(E_{\varepsilon,\eta}(s^* | t = 1, X_i) - E_{\varepsilon,\eta}(s^* | t = 0, X_i) \right) m \right) d\Xi(X_i) = m\beta',$$

where Ξ is a cumulative distribution function describing variation in the vector X, which is unobservable to a fully uninformed principal. This equation says that if no agents are treated, the total gains are zero. If all agents are treated, the total gains are equal to the average treatment effect of the intervention. If a partial measure $m \in (0,1)$ is treated, the total gains are proportional to roll-out m, and the marginal impact of enhancing roll-out is always the average impact $\beta' \bar{X}$.

Marginal treatment impact under decentralization A supervisor observes each agent's characteristics (X_i, θ_i, ζ_i), and selects agents to treat according to the value of the expected index

$\mathbb{E}\{v_i|X_i,\theta_i,\zeta_i\}$ as given by,

$$\begin{aligned}\mathbb{E}\{v_i|X_i,\theta_i,\zeta_i\} &= \underbrace{\left(\beta' X_i + \mathbb{E}\{\eta_i|X_i,\theta_i\}\right)}_{\mathbb{E}\{\beta_i(\cdot)|\theta_i,X_i\}} + \underbrace{\left(\psi' X_i + \zeta_i\right)}_{\psi_i(\cdot)} \\ &= \underbrace{\left(\beta' + \psi'\right)}_{\Gamma'} X_i + \underbrace{\left(\mathbb{E}\{\eta|X_i,\theta_i,\zeta_i\} + \zeta_i\right)}_{u_i}. \end{aligned} \tag{6}$$

A key hurdle is that we do not have a direct measure of $\mathbb{E}\{v_i|X_i,\theta_i,\zeta_i\}$, but we only observe the supervisor selection decision D_i^S. To recover Γ, we further assume that η_i, ξ_i, and ζ_i are mean zero, normally distributed random variables with variances σ_η^2, σ_ξ^2, and σ_ζ^2, respectively. This implies that D_i^S takes the familiar form of a probit model:[7]

$$Pr\left\{D_i^S = 1|X_i\right\} = \Phi\left(\frac{1}{\sigma_u}\Gamma' X_i - c_p(m)\right)$$

where Φ is the cumulative Normal $\left(0, \sigma_u = \sigma_\eta^2 + \sigma_\xi^2 + \sigma_\zeta^2\right)$ distribution of u. Under these assumptions, $\mathbb{E}\{\eta_i|u_i\} = \frac{\sigma_\eta^2}{\sigma_u^2}u_i$, implying $\mathbb{E}\left\{\eta_i|D_i^S,X_i,m\right\} = \frac{\sigma_\eta^2}{\sigma_u^2}\frac{\phi\left(\frac{1}{\sigma_u}\Gamma' X_i - c_p(m)\right)}{D_i^S - \Phi\left(\frac{1}{\sigma_u}\Gamma' X_i - c_p(m)\right)} \equiv \frac{\sigma_\eta^2}{\sigma_u^2}\lambda\left(D_i^S,X_i,m\right)$. Using these expressions after taking conditional expectations in equation 5, we get[8]

$$\begin{aligned}\mathbb{E}\left\{s_i^*|X_i,D_i^S,t_i,m\right\} &= \alpha' X_i + \left(\beta' X_i\right)\times t_i + \mathbb{E}\left\{\varepsilon_i|X_i,D_i^S\right\} + \mathbb{E}\left\{\eta_i|X_i,D_i^S,m\right\}\times t_i \\ &= \alpha' X_i + \left(\beta' X_i\right)\times t_i + \frac{\sigma_\eta^2}{\sigma_u^2}\lambda\left(D_i^S,X_i,m\right)\times t_i. \end{aligned} \tag{7}$$

Note that ε_i and η_i are independent of X_i *by definition* and t_i by way of the randomized experiment. Thus we can estimate equation 7 via a two-step procedure using OLS. The first step allows to estimate the selection model that will yield $\lambda\left(D_i^S,X_i,m\right)$ and the second step yields estimates for the coefficients in equation 7.

Equation 7 is the crucial resource to estimate the marginal treatment impact of the intervention under different scenarios of decentralization and informational advantage. To see how this, consider first the simplest case where there are no agent traits that the principal nor the supervisor can observe so the vector X_i is constant. The expected index on which the supervisor selects is

[7] In our estimation, c_p is not separately identified from including a constant vector in X_i and thus we normalize it to zero. We revisit c_p in section 7.

[8] We do not impose any restrictions on u_i and ε_i and so also include $\lambda(D_i,X_i,m)$ as a main effect without any interaction with t_i. This parameter (along with α) is not of direct interest to us and is not required for identification, but may improve the efficiency of the other estimates.

$\mathbb{E}\{v_i|\theta_i,\zeta_i\} = \mathbb{E}\{\eta|\theta_i,\zeta_i\} + \zeta_i = u_i$. Given the 50 percent roll-out in the experiment, we know that under decentralization, the total treatment impact of 50 percent rollout is $\beta_0 + \beta_1$ from OLS estimation of equation 4. In order to trace the marginal treatment impact at any other roll-out m, we only need to consult the value of u_i at the m percentile in the Normal distribution of u_i.[9] Thus, it is possible to trace the total treatment gain from following the supervisor's selection criterion for all m.

As the expression $\mathbb{E}\{\eta|\theta_i,\zeta_i\} + \zeta_i = u_i$ makes clear, we cannot tell whether a supervisor's selection is due to information on unobservables that affect true responsiveness to treatment (η) as opposed to unobservables that make the supervisor select an agent for other reasons (ζ). But if $\beta_1 > 0$ we know the supervisor gets a precise enough signal on η, and places enough weight on it, so that even if she is biased in her choices, her selection yields higher treatment impact than selecting the AEAs at random.

In most situations, supervisors will know characteristics of their agents, and so the expected index $\mathbb{E}\{v_i|X_i,\theta_i,\zeta_i\}$ on which supervisors select will indeed be a function of X_i. In this situation, each expansion of roll-out will imply extending treatment to new agent types, where the type space as seen by the supervisor is some unidimensional path in a higher dimensional space of traits X_i and the supervisor-only observed u_i. The analyst does not observe u_i, but can form an expectation of it conditional on an agent with traits X_i being selected. Knowing traits X_i and a conditional expectation on u_i for an agent being selected at a given level of roll-out, Equation 7 delivers the treatment impact. Thus, it is possible to derive the total treatment gain from following the supervisor's selection criterion for all m.

Further uses of the model: evaluating supervisors, and the potential for sophisticated centralization We have now described ways to obtain marginal treatment impacts at varying roll-outs for the cases of an uninformed centralized principal and an informed supervisor. But the selection model laid out in this section can be put to other uses. First, it is possible to evaluate the supervisors in a more complete way than simply saying whether they have an informational advantage over the principal. We can ask the extent to which their advantage is related to their knowledge of elements that are potentially observable to the principal (X_i) versus things the principal cannot expect to learn (η). Moreover, the analyst can econometrically evaluate the extent to which supervisors make optimal use of observable data in X_i.

Second, with gains in the ability to gather and process data, a principal could learn some traits of its agents, captured by X_i. This opens up the consideration of a class of counterfactuals, with a natural

[9] We do not recover separate values for σ_u and σ_η, since all parameters are scaled by σ_u in the probit regression.

one involving the marginal treatment impact for varying m for a decisionmaker that knows X_i but does not observe θ_i. Thus, we can ask the question of whether a sophisticated centralized principal can emulate or surpass the performance of supervisors despite its informational disadvantage. We perform these exercises in Section 7.

Discussion In a literal sense, we have presented a heterogeneous treatment effect model where supervisors have private information about the treatment effects. Equation 7 shares the same functional form as a Heckman (1974) "Heckit" selection model. However, in most settings where the Heckit is applied, $t_i = D_i^S$. In settings that mirror ours where D_i^S is assigned according to ε_i or η_i, inclusion of the $\lambda(\cdot)$ control function in estimation is required for identification because of non-random censoring of potential outcomes, the raison d'etre for the literature on selection correction. However, control functions require credible instruments; without an instrumental variable that could be excluded from one equation or the other, if one instead assumes a uniform distribution for u_i (which in turn implies a linear probability model for D_i^S), $\lambda(\cdot)$ would be collinear with the vector X_i (Olsen, 1980). Even in contexts where there are credible instruments that generate experimental variation, selection models have been used to extrapolate treatment effect heterogeneity among never-takers and always-takers from instrument-implied local average treatment effects (see Heckman and Vytlacil (2005) and Kline and Walters (2017)).

In our context, however, t_i is independently and randomly assigned, and not equal to D_i^S. While supervisor preferences are elicited, they are not used to determine assignment in our main sample. This means that we neither have censored potential outcomes nor always-takers and never-takers. Instead, we have a randomized experiment with full compliance as well as information about supervisor preferences that were never implemented, and so we are able to credibly estimate treatment effects along the full distribution of η_i, an exercise that requires no extrapolation. Because we observe treatment effects for non-selected AEAs (i.e., those with $D_i^S = 0$), even if mispecified, $\lambda\left(D_i^S, X_i\right)$ is just a transformation of D_i^S and X_i, and with inclusion of controls, its independent variation is driven primarily by D_i^S.[10]

[10]Stated differently, if we had continued to assume a linear conditional expectation function $\mathbb{E}\{\eta_i|u_i\} \propto u_i$ but instead assumed u_i were uniformly distributed, the coefficients on $\lambda(\cdot)$ would be numerically equivalent to substituting $\lambda(\cdot)$ for $\frac{D_i^S}{2}$, a result that follows immediately from the Frisch-Waugh theorem. From a purely statistical standpoint, D_i^S is indistinguishable from any other covariate such as gender or education—its unique content emerges from its implications about economic theory.

5 Data

We collected two main sources of data. The first is a survey of AEAs. Each AEA and supervisor filled out answers on his own with survey enumerators available to answer questions. The survey contains questions regarding the AEAs' demographics, work history, and measures of personality such as the Big-5 inventory, digit span test measuring cognitive ability, the Perry public service motivation index, and Moore's moral disengagement scale. The second source of data we have is two rounds of farmer phone surveys. We called farmers who were beneficiaries of the AEAs and asked questions about their interactions with the AEAs such as how often they saw the AEA and how satisfied they were with his work.

The timeline of events is as follows. In March of 2014, the ALAT-level supervisors chose which AEAs they would like to prioritize for receiving a phone with the objective of expanding effort in service to farmers. In April, we were given a list of the names of all farmers who were beneficiaries of the AEAs and their phone numbers when applicable. The first round of phones was distributed to the AEAs between April 30, 2014 and July 16, 2014. Individuals from the central ministry office travelled to each of 19 CDAs to meet with the AEAs who were scheduled to receive phones, distribute the phones to them, and teach them how to use the phones. It took over two months because this involved 19 meetings spread across the country.

After the first round of phones was distributed, we conducted two types of data collection. From July 7 through September 7, 2014 we conducted the first round of farmer phone surveys. Additionally, during September 2014, we conducted a survey of all AEAs as well as the supervisors. We treat variables such as sex, age, years of education, and the personality indices as being fixed characteristics which were not affected by the roll-out of the phones. On the other hand, we treat variables such as the AEAs' perceptions of whether their supervisors know where they are during the working week as potentially being affected by the roll-out of the phones. In the control group, which did not yet receive phones, these variables are treated as baseline measures, not impacted by the roll-out of the phones.

After completing the first round of surveying, the second round of phones was distributed between February 10 and March 13, 2015. We then conducted a second round of farmer phone surveys between March 24 and May 7, 2015. The Ministry of Agriculture planned to distribute phones to all AEAs who had not yet received one before the end of 2015.

The ministry did not give any phones to AEAs who were not on our randomized list. There were a few cases in which phones broke down or sick AEAs were not able to pick up their phones. For this reason we look at intent-to-treat (ITT) estimates using our initial random group assignment.

In early 2014, we were given full information, including job title, job location, and client names and phone numbers for 368 agricultural extension agents - 139 supervisors and 229 AEAs. In late 2014, we were able to interview 301 of these – 119 supervisors and 182 AEAs. We interviewed 79% of the AEAs in our original administrative data, 15% no longer worked for DEAg, and 6% were absent the day of the surveying.

The job description of an AEA involves working with 80 farmers. Thus, it is no surprise that the median extension agent in our data listed the names of 80 farmers with whom he worked; the mean of the distribution is 75 with a standard deviation of 26. The median extension agent in our data listed phone numbers for 78% of the farmers he served, while the mean share listed is 73%. These numbers vary very little for AEAs versus supervisors.

We conducted two rounds of farmer phone surveys, but we wanted to leave open the possibility of conducting three rounds. For AEAs and supervisors in big ALATs who listed 75 or more farmer phone numbers, we randomly chose 75 farmers to call and then randomly divided them to call 25 farmers in each of three rounds. For those who listed fewer than 75 farmer phone numbers, we randomly divided their farmers into thirds to call in each of the three rounds. Similarly, for AEAs and supervisors in small ALATs who listed 24 or more farmer phone numbers, we randomly chose 24 farmers to call and then randomly divided them to call 8 farmers in each round. For those who listed fewer than 24 farmer phone numbers, we randomly divided their farmers into thirds to call in each of the tree rounds.

In total, we made 2,635 phone calls in the first round and 2,642 in the second round for the 182 AEAs who responded to the AEA survey. 68% of those phone calls led to completed surveys.[11] Conditional on completing the survey, 70% of farmers confirmed that the AEA that had provided their number worked with them and thus were asked more detailed questions about their interaction with that AEA.[12] This leads to 2,519 usable phone calls.

Table 1 presents sample means and a randomization check of the cellphone assignment for various AEA characteristics.[13] The table distinguishes between small single agent ALATs (columns 1 and

[11] In 18% of cases, we reached voicemail on all five tries, 7% of cases were wrong numbers, 4% were out-of service phone numbers, and 2% of farmers did not agree to complete the survey.

[12] We first asked the farmers to talk about any AEAs with whom they worked and did not offer up the name of the AEA we had on record for them. We only asked the farmer about the specific name we had on record if either the farmer worked with an AEA whose name he couldn't remember or if he did not list the name of the AEA we had on record on his own.

[13] In appendix table **??** we also check for balance on a set of ALAT-level characteristics extracted from the population and agricultural censuses. We look at 27 variables, and only 2 are significantly different. Large ALATs are located

2) and large multi-agent ALATs (columns 3 and 4). On average, AEAs are 37 years old, and 76% of them are male. The AEAs were able to recall an average of 5.2 digits in the memory digit span test, which is a commonly-used measure of cognitive ability.[14] AEAs are also required to travel on average 2.3 kms to visit a given farmer. Overall the results in Table 1 suggest that the treatment, which was randomized at the ALAT level, was done so in a balanced way.

6 Results

In this section, we begin by estimating the impact of the cell phones on AEA performance. According to the model, under certain conditions (cell phones improve monitoring and there are sufficiently many agents who respond positively to it), the increase in monitoring induced by the phones should boost the effort levels of the AEAs and thus increase the number of farmers visited. Subsequently, we test whether the impact of cell phones was higher among the AEAs who were selected by the supervisors, which would be the case if supervisors were able to target the AEAs with highest responsiveness to treatment. Finally, we estimate heterogeneous treatment effects, which we use to evaluate impacts under various counterfactual scenarios with different levels of roll-out.

6.1 Increased Monitoring and Performance

As we discussed in Section 2, the primary task of an AEA is to visit farmers. In columns 1-5 of Table 2, we estimate the impact of the phone on whether the farmer reported having been visited by his AEA in the last week. Except for in column 4, our estimation sample includes all AEAs in the small and large ALATs, excluding those randomized into the partial treatment cells (cells C and F). In column 1, we present the estimates without any additional controls. In column 2, we add a set of basic controls (e.g. age, gender, etc.), and in column 3, we further augment the specification to include controls measuring AEA personality type (e.g. Big 5 and Digit Span). In column 4, we re-estimate the specification presented in column 3, excluding the small ALATs.

We find that the increase in monitoring leads AEAs to visit their farmers more often. They are

in more populated districts with both larger urban and rural populations. The large ALATs are located in districts in which basic needs are met for more of the population. The results in the table again suggest that the randomization led to balance across treatments.

[14]For the digit span test, the enumerator read out loud a random number that the AEA was then required to recite back. The test began with two numbers and then increased incrementally in the number of digits until the AEA could no longer recall the number correctly.

approximately 6 percentage points more likely to have visited a given farmer in the past week, which is an increase of 22% over the control group. As expected given the random assignment, the estimated impact is robust across the various specifications, and when we restrict the estimation to only big ALATs, which is our main sample moving forward. Overall, the demographic and personality-based controls have little predictive power.[15]

ALAT-level supervisors are in charge of both supervising the AEAs in their ALAT as well as serving their own farmers. In column 5, we test the impact of the phone on the visits to those farmers who are served by a supervisor. We find a small and insignificant impact (point estimate = -0.001; clustered standard error = 0.035). This suggests that the impact of the phone is related to the greater monitoring ability it gives supervisors and not due to productivity-enhancing functions of the phone (e.g., ease in communication), which would have the same effect on both supervisors and AEAs. As a further check, AEAs were asked whether they agreed with the statement that their supervisor usually knows where they are during the work week. In column 6, we see that having a phone significantly increased the extent to which they agreed with this statement.

While the treatment led to more visits, this does not necessarily imply that the AEAs are exerting more effort. AEAs could be making more visits but making them shorter. In column 7, we test for this possibility but do not find evidence to support the idea. The point estimate, which suggests that treated AEAs spend only 1.9 percent less time, or approximately 2 minutes, is small and statistically insignificant.

6.2 Do Supervisors Have Useful Information?

Recall that our model assumes agents differ in their responsiveness to enhanced monitoring and that supervisors know this information. If supervisors wish to increase the number of farmers visited, then when tasked with the responsibility of assigning increased monitoring, they should target the agents for whom a larger increase in performance ought to be expected. Our research design allows us to test this precisely.

Prior to the randomization, we asked the local supervisors to prioritize which half of their AEAs should receive the phones. Given this information, we test for the value of information using a simple difference-in-differences estimator for our sample of large ALATS: We compare the performance of AEAs who were selected and received the phone against those who were selected but did

[15] In results not shown here we look separately at short run versus long run impacts of the phone, and find that they are quite similar. The impact of the phone does not diminish over time.

not receive the phone, net of the difference in performance between those who were not selected and received the phone against those who were not selected and did not receive the phone (i.e. $(B-A)-(E-D)$).

From Table 3, we see that the effects of the phones on performance are entirely driven by the effects of the agents prioritized to receive the phone prior to the randomization. These agents are 11 percentage points more likely to have visited their farmers in the last week. Compared to the prioritized agents in the control, this effect represents a substantial increase of 63 percent. From column 2, we also see that prioritized agents in the control group are 3.3 percentage points less likely to have visited their farmers relative to the non-selected, although this difference is not statistically significant.

In sum, we find strong evidence that the phones do have an impact on AEA behavior and that supervisors possess useful information regarding which AEAs' performance will improve most after receiving a phone. This of course begs the question of what characteristics the supervisors used to create their prioritized list and the extent to which supervisors used information on characteristics analysts could hope to obtain. The next subsection answers these questions.

6.3 Heterogeneous treatment effects

In Table 4, we present estimates from a Probit regression, in which the dependent variable is an indicator for whether the AEA was prioritized by the local supervisor. Based on standard observable characteristics, we find that supervisors tended to prioritize AEAs who were younger, more likely to be married, or had to travel further distances to visit their farmers. In terms of their personality traits, supervisors were more likely to select AEAs with lower levels of the Big-5 Stability trait. This trait is seen by psychologists to capture an individual's tendency to avoid emotional swings and negative feelings (its negative is often labeled Neuroticism). It is conceivable that individuals with higher Stability scores are more likely to stay motivated and attain their goals. Despite the richness of our data, our ability to predict the choices of the supervisors is fairly low: the highest pseudo R^2 is only 14.7%. This opens the possibility that supervisors are selecting AEAs based on unobservable but productive characteristics (η) or unobservable and idiosyncratic characteristics (ψ), features that are not captured by demographic traits or even indicators of cognitive and non-cognitive ability.

In Table 5, we present a series of second stage estimates based on Equation 7. In column 1, we present a specification without any additional controls or interaction terms, whereas in columns 2 and 3 we include additional controls along with their interactions with the treatment indicator. For columns 2-3, the first stage regressions correspond to the ones presented in Table 4.

The key finding in Table 5 concerns the Inverse Mills Ratio variable and particularly its interaction with treatment. The inverse mills ratio captures the expected unobservable traits that recommended an AEA for selection by the supervisor. Because no controls were included, the inverse mills coefficient in column 1 replicates the findings from Table 3 that supervisors are selecting individuals with higher treatment effects. When we allow the effects of the treatment to vary by the characteristics that we found were predictive of the likelihood of selection, we find that the inverse mills ratio is still highly predictive of responsiveness to treatment, direct evidence that the unobservable reasons supervisors are selecting AEAs are productive rather than non-germane. In addition to the unobservable traits, the treatment effect also varies by the IQ of the AEAs: those who performed worse on the digit span test exerted more additional effort in response to the treatment.

The results so far suggest several questions. What is the basis for the supervisors' informational advantage? At a given cost of decentralization, does the informational advantage justify the cost? To answer these questions, one needs to know two elements. One, what is the extent of roll-out that is anticipated. Under decentralization, the anticipated extent of roll-out should be whatever is optimal, and this motivates the need to identify that optimal extent. Second, how much information does the central authority have? In the next section we apply the framework introduced in Section 4 to provide answers to these questions.

7 Counterfactuals

In this section, we exploit our heterogeneous treatment effects model to compute counterfactual treatment effects under alternative selection rules. This allows us to assess the benefits of decentralization relative to centralization under different informational assumptions.

The first step is to define a counterfactual aggregate benefit under an arbitrary selection rule D_i^{CF} as:

$$\begin{aligned}\Delta Y^{CF} &= \mathbb{E}\left\{\underbrace{\beta_i\left(X_i,\eta_i\right)}_{\text{how much?}}\times\underbrace{D_i^{CF}}_{\text{who?}}\right\}\\ &= \int\mathbb{E}\left\{\beta\left(X_i,\eta_i\right)|D_i^{CF}=1\right\}Pr\left\{D_i^{CF}=1\right\}dX_i \qquad (8)\end{aligned}$$

In keeping with the rest of our notation, we write our arbitrary selection rule as a threshold problem, $D_i^{CF}(X_i, u_i) = 1\left[\tilde{\Gamma}' X_i + \tilde{u}_i \geq c_p\right]$; because we have not made any distributional assumptions about $\tilde{u}_i$, this does not impose additional assumptions. Note that the assumed cost c_p is not directly observable, and the threshold problem is not a unique representation of the selection rule—any monotonic transformation of the latent index and c_p will yield the same choices. However, we are not trying to directly obtain either of these objects: only the consequences $Pr\left\{D_i^{CF} = 1|X_i, c_p\right\}$ and $\mathbb{E}\left\{\beta(X_i, \eta_i)|D_i^{CF}\right\}$, which map into the scale of rollout m and the aggregate counterfactual impact $\Delta Y^C F$.

One example of a selection rule is the one implicitly applied by supervisors, $D_i^S(X_i, u_i)$, which anchors our portrait of what can be achieved under decentralization. Note that from our estimation of Equation 7, we have recovered $\mathbb{E}\{\beta_i(\cdot)|X_i, u_i\}$, and under distributional assumptions, the selection rule under decentralization, $D_i^S(X_i, u_i)$. Given this, we can use Equation 8 to trace out the expected treatment effects of the cell phones under decentralization for any given threshold c_p or, by extension, any roll-out m. But, we can impose any other selection rule capturing different counterfactual scenarios corresponding to different forms of centralized assignment and trace out the expected treatment effects for all roll-out levels in each senario.

Uninformed Principal A natural, if extreme, benchmark, is that of a principal who does not have any information about how best to target roll-out. In this situation, the selection rule is random allocation. At a roll-out level m, a fraction m of all agents receive a cell phone, and the expected total treatment effect is m% of the average treatment effect (as in the theory section, we are considering a large number of agents who can then be approximated by a continuum of mass 1). The dotted line in Figure 3 plots this counterfactual selection rule at various roll-out levels. For instance, if the principal decided to allocate the phones to everyone then the expected aggregate treatment of the program would be 6 percentage points, which corresponds to the average treatment effect. If instead she decided to treat only half of the AEAs, then we would expect an aggregate treatment effect of only 3 percentage points. Thus, it is easy to see that with a random selection rule, we get a set of counterfactuals that traces a straight line from zero to the average treatment effect.

Supervisor We can contrast the random allocation rule with the aggregate benefits based on the supervisor's selection rule. In this case, the selection rule is given by $Pr\left\{D_i^S = 1|X_i\right\} = \Phi\left(\frac{1}{\sigma_u}\Gamma' X_i - c_p\right)$ and the expected aggregate treatment effect is $\Delta\mathbb{E}\left\{s_i^*|X_i, D_i^S, T_i\right\} = \beta' X_i + \frac{\sigma_\eta^2}{\sigma_u^2}\lambda\left(D_i^S, X_i\right)$. This counterfactual is depicted in Figure 3 with the solid line. Note that by construction, the curve must cross three points: the origin, 0.06 at 100% roll-out, and 0.078 at 53.8% roll-out which corresponds to share of AEA that received the phones under the actual research design. The difference between the supervisor counterfactual and the random allocation rule measures the benefits of decentralization at each level of roll-out under the assumption that the principal does not possess any information. As we can see from the figure, the difference between the random allocation and the supervisor rule is maximized at a roll-out threshold of 63% where the additional treatment effect is over 5 percentage points. The optimal scale of roll-out under decentralization is not 63%, however, but 81.8%, at which level the total treatment effect is 9.5 percentage points.

What underlies the informational advantage of the supervisor over an uninformed principal? One way to tackle this question is to ask how much of the supervisor's advantage is predicated on the use of information on observables X_i versus information on unobservables η_i. The dot-dash line in Figure 3 traces out the counterfactual treatment effect under the assumption that the supervisor does not use his signal of η. In other words, the dot-dash line tells us what the treatment effects would be under a supervisor who cannot use information on unobservables. In this case, the selection rule and expected treatments are only computed based on the observable (to the econometrician) traits, setting λ=0. The dot-dash curve is much closer to the one under random assignment. This suggests that most of the supervisor's informational advantage is driven by access to information that is likely hard to collect for a centralized authority lacking personal contact with the agents.

Giving Centralization A Chance: Counterfactual Treatment Effects With A Partially Informed Principal

Minimally informed principal: Assignment based on job difficulty Thus far, we have assumed that the principal does not have any prior information about how AEAs will respond to the program, which is extreme although it may not be a wholly unreasonable approximation to the situation facing the leadership of government programs in low state capacity contexts. This does not suggest however that adopting a sensible heuristic might not outperform a random assignment mechanism, which would of course affect the centralization versus decentralization calculus. One such heuristic might be to simply allocate the phones to the AEAs who have to travel the farthest in order to visit their farmers. This of course requires some information on the work environment of agents, and it constitutes the case we associate with a minimally informed principal. This counterfactual is displayed in Figure 4 with a dashed line. We find that this method generally outperforms random assignment (a 2.4. p.p. advantage at 50 percent coverage), but it cannot beat the supervisor at any roll-out level.

Significantly informed principal: Assignment based on baseline performance We consider a second type of partially informed principal that has the capacity to gather information on individual agent characteristics, and can map them onto their baseline productivity. To this end, we run a simple prediction model in which, among the AEAs in the control ALATs, we regress the share of farmers visited on our set of basic and cognitive controls. Using the estimated coefficients, we can then compute an AEA's expected productivity based on his or her observable traits. Given this information, a sensible centralized policy would be to assign cell phones starting with the AEAs who had the lowest predicted productivity, and as roll-out increases, expand coverage to AEAs with higher productivity. As we see in Figure 4, under this approach centralization would dominate decentralization at virtually all levels of roll-out. It is worth noting that the data requirements to estimate our performance-prediction model are not trivial and often beyond the capacity of government programs in several developing countries.

A sophisticated principal: Experimentation and assignment based on response to treatment For all its data demands, the approach in the previous section that assigns phones based on performance prediction does not exhaust the possibilities open to a central authority who has the capacity to gather and analyze data. The key shortcoming of this approach is that baseline performance is not always a great predictor of responsiveness to treatment. While baseline performance can reflect individual heterogeneity in, say, linear terms of the effort cost function, response to treatment depends on other cost drivers, such as the disutility from receiving a reprimand. In addition, baseline effort may reflect endogenous adjustments of monitoring effort by supervisors that tend to mitigate effort heterogeneity. To overcome these difficulties, a sophisticated principal can conduct a pilot experiment at a low roll-out level and establish a map between agent observable characteristics and response to treatment. Then it is possible to construct an assignment rule $D_i^{CF}(X_i)$ that first allocates phones starting to agents who are predicted to have the highest response to treatment and work downwards to treat progressively less responsive agents, tracing out the total treatment effect for each roll-out level (one can assume the 0-100 roll-out in the exercise corresponds to still-to-be-treated agents). As shown in Figure 4, this optimal approach outperforms all others by a wide margin. The largest gap relative to the decentralized approach is above 2 percentage points and occurs at a roll-out level of 45.7%. A sophisticated principal would be more interested in setting roll-out at its optimal scale. The maximum total treatment effect for an 'experimenting principal' is 10.93 percentage points and is achieved at a roll-out of 70.6%. Note that relative to "blind centralization," which treats everyone and attains a total treatment effect of 6 percentage points, this arrangement saves on almost a third of the phones and attains almost twice the total treatment impact. Relative to decentralization, the sophisticated centralized approach saves on roughly 10% of phones and attains roughly an additional 1.5 additional percentage points in the total treatment effect.

Conclusions

One of the primary benefits of decentralization is that lower-level agents are presumably better informed than their principals about how to implement a particular task. But, the importance of this superior information will often depend on the scale of the task at hand. Because decentralization is costly, the decision to devolve decision-making powers to lower-level agents requires knowing not only the value of their information, but this value at different scales of roll-out. Despite the fact that the informational advantage of agents is a maintained assumption in much principal-agent theory, evidence of the presence and extent of that advantage has been scarce. The same is true about methods to place organizational choices such as decentralization in a context in which scale of implementation alters the treatment effects of an intervention under the respective organizational modes at hand.

In this paper, we establish that lower-level elements in a government hierarchy do have superior information that is pertinent to the way in which an intervention should be implemented and develop an approach to trace the total treatment effects of the intervention at all levels of roll-out. The context is an initiative by the federal government in Paraguay to introduce a new monitoring device that enabled supervisors in rural areas to track their agricultural extension agents (AEAs). Our experimental design randomly assigned monitoring devices across agents and elicited the preferences of their supervisors as to which agents should be prioritized for monitoring. Crucially, in the main sample, treatment assignment was kept independent of supervisor recommendations. This allows us to establish that supervisors have valuable knowledge because the AEAs selected by them are far more responsive to treatment. We find that the informational advantage of supervisors is tied to information other than observables that analysts might reasonably collect, and argue theoretically that the value of this information advantage varies with the scale of anticipated roll-out for the new technology. In addition, we estimate the full schedule of marginal treatment effects as roll-out scale is expanded from 0 to 100 percent. We do this for the selection rule that supervisors are seen to have used as well as several other counterfactual assignment rules. Our counterfactual assignment rules approximate what principals with varying levels of information might achieve when targeting agents for treatment in centralized fashion. While a minimally informed principal cannot attain results as good as under decentralization, reasonably informed principals can. In the best case scenario, a principal who can conduct a pilot RCT to obtain predictors of individual response to treatment can vastly outperform supervisor choices; such a principal would substantially reduce roll-out scale and still attain larger aggregate gains in agent performance.

Overall our findings suggest that as information and communication technologies continue to improve the capabilities of government and organizations more generally, the informational benefits that lower level agents bring become less clear. Although studies have shown that innovation in information technologies can lead to more decentralization (e.g. Bloom et al. (2009), Bresnahan et al. (2002)), our findings suggest the opposite may occur, particularly if these technologies primarily serve to reduce the information frictions between agents and principals.

Of course, the value of the information that lower-level agents possess is specific to the task and context, which may raise concerns of external validity. But while our findings may not be generalizable, our method is, as it can be easily exported to other settings.

References

Acemoglu, D., Aghion, P., Lelarge, C., Van Reenen, J., and Zilibotti, F. (2007). Technology, information, and the decentralization of the firm. *The Quarterly Journal of Economics*, 122(4):1759–1799.

Aghion, P. and Tirole, J. (1997). Formal and real authority in organizations. *Journal of Political Economy*, 105(1):1–29.

Aker, J. C. and Ksoll, C. (2017). Call me educated: Evidence from a mobile monitoring experiment in Niger. Center for Global Development Working Paper 406.

Alatas, V., Banerjee, A., Hanna, R., Olken, B. A., and Tobias, J. (2012). Targeting the poor: Evidence from a field experiment in Indonesia. *American Economic Review*, 102(4):1206–40.

Alderman, H. (2002). Do local officials know something we don't? Decentralization of targeted transfers in Albania. *Journal of Public Economics*, 83(3):375 – 404.

Banerjee, A., Chattopadhyay, R., Duflo, E., Keniston, D., and Singh, N. (2015). Improving police performance in Rajasthan, India: Experimental evidence on incentives, managerial autonomy and training. Unpublished Working Paper.

Banerjee, A., Glennerster, R., and Duflo, E. (2008). Putting a band-aid on a corpse: Incentives for nurses in the Indian public health care system. *Journal of the European Economic Association*, 6(2-3):487–500.

Bloom, N., Garicano, L., Sadun, R., and Van Reenen, J. (2009). The distinct effects of information technology and communication technology on firm organization. NBER Working Papers 14975, National Bureau of Economic Research, Inc.

Bloom, N., Sadun, R., and Van Reenen, J. (2012). The organization of firms across countries. *The Quarterly Journal of Economics*, 127(4):1663–1705.

Bresnahan, T., Brynjolfsson, E., and Hitt, L. M. (2002). Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *The Quarterly Journal of Economics*, 117(1):339–376.

Callen, M., Gulzar, S., Hasanain, A., Khan, Y., and Rezaee, A. (2015). Personalities and public sector performance: Evidence from a health experiment in Pakistan. NBER Working Paper 21180.

de Rochambeau, G. (2017). Monitoring and intrinsic motivation: Evidence from Liberia's trucking firms. Unpublished Manuscript.

Dessein, W. (2002). Authority and communication in organizations. *The Review of Economic Studies*, 69(4):811–838.

Dhaliwal, I. and Hanna, R. (2017). The devil is in the details: The successes and limitations of bureaucratic reform in India. *Journal of Development Economics*, 124:1–21.

Duflo, E., Greenstone, M., Pande, R., and Ryan, N. (2016). The value of regulatory discretion: Estimates from environmental inspections in India. Working Paper 20590, National Bureau of Economic Research.

Duflo, E., Hanna, R., and Ryan, S. (2012). Incentives work: Getting teachers to come to school. *American Economic Review*, 102(4):1241–78.

Heckman, J. J. and Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*, 73(3):669–738.

Hussam, R., Rigol, N., and Roth, B. (2017). Targeting high ability entrepreneurs using community information: Mechanism design in the field. Unpublished Manuscript.

Khan, A. Q., Khwaja, A. I., and Olken, B. A. (2016). Tax farming redux: Experimental evidence on performance pay for tax collectors. *Quarterly Journal of Economics*, 131(1):219–271.

Kline, P. and Walters, C. R. (2017). Through the looking glass: Heckits, LATE, and numerical equivalence. Unpublished Manuscript.

Mookherjee, D. (2006). Decentralization, hierarchies, and incentives: A mechanism design perspective. *Journal of Economic Literature*, 44(2):367–390.

Olsen, R. J. (1980). A least squares correction for selectivity bias. *Econometrica*, 48(7):1815–1820.

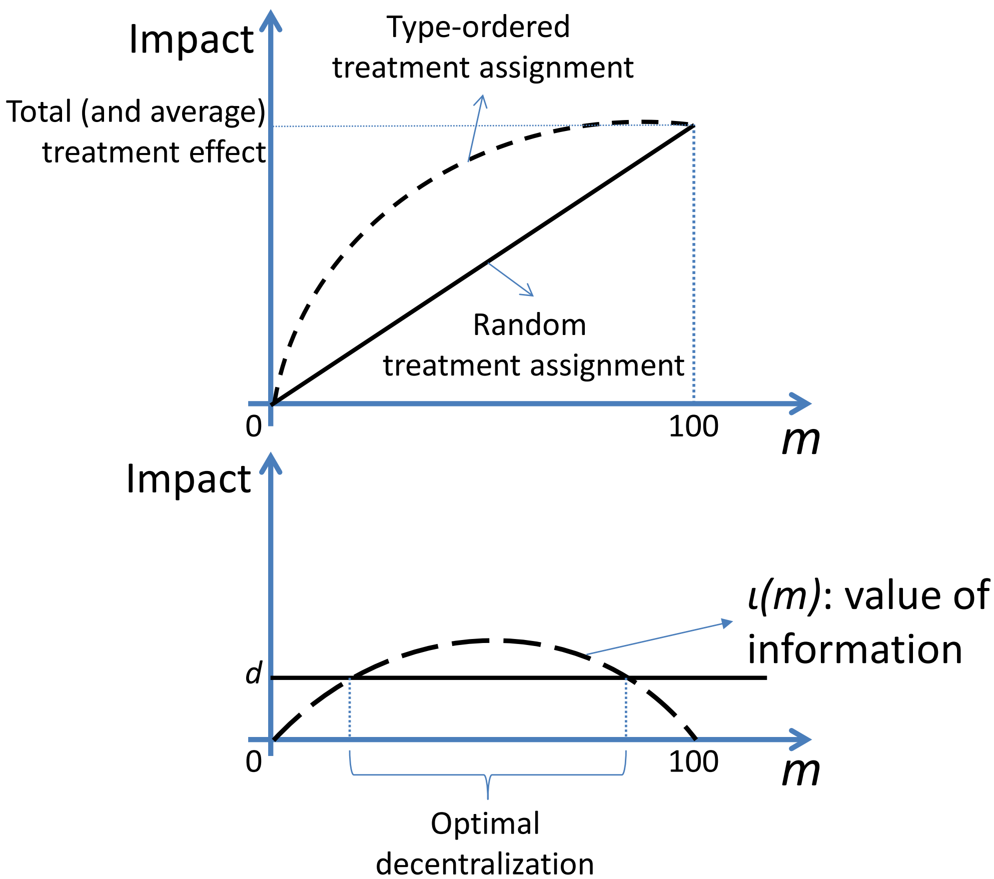

Figure 1: Treatment effects, roll-out extent, and the value of information

	Control Group	100% Coverage	50% Coverage
Selected	A	B	C
Not Selected	D	E	F

Figure 2: Experimental Design

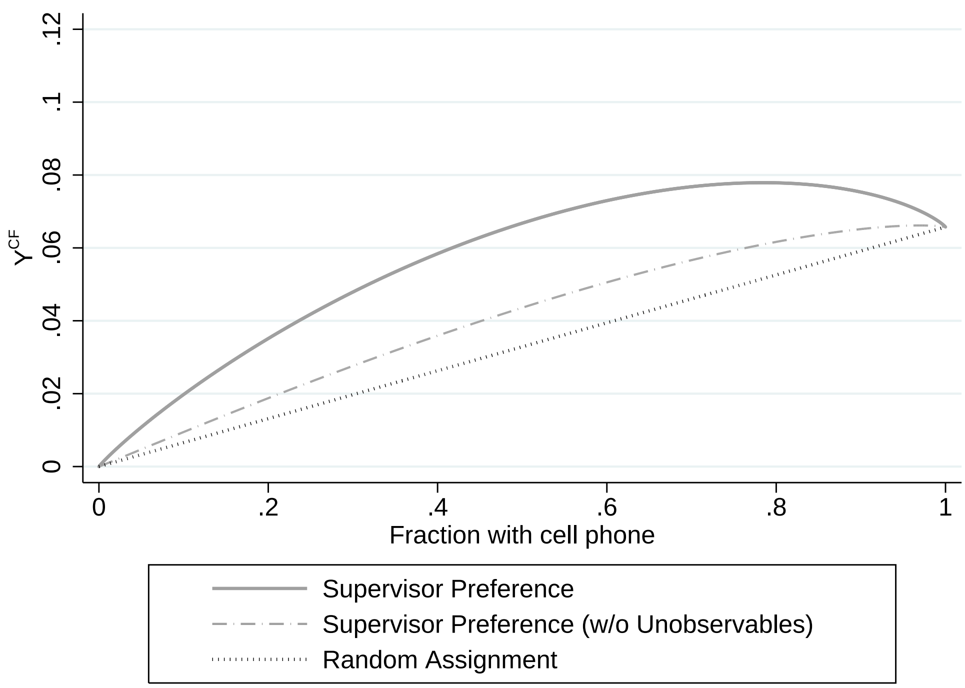

Figure 3: Supervisor versus Random Assignment

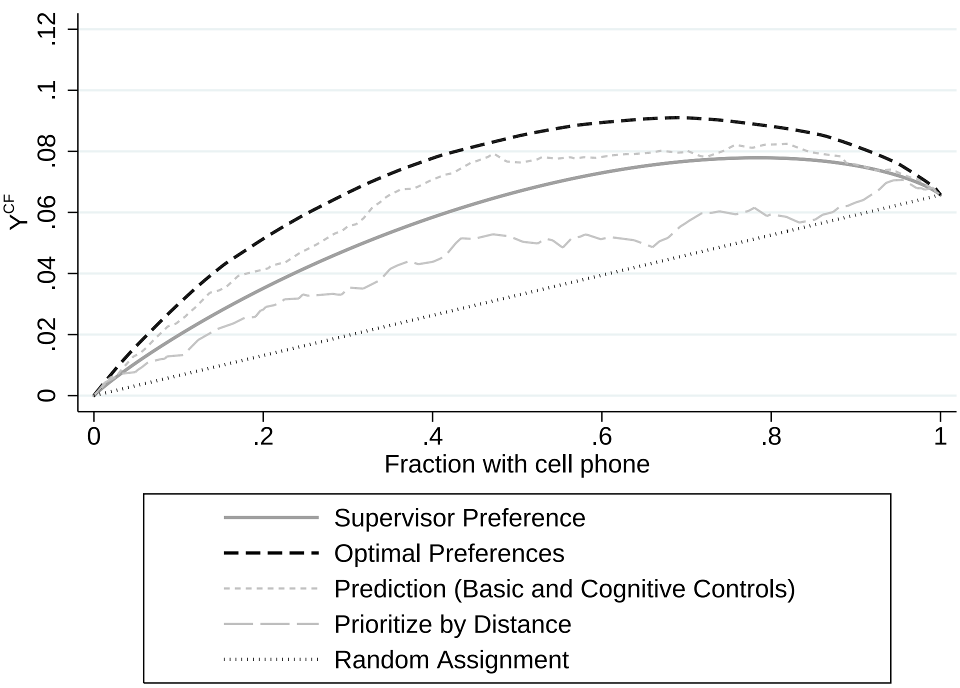

Figure 4: Supervisor versus Alternative Allocation Rules

Table 1: Covariate Balance

	Small ALATs		Large ALATs	
	(1) Control	(2) Difference (T-C)	(3) Control	(4) Difference (T-C)
Male	0.800	-0.153	0.754	-0.138
	[0.422]	{0.294}	[0.434]	{0.380}
Age	34.100	3.165	37.913	4.318
	[12.004]	{0.459}	[10.754]	{0.112}
Married	0.300	-0.006	0.435	0.065
	[0.483]	{0.973}	[0.499]	{0.474}
AEA is on permanent contract	0.400	0.225	0.941	0.020
	[0.516]	{0.253}	[0.237]	{0.693}
Average Distance	2.298	-0.158	2.237	0.000
	[0.808]	{0.559}	[0.616]	{0.999}
Digit Span	5.400	0.176	5.174	0.288
	[1.265]	{0.678}	[1.070]	{0.318}
Big 5 — Stability	0.216	-0.184	-0.063	0.094
	[0.906]	{0.574}	[0.894]	{0.631}
Big 5 — Openness	0.076	-0.103	-0.100	0.233
	[0.760]	{0.720}	[0.796]	{0.167}
Perry: Public Service Motivation Index	-0.127	0.021	-0.077	0.220
	[0.553]	{0.925}	[0.562]	{0.380}
Moore: Moral Disengagement Index	0.284	-0.002	-0.035	-0.074
	[1.115]	{0.997}	[0.873]	{0.759}
Selected			0.594	-0.017
			[0.495]	{0.791}
Number of AEAs	10	34	69	26
Number of ALATs	10	32	22	11
p-value from Joint Test		0.157		0.792

[a] Standard deviations reported in square brackets and p-values from a wild bootstrap procedure with 10,000 replication draws reported in curly braces. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Average Effects of Receiving a Cell Phone on Productivity

	Farmer was visited in the last week					Supervisor knows	Log Length of Meeting (mins)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treated	0.065**	0.065**	0.066**	0.060	-0.009	0.205*	-0.016
	(0.032)	(0.028)	(0.028)	(0.032)	(0.037)	(0.107)	(0.061)
	{0.040}	{0.031}	{0.027}	{0.103}	{0.817}	{0.051}	{0.812}
Selected	0.006	0.013	0.018	0.015		-0.149	-0.068*
	(0.046)	(0.042)	(0.044)	(0.045)		(0.137)	(0.041)
		{0.769}	{0.699}	{0.756}		{0.284}	{0.099}
Male		0.022	0.020	0.020	-0.010	0.109	-0.021
		(0.012)	(0.013)	(0.015)	(0.020)	(0.071)	(0.019)
		{0.111}	{0.204}	{0.283}	{0.636}	{0.141}	{0.331}
Age		0.026*	0.027	0.022	0.003	-0.054	0.049**
		(0.014)	(0.016)	(0.018)	(0.019)	(0.076)	(0.019)
		{0.097}	{0.136}	{0.279}	{0.865}	{0.491}	{0.018}
Married		-0.004	-0.001	0.011	-0.019	-0.010	-0.026
		(0.015)	(0.015)	(0.016)	(0.017)	(0.069)	(0.022)
		{0.836}	{0.975}	{0.549}	{0.299}	{0.885}	{0.278}
Distance to Farmers (log)		0.015	0.019	0.017	-0.039**	-0.055	-0.009
		(0.016)	(0.015)	(0.018)	(0.015)	(0.068)	(0.025)
		{0.383}	{0.250}	{0.419}	{0.024}	{0.463}	{0.726}
Digit Span			0.011	0.014	0.031	0.054	0.014
			(0.013)	(0.014)	(0.020)	(0.057)	(0.020)
			{0.438}	{0.384}	{0.191}	{0.353}	{0.514}
Big 5 — Stability			0.012	0.010	0.034	0.117**	0.018
			(0.015)	(0.015)	(0.020)	(0.047)	(0.024)
			{0.492}	{0.595}	{0.109}	{0.016}	{0.511}
Big 5 — Openness			-0.012	-0.012	-0.026	0.145*	-0.020
			(0.012)	(0.012)	(0.019)	(0.079)	(0.019)
			{0.332}	{0.360}	{0.184}	{0.063}	{0.323}
Servicer	AEA	AEA	AEA	AEA	Supervisor	AEA	AEA
Mean of Dep. Var	0.293	0.293	0.293	0.294	0.294	4.632	4.411
R^2	0.005	0.012	0.015	0.012	0.031	0.225	0.015
Number of Phone Surveys	1865	1865	1865	1600	1173	125	1842
Number of ALATs	72	72	72	33	107	62	72
Includes small ALATs	✓	✓	✓		✓	✓	✓

[a] All control variables standardized to have mean zero and standard deviation 1. Observations that are missing controls are imputed to be zero and an indicator variable for the missing control is included.

[b] Regressions also include separate wave indicators for selected and unselected AEAs and small ALATs, respectively.

[c] Outcome in column 6 is an indicator for whether the AEA strongly agrees that the ALAT supervisor knows his location.

[d] Liang and Zeger (1986) cluster robust standard errors and p-values from a wild bootstrap procedure with 10,000 replication draws reported in parentheses and curly braces, respectively. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Do supervisors have an informational advantage?

	Farmer was visited in the last week			
	(1)	(2)	(3)	(4)
Treated	0.0582*	-0.0285	-0.0332	-0.0261
	(0.034)	(0.043)	(0.037)	(0.040)
	{0.0924}	{0.532}	{0.387}	{0.546}
Treated × *Selected*		0.147**	0.154**	0.143**
		(0.058)	(0.052)	(0.054)
		{0.0304}	{0.0157}	{0.0430}
Selected	0.00584	-0.0403	-0.0387	-0.0316
	(0.046)	(0.045)	(0.036)	(0.039)
	{0.907}	{0.430}	{0.321}	{0.482}
R^2	0.004	0.009	0.015	0.017
Number of Phone Surveys	1600	1600	1600	1600
Number of ALATs	33	33	33	33
Includes basic controls			✓	✓
Includes cognitive controls				✓

[a] Column 1 includes a separate indicator for observations in cell C in wave 1; in wave 1, treatment corresponds with selected. This estimate is an unbiased estimate of the population average treatment effect. The remaining columns use all data collected from large ALATs. [b] Basic controls include gender, age, marital status, and average distance to farmers. Cognitive controls include digit span, the big 5 stability metatrait, and the big 5 plasticity metatrait. [c] All control variables standardized to have mean 0 and standard deviation 1. Observations that are missing controls are imputed to be zero and an indicator variable for the missing control is included. [d] All regressions also include separate wave indicators for selected and unselected AEAs. [e] Liang and Zeger (1986) cluster robust standard errors and p-values from a wild bootstrap procedure with 10,000 replication draws reported in parentheses and curly braces, respectively. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: First Stage Probit Regressions

	(1)	(2)
Constant	0.263***	0.253***
	(0.093)	(0.100)
	{0.005}	{0.009}
Male	-0.108	-0.099
	(0.155)	(0.174)
	{0.530}	{0.627}
Age	-0.558***	-0.580***
	(0.194)	(0.181)
	{0.005}	{0.004}
Married	0.470**	0.447**
	(0.222)	(0.213)
	{0.031}	{0.044}
Average Distance	0.179	0.181*
	(0.084)	(0.086)
	{0.110}	{0.086}
Digit Span		-0.115
		(0.166)
		{0.530}
Big 5 — Stability		-0.210*
		(0.117)
		{0.097}
Big 5 — Openness		0.128
		(0.148)
		{0.395}
Psuedo R^2	0.124	0.147
Number of AEAs	95	95
Number of ALATs	33	33

[a] All covariates are first standardized to have population mean 0 and standard deviation 1. Observations that are missing relevant covariates are imputed to be zero. [b] Covariates in the probit regressions in columns 1–6 are then demeaned by the relevant choice set. That is, covariates in columns 1–3 are demeaned by ALAT and covariates in columns 4–6 are demeaned by CDA.

[c] Liang and Zeger (1986) cluster robust standard errors and p-values from Kline and Santos score bootstrap procedure with 10,000 replication draws reported in square brackets and curly braces, respectively. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Treatment Effect Heterogeneity on Observable and Unobservable Characteristics

	(1)	(2)	(3)
Main Effects:			
Inverse Mills Ratio	-0.025	-0.022	-0.019
	(0.028)	(0.020)	(0.023)
	{0.437}	{0.324}	{0.486}
Average Treatment Effect	0.059	0.057*	0.066*
	(0.035)	(0.025)	(0.028)
	{0.107}	{0.054}	{0.056}
Interactions with Treatment:			
Inverse Mills	0.091**	0.069**	0.068**
	(0.036)	(0.021)	(0.026)
	{0.030}	{0.013}	{0.033}
Age		-0.059	-0.074
		(0.036)	(0.035)
		{0.141}	{0.124}
Married		-0.041	-0.047
		(0.030)	(0.026)
		{0.255}	{0.143}
Average Distance to Farmers (log)		0.064	0.057
		(0.035)	(0.038)
		{0.166}	{0.320}
Male		-0.001	0.010
		(0.025)	(0.029)
		{0.982}	{0.782}
Digit Span			-0.074**
			(0.030)
			{0.034}
Big5 — Stability			-0.041
			(0.048)
			{0.535}
Big5 — Openness			0.036
			(0.025)
			{0.327}
R^2	0.00902	0.0221	0.0285
Number of Phone Surveys	1600	1600	1600
Number of ALATs	33	33	33
Basic Controls		✓	✓
Cognitive Controls			✓

[a] The inverse mills ratio is the generalized residual—the expected value of the error term—from a probit regression of being selected on the corresponding controls from the column. [b] Main effects on basic controls and cognitive controls omitted for space. [c] All control variables standardized to have mean 0 and standard deviation 1. Observations that are missing controls are imputed to be zero and an indicator variable for the missing control is included. [d] All regressions also include separate survey wave indicators for selected and unselected AEAs.

[e] Liang and Zeger (1986) cluster robust standard errors and p-values from a wild bootstrap procedure with 10,000 replication draws reported in parentheses and curly braces, respectively. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.