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

July 2018

Abstract

Standard models of hierarchy assume that agents and middle managers are better informed than principals about how to implement a particular task. We estimate the value of the informational advantage held by supervisors (middle managers) 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 that, before randomization of treatment, elicited from supervisors which AEAs they believed should be prioritized for treatment. We find that supervisors did have valuable information—they prioritized AEAs who would be more responsive to the monitoring treatment. We develop a model of monitoring under different allocation rules and rollout scales (i.e., the share of AEAs to receive treatment). We semi-parametrically estimate marginal treatment effects (MTEs) to demonstrate that the value of information and the benefits to decentralizing treatment decisions depend crucially on the sophistication of the principal and on the scale of rollout.

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

*We would like to thank José Molinas, the Minister of Secretaría Técnica de Planificación del Desarrollo Económico y Social (STP), whose initiative, support, and guidance made this project possible. Maureen Stickel and Patricia Paskov provided excellent research assistance. We also gratefully acknowledge the IGC and JPAL-Governance Initiative for their generous financial support. We thank Anukriti, Rachel Heath, Melanie Khamis, Adriana Kugler, Annemie Maertens, Shing-Yi Wang, and seminar participants at the ASSA and Banff meetings, SITES summer school, UC Berkeley, UW Madison, and Yale for thoughtful comments.

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 – often 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 giving 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, who 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 over random assignment that supervisors possess in targeting the cell phones.

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 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 15.4 percentage points compared to a statistically insignificant 3.6 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 in general both commonly observed demographic traits (e.g., gender) and even harder-to-measure characteristics such as their cognitive ability or personality type do a poor job of explaining supervisors' targeting decisions. Among the few observable traits that predict targeting, the AEAs party affiliation is one of the most robust. Supervisors are much less likely to place members of the incumbent party under additional monitoring, suggesting that non-benevolent motives may have influenced, at least in part, their targeting decisions. Nevertheless, 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 and use these to estimate 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 53 percent coverage. At this coverage, the supervisor allocation increases the share of farmers visited by 6.9 percentage points versus only a 3.3 percentage point increase under random assignment. A slightly more effective approach compared to random assignment 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 2.0 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) find that firms are more decentralized when located in regions that are judged to contain more trustworthy people by those in the headquarters location. The authors view trust as a proxy for congruency of preferences.

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

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

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.

We complement this study in some important ways. Our experiment was designed to identify who the supervisors would target for monitoring without having to rely on strong functional form assumptions. As a result, we can experimentally identify the decentralized counterfactual to a centralized approach. Moreover, that counterfactual depends both on supervisors' informational advantage and on potential preference biases, which we allow for but are absent from the targeting rules in Duflo et al. (2016). Thus, we incorporate elements that are crucial to the evaluation of the relative merits of decentralization.

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, she 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-means testing against community-based targeting of a social program. They find that proxy-means 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. In contrast with the standard MTE setup, our selection equation does not model an AEA's selfselection into treatment but rather the selection by a supervisor. Crucially, only those AEAs who were randomized into treatment were in fact treated. Thus, when we compute the MTEs we do not have to extrapolate to subgroups of "always-takers" and "never-takers" because we only have compliers by design. In this respect, our approach implements a variant of the selective trial designs proposed by Chassang et al. (2012). In that paper, the authors recast randomized control trials into a principal-agent problem and show theoretically how one can recover the MTEs necessary to forecast alternative policies and treatment assignments by eliciting subjects' willingness to pay for the treatment. Instead of eliciting our agents' willingness to pay for the treatment, we elicit the targeting preferences of the supervisor, who in our context is the relevant decision maker.

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 group meetings. Meetings often take place during farm visits in which problems with the farm are diagnosed and dealt with. Meetings are also used to talk about technical topics 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 the hierarchical agency model we will lay out in the next section, we view the ministerial leadership introducing the new technology as the principal, we will refer to the ALAT-level supervisors as "supervisors", and the AEAs as the "agents."

3 Model

Consider a hierarchy composed by a principal, a supervisor, and a continuum of agents with mass 1. The supervisor is responsible for monitoring the agents. In such a hierarchy there are two possible agency problems: that between the agent and his supervisor and that between the supervisor and the principal. We will focus mainly on the problem between agent and supervisor, and analyze how it changes when the agents are placed under a new monitoring technology. The question will be whether the principal can obtain better results by relying on supervisors in deciding how to deploy the technology.

Agents and monitoring Each agent caters 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 choose a share $s_i \in [0,1]$ of farmers to visit. The agent obtains an intrinsic motivation $m_i s_i$ from visiting a share s_i of farmers, but also incurs a cost $a_i s_i + b_i \frac{s_i^2}{2}$. The share s_i is a measure of agent effort, and because it directly constitutes a measure of service provision (visits to farmers), the principal cares about it. From now on, we will refer to s_i as effort and assume that it is noncontractible.

The supervisor operates a monitoring technology such that with probability $q_i \in (0,1)$, she learns s_i and reprimands the agent in proportion to the amount by which his effort falls short, $1-s_i$. The agent gets a disutility from being reprimanded equal to $(1-s_i)r_i$, with $r_i > 0$.[2] While monitoring

[2] Alternatively, one may assume that the supervisor draws a farmer at random, and finds he has not been visited with probability $1-s_i$, in the event of which she proceeds to reprimand the agent with a fixed intensity r_i. It is also possible to extend the model to make q_i a function of monitoring effort by the supervisor. The choice of monitoring effort

allows the supervisor to obtain information about the agent's effort, it can potentially weaken intrinsic motivation. When monitored, the intrinsic motivation payoff of agent i becomes $(m_i - g_i)s_i$, which is potentially negative. It reflects the fact that agents may feel aggrieved to an extent $g_i \geq 0$ when under close supervision. In sum, wages and effort costs accrue to the agent regardless of supervision, while reprimand and intrinsic motivation payoffs accrue in relation to monitoring intensity q_i. Thus, agent i can be seen to maximize utility,

$$u_i(s_i) = w - a_i s_i - b_i \frac{s_i^2}{2} + q_i \left[(m_i - g_i) s_i - (1 - s_i) r_i \right] + (1 - q_i) m_i s_i,$$

or, collecting terms,

$$u_i(s_i) = \omega_i + \mu_i s_i - b_i \frac{s_i^2}{2} + q_i s_i \rho_i,$$

where $\omega_i \equiv w - r_i q_i, \mu_i \equiv m_i - a_i, \rho_i \equiv r_i - g_i$. Agent i chooses the share s_i of farmers to visit to maximize utility $u_i(s_i)$, and he does so after learning the level of monitoring intensity q_i he is under. Because $u_i(s_i)$ is concave, agent i's optimal effort is $s_i^*(q_i) = \max\left\{0, \frac{q_i \rho_i + \mu_i}{b_i}\right\}$. Since b_i only affects effort through ratios involving ρ_i and μ_i, parameters that can be scaled arbitrarily, we normalize $b_i = 1$, yielding,

$$s_i^*(q_i) = \max\left\{0, \min\left\{q_i \rho_i + \mu_i, 1\right\}\right\}. \tag{1}$$

The term μ_i – a proxy for net-of-cost intrinsic motivation – is individual-specific and for some agents potentially negative. Even more important for our purposes, the term ρ_i, which captures both the agent's distaste for being reprimanded (which raises effort) and his resentment at being monitored (which lowers effort) is also potentially negative for some agents. We will assume ρ_i to be drawn from a continuous distribution $F(\rho_i)$ over a support $[\rho_l, \rho_h]$, where $\rho_h > 0$ but ρ_l is potentially negative.

New technology and treatment effects We assume that q_i can take one of two levels $\{q_l, q_h\} \in (0,1)$, with $q_h \equiv q_l + t_i \Delta q$, $\Delta q > 0$, where q_l denotes a status quo level of monitoring, and $t_i \in \{0,1\}$ reflects whether agent i is "treated" to a new monitoring technology.[3] In order to characterize treatment effects neatly and avoid awkward truncation issues, in what follows we will assume that

remains unmodeled here, in order to stick with the simplest formulation that will deliver the results of interest. Such an extension could also involve an agency problem in the supervisor's choice of monitoring effort without affecting the essence of our results. The only tension between supervisor and principal that may arise in our simpler setting relates to the deployment of the monitoring technology to be described below.

[3]Here we assume treatment only affects the agent's problem by raising monitoring intensity, although it could in principle also affect μ_i via the agent's cost a_i. This is plausible as some technologies, like GPS phones, can be productivity-enhancing. However, as we will show later, the data do not support that possibility.

$min\{q_h\rho_l, q_l\rho_l\} + \mu_i > 0$ and $q_h\rho_h + \mu_i < 1$, which guarantees interior solutions for s_i.

While μ_i and ρ_i both affect the level of effort, only ρ_i affects the response of effort to a change in monitoring technology. Thus, in the remainder of this section we will refer to different levels of ρ as agents' "types." Under increased monitoring, an agent of type ρ increases his effort by $\rho\Delta q \equiv T(\rho)$, which captures the treatment impact of the new technology for that agent. Note that since ρ_l can be negative, $T(\rho)$ can be negative for some types. To deploy 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." Given a continuum of agents, the total (and average) treatment impact is $\int_{\rho_l}^{\rho_h} T(\rho)f(\rho)d\rho$, achieved at a total (and average) cost c. If the new technology is deployed on all agents with type above some level k, the total treatment impact over all agents is $\int_{k}^{\rho_h} T(\rho)f(\rho)d\rho$, achieved at total cost $c(1-F(k))$. Note that our definition of the total treatment abstracts from spillover effects across agents. These effects could be modeled, but we keep the theory consistent with our empirical approach to measuring marginal treatment effects, which will likewise abstract from spillovers.

Inspection of the expression for the treatment impact $\int_{\rho_l}^{\rho_h} T(\rho)f(\rho)d\rho$ yields the following:

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

This remark, while somewhat obvious, highlights the conditions under which a new technology rolled out to all agents, or a representative sample of them, would yield positive results when assessed through a standard impact evaluation that estimates average treatment effects. In addition, the definition of $T(\rho_i)$ implies that equilibrium agent effort s_i^* (weakly) increases in monitoring technology q_i for all agents with $\rho_i > 0$, and an improvement in monitoring technology (an increase in q_i) has a more positive effect on the effort of agents with a higher type ρ_i.

The value of information and optimal decentralization To isolate a central advantage of decentralization, we assume that the principal knows the distribution of types $F(\rho)$, but does not know the type of any specific agent. The supervisor, in contrast, knows both $F(\rho)$ and agents' individual types – this constitutes the information advantage of the supervisor vis-a-vis the principal. Both principal and supervisor know all other model parameters. The thought experiment of interest is whether, given a new monitoring technology, the principal would want to delegate to supervisors the choice of which agents to treat with it.

Centralization We take a centralized regime to be one in which the principal makes all decisions without any further input beyond what she already knows, namely $F(\rho)$.[4] Thus, she can only make a general decision about whether to adopt the new technology or not and cannot determine whether any one agent is more profitably treated than another. We denote the scale of adoption with m (for the *measure* of the treated, not to be confused with the individual-specific intrinsic motivation m_i used earlier). If roll-out has scale m, and treated agents are selected at random, the total treatment impact will be $m \int_{\rho_l}^{\rho_h} T(\rho) f(\rho) d\rho$, which increases linearly in m as illustrated by the strictly increasing diagonal line in Figure 1 depicting the impact of random treatment assignment. The cost will be mc. The principal will adopt whenever $\int_{\rho_l}^{\rho_h} T(\rho) f(\rho) d\rho \geq c$ (breaking indifference in favor of adoption), i.e., whenever the average treatment effect of the new technology is larger than its marginal and average cost. If this condition is met, a roll-out at 100% would produce a total treatment impact equal to the average treatment effect, recommending not only adoption, but also adoption at full scale.

Decentralization In the decentralized regime, the principal can pay a cost $d \geq 0$ to delegate to the supervisor the decision over which agents to place under the new monitoring technology.[5] For simplicity, we focus on a well-meaning supervisor who deploys the new technology to maximize agent output. Our empirical approach allows for potential supervisor bias. If the marginal cost of the new technology is lower than the treatment effect for the type with highest type ρ_h, a benevolent supervisor will place agents under the new monitoring system starting from the highest type ρ_h and work downwards. How far down he goes depends on the scale of roll-out for the new technology. If the supervisor chooses which agents to treat but the scale of roll-out is fixed, he will choose the highest types to fill the quota. Thus, if the supervisor is told to place a share m of agents under the new technology, he will treat every agent with type $\rho \in \left[\rho_m \equiv F^{-1}(1-m), \rho_h\right]$. This implies,

Remark 2. *If supervisors know agents' types and assign treatment with a benevolent intent of maximizing visits to farmers, treatment effects on those agents selected by supervisors will be higher*

[4] We equate centralization with a regime where the principal makes all decisions based upon her own information, and decentralization to one where the principal delegates decisions to supervisors, or, equivalently, one where supervisors submit information that mechanically drives the principal's decisions. Thus, we abstract from the interesting distinctions made by Dessein (2002) between delegation and strategic communication.

[5] 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 the supervisor's recommendations and/or technology deployment decisions. In some empirical settings, like the one discussed in the introduction on screening candidates for income support, the costs of decentralization are fixed and likely large. The reason is that identifying the best units to treat–even if they are just a few–may require deploying a nation-wide organizational operation. In other settings costs may be small, and in others even negative, since centralization may at times be costlier. In the latter cases, there will be no tension – decentralization is both informationally advantageous *and* cheaper – and therefore of less analytical interest to us. In the case in which decentralization costs are variable rather than fixed there will be quantitative differences in terms of the roll-out rates that make centralization preferred to decentralization, but the basic point will remain that marginal treatment effects and roll-out rates matter.

than treatment effects on agents selected at random.

If, against our assumptions, the supervisor is not well informed, then the treatment effects among agents selected by the supervisor could be similar to those among agents selected at random. If the supervisor has mistaken views or is not benevolent, treatment effects among agents selected by him could be even lower than among those selected at random: a supervisor who ranks types to be treated in an inverse way (i.e., starting with ρ_l and working upwards) would in fact minimize the impact of technology adoption.

Getting back to the case of an informed and benevolent supervisor, it is helpful to consider the situation in which the supervisor also has control over the scale of adoption. In this situation, he will choose the lowest treated type k to maximize,

$$\int_k^{\rho_h} T(\rho) f(\rho) d\rho - c(1 - F(k)),$$

which 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 Consider the case where the supervisor, under decentralization, has authority over the selection of agents to be treated but not over the scale of technology adoption.[6] Given a scale of adoption m, the principal will choose to decentralize if and only if $\int_{\rho_m}^{\rho_h} T(\rho) f(\rho) d\rho - cm - d \geq \left(\int_{\rho_l}^{\rho_h} T(\rho) f(\rho) d\rho - c \right) m$, or equivalently, iff,

$$\iota(m) \equiv \int_{F^{-1}(1-m)}^{\rho_h} T(\rho) f(\rho) d\rho - m \int_{\rho_l}^{\rho_h} T(\rho) f(\rho) d\rho \geq d, \tag{2}$$

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 $\rho_m = \rho_h$, 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 because the new technology would be applied to very few agents, the value of the informational gain from decentralization is zero and does not justify paying a fixed positive cost d to decentralize. On the other extreme, where $m = 1$ and $\rho_m = \rho_l$, the difference in value again goes to zero because the advantage of treating the more responsive agents first is completely diluted. Since all agents

[6] A realistic example fitting our empirical setting is when a new technology is acquired by government and is made available to an agency in a fixed amount.

will be treated, there is no need to decentralize.

For every value m strictly between 0 and 1, the total treatment effect attained by a supervisor who treats the most responsive agents first is larger than that which can be attained by assigning treatment at random. As illustrated in Figure 1, the value of information $\iota(m)$ is decreasing in m near $m = 1$ (or, equivalently, increasing in ρ_m near ρ_l). This is true up to a type $\bar{\rho}$ for whom the marginal treatment effect $T(\bar{\rho})$ is equal to the average treatment effect $\bar{T} \equiv \int_{\rho_l}^{\rho_h} T(\rho) f(\rho) d\rho$.[7] The value of information $\iota(m)$ then increases in m as m approaches 0 (or equivalently decreases in ρ_m as ρ_m approaches ρ_h. Given these considerations, and recalling continuity of F, standard intermediate value theorem arguments imply:

Proposition 1. *(i) If $0 < \iota(\bar{\rho}) < d$, then decentralization is never optimal and if $d = 0$, decentralization is always optimal. (ii) If $0 < d < \iota(\bar{\rho})$, there exist two values $m' < m''$ in $[0,1]$ such that for any scale of roll-out of the new technology $m \in [m', m'']$ the principal prefers decentralization to centralization; whereas 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, which in turn depends crucially on the scale at 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. To see this, suppose an intervention satisfies $\int_{\rho_l}^{\rho_h} T(\rho) f(\rho) d\rho < c$, yielding an average treatment impact below marginal and average cost, so adoption at a 100% scale would yield a loss. But suppose also that treatment impact is larger than cost for a set of highest types, so that $T(\rho) > c$ for all types in $\rho \in (\rho', \rho_h]$, with $T(\rho') = c$, and that $\int_{\rho'}^{\rho_h} T(\rho) f(\rho) d\rho > c(1 - F(\rho')) + d$. In other words, there is a set of types for whom treatment effects are larger than the marginal cost by more than the cost of decentralization. In this situation the principal would gain by delegating to the supervisor the adoption decision if the latter will treat only those types in $(\rho', \rho_h]$. This suggests that impact evaluation cannot abstract from the extent of roll-out and its implementation mode, i.e., centralization versus decentralization, since the implementation mode affects who gets treated. In other words, determining whether a technology is valuable – presumably the ultimate goal of an impact evaluation – requires assessing 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

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

average? Second, do supervisors have valuable knowledge (net of potential limitations in benevolence) about which agents ought to be treated given partial roll-out? If so, we should observe treatment effects among those selected by supervisors to be larger than among agents who were selected into treatment at random. Third, could the scale of roll-out alter the relative advantage of decentralization vs centralization? Answering these questions will require developing a method for ascertaining the marginal treatment impact on different types of agents while designing 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 (AEAs). Some ALATs have a single AEA, but 48 ALATs have at least 2 AEAs. We asked the supervisors of the latter group to indicate which half of his AEAs should receive the phones first given the program's objective to increase worker performance. We refer to these AEAs as "selected." 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, C, and D in Figure 2. The ALATs in cells B and D (a quarter of the ALATs), serve as the treatment group. In these ALATs all AEAs, both selected and non-selected, received the GPS-enabled cell phone which increased monitoring. The ALATs in cells A and C, (half of the ALATs), serve as our control group as no AEAs received the phones in these groups. The average difference in performance between AEAs in cells B and D and AEAs in cells A and C estimates the average impact of treatment. And, the difference-in-differences computed as the performance by AEAs in cells $(B-A)-(D-C)$ estimates whether the impact on selected AEAs is larger than the impact on non-selected AEAs. This difference-in-differences allows us to determine whether supervisors had valuable information on how to direct treatment. A third group of ALATs (cells E and F) received partial treatment. Only those AEAs who had been selected by their supervisors for treatment were treated immediately (cell E). This design helped make the elicitation of supervisors' preferences credible and relevant. Eight months after the delivery of these phones, a second wave of phones were delivered to the AEAs in group F.

The difference in performance between the AEAs in cell F and in cell C provides a test of whether allocating phones to the selected AEAs can also affect the performance of non-selected AEAs in the same ALAT. This would be the case if the supervisors also responded to the treatment of those in cell E by monitoring more intensively the AEAs without the cell phone in cell F. Unfortunately, after the randomization we discovered that several AEA characteristics were not balanced across these treatment arms. Consequently, our estimates of spillover effects vary considerably depending on the model specification. For this reason, we have decided to restrict our main analysis to the shaded cells (A, B, C, and D) where, as we demonstrate below, the randomization functioned properly.

4.1 Taking the Theory to the Data

Recall that the performance of AEAs whenever interior is given by equation (1):

$$s_i^* = q_i \rho_i + \mu_i.$$

To operationalize this equation, recall that the level of monitoring for each AEA, q_i, is a function of the monitoring technology t_i, according to the expression $q_i = q_l + \Delta q t_i$ 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 AEAs respond to exogenous changes in q_i, we normalize $q_l = 0$ and can rewrite the expected disutility of being reprimanded (net of monitoring grievance) to be $q_i \rho_i = \beta_i t_i$, where $\beta_i = \Delta q \rho_i$.

The central goal of our approach is to model various selection criteria and estimate the marginal treatment effects under each 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 (μ_i, β_i), in an individual AEA's effort function in equation $s_i^* = \mu_i + \beta_i t_i$. In particular, we will map these parameters into a vector of fixed characteristics (X_i) and two independently random characteristics (ε_i, η_i), to write: $\mu_i(X_i, \varepsilon_i)$ and $\beta_i(X_i, \eta_i)$. While the vector X_i may be observable to both the principal and supervisor, the elements (ε_i, η_i), may only be partially observed 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: $\mu_i = \mu' X_i + \varepsilon_i$ and $\beta_i = \beta_0$. An individual AEA's effort function becomes

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

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 AEAs in both treated cells B and D

relative to AEAs in the control cells A and C. 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 score bootstrap procedure procedure to account for the fact that we have relatively few clusters (Wu, 1986; Kline and Santos, 2012). In estimating Equation 3, we can also include the single AEA ALATs, which were assigned phones at random. In this randomization, one-third of AEAs 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 AEA.

Average treatment effect by supervisor's choice To test whether supervisors are able to select those AEAs whose effort would most increase when 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

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

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 AEAs in the treatment and control groups (cells $B - A$) net of the difference in non-selected AEAs in the treatment and control group (cells $D - C$). 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 μ' and $\beta = (\beta_0, \beta_1)$ via ordinary least squares since t_i is randomly assigned. This is because the supervisor's selection D_i^S is elicited in a way that does not affect treatment assignment in cells A, B, C, and D.

4.2 Estimating the Marginal Treatment Effects of the Program

A strictly positive value for β_1 in 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 the average treatment effect be at scales other than 50 percent? We asked supervisors to select half of their AEAs but this pilot implementation does not directly tell us what β_1 would be at different selection shares. In this section we develop a method for tracing out the 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 AEAs to treatment. In addition, this framework will allow us to consider 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 AEAs according to a suitably defined latent index model.

When we implement empirically the study of supervisors choices, how worthy of treatment a particular AEA is in the eyes of the supervisor will be seen as a function of observables X_i and unobservables u_i according to the function $\Gamma' X_i + u_i$. In what follows we develop some structure to link this empirical object to the theory.

In the case of decentralization, supervisors select AEAs according to some value they perceive from treating supervisor i,

$$v_i = \beta_i(X_i, \eta_i) + \psi_i(X_i, \zeta_i), \tag{5}$$

where v_i is AEA 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 treating AEA i that depends on X_i and an independent, idiosyncratic preference term ζ_i. A benevolent supervisor would only select AEAs 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 may 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 element (ξ), the supervisor faces uncertainty. A risk neutral supervisor will assign monitoring technology to AEAs depending on the expected value $\mathbb{E}\{v_i | X_i, \theta_i, \zeta_i\}$. The expectation is taken over ξ, and conditional on ζ, since to the supervisor the former represents noise while the latter may capture preferences.

Given a selection criterion (such as v_i), and a well defined measure of diversity across AEAs as given by a joint distribution over (X_i, θ_i, ζ_i), it is possible for the supervisor to rank order all AEAs 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 scale m under a selection criterion based on v_i implies treating all AEAs 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 AEAs. 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 for the supervisor to want to treat more AEAs, the putative cost of treatment must be lower; for the supervisor to treat no AEAs, the putative marginal cost of treating a single AEA must exceed the benefit of treating the most valuable AEA; and that for the supervisor to treat all AEAs, the expected desirability of treating the least valuable AEA must cover the putative 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 AEA-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 AEA. 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 $\mu_i(\cdot)$, $\beta_i(\cdot)$, and $\psi_i(\cdot)$ on X_i. We parameterize each of these linearly. Slightly abusing notation, and anticipating our assumption that η is mean zero, we can re-write equations (3) and (5 respectively as,

$$\begin{aligned} s_i^* &= \underbrace{(\mu' X_i + \varepsilon_i)}_{\mu_i(\cdot)} + \underbrace{(\beta' X_i + \eta_i)}_{\beta_i(\cdot)} t_i \\ &= \mu' X_i + (\beta' X_i) \times t_i + \varepsilon_i + \eta_i \times t_i. \end{aligned} \tag{6}$$

and

$$v_i = \underbrace{(\beta' X_i + \eta_i)}_{\beta_i(\cdot)} + \underbrace{(\psi' X_i + \zeta_i)}_{\psi_i(\cdot)}. \tag{7}$$

Marginal treatment impact under an uninformed principal An uninformed principal knows nothing about individual values of β_i, so she can only select which AEAs should be placed under the new technology at random. Using equation 6, given a scale of roll-out m (the share of AEAs to be treated), the total treatment effect on expected performance is

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

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 AEAs are treated, the total gains are zero. If all AEAs 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 AEA's characteristics (X_i, θ_i, ζ_i), and selects AEAs to treat according to the value of the expected index $\mathbb{E}\{v_i|X_i, \theta_i, \zeta_i\}$ as given by,

$$\begin{aligned} \mathbb{E}\{v_i|X_i,\theta_i,\zeta_i\} &= \underbrace{(\beta' X_i + \mathbb{E}\{\eta_i|X_i,\theta_i\})}_{\mathbb{E}\{\beta_i(\cdot)|\theta_i,X_i\}} + \underbrace{(\psi' X_i + \zeta_i)}_{\psi_i(\cdot)} \\ &= \underbrace{(\beta' + \psi')}_{\Gamma'} X_i + \underbrace{(\mathbb{E}\{\eta|X_i,\theta_i,\zeta_i\} + \zeta_i)}_{u_i}. \end{aligned} \tag{8}$$

This equation is important for our linking the theory with the empirics of AEA selection by supervisors. The AEA observables in X_i matter both because they affect response to treatment (through β', as in the theory), but also because supervisors may have biases (through ψ'). Unobservables in u_i may also reflect components that affect response to treatment (through η) and biases (through ζ).

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. Given all of these distributional assumptions, the variable u_i can be characterized as drawn from Φ, a cumulative Normal $(0, \sigma_u^2 = \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\xi^2}\sigma_\eta^2 + \sigma_\zeta^2)$ (this stems from the fact that the supervisor is Bayesian and updates his expectation of η upon observing θ). This, in turn, implies that D_i^S takes the familiar form of a probit model:[8]

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

Under these assumptions, standard arguments yield $\mathbb{E}\{\eta_i|u_i\} = \frac{\sigma_{\eta u}}{\sigma_u^2}u_i$, where $\sigma_{\eta u} = \left(\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\xi^2}\sigma_\eta^2\right)$. It follows that $\mathbb{E}\{\eta_i|D_i^S, X_i, m\} = \frac{\sigma_{\eta u}}{\sigma_u}\frac{\phi(\frac{1}{\sigma_u}(\Gamma' X_i - c_p(m)))}{D_i^S - \Phi(\frac{1}{\sigma_u}(\Gamma' X_i - c_p(m)))} \equiv \frac{\sigma_{\eta u}}{\sigma_u}\lambda(D_i^S, X_i, m)$. Using these expressions after taking conditional expectations in equation 6, we get [9]

$$\begin{aligned} \mathbb{E}\{s_i^*|X_i, D_i^S, t_i, m\} &= \mu' X_i + (\beta' X_i) \times t_i + \mathbb{E}\{\varepsilon_i|X_i, D_i^S\} + \mathbb{E}\{\eta_i|X_i, D_i^S, m\} \times t_i \\ &= \mu' X_i + (\beta' X_i) \times t_i + \frac{\sigma_{\eta u}}{\sigma_u}\lambda(D_i^S, X_i, m) \times t_i. \end{aligned} \tag{9}$$

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 9 via a two-step procedure using OLS. The first step allows us to estimate the selection model that will yield $\lambda(D_i^S, X_i, m)$ and the second step yields estimates for the coefficients in equation 9.

Equation 9 is the crucial resource to estimate the marginal treatment impact of the intervention under different scenarios of decentralization and informational advantage. To see this, consider first the simplest case where neither the principal nor the supervisor can observe any AEA traits so the vector X_i is constant. The expected index on which the supervisor selects is $\mathbb{E}\{v_i|\theta_i, \zeta_i\} = \mathbb{E}\{\eta|\theta_i, \zeta_i\} + \zeta_i = u_i$. Given the 50 percent roll-out in the experiment, we know that under decentralization, the total treatment impact of 50 percent roll-out 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.[10] 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 AEA for other reasons (ζ). But if $\beta_1 > 0$ we

[8] 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.

[9] 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 μ) is not of direct interest to us and is not required for identification, but may improve the efficiency of the other estimates.

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

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 AEAs, 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 AEA 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 AEA with traits X_i being selected. Knowing traits X_i and a conditional expectation on u_i for an AEA being selected at a given level of roll-out, Equation 9 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 AEAs, captured by X_i. This opens up consideration to a new class of counterfactuals, with a natural one involving the marginal treatment impact for varying m for a decision maker that knows X_i but does not observe θ_i. Thus, we can ask whether a sophisticated centralized principal can emulate or surpass the performance of supervisors despite her informational disadvantage. We perform these exercises in Section 7.

Discussion We have presented a heterogeneous treatment effect model where supervisors have private information about the treatment effects. Equation 9 shares the same functional form as the "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 and then uses OLS in the firs stage 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 (Heckman and Vytlacil, 2005; 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 not 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 misspecified, $\lambda(D_i^S, X_i)$ 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.[11]

5 Data

We collected two main sources of data. The first is a survey of AEAs. Each AEA and supervisor independently filled out answers on a paper questionnaire with survey enumerators available to answer any questions. The survey contains questions regarding the AEAs' demographics, work history, and measures of personality such as the digit span test measuring cognitive ability, the Perry public service motivation index (Perry, 1996), the moral disengagement scale (Moore et al., 2012), and the Big-5 inventory (John et al., 2008), which we combine into two higher-order personality traits called Stability and Plasticity. Stability combines Neuroticism, Agreeableness, and Conscientiousness and therefore keeps track of traits that are usually found in the literature to predict earnings and job attainment, such as the tendency to remain emotionally stable and motivated

[11]Thus, if one wanted to stick with OLS in the first stage rather than a Probit, while continuing to assume a linear conditional expectation function $\mathbb{E}\{\eta_i|u_i\} \propto u_i$, the coefficient on $\lambda(\cdot)$ in the second stage would be numerically equivalent to estimating an OLS regression in one step with $\frac{D_i^S}{2}$ in place of $\lambda(\cdot)$, a result that follows immediately from the Frisch-Waugh theorem. Its unique role in our context derives from its tagging supervisors' choices so that it may reflect their perceptions of how AEAs respond to treatment.

and be organized and thorough. Plasticity, which aggregates Extraversion and Openness, is a measure of a person's gregariousness and openness to new experiences. These two meta-traits tend to account for much of the shared variance among the lower order dimensions (DeYoung, 2006).

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 traveled across the country to meet with the AEAs who were scheduled to receive phones, distribute the phones to them, and teach them how to use the phones. This process took over two months because it 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 the survey of all AEAs as well as their supervisors. We treat AEA characteristics such as sex, age, years of education, and the personality indices as being fixed and 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, those ALATs where no AEAs received phones, these responses should not be 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 but in the end did not do so.

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 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 AEA 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 AEA 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 multi-AEA 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 single-AEA 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 called 2,635 farmers in the first round and 2,642 in the second round for the 182 AEAs who responded to the AEA survey. Of those, 68% led to completed surveys.[12] 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.[13] 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. The table distinguishes between treated and control small single-AEA ALATs (columns 1 and 2) and treated and control large multi-AEA 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]

[12]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.

[13]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.

[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 a number that was two digits long and then increased incrementally in the number of digits until the AEA could no longer recall a number correctly on both of two chances.

AEAs are also required to travel on average 12 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 in a balanced way.[15]

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 AEAs 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 scales 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) through (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. In columns (1) through (3), our estimation sample includes all AEAs in the small and large ALATs, excluding those randomized into the partial treatment cells (cells E 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 and gender), and in column (3), we further augment the specification to include controls measuring AEA personality type (e.g., Big 5 meta-traits and Digit Span). In column (4), we re-estimate the specification presented in column (3), excluding the single-AEA ALATs.

We find that the increase in monitoring leads AEAs to visit their farmers more often. They are approximately 6 percentage points more likely to have visited a given farmer in the past week,

[15] In Appendix Table A1, we also check for balance on a set of ALAT-level characteristics extracted from the population and agricultural censuses, as well as the 2013 presidential elections. We look at 18 comparisons, and only one shows significant imbalance across treatment and control. The results in the table again suggest that the randomization led to balance across treatments. The most noticeable difference between small and large ALATs is that large ALATs are located districts with both larger urban and rural populations, and a lower share of their population is rural.

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 multi-AEA ALATs, which will be our main sample moving forward. Overall, the demographic and personality-based controls have little predictive power.[16]

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 = -0.008; clustered standard error = 0.036). 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 AEAs 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 1.6 percent less time on each visit, or approximately one and a half minutes, is small and statistically insignificant.

In Appendix Table A2, we examine the effects of the treatment on other dimensions of performance. We consider four additional measures: 1) how satisfied the farmer is with the AEA (1=very, 2=somewhat, 3=not at all); 2) an indicator for whether the farmer thought the AEA conducted helpful training sessions; 3) an indicator for whether the farmer did not find the AEA helpful at all; 4) and the first principal component for the three measures, with higher values indicating worse performance. All four measures are significantly correlated with AEA visits (see Appendix Table A3). Farmers who receive more visits from their AEAs are also more likely to think the AEA conducts helpful training sessions, are more satisfied with their AEAs, and more likely to find the AEA helpful in some way. In general, we find that the additional monitoring improved performance along these dimensions as well, although the estimates are measured with less precision. Based on our principal component measure, the treatment improved aggregate performance by 0.14 of a standard deviation (standard error = 0.07).

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

6.2 Do Supervisors Have Useful Information?

Recall that our model assumes AEAs 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 AEAs for whom a larger increase in performance ought to be expected. Our research design allows us to test this precisely.

Prior to the randomization, supervisors identified which half of their AEAs they believed should receive the phones. Given these selections, 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 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., $(B - A) - (D - C)$).

From Table 3, we see that the effects of the phones on performance are entirely driven by the effects on the AEAs prioritized to receive the phone prior to the randomization. These AEAs increased the share of farmers visited in the last week by approximately 15 percentage points . Compared to the prioritized AEAs in the control, this effect represents a substantial increase of 54 percent. From column (2), we also see that prioritized AEAs 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, married, and had to travel further distances to visit their farmers (although this last characteristic is only significant at a 89 percent level of confidence). In terms of their personality traits, supervisors were more likely to select AEAs with lower levels of the Big-5 Stability meta-trait. Individuals with higher stability scores may be more likely to stay motivated and have better relationships with their supervisors.

Interestingly, we also find that supervisors of the large ALATs, who except for one supervisor are all registered with the incumbent political party, are significantly less likely to place AEA's who are members of the incumbent party under increased monitoring. This suggests that either supervisors are acting non-benevolently or, as we will subsequently test, that party affiliation serves as a marker for those who are less likely to respond to treatment. 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 18.5%. This opens the possibility that supervisors are also 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. Ultimately, the only way to determine whether there could be an advantage to decentralization, is to rely on our experimental design, and ask whether supervisors select AEAs who will be more responsive to treatment.

In Table 5, we present a series of second stage estimates based on Equation 9. 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) and (3), the first stage regressions correspond to the ones presented in Table 4.

The key finding in Table 5 concerns the inverse Mills ratio 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 coefficient on the inverse Mills ratio interacted with treatment 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 (columns 2 and 3), 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 cognitive ability of the AEAs; those who performed worse on the digit span test exerted more additional effort in response to the treatment. Moreover, once we account for these differential effects, we do not find statistical evidence that members of the incumbent party respond less to the treatment. This suggests that non-benevolent motives may have influenced the supervisors' targeting.

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. First, what is the scale of roll-out anticipated. Under decentralization, the anticipated scale of roll-out should be whatever is optimal, and this motivates the need to identify that optimal level. 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:

$$\begin{aligned} \Delta Y^{CF} &= \mathbb{E}\{\underbrace{\beta_i(X_i,\eta_i)}_{\text{how much?}} \times \underbrace{D_i^{CF}}_{\text{who?}}\} \\ &= \int \mathbb{E}\{\beta(X_i,\eta_i)|D_i^{CF}=1\}Pr\{D_i^{CF}=1\}dX_i \end{aligned} \tag{10}$$

In keeping with the rest of our notation, we write our arbitrary selection rule as a threshold problem, $D_i^{CF}(X_i,u_i) = 1[\tilde{\Gamma}'X_i + \tilde{u}_i \geq c_p]$; 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 $Pr\{D_i^{CF}=1|X_i,c_p\}$ and $\mathbb{E}\{\beta(X_i,\eta_i)|D_i^{CF}\}$, which map into the scale of rollout m and the aggregate counterfactual impact ΔY^{CF}.

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 9, 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 10 to trace out the expected treatment effects of the cell phones under decentralization for any given threshold c_p or, by extension, any scale of 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 scenario.

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 AEAs receives a cell phone, and the expected total treatment effect is m% of the average treatment effect (considering, as in the theory section, a large number of AEAs who can then be approximated by a continuum). 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 6.4 percentage points, which corresponds to the average treatment effect in column (3) in Table 5. If instead she decided to treat only half of the AEAs, then we would expect an aggregate treatment effect of only 3.2 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. In Table 6, we present our estimated treatment effects at different roll-out levels for the various allocation rules we consider. The number displayed in bold represents the largest treatment effect under a given allocation rule.

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 $Pr\{D_i^S = 1|X_i\} = \Phi(\frac{1}{\sigma_u}(\Gamma' X_i - c_p))$ and the expected aggregate treatment effect is $\Delta\mathbb{E}\{s_i^*|X_i, D_i^S, T_i\} = \beta' X_i + \frac{\sigma_{\eta u}}{\sigma_u}\lambda(D_i^S, X_i)$. This counterfactual is depicted in Figure 3 with the solid line. Note that by construction, the curve must cross three points: the origin, 0.064 at 100% roll-out, and 0.070 at 53.8% roll-out which corresponds to the share of AEAs 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 53% where the additional treatment effect is over 3.5 percentage points. The optimal scale of roll-out under decentralization is not 53%, however, but 77%, at which level the total treatment effect is 7.7 percentage points. The total treatment effect starts to decline at a roll-out scale of 77% as we begin to assign to treatment individuals for whom the treatment effect is negative. The existence of individuals for whom the treatment effect is negative is consistent not only with our model (as ρ is allowed to be negative), but with the findings in de Rochambeau (2017), who showed that the introduction of a new monitoring device for truck drivers in Liberia lowered the productivity of the intrinsically motivated.

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 $\lambda = 0$. The dot-dash curve is much closer to the one under random assignment. This suggests that, in our setting, 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 AEAs.

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

Minimally informed principal: Assignment based on distance traveled 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 requires some information on the work environment of AEAs, and it constitutes the case we associate with a minimally informed principal. This counterfactual is displayed in Figure 4 with a dashed light gray line. We find that this method generally outperforms random assignment (a 2.0 p.p. advantage at 50 percent coverage), but it cannot beat the supervisor at any roll-out level.

Significantly informed principal: Assignment based on predicted baseline performance We consider a second type of partially informed principal that has the capacity to gather information on individual AEA 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 (see Appendix Table A4 for the estimation results). 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 baseline 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 that 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.

To overcome these difficulties, a sophisticated principal can conduct a pilot experiment at a low roll-out level and establish a map between AEA observable characteristics and response to treatment. Then it is possible to construct an assignment rule $D_i^{CF}(X_i)$ that allocates phones starting with those AEAs who are predicted to have the highest response to treatment and work downwards to treat progressively less responsive AEAs, tracing out the total treatment effect for each roll-out level. Note that we are priviliging principals in the "sophisticated" case because they would need to have digit span and Big-5 measures for all of their workers, which may be as much of a data constraint as running the pilot RCT.

As shown in Figure 4, this approach outperforms all others by a wide margin. The largest gap relative to the decentralized supervisor-choice approach is above 1.7 percentage points and occurs at a roll-out level of 38.4%. 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 9.0

percentage points and is achieved at a roll-out of 70%.

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 1.5 times the total treatment impact. Relative to the decentralized supervisor choice, the sophisticated centralized approach distributes roughly 10% fewer phones and attains roughly 1.3 additional percentage points in total treatment effect.

8 Conclusions

One of the primary benefits of decentralization is that mid-level supervisors 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 supervisors 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 middle managers is a maintained assumption in much principal-agent theory, evidence of the presence and extent of that advantage has been scarce. We also have little evidence on the effects of decentralization in a context in which the scale of implementation affects the average treatment effects.

In this paper, we establish that middle managers in the government hierarchy do have information which can improve targeting of an intervention. We develop an approach to trace out 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 enables supervisors in rural areas to track their agricultural extension agents.

Our experimental design randomly assigned monitoring devices across AEAs and independently elicited the preferences of their supervisors as to which AEAs 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 AEAs for treatment in a centralized fashion. In our setting, impacts resulting from treatment decided upon by a minimally informed principal are not as high as those attained under decentralization. However, a reasonably well-informed principal can approach the level of impact of decentralization. And in the best case scenario for centralization, a principal who can conduct a pilot RCT to obtain predictors of individual response to treatment can outperform supervisor choices; such a principal would substantially reduce the roll-out scale and still attain larger aggregate gains in AEA 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 (Bresnahan et al., 2002; Bloom et al., 2009), our findings suggest the opposite may occur, particularly if these technologies primarily serve to reduce the information gap between principals and agents (or middle managers such as supervisors).

Of course, the value of the information that supervisors 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. Our approach is designed for settings in which spillovers across treatment units are minimal. Thus, it would be interesting to extend our framework to incorporate the potential effects of spillovers in the calculus to decentralize. We view this as a potential avenue for future research.

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.

Chassang, S., Padró I Miquel, G., and Snowberg, E. (2012). Selective trials: A principal-agent approach to randomized controlled experiments. *American Economic Review*, 102(4):1279–1309.

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.

DeYoung, C. G. (2006). Higher-order factors of the big five in a multi-informant sample. *Journal of Personality and Social Psychology*, 91(6):1138–1151.

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.

John, O. P., Naumann, L. P., and Soto, C. J. (2008). *Paradigm Shift to the Integrative Big-Five Trait Taxonomy: History, Measurement, and Conceptual Issues*. Guilford Press, New York, NY.

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 Santos, A. (2012). A score based approach to wild bootstrap inference. *Journal of Econometric Methods*, 1(1):23–41.

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.

Moore, C., Detert, J. R., Trevino, L. K., Baker, V. L., and Mayer, D. M. (2012). Why employees do bad things: Moral disengagement and unethical organizational behavior. *Personnel Psychology*, 65(1):1–48.

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

Perry, J. L. (1996). Measuring public service motivation: An assessment of construct reliability and validity. *Journal of Public Administration Research and Theory*, 6(1):5–22.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics*, 14(4):1261–1295.

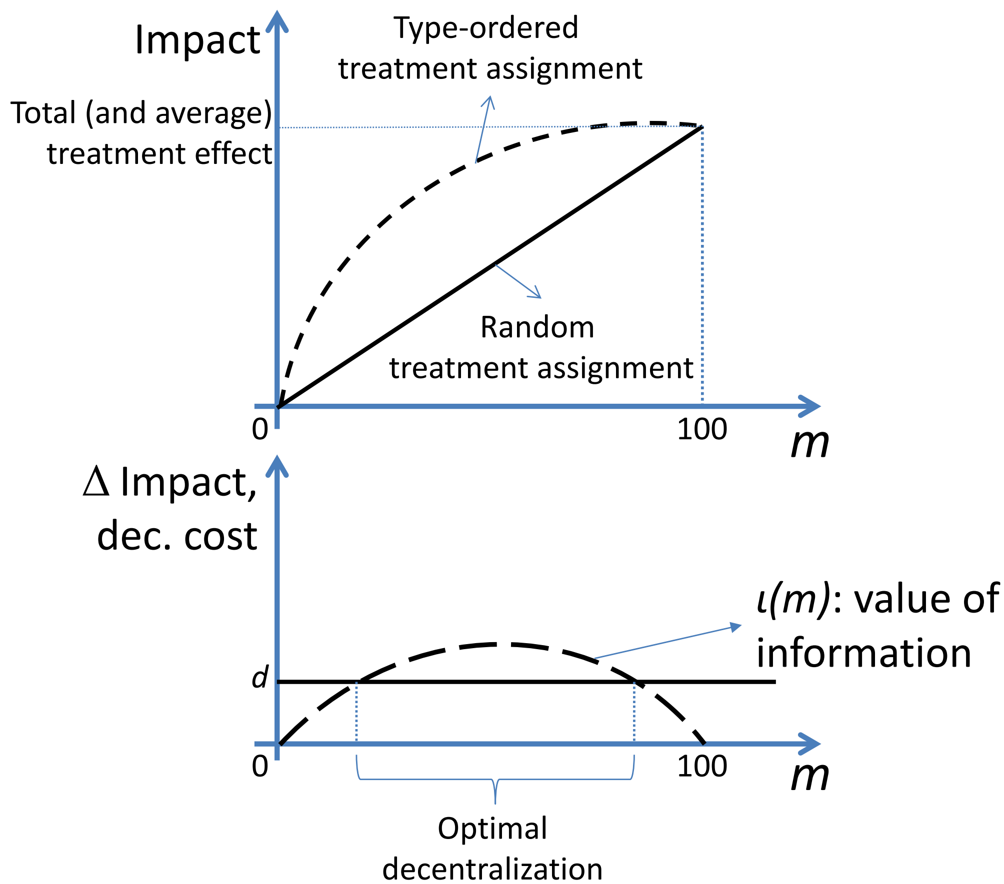
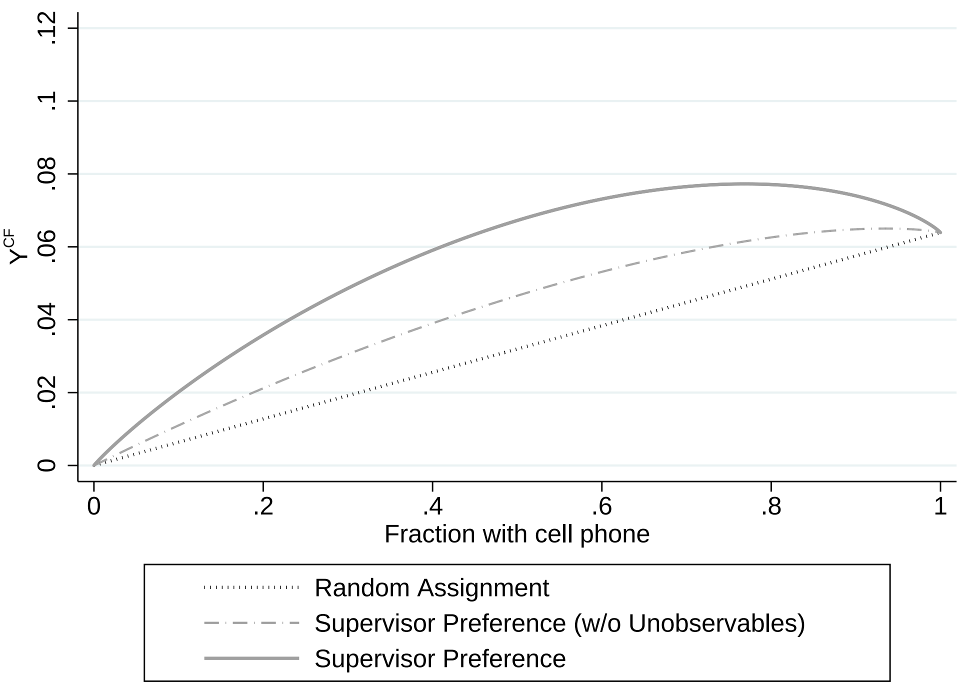


Figure 1: Treatment effects, roll-out extent, and the value of information

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

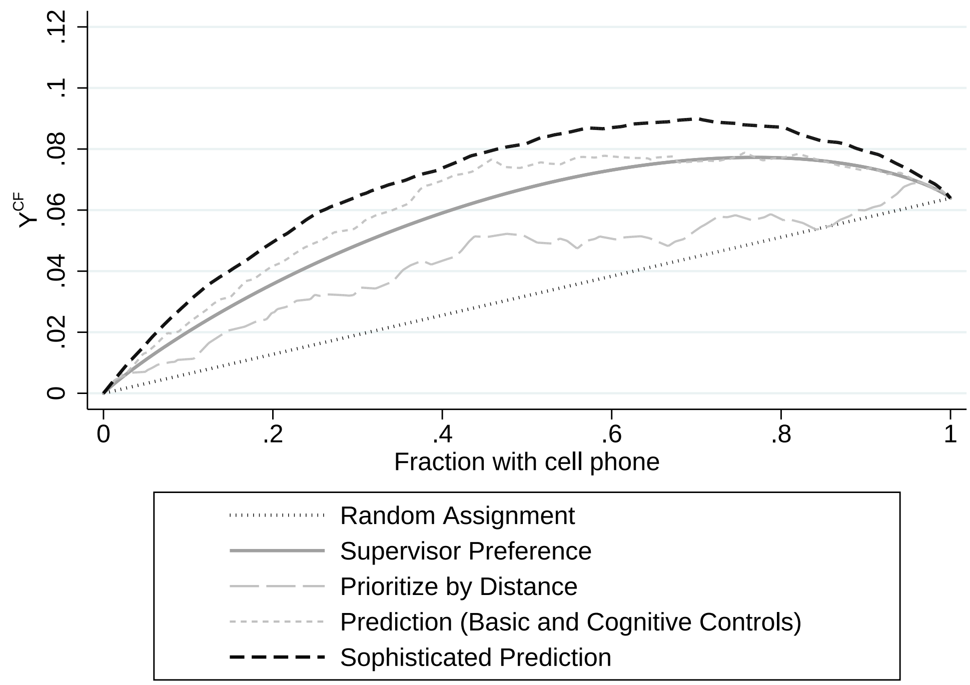
The columns correspond to ALATs, and the rows correspond to AEAs. For the ALATs in the column labeled Coverage 100% (cells B and D), every AEA received a cell phone independently of whether or not they had been selected by their supervisor. For the ALATs in the column labeled Coverage %50 (cells E and F), only the selected AEAs received cell phones (i.e. those in cell E). The control group contains cells A and C, where none of the AEAs received a cell phone.

Figure 2: Experimental Design



The *y*-axis shows the total treatment effect at different scales of roll-out under different assignment rules. With supervisor preference the treatment assignment is what would be achieved under decentralization if supervisors made the assignment decision based on all the information they had; supervisor preference without unobservables refers to the case in which supervisors made their decision based purely on the observable AEA characteristics; and under random assignment the treatment assignment is made randomly.

Figure 3: Supervisor versus Random Assignment



The *y*-axis shows the total treatment effect at different scales of roll-out under different assignment rules. Under random assignment the treatment assignment is made randomly; with supervisor preference the treatment assignment is what would be achieved under decentralization if supervisors made the assignment decision based on all the information they had; prioritize by distance is what would happen if treatment were assigned first to those AEAs whose beneficiaries live further from the local ALAT office; prediction (basic and cognitive controls) uses the control group to predict baseline performance using the observable variables and then treats first those AEAs who are predicted to be the worst performers in the baseline; sophisticated prediction runs a pilot experiment at low roll-out to establish a map between treatment response and observables and then treats first those AEAs who are predicted to have the highest treatment response.

Figure 4: Supervisor versus Alternative Allocation Rules

Table 1: Covariate Balance Across AEAs

	Small ALATs		Large ALATs	
	(1) Control	(2) Difference (T-C)	(3) Control	(4) Difference (T-C)
Male	0.611	0.139	0.750	-0.135
	[0.502]	{0.327}	[0.436]	{0.397}
Age	36.889	0.153	37.838	4.393
	[11.386]	{0.964}	[10.815]	{0.106}
Married	0.278	0.056	0.441	0.059
	[0.461]	{0.684}	[0.500]	{0.519}
Average Distance	10.425	3.156	12.678	-1.355
	[8.410]	{0.295}	[9.206]	{0.449}
Incumbent Party	0.611	0.014	0.559	-0.020
	[0.502]	{0.927}	[0.500]	{0.893}
Digit Span	5.333	0.333	5.191	0.270
	[0.840]	{0.255}	[1.069]	{0.340}
Big 5 — Stability	-0.083	0.345	-0.057	0.112
	[1.126]	{0.283}	[1.126]	{0.646}
Big 5 — Plasticity	-0.496	0.713**	-0.140	0.325
	[1.208]	{0.046}	[1.111]	{0.161}
Perry: Public Service Motivation Index	-0.493	0.465	-0.141	0.337
	[1.455]	{0.311}	[0.875]	{0.377}
Moore: Moral Disengagement Index	0.245	0.172	-0.019	-0.073
	[0.975]	{0.617}	[0.896]	{0.765}
Selected			0.603	-0.026
			[0.493]	{0.685}
Number of AEAs	18	24	68	26
Number of ALATs	17	23	22	11
p-value from Joint Test		0.812		0.560

"Control" and "Treatment" for small ALATs refer to ALATs that received cell phones in round 3 and rounds 1 and 2, respectively. The number of AEAs and ALATs in columns (1) and (3) correspond to the respective numbers in the control group. Those in columns (2) and (4) correspond to the respective numbers in the treatment group. The fraction of selected AEAs exceeds 50% because when ALATs had an odd number of AEAs, supervisors were told to round up. The joint test in the bottom row runs a regression of treatment assignment on all listed covariates. The joint test p-value is from a wild bootstrapped F-test imposing the null hypothesis that all coefficients equal zero. Standard deviations reported in square brackets and p-values from a Wu (1986) wild bootstrap procedure with 100,000 replication draws reported in curly braces. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, corresponding to the bootstrapped p-values.

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	0.069**	0.067**	0.067**	0.057*	-0.008	0.185*	-0.016
	(0.031)	(0.027)	(0.027)	(0.030)	(0.036)	(0.103)	(0.060)
	{0.024}	{0.017}	{0.021}	{0.089}	{0.831}	{0.065}	{0.817}
Selected	0.011	0.015	0.018	0.011		-0.080	-0.034
	(0.033)	(0.027)	(0.029)	(0.029)		(0.138)	(0.036)
	{0.754}	{0.595}	{0.592}	{0.744}		{0.567}	{0.377}
Male		0.040	0.040	0.032	-0.029	0.301*	-0.048
		(0.027)	(0.030)	(0.033)	(0.046)	(0.147)	(0.042)
		{0.210}	{0.273}	{0.466}	{0.551}	{0.053}	{0.288}
Age		0.003*	0.003*	0.003	-0.000	-0.008	0.004**
		(0.001)	(0.001)	(0.002)	(0.002)	(0.008)	(0.002)
		{0.060}	{0.081}	{0.128}	{0.936}	{0.368}	{0.029}
Married		-0.004	-0.003	0.020	-0.041	0.014	-0.053
		(0.030)	(0.030)	(0.031)	(0.034)	(0.141)	(0.043)
		{0.895}	{0.924}	{0.560}	{0.257}	{0.925}	{0.269}
Distance to Farmers (log)		0.038	0.036	0.032	-0.069**	-0.139	-0.018
		(0.027)	(0.028)	(0.034)	(0.028)	(0.119)	(0.047)
		{0.200}	{0.246}	{0.420}	{0.026}	{0.279}	{0.726}
Incumbent Party		-0.021	-0.020	-0.040	0.058	0.254*	0.001
		(0.025)	(0.025)	(0.025)	(0.057)	(0.151)	(0.052)
		{0.420}	{0.452}	{0.154}	{0.355}	{0.098}	{0.984}
Digit Span			0.010	0.013	0.030	0.051	0.013
			(0.012)	(0.013)	(0.018)	(0.052)	(0.018)
			{0.440}	{0.368}	{0.158}	{0.344}	{0.521}
Big 5 — Stability			0.012	0.011	0.035	0.141***	0.018
			(0.015)	(0.016)	(0.020)	(0.050)	(0.024)
			{0.513}	{0.592}	{0.104}	{0.006}	{0.523}
Big 5 — Plasticity			-0.011	-0.011	-0.023	0.111*	-0.017
			(0.011)	(0.010)	(0.017)	(0.064)	(0.017)
			{0.340}	{0.330}	{0.197}	{0.082}	{0.347}
Servicer	AEA	AEA	AEA	AEA	Supervisor	AEA	AEA
Mean of Control Dep. Var	.271	.271	.271	.274	.308	4.593	4.414
R^2	0.006	0.013	0.014	0.012	0.032	0.199	0.013
Number of Phone Surveys	1842	1842	1842	1584	1173		1819
Number of AEAs	136	136	136	94	107	126	132
Number of ALATs	71	71	71	33	107	65	71
Includes Small ALATs	✓	✓	✓		✓	✓	✓

Big 5 stability and plasticity measures normalized to have mean zero and unit variance. Outcomes in all columns other than (6) are from the farmer phone survey. These regressions include unreported indicators for small ALATs, survey wave, and an interaction of the two. Outcome in column (6) is from the AEA survey and is from a regression of a five-value index ranging from (1) “Strongly Disagree” to (5) “Strongly Agree.” Cluster robust standard errors and p-values from a Wu (1986) wild bootstrap procedure with 100,000 replication draws reported in parentheses and curly braces, respectively. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, corresponding to the bootstrapped p-values.

Table 3: Do Supervisors Have an Informational Advantage?

	Farmer was visited in the last week			
	(1)	(2)	(3)	(4)
Treated	0.0607*	-0.0233	-0.0397	-0.0361
	(0.034)	(0.043)	(0.035)	(0.038)
	{0.0723}	{0.608}	{0.278}	{0.389}
Treated × *Selected*		0.142**	0.161**	0.154**
		(0.058)	(0.051)	(0.050)
		{0.0349}	{0.0149}	{0.0276}
Selected	0.0113	-0.0332	-0.0445	-0.0409
	(0.033)	(0.036)	(0.027)	(0.028)
	{0.756}	{0.480}	{0.189}	{0.273}
R^2	0.004	0.009	0.017	0.018
Number of Phone Surveys	1584	1584	1584	1584
Number of AEAs	94	94	94	94
Number of ALATs	33	33	33	33
Includes Basic Controls			✓	✓
Includes Cognitive Controls				✓

Regressions also include survey wave indicators. Basic controls include gender, age, marital status, and average distance to farmers. Cognitive controls include digit span, the Big 5 stability meta-trait, and the Big 5 plasticity meta-trait. Cluster robust standard errors and p-values from a Wu (1986) wild bootstrap procedure with 100,000 replication draws reported in parentheses and curly braces, respectively. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, corresponding to the bootstrapped p-values.

Table 4: First Stage Probit Regressions

	(1)	(2)
Male	-0.519	-0.511
	(0.349)	(0.395)
	{0.165}	{0.257}
Age	-0.036**	-0.037**
	(0.018)	(0.017)
	{0.029}	{0.027}
Married	0.863**	0.791*
	(0.459)	(0.441)
	{0.045}	{0.064}
Average Distance	0.331	0.318
	(0.173)	(0.197)
	{0.121}	{0.115}
Incumbent Party	-0.808**	-0.840**
	(0.359)	(0.367)
	{0.044}	{0.043}
Digit Span		-0.111
		(0.142)
		{0.474}
Big 5 — Stability		-0.234*
		(0.124)
		{0.085}
Big 5 — Plasticity		0.122
		(0.154)
		{0.431}
Pseudo R^2	0.159	0.185
Number of AEAs	94	94
Number of ALATs	33	33

Coefficients are from a probit regression of an indicator for the AEA being selected on AEA characteristics. Cluster robust standard errors and p-values from a Kline and Santos (2012) wild bootstrap procedure with 100,000 replication draws reported in parentheses and curly braces, respectively. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, corresponding to the bootstrapped p-values.

Table 5: Treatment Effect Heterogeneity on Observable and Unobservable Characteristics

	(1)	(2)	(3)
Main Effects:			
Inverse Mills Ratio	-0.021	-0.019	-0.016
	(0.023)	(0.016)	(0.016)
Average Treatment Effect	0.062*	0.058*	0.064*
	(0.035)	(0.026)	(0.028)
	{0.093}	{0.067}	{0.064}
Interactions with Treatment:			
Inverse Mills	0.088**	0.064**	0.064**
	(0.036)	(0.022)	(0.025)
	{0.034}	{0.018}	{0.035}
Male		-0.021	-0.004
		(0.051)	(0.066)
		{0.721}	{0.961}
Age		-0.003	-0.005
		(0.003)	(0.003)
		{0.325}	{0.208}
Married		-0.084	-0.094
		(0.061)	(0.052)
		{0.241}	{0.135}
Average Distance to Farmers (log)		0.101	0.104
		(0.068)	(0.077)
		{0.252}	{0.362}
Incumbent Party		-0.081*	-0.069
		(0.038)	(0.048)
		{0.056}	{0.153}
Digit Span			-0.064**
			(0.027)
			{0.034}
Big5 — Stability			-0.038
			(0.049)
			{0.562}
Big5 — Plasticity			0.035
			(0.025)
			{0.317}
R^2	0.009	0.023	0.028
p-value for Observable Interactions		0.209	0.147
p-value for Observable Interactions (Not Wild)		0.031	0.001
Number of Phone Surveys	1584	1584	1584
Number of AEAs	94	94	94
Number of ALATs	33	33	33
Basic Controls		✓	✓
Cognitive Controls			✓

Left-hand side variable is whether the farmer was visted in the last week. 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. Regressions also include survey wave indicators. Main effects of basic and cognitive controls omitted for space. The p-value for all observable interactions reported in the bottom rows is the implied p-value from a wild bootstrapped F-test for coefficients on treatment interacted with age, married, average distance, male, digit span, and Big 5 measures. Cluster robust standard errors and p-values from a Wu (1986) wild bootstrap procedure with 100,000 replication draws reported in parentheses and curly braces, respectively. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, corresponding to the bootstrapped p-values.

Table 6: Treatment Effects by Roll-Out Levels and Allocation Rules

Rollout	Random	Supervisor	Distance	Prediction	Sophisticated
0.25	0.016	0.043	0.032	0.049	0.059
0.54	0.034	0.070	0.051	0.075	0.085
0.70	0.045	0.077	0.054	0.076	**0.090**
0.75	0.048	0.077	0.058	0.078	0.088
0.76	0.048	0.077	0.058	**0.079**	0.088
0.77	0.049	**0.077**	0.057	0.077	0.088
0.97	0.062	0.069	**0.069**	0.069	0.070
1.00	**0.064**	0.064	0.064	0.064	0.064

This table displays estimated treatment effects at the different roll-out levels shown in the first column for the various allocation rules we consider. The number displayed in bold represents the largest treatment effect under a given allocation rule. The allocation rules shown are random - treatment assignment is made randomly; supervisor - treatment assignment is what would be achieved under decentralization with supervisors making the assignment decision; distance - treatment is assigned first to those AEAs whose beneficiaries live further from the local ALAT office; prediction - uses the control group to predict baseline performance using the basic and cognitive controls and then treats first those AEAs who are predicted to be the worst performers in the baseline; and sophisticated - runs a pilot experiment at low roll-out to establish a map between treatment response and observables and then treats first those AEAs who are predicted to have the highest treatment response.

Table A1: Covariate Balance Across ALATs

	Small ALATs		Large ALATs	
	(1) Control	(2) Difference (T-C)	(3) Control	(4) Difference (T-C)
# of rural hhds	2298	320	3218	879
	[1978]	{0.633}	[1696]	{0.407}
Share of hhds that are rural	0.809	-0.094*	0.708	-0.039
	[0.104]	{0.082}	[0.217]	{0.597}
Average hhd size	4.71	0.01	4.70	0.14
	[0.34]	{0.950}	[0.35]	{0.302}
Land per farm (hectares)	41.35	7.63	43.40	0.68
	[43.49]	{0.711}	[42.48]	{0.973}
Cropland per farm (hectares)	12.58	-3.23	6.88	2.30
	[21.46]	{0.608}	[8.15]	{0.797}
Share of farmers working with DEAg AEAs	0.082	-0.020	0.096	0.014
	[0.109]	{0.562}	[0.072]	{0.878}
Corn yield (metric tons) per hectare	2.14	0.16	1.97	0.26
	[1.49]	{0.748}	[0.90]	{0.601}
Share of farms with running water	0.466	0.035	0.476	-0.051
	[0.247]	{0.673}	[0.210]	{0.490}
Colorado (winner) vote share	0.455	-0.007	0.472	-0.023
	[0.093]	{0.802}	[0.083]	{0.544}
Number of ALATs	17	21	22	11
p-value from Joint Test		0.299		0.178

"Control" and "Treatment" for small ALATs refer to ALATs that received cell phones in round 3 and rounds 1 and 2, respectively. The first three variables come from the 2002 census, the next five come from the 2008 agricultural census, and the final variable comes from the 2013 presidential elections. The number of AEAs and ALATs in columns (1) and (3) correspond to the respective numbers in the control group. Those in columns (2) and (4) correspond to the respective numbers in the treatment group. The joint test in the bottom row runs a regression of treatment assignment on all listed covariates. The joint test p-value is from a wild bootstrapped F-test imposing the null hypothesis that all coefficients equal zero. Standard deviations reported in square brackets and p-values from a Wu (1986) wild bootstrap procedure with 100,000 replication draws reported in curly braces. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, corresponding to the bootstrapped p-values.

Table A2: Average Effects of Receiving a Cell Phone on Other Measures of Performance

	(1) Satisfied	(2) Received Training	(3) Good for Nothing	(4) PCA
Treated	-0.085	0.056	-0.037	-0.145*
	(0.049)	(0.031)	(0.022)	(0.072)
	{0.119}	{0.117}	{0.134}	{0.079}
Selected	0.072	-0.014	0.017	0.071
	(0.052)	(0.023)	(0.021)	(0.065)
	{0.250}	{0.564}	{0.461}	{0.336}
Male	-0.062	-0.001	-0.012	-0.048
	(0.058)	(0.030)	(0.025)	(0.081)
	{0.352}	{0.987}	{0.669}	{0.601}
Age	-0.003	0.002	-0.003*	-0.007*
	(0.002)	(0.002)	(0.001)	(0.004)
	{0.203}	{0.173}	{0.064}	{0.094}
Married	0.063	-0.022	0.024	0.085
	(0.036)	(0.029)	(0.021)	(0.065)
	{0.115}	{0.484}	{0.286}	{0.228}
Distance to Farmers (log)	-0.004	-0.028	0.025	0.053
	(0.053)	(0.027)	(0.023)	(0.073)
	{0.952}	{0.365}	{0.323}	{0.510}
Incumbent Party	0.151**	-0.037	0.031	0.156**
	(0.052)	(0.024)	(0.025)	(0.067)
	{0.027}	{0.136}	{0.242}	{0.038}
Digit Span	-0.026	0.001	-0.006	-0.023
	(0.015)	(0.011)	(0.010)	(0.027)
	{0.135}	{0.900}	{0.588}	{0.443}
Big 5 — Stability	0.008	0.000	-0.010	-0.007
	(0.019)	(0.012)	(0.008)	(0.027)
	{0.687}	{0.980}	{0.216}	{0.811}
Big 5 — Plasticity	0.035	-0.014	0.012	0.049
	(0.024)	(0.012)	(0.012)	(0.034)
	{0.194}	{0.270}	{0.325}	{0.196}
Servicer	AEA	AEA	AEA	AEA
Mean of Control Dep. Var	1.458	.765	.183	.011
R^2	0.027	0.019	0.026	0.020
Number of Phone Surveys	1838	1841	1841	1838
Number of AEAs	135	136	136	135
Number of ALATs	71	71	71	71
Includes Small ALATs	✓	✓	✓	✓

The outcome measure in the fourth column is the first principle component from a polychoric PCA of the outcome variables in the first three columns. These regressions include unreported indicators for small ALATs, survey wave, and an interaction of the two. Cluster robust standard errors and p-values from a Wu (1986) wild bootstrap procedure with 100,000 replication draws reported in parentheses and curly braces, respectively. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, corresponding to the bootstrapped p-values.

Table A3: Correlation Matrix of Performance Measures

	(1)				
	Share visited	Satisfied	Received Training	Good for Nothing	PCA
Share Visited	1				
Satisfied	-0.311**	1			
Received Training	0.375***	-0.522***	1		
Good for Nothing	-0.372***	0.495***	-0.783***	1	
PCA	-0.415***	0.774***	-0.908***	0.885***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The sample includes only AEAs in the control group. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Predictors of Productivity in the Control Group

	(1)
Male	0.040
	(0.040)
	{0.585}
Age	0.002
	(0.002)
	{0.516}
Married	0.070
	(0.037)
	{0.142}
Average Distance to Farmers (log)	0.019
	(0.034)
	{0.626}
Incumbent Party	0.014
	(0.031)
	{0.657}
Digit Span	0.032**
	(0.013)
	{0.014}
Big5 — Stability	0.008
	(0.016)
	{0.710}
Big5 — Plasticity	-0.010
	(0.010)
	{0.356}
R^2	0.018
p-value for Model	0.156
p-value for Model (Not Wild)	0.000
Number of Phone Surveys	1091
Number of AEAs	68
Number of ALATs	22

The sample is all AEAs in the control group in large ALATs. Regressions also include survey wave indicators. The p-value for the model reports the implied p-value from a wild bootstrapped F-test for the null that all reported coefficients are equal to zero. Cluster robust standard errors and p-values from a Wu (1986) wild bootstrap procedure with 100,000 replication draws reported in parentheses and curly braces, respectively. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, corresponding to the bootstrapped p-values.