Intrinsic Incentives: A Field Experiment on Leveraging Intrinsic Motivation in Public Service Delivery

Scott S. Lee*

August 18, 2017

(Please click here for latest version.)

Abstract

Although extrinsic and intrinsic motivation likely jointly explain the effort of many agents engaged in public service delivery, incentives typically only appeal to the former. In the context of a rural health worker program in India, I develop and test a novel, mobile phone-based self-tracking app designed to increase agents’ intrinsic returns to effort. At ten months of follow-up, the self-tracking app leads to a 27% increase in performance as measured by the main job task (home visits). Moreover, the app is most effective when it leverages pre-existing intrinsic motivation: it produces a 46% increase in performance in the top tercile of intrinsically motivated workers, but no improvement in the bottom tercile. Supplementary evidence indicates that the treatment effect on performance is mediated primarily by making effort more intrinsically rewarding, and not by other mechanisms such as providing implicit extrinsic incentives.

*Harvard Business School and Harvard Medical School, ssl@mail.harvard.edu. I thank the Chief Medical Office of Kaushambi District, Dr. Nandini Sharma/Maulana Azad Medical College, and Dimagi, Inc. for hosting this research; Brian DeRenzi, Andrew Ellner, and Neal Lesh for ongoing collaboration as co-Investigators; and Sapana Gandhi, Sangya Kaphle, Robert Racadio, Sugandha Singh, and Jeremy Wacksman for excellent research assistance. I am grateful to Nava Ashraf, Oriana Bandiera, Iqbal Dhaliwal, Paul Farmer, Rema Hanna, Michael Kremer, Matthew Rabin, Andrew Weiss, and seminar participants at Harvard Business School, Harvard Department of Economics, and Harvard Medical School for helpful comments. Generous financial support has been provided by the Massachusetts General Hospital Consortium for Affordable Medical Technologies, Child Relief International, and the Harvard Business School Doctoral Program.
1 Introduction

Public services—governance, education, health care, national defense—rely on agents for their provision. In developing countries, the effort of these agents is often a binding constraint, prevailing over other factors such as the agents’ ability and market demand (Das & Hammer, 2007, 2014; Leonard et al., 2013; Maestad et al., 2010). The standard agency model offers both an explanation and a solution: agents dislike effort, and they can be persuaded with incentives to exert it. In line with this view, recent field experiments have shown that monetary and non-monetary incentives can improve the performance of agents engaged in public service delivery (Ashraf et al., 2014; Basinga et al., 2011; Duflo et al., 2012; Miller et al., 2012; Muralidharan & Sundararaman, 2011).1

What this perspective neglects, however, is the converse scenario: agents who exert effort despite having little extrinsic incentive to do so. In settings in which forty percent of health workers are absent from their posts on any given day (Chaudhury et al., 2006), what explains the presence of the remaining sixty percent? While other extrinsic factors (e.g., monitoring, social pressure, status-seeking) likely contribute, the persistence of effort in the face of weak incentives—and the decision to select into pro-social jobs in the first place—suggests a role for intrinsic motivation.2

That agents can be intrinsically motivated is not new in economics (Benabou & Tirole, 2003; Fehr & Schmidt, 1999). But this motivation is typically taken as given—a fixed trait that organizations may wish to select for but cannot influence (Besley & Ghatak, 2005). If it can be influenced, it is only for the worse, as proposed by theories of motivational crowd-out (Deci et al., 1999; Benabou & Tirole, 2006).3 In contrast, the business literature has long posited that managers can leverage intrinsic motivation by manipulating job attributes such as autonomy (Deci & Ryan, 1985), purpose (Weick, 1995), and organizational culture (Schein, 1985).4 But these attributes are

---

1For evidence challenging the effectiveness of financial incentives in public service delivery, see Banerjee et al. (2008), Glewwe et al. (2010), and, in the US context, Fryer (2013). For theoretical contributions on why financial incentives in public service delivery may fail, see Benabou & Tirole (2003, 2006).

2For empirical evidence that supports the hypothesis that those with pro-social preferences select into pro-social jobs, see Kolstad & Lindkvist (2013) and Lagarde & Blauuw (2013).

3A small but notable exception is theoretical and empirical work on the role of delegation and empowerment in enhancing intrinsic motivation. See, e.g., Aghion & Tirole (1997); Rasul & Rogger (2015).

4For more recent evidence from laboratory experiments, see Grant (2007) and Ariely et al. (2008).
generally conceptualized as static elements of job design, as opposed to incentives that interact with heterogeneous preferences.

To provide a richer view into the relationship between intrinsic incentives and intrinsic motivation, this paper explores the potential role of an intrinsic incentive technology in enhancing the effort of agents engaged in public service delivery. I define an intrinsic incentive as any variable in the principal’s choice set that modifies the agent’s marginal intrinsic utility of effort, analogous to how an extrinsic incentive modifies the marginal extrinsic utility of effort. In the setting of a rural health worker program in India, I develop a novel mobile phone technology—a “self-tracking” app—designed to act as an intrinsic incentive by delivering information that makes effort more intrinsically rewarding. The app comprises a set of graphs that a health worker can access on her phone to view her performance with respect to the job’s primary task: visiting pregnant women in their homes. As a counterfactual, I develop an analogous app—a “generic encouragement” app—designed to be lower-powered in that it provides generic messages of encouragement that are independent of the agent’s effort.

I test these incentives head-to-head by randomly assigning each of 145 health workers to receive one of the two apps on their work phone, and then tracking both app usage and performance on a daily basis for ten months. The experiment yields four main findings. First, both intrinsic incentive technologies are demanded. Across the two treatments, despite receiving minimal encouragement and no directive to do so, the average health worker accesses the software application approximately once every three days. Second, turning to effects on effort, compared to the generic encouragement app, the self-tracking app leads to a 27% increase in performance as measured by home visits. Third, the self-tracking app is most effective when it leverages pre-existing intrinsic motivation; it produces a 46% increase in performance in the top tercile of intrinsically motivated workers, but no improvement in the bottom tercile, indicating that, in this setting, intrinsic incentives and

---

5 The technology is intended to leverage intrinsic motivation both in the sense used in the psychological literature (motivation to do a task for its own sake—effort as its own reward) and in the sense of prosocial preferences (motivation to do a task due to its positive social externalities—impact as its own reward), both of which are distinguishable from extrinsic motivation in that no benefit external to the task and its output is needed to justify effort. Throughout this paper, unless otherwise stated, I use “intrinsic” motivation to encompass both intrinsic and prosocial preferences.

6 The conceptual distinction between the two lies in the slope with which they are expected to modify agents’ marginal intrinsic utility of effort, analogous to how a piece rate of $x$ dollars per piece is higher-powered than one that pays $\frac{x}{2}$ dollars per piece.
intrinsic motivation are complements. Fourth, despite these gains in client visits, no aggregate impact on the health of the pregnant clients and their children is observed. Finally, in terms of mechanisms, supplementary evidence suggests that the treatment effects of the self-tracking app are not mediated by effects on the production function or on extrinsic preferences. Rather, they appear to increase effort by making effort more intrinsically rewarding.

This paper contributes to two nascent literatures. First, recent empirical work has evaluated information incentives as a tool for motivating prosocial behavior. In particular, various forms of relative performance feedback have been found to be effective in improving home energy conservation (Allcott, 2011), student learning (Tran & Zeckhauser, 2012; Azmat & Iriberri, 2010), and physician quality (Kolstad, 2013). Information incentives, however, can function as extrinsic or intrinsic rewards, and when the targeted task (such as energy conservation or school performance) confers financial benefits, an information incentive that increases effort in the task is likely to be operating at least in part via an extrinsic channel. In contrast, the self-tracking app studied in this experiment is designed to increase only intrinsic returns to effort. In this regard, this intrinsic information incentive conceptually shares more in common with other intrinsic rewards, such as task meaning and organizational mission, than with other information incentives such as performance feedback.

The second literature examines the interaction between incentives and psychological traits such as intelligence and personality in prosocial behavior, in the context, for example, of civil servants (Dal Bo et al., 2013), health agents (Ashraf et al., 2014, 2015), and taxpayers (Dwenger et al., 2014). In the psychologically richest of the analyses, Callen et al. (2015) find that personality traits predict job performance among health officials in Pakistan, and that the experimental response to a novel monitoring technology varies with these personality traits. I extend this approach by (a) testing an intrinsic rather than extrinsic incentive, (b) testing a specific, theoretically guided

---

7 Kolstad (2013) finds that cardiac surgeons respond to physician report cards in ways that cannot be explained solely by profit maximization. He defines this reduced-form residual as an “intrinsic incentive” effect, but is silent about what in the physicians’ utility function drives this effect. As such, the results are consistent with physicians behind motivated not only intrinsically, but also by non-financial extrinsic preferences such as those for prestige, recognition, and career promotion.
interaction—that between intrinsic incentives and intrinsic motivation—and (c) elucidating psychological mechanisms.

The paper is organized as follows. Section 2 presents a simple principal-agent framework in which the principal can offer extrinsic and intrinsic incentives, and agents have extrinsic and intrinsic preferences. Section 3 describes the policy and program context. Sections 4 and 5 present the experimental design and results, respectively. Section 6 concludes.

2 Framework

In this section, I develop a simple framework for analyzing the potential role of intrinsic incentives in public service delivery. I take extrinsic incentives as given and instead focus on the principal’s ability to stimulate effort by increasing intrinsic—i.e., direct hedonic—returns to effort. The framework assumes that an agent may find an effort task tasteful or distasteful for intrinsic reasons. Modifying or augmenting some aspect of the task may make it more tasteful, thereby increasing its marginal intrinsic returns and, by extension, equilibrium effort. The possibility that the principal may be able to control this modification paves the way for intrinsic incentives. I show that, like an extrinsic incentive, an intrinsic incentive can be high- or low-powered, and the incentive interacts with the agent’s intrinsic motivation to determine the overall impact on marginal intrinsic utility. This interaction generates simple testable implications that I then take to the experimental data.

2.1 Principal’s choice

Consider a single-principal, single-agent framework. The principal (e.g., a government, nonprofit organization, hospital) is invested in the production of a social good $Y$ (e.g., population health). The agent can contribute to the production of $Y$ with effort $e_i \geq 0$. Assume that the production function $Y(e_i)$ is monotonically increasing in $e_i$.

To motivate the agent to exert effort, the principal may use extrinsic and intrinsic incentives. The former provides a reward contingent on output, whereas the latter enhances the hedonic returns to effort itself. Denote by $\alpha \geq 0$ and $\beta \geq 0$ the extrinsic incentive contract, where $\alpha$ is a fixed
reward and $\beta$ is paid linearly for each unit of $Y$ produced.\(^8\) Because $\alpha$ does not affect equilibrium effort, for notational convenience, assume that $\alpha = 0$.

Denote by $\psi(\cdot)$ an intrinsic incentive function. $\psi(\cdot)$ will enter into the agent’s utility function as described in Section 2.2 below. Here, I comment on its conceptual underpinnings. The $\psi(\cdot)$ function is left purposefully vague, because it is likely to have many parameters. Definitionally, $\psi(\cdot)$ is a function that alters the marginal intrinsic utility of effort. It is analogous to $\beta$, but whereas $\beta$ is non-negative, the value of $\psi(\cdot)$ is unbounded—i.e., it can increase or decrease marginal intrinsic returns. Theories of motivational crowd-out (Deci et al., 1999), for example, imply $\frac{\partial \psi}{\partial \beta} < 0$—i.e., an increase in extrinsic returns to effort reduces its intrinsic returns.\(^9\)

Potentially any job attribute could affect marginal intrinsic utility of effort: the nature of the effort task, the technology of production, the degree of monitoring vs. autonomy, organizational norms and culture, and so forth. For the current purposes, let the principal’s choice variable in the $\psi(\cdot)$ function be the psychological salience and observability to the agent of the effort task and its social impact. The principal alters the agent’s information environment to achieve this effect. Much more is described about this information technology in Section 4.1. The basic concept is that the principal can provide a technology by which the agent is better able to self-observe effort and output and, in so doing, experience greater marginal utility (or disutility) of effort.

The principal chooses intrinsic incentive regime $j \in \{h, l\}$, corresponding to “high-powered” and “low-powered,” respectively. The high-powered (low-powered) regime enables the agent to access information that makes the effort task and its social impact highly (minimally) salient. Assume that $j$ maps onto $\psi(j)$ such that $\psi_h = 1$ and $\psi_l = 0$. Assume also that providing $j$ carries zero marginal cost, in terms of both the direct supply cost and the shadow cost of reputation.\(^{10}\) Thus,

---

\(^8\)The conventional interpretation for $\beta$ would be a variable wage, but the conceptual intuition can be extended to other extrinsic goods such as social status and job security.

\(^9\)I abstract away from this relationship in this framework because it is not empirically relevant in this experiment; as I describe in Section 4, there is no variation in $\beta$ in the sample and hence no way to identify $\psi(\beta)$.

\(^{10}\)Benabou & Tirole (2003) describe this shadow cost as follows: “A teacher or a manager who makes very complimentary comments to every pupil or employee may lose her credibility....[W]hen disclosing soft information to several agents the principal must realize that they will see through her ulterior motivation, and believe her only if she builds a reputation for not exaggerating claims.” Why there is no shadow cost of $j$ in the experimental intervention is explained in more detail in Section 4.1, but in brief, because $j$ only contains objective information, it does not rely on scarcity for its value, whereas recognition, praise, or positive feedback does.
the only cost incurred by the principal is $\beta Y$, but, as is made explicit below, $Y$ is affected by $\psi(j)$ via the latter term’s effect on optimal effort.

Let the principal’s utility be $V(Y - \beta Y)$, where, given risk neutrality, $V' > 0$ and $V'' = 0$. The principal chooses $\{e, \beta, j\}$ to maximize her expected utility, subject to the agents’ individual rationality and incentive compatibility constraints.

### 2.2 Agent’s choice

I now turn to the agent’s choice of $e$. A risk-neutral agent $i$ with preferences $(\theta^E_i, \theta^I_i)$ chooses effort $e_i$ to maximize his expected utility:

$$U_{ij} = \theta^E_i [\beta Y(e_i)] + \theta^I_i [\gamma_i + (1 + \psi_j)e_i] - \frac{e_i^2}{2}$$

Equation 1 has three terms: an extrinsic payoff term, an intrinsic payoff term, and an effort cost function. The two payoff functions are weighted by extrinsic ($\theta^E_i$) and intrinsic ($\theta^I_i$) preference parameters, respectively. Assume $\theta^E_i$ and $\theta^I_i$ are independent and that $\theta^E_i$ is distributed over $[0, 1]$. In contrast, assume $\theta^I_i$ is distributed over $[-1, 1]$. Similar to the discussion of $\psi(\cdot)$ above, whereas preferences for extrinsic rewards can logically only be weakly positive, intrinsic preferences can be positive or negative since effort can be intrinsically utility-enhancing or -undermining. The convex effort cost function, $\frac{e_i^2}{2}$, captures disutility of effort in a baseline state absent intrinsic preferences, around which the intrinsic payoff term allows for heterogeneity.\(^\text{11}\) That is, $\theta^I_i$ is meant to account for heterogeneity in intrinsic preferences regarding task-specific effort, which can increase or decrease the cost of effort relative to the baseline of no intrinsic preferences.

To return to the utility function, the first term—the extrinsic payoff function—is simply the principal’s linear payment contract weighted by the agent’s preference for extrinsic rewards, $\theta^E_i$.

As with the payment contract, the intrinsic payoff term has two components—one that is effort-independent and another that is effort-dependent. The effort-independent parameter, $\gamma_i$, can be thought of as an endowment of intrinsic (dis)utility—e.g., the warm glow experienced from having

\(^{11}\)For example, an agent with a fear of blood would find being a surgeon particularly distasteful, and, all else equal, surgical effort would be more costly for this person than for someone without this phobia. Likewise, an individual with anti-abortion political views may have an intrinsic aversion to working in an abortion clinic, and so forth.
a prosocial job, or the effort-independent disutility of having a job that runs counter to one’s tastes. Since this endowment affects equilibrium utility but not effort choice, for simplicity, assume $\gamma_i = 0$.

The second component of the intrinsic payoff term, $[(1 + \psi_j)e_i]$, captures the practical intuition that, in some circumstances (e.g., volunteering), even when there is no extrinsic payoff ($\alpha = \beta = 0$), agents are still willing to supply $e_i^* > 0$, implying that, over the interval $[0, e_i^*)$, marginal intrinsic benefit of effort exceeds marginal effort cost. In the absence of any intrinsic incentive (i.e., $\psi = 0$), the agent derives marginal intrinsic utility $\theta^I_i$ per unit effort. With the high-powered intrinsic incentive ($j = h$), the marginal intrinsic utility increases in absolute magnitude to $2\theta^I_i$. In contrast, marginal intrinsic utility is negative if $\theta^I_i < 0$ and 0 if $\theta^I_i = 0$.

The agent maximizes Equation 1 with respect to $e_i$, such that:

$$e^*_{ij} = \theta^E_i \beta + \theta^I_i (1 + \psi_j)$$

(2)

This simple first-order condition illustrates the complementarity between intrinsic incentives and intrinsic motivation that this experiment tests. It implies that intrinsic incentives will lead to higher effort for intrinsically motivated agents, but may have no effect or even reduce effort for those who are intrinsically unmotivated. Thus, the aggregate effect of an intrinsic incentive depends on the distribution of intrinsic motivation in the agent population.

3 Context

3.1 Program context

In 2005, the Government of India launched the Accredited Social Health Activist (ASHA) program, a nationwide effort to improve health services at the community level, especially in rural regions.\textsuperscript{12} ASHAs are female community health workers who are selected by local village councils to provide for the health needs of the villages in which they reside.\textsuperscript{13} Each ASHA is assigned a discrete and

\textsuperscript{12}As of 2014, 828,000 ASHAs had been recruited across India\textsuperscript{14}Government of India (2015).

\textsuperscript{13}Job qualifications include: female gender; married, widowed, or divorced status (due to patrilocality); grade 8 education or higher; age between 25 to 45 years; and preferably, literacy.
non-overlapping catchment area, which typically corresponds to a village or group of villages with a population of roughly 1,000 people.

ASHAs’ primary responsibilities revolve around supporting and counseling women during pregnancy, childbirth, and the postpartum period. The typical sequence of serving a pregnant client proceeds as follows. When the ASHA learns of a new pregnancy in her village, she visits the woman and offers to support the woman. If the woman accepts, the ASHA registers the client using a phone-based record-keeping tool.

Over the course of the pregnancy, ASHAs are expected to visit the client at home at least monthly. During these visits, ASHAs carry out a variety of tasks: counseling on nutrition, physical activity, and other day-to-day aspects of pregnancy; counseling on identifying pregnancy-related danger signs requiring urgent medical attention; encouraging the client to obtain facility-based antenatal care; providing iron and folic acid supplements; working with the client to develop a birth plan, which includes calculating the estimated delivery date, advising the client on local health facilities for delivery, identifying means of transport, and engaging family support; and updating the client’s maternal health card. The ASHA records and submits these follow-up visits also using her phone-based tool.

At the time of labor, the ASHA accompanies the client to a health center or hospital and remains with her throughout labor and delivery; the ASHA’s payment, as discussed below, is contingent upon this presence. After the mother is discharged, the ASHA visits the mother and the child several times over the ensuing six weeks to monitor their health and counsel on newborn care, breastfeeding, family planning, and immunizations. At six weeks postpartum, the ASHA “discharges” the client from care.

14 In 2006, an expert group convened by The Lancet identified one “overwhelming priority strategy” for reducing maternal deaths: “promoting delivery in primary-level institutions (health centers), backed up by access to referral-level facilities,” as opposed to home delivery (Campbell & Graham, 2006).

15 The mobile phone tool is not a feature of the national ASHA program but, rather, the program site in which this experiment takes place. See Section 3.2.
The ASHA job is typically not a full-time position, and ASHAs do not receive salaries.\textsuperscript{16} Instead, they are paid piece rates for discrete activities such as facilitating institutional delivery, assisting with polio immunization campaigns, and mobilizing men and women to undergo sterilization.\textsuperscript{17} ASHAs’ chief source of income arises from a federally sponsored conditional cash transfer scheme designed to encourage institutional delivery, called Janani Suraksha Yojana (JSY). In this scheme, pregnant women are paid INR 1,400 (USD 23 at 2015 exchange rates) for delivering in an accredited public or private health facility. In addition, an ASHA who accompanies the woman to the hospital for delivery is paid INR 600 (USD 10). ASHAs are not paid for visits to the client during the antenatal period. Thus, in the absence of intrinsic motivation, antenatal visits are rational only inasmuch as they increase the probability of institutional delivery.

3.2 Program site

With a population of 1.6 million, Kaushambi District is one of 19 (out of 70 total) districts in Uttar Pradesh designated by the state government as “high-focus” in view of its poor development indicators. Its maternal mortality ratio of 442 deaths per 100,000 live births is nearly twice the national average and 30\% higher than the state average (Government of India, 2011a); the district’s neonatal mortality rate is twice the national average; and the district has the second-highest scheduled caste population share in the state (Government of India, 2011a).\textsuperscript{18} This experiment takes place in Muratganj, one of eight sub-districts in Kaushambi, with a population of 193,355 (Government of India, 2011b). At the time of the experiment launch in 2014, Muratganj had 145 ASHAs, all of whom had been recruited and trained in 2006-2007 when the ASHA program was rolled out in the District.

\textsuperscript{16}In the study population in the pre-experimental period, the average ASHA earned USD 372 in total annual ASHA payments. For context, auxiliary nurse-midwives (ANMs), the supervisory cadre directly above ASHAs and the actual providers of facility-based antenatal care services, earn approximately USD 2,000 per year. Anganwadi workers (AWWs)—child health and nutrition workers who are positioned laterally to ASHAs and have similar job qualifications but work full-time—earn approximately USD 880 per year.

\textsuperscript{17}For example, if an infant registered by the ASHA completes his or her complete course of routine immunizations, the ASHA receives a payment of INR 150 (USD 2.50).

\textsuperscript{18}Scheduled castes are castes designated in the Constitution of India as historically disadvantaged. Nationwide, 17\% of the Indian population belongs to a scheduled caste; in Kaushambi, the proportion is 36\%.
In 2012, a nongovernmental organization established a partnership with the Kaushambi district health office to equip the ASHAs with mobile phones to facilitate their work. The phones contain a software application called CommCare, through which the ASHAs register clients and document home visits as described above. These records, which are synchronized with a cloud-based server, provide the data on which the self-tracking app functions. By the time of the experiment launch, the ASHAs had been using CommCare for 15 months. As all of the ASHAs in the experiment were trained to use CommCare, this experiment is not designed to evaluate the underlying work tool.

Table 1, Panel A shows descriptive statistics for the performance of the 72 ASHAs in the control (generic encouragement) treatment over the pre- and post-intervention periods. In contrast to the expectation that ASHAs should visit all clients on a monthly basis, the average control ASHA visits 46% of her 13.1 pregnant clients in the average month. Panel B shows descriptive statistics for a limited set of client characteristics and health outcomes that ASHAs report. Of note, the 77% institutional delivery rate is substantially higher than the 22% rate for rural Uttar Pradesh reported in the most recent (2005-2006) wave of the National Family Health Survey (International Institute for Population Sciences & Macro International, 2007); though possibly overstated due to selection (both in registration of pregnant women and in reporting of outcomes), this statistic is consistent with a broad increase in institutional delivery that has been observed across India since the introduction of the JSY conditional cash transfer scheme in 2005 (Lim et al., 2010).

4 Experimental design

4.1 Experimental interventions

I create a novel mobile phone-based “self-tracking app” designed to enhance the intrinsic utility that ASHAs derive from providing care to pregnant women. In this paper, I mean “intrinsic” both in the sense used in the psychological literature (motivation to do a task for its own sake—effort as its own reward) and in the sense of prosocial preferences (motivation to do a task for its positive social externalities). Both types of motivation stand in contrast to extrinsic motivation,

---

19 The application is developed by a US-based company called Dimagi, Inc.
20 Credit for the technological development of the app is due to Brian DeRenzi.
which may be rooted in individualistic (income, job security, career advancement) or social (status, reputation, recognition) concerns, but which in either case regards effort as palatable only for the extrinsic rewards it earns. A key challenge in assessing the mechanism of an intrinsic incentive is that it may also operate via these extrinsic channels. This is a partly a matter of design and partly an empirical question; I address both in the analysis of effects and mechanisms in Section 5.

Another challenge is that an incentive may function as a production technology; i.e., it may increase output not only by increasing effort but also by increasing the productivity of effort, \(Y(e)\). For example, an app that increases the frequency with which an ASHA interacts with her work phone may make her more adept at using the phone, which in turn may increase her productivity.\(^{21}\) To mitigate this confound, I create an analogous app that mimics the user interface of the self-tracking app, but replaces its content with information that is, \textit{a priori}, expected to be lower-powered as an intrinsic incentive. Because the two apps’ interfaces are identical, treatment effects on take-up and performance can be attributed to the information content of the apps as opposed to putative motivational or learning effects of the technology with which the information is delivered. There remains the possibility that the information content could alter the production function; I address this in Section 5.

4.1.1 High-powered intrinsic incentive: Self-tracking

The self-tracking app enables ASHAs to access data visualizations of their work performance.\(^{22}\) The data are compiled from the ASHAs’ own submissions via their phone-based reporting tool. With one exception (the relative performance graph described below), all of the information contained in the visualizations is generated by the ASHA herself, which highlights the notion of “self-tracking,” as opposed to performance feedback, in which the agent is provided with information that is not observable to her.

\(^{21}\)Such an app would need not deliver performance-related information—e.g., a mobile phone game that confers learning benefits (e.g., how to use the phone) that could increase job-specific ability (e.g., how to fill out forms on the phone).

\(^{22}\)The app is similar in concept to recently developed consumer-oriented mobile apps and wearable technologies that enable users to track and visualize data related to personal activities—e.g., Fitbit, Runkeeper, Apple Watch.
The app has four screens: a menu, a relative performance page, a calendar page, and historical trends page.

Figure 1: Self-tracking app: menu page (Hindi sample and English mock-up).

Menu (Figure 1). The menu page serves as the entry point for accessing the tool. It is accessed by pressing a predefined key on the phone. The menu displays a date interval spanning the first day of the current month to the current day. The calendar month is the primary performance interval for ASHAs; the official ASHA program guideline is that ASHAs should visit all of their pregnant clients at least once per month. Below the date interval are three rows that link to the three other pages, which the ASHA can select using the phone’s navigation buttons.

Figure 2: Self-tracking app: relative performance page.

Relative performance (Figure 2). This page features an ordered bar graph of the number of unique pregnant clients visited in the current month, for the ASHA and 15 other ASHAs. The peer ASHAs are chosen randomly from among the 73 ASHAs who belong to the self-tracking treatment condition. The ASHAs are not told the identities of the peers. Anonymity is important for both ethical and theoretical reasons—ethically, to avoid harmful repercussions that may result from
publicizing individual ASHAs’ performance, and theoretically, to unlink the informational effect of social comparison from external peer effects, such as these very repercussions. Each ASHA’s peer group is randomly chosen, such that all peer groups are asymptotically identical but individually unique.\textsuperscript{23} In order to avoid serial correlation effects, all peer groups are redrawn at the beginning of each month. The graph updates in response to changes in the ego ASHA’s performance, as well as the performance of peer ASHAs.

The relative performance graph is designed to provide an anonymous social benchmark that makes the information signal about performance more useful.\textsuperscript{24} Conceptually, the relationship between social comparison and intrinsic and extrinsic motivation is nuanced. On one hand, publicly identifiable social comparison would be expected to interact with externally-oriented social preferences such as tastes for status and recognition. Private social comparison, however, does not engage these external preferences; status, for example, cannot be conferred on an anonymous entity.\textsuperscript{25} On the other hand, private social comparison can interact with internally-oriented social preferences, such as competitiveness and taste for winning, or preferences that may not involve social interaction but may be mediated by social information, such as social norms regarding private behavior. If these preferences are in relation to an abstract reference group and do not implicate social interaction, they more closely align with standard definitions of intrinsic rather than extrinsic motivation.\textsuperscript{26}

\textsuperscript{23}In consequence, each peer group has, in expectation, the same average performance, but any given peer group may be higher- or lower-performing than average (and this may fluctuate on a day-to-day basis).

\textsuperscript{24}In a different study population (such as students taking an exam), analogous information could be conveyed by disclosing only the group mean, which is also anonymous. In the ASHA population, piloting exercises revealed that the concept of “average” was difficult to convey to many ASHAs, and thus the visual display format was adopted.

\textsuperscript{25}In the absence of peers, motivation around managing one’s self-image could affect effort—e.g., an ASHA who is motivated to do ASHA work because she self-identifies as a prosocial type. However, whereas social image motivation (in which an agent has preferences over how others attribute her behavior) implicates social preferences, self-image motivation can exist in the absence of social preferences, and thus I classify this as a type of intrinsic, not extrinsic, motivation, consistent with Benabou & Tirole (2006); Bénabou & Tirole (2011).

\textsuperscript{26}To illustrate, consider a competitive athlete. While preferences for status and recognition may play a role in driving the athlete’s effort, there may also be a private component—a desire to excel—that is intrinsic. The athlete might set a performance goal that is absolute (running a mile in less than five minutes), relative to her own performance (setting a personal record), or relative to others’ performance (setting a world record). In each case, good performance can be an end in itself, not an instrument for attaining other rewards. Relative social benchmarks, a type of social norm, have been shown to exert substantial influence on behavior even when the behavior is private and does not entail social interaction, such as home energy consumption (Allcott, 2011; Schultz et al., 2007).
Calendar (Figure 3). The remaining two pages are more straightforward. The calendar page displays a calendar in the standard seven-day format, which is the local convention. On each day, if any visits to pregnant clients occur, the number of visits is indicated in a circled number. A counting rule restricts the number of times an ASHA can get “credit” for visiting a given client in a given day to one, but there is no restriction on the number of times an ASHA can report visiting a client in a given month.

Historical (Figure 4). This page displays a line graph plotting the total number of clients visited each month for the current month (to date) and the five preceding months, in a rolling manner.

Taken together, the self-tracking app provides the ASHA with the ability to access personalized information about her performance. The data elements conveyed include: total caseload, number of clients visited in the current month, number of clients visited each day in the current month, number

---

27 Sundays and national holidays are shaded red; the current day is shaded black.
of visits conducted in the current month, number of clients visited monthly over the previous six months, and visit performance relative to an abstract peer group. All of these elements are meant to make the effort task and its prosocial impact more psychologically salient to the ASHA, thereby, in theory, amplifying the marginal intrinsic utility of effort.

4.1.2 Low-powered intrinsic incentive: Generic encouragement

The theoretical framework states that intrinsic incentive regimes \( h \) and \( l \) produce \( \psi_h = 1 \) and \( \psi_l = 0 \), respectively. In view of the latter, a counterfactual to the self-tracking app could be to provide no experimental intervention to the control group. However, this would introduce a mechanism confound; any treatment effect could be due to the interactive nature of the app, rather than its information content. The intent of the experiment is to focus on the incentive effect of information itself. To this end, I develop a counterfactual intervention that preserves the technological interface of the self-tracking app, but provides generic information that is putatively less effective at making the effort task and its prosocial impact salient.

The low-powered app also has four pages; sample pages are shown in Figure 5. The content includes generic encouragement messages. A different set of pages is generated on a daily basis, out of an inventory of 52 sets. The menu page points to three further pages: “Responsibilities of an ASHA,” “Advice for Healthy Mothers & Babies,” and “Inspiring Quotes.” Each page includes a statement accompanied by a picture that illustrates the statement. Content for the first two sections is drawn from ASHA training materials. The “Inspiring Quotes” section contains quotes drawn from Hindi-language websites, with many attributed to well-known South Asian cultural figures such as Gandhi and Mother Theresa.

4.1.3 Audio service

In the study sample, 28% of the ASHAs are illiterate.\(^{28}\) To ensure that the interventions would be useful to these ASHAs, the research team developed an automated audio version of each interven-

\(^{28}\)Each ASHA was asked to read a Hindi sentence which stated, “The woman went to the market to buy vegetables.” Twenty-eight percent were able to read no words or only a few words; the remainder were able to read all of the words and are classified as literate.
tion. Analogous to the act of accessing the app on the phone, an ASHA calls a designated phone number from her work phone. An audio recording specific to the phone number is played back to the caller. There is no limit on how often the ASHA can utilize the audio service.

While there are differences in the user experience of the audio and the visual software interface (e.g., navigation is possible only in the latter), the underlying information content is the same. Thus, the theoretical framework is unaffected, and in the empirical analysis, unless otherwise stated, I combine usage of the visual and audio systems to yield a composite measure of take-up.

4.2 Randomization, implementation, and data sources

The 145 ASHAs in Muratganj were randomly assigned to one of the two intrinsic incentive treatments: self-tracking and generic encouragement. Randomization was conducted in May 2014, one month before the launch of the experiment, and was stratified by six variables: Hindi literacy, total client visits conducted over the prior 4 and 12 months, respectively, and scores on three psychometric scales for extrinsic, intrinsic, and prosocial motivation, respectively.\textsuperscript{29,30} In June 2014, the self-tracking and generic encouragement apps were installed on ASHAs’ phones, and all 145 ASHAs were trained in their use. All ASHAs were told that two different phone-based tools were being piloted, and that the eventual plan was to make both available if so desired by ASHAs. No complaints were raised to the research team or the implementing NGO during the training or at any point thereafter regarding the randomization.

Once the experiment launched, care was taken to preserve the intent of the intervention: to alter the intrinsic utility of effort without altering real or perceived extrinsic returns to effort. To avoid a potential monitoring effect, no efforts were made to affect demand through, e.g., routine follow-up visits, marketing, or interactions during other CommCare activities such as trainings.\textsuperscript{31}

\textsuperscript{29}Literacy was directly measured during the baseline survey by asking the ASHAs to read a Hindi sentence. The two visit measures were included to capture both short- and medium-term baseline work performance. The psychometric scales are described in detail in Section A.

\textsuperscript{30}I used the “T-min-max” method with 1,000 draws to carry out the randomization. For a discussion of this randomization method, see Bruhn & Mckenzie (2009).

\textsuperscript{31}The one partial exception was an automated SMS system created by the research team, in which a text message is sent to each ASHA every Monday stating either (for self-tracking app users), “Your visits information is available. Please press the shortcut button to access your information,” or (for encouragement app users), “Your advice and encouragement is available. Please press the shortcut button to access your encouragement.” ASHAs were told that
Instead, the research team interacted with ASHAs only when an ASHA called a research assistant to request assistance with the self-tracking or encouragement app. Such assistance variously involved re-installing the software in the case of accidental deletion, re-saving the phone number used to access the audio service, and addressing questions regarding interpreting the information in the app.

For the first eight months of the experiment, the research team had no contact with ASHAs other than through these troubleshooting visits. At nine months of follow-up, a midline survey was administered, in which 142 out of 145 ASHAs participated. In the course of each ASHA’s interview, the research team rechecked all phone settings related to the use of the visual and audio services, fixed settings as necessary, and documented any steps taken.

In addition to the baseline and midline survey data, this experiment relies on performance data reported by the ASHAs through their phone-based record-keeping tool, and app usage data measured directly. Client visits and app sessions are timestamped with start and end times. Client visits are tied to individual clients, allowing differentiation of initial registration visits and follow-up visits. These and other data sources are described in further detail in Appendix Section A.

5 Analysis

5.1 Randomization balance

Table 2 reports tests of equality for variables measured at baseline across five domains: job performance, job-specific ability, psychological traits, ASHA demographic characteristics, and village characteristics. Of the 18 variables tested, one (whether ASHA reports that her work village is predominantly Muslim) is significant at the 10% level, as would be expected by chance. Jointly, the variables are not significant ($p = 0.82$). In addition, I run $t$-tests for differences in means between the two treatment conditions for all 263 numeric variables in the baseline survey. Of these, 4.9% of the differences are significant at the 10% level; 1.9% at the 5% level; and 0.4% at the 1% level. Repeating this exercise for daily home visits for all 455 days in the pre-experimental period, these messages were sent automatically to all ASHAs. SMS use is low overall in this setting, and we find no significant Monday fixed effect for app usage.
t-tests reveal that 6.9%, 2.5%, and 0.5% of daily visit counts are significantly different at the 10%, 5%, and 1% levels, respectively. Taken together, these results indicate failure to reject the null hypothesis that the two treatment groups are identical on observables. In the analysis that follows, to the extent possible, baseline differences are controlled for, either with explicit covariates or with fixed-effect estimators that implicitly control for all time-invariant ASHA characteristics.

5.2 Empirical strategy and average treatment effects

To identify causal effects, I exploit the fact that performance are observed at the daily (or, in the case of earnings, monthly) level. This allows for panel analysis with ASHA fixed effects that control for all time-invariant ASHA characteristics. In addition, the availability of pre-experimental data allows for differencing out time-variant, ASHA-specific trends. I estimate the general equation:

\[ y_{it} = \beta_0 + \beta_1 A_t + \beta_2 A_t * T_i + \alpha_i + Z_m \gamma_m + u_{it} \]

where \( y_{it} \) is an outcome of interest for ASHA \( i \) at time \( t \); \( A_t \) is an indicator for whether time \( t \) is after the launch of the experiment; \( T_i \) is an indicator for treatment assignment that takes value 0 in the generic encouragement condition and 1 in the self-tracking condition; \( \alpha_i \) is a vector of ASHA fixed effects; \( Z_m \) is a vector of month-year fixed effects; and \( u_{it} \) is the error term. I assume that errors are serially correlated and thus present standard errors clustered at the ASHA level throughout.

Randomization ensures, in expectation, that \( T_i \perp u_{it} \), and that trends in \( Y_{it} \) between the two treatment groups, in the absence of treatment, would be parallel. Furthermore, spillovers between treatments is unlikely given that ASHAs live in different villages and cover defined, non-overlapping catchment areas and thus are not likely to experience spillovers via either market demand or through direct interactions. Finally, there is virtually no attrition during the follow-up period of the experiment. Of the 145 ASHAs, all participate in app training, and at ten months of follow-up, only 3 ASHAs (one in the generic encouragement condition and two in the self-tracking condition) are no longer working (all three due to outmigration). Under these identifying assumptions, \( \beta_2 \)
measures the average causal effect of the self-tracking app relative to the generic encouragement app.

### 5.2.1 Average treatment effects on take-up and client visits

Table 3 presents average treatment effects on take-up of the interventions, on client visits, and on earnings. Columns 1-3 show results for take-up of the main visual software, the complementary audio service, and total use of the two modalities combined. All estimates are in units of sessions per day. Three findings are noteworthy. First, despite an experimental protocol that provides little explicit encouragement to use the interventions, demand for both apps is high. The average ASHA in the self-tracking (generic encouragement) treatment uses her app once every 3.75 (3.11) days over the course of the experimental period. To contextualize this, client visits occur every 3.83 and 4.85 days in the self-tracking and generic encouragement conditions, respectively, during the experimental period. Thus, in both conditions, app usage is higher than visit frequency, suggesting that use of the phone for visits is not exclusively driving use of the apps. Second, that take-up of the two apps is similar (with the point estimate favoring the encouragement app) suggests that any treatment effects favoring the self-tracking app cannot be explained by differences in use of the phone (e.g., a learning-by-doing effect that makes frequent phone users more efficient at filling out forms).

The remaining columns of Table 3 examine two measures of performance: visits and earnings. Partitioning visits into the initial registration visit and follow-up visits to a given client, the experiment reveals no effect of the self-tracking app on the former (Column 4), but a precisely estimated 36.6% increase in reported follow-up visits in the self-tracking treatment (Column 5). Driven by this effect, self-tracking CHAs report 26.8% more total visits than their counterparts (Column 6). To investigate the time pattern of this effect, Column 7 separates the 10-month experimental period into first and second halves. The treatment effect is concentrated in the first half of the experimental period. Figure 6 illustrates this finding graphically by plotting the treatment effect.

---

32 That take-up of the visual software is higher for the generic encouragement app than for the self-tracking app is likely because the encouragement app provides new content daily, whereas the self-tracking app is informationally static in the absence of visits, which occur less than daily for all ASHAs.
on total visits for each month of the experiment. The effect is largest in the first four months of the experiment, and then steadily declines over the next six months, though all point estimates remain positive.

Finally, reinforcing these results as well as the intuition that follow-up visits reflect intrinsic effort, the self-tracking app has no impact on earnings (Column 8). Thus, assuming rational expectations and accurate measurement, the increase in follow-up visits in the self-tracking app treatment cannot be explained by monetary preferences alone.

5.3 Heterogeneous treatment effects

The theoretical framework in Section 2 predicts complementarity between intrinsic incentives and intrinsic motivation and that, in the presence of intrinsic aversion, an intrinsic incentive may dampen effort. In this section, I assess both predictions.

Table 5 reports the estimates of

\[
\ln(y_{it}) = \beta_0 + \beta_1 A_t + \beta_2 A_t T_i + \beta_3 A_t \ln(\theta^k_i) + \beta_4 A_t T_i \ln(\theta^k_i) + \alpha_i + Z_m \gamma_m + u_{it} \tag{4}
\]

where \(\ln(y_{it})\) is log total visits reported by ASHA \(i\) on day \(t\); \(A_t\) is the post-intervention indicator; \(T_i\) is the treatment indicator; \(\ln(\theta^k_i)\) is the log motivation of ASHA \(i\) with respect to psychometric dimension \(k \in \{\text{extrinsic, intrinsic, prosocial, social desirability, competitive}\}\); \(\alpha_i\) is a vector of ASHA fixed effects; \(Z_m\) is a vector of month-year dummies; and \(u_{it}\) is the error term. The coefficient of interest is \(\beta_4\), which is the marginal elasticity of effort with respect to psychometric trait \(\theta^k\) in the self-tracking treatment, relative to that in the generic encouragement treatment. In other words, it is the difference in elasticity between two curves—those between total visits and \(\theta^k\) in the self-tracking treatment and in the generic encouragement treatment, respectively—and it measures the self-tracking app’s degree of complementarity with a given motivational trait relative to that of the generic encouragement app.

In Table 5, each row is a fixed-effects regression of Equation 4, where only the row variable changes. Column 1 (which reports estimates of \(\beta_3\) in the above equation) shows that, in the
control condition, intrinsic and prosocial motivation have negative elasticities of effort as measured by total visits. That is, a 1% increase in intrinsic (prosocial) motivation is associated with a 0.11% (0.09%) decrease in total visits per day. This itself is not remarkable, as motivation is not exogenous and could be correlated with other traits (e.g., ability) that affect performance. More important is the finding in Column 2 that the intrinsic and prosocial motivation elasticities of effort are significantly higher in the self-tracking condition than in the generic encouragement condition. In other words, the self-tracking condition is more effective at eliciting performance the more intrinsically/prosocially motivated an ASHA is; it leverages intrinsic/prosocial motivation. This relationship does not hold for extrinsic motivation, and as two additional placebo tests, it does not hold for social desirability or competitive motivation. 33 The two placebo traits are notable in that they are plausible confounds: the self-tracking app may leverage social desirability or competitiveness, but this is not observed in the data.

Table 6 further characterizes these results by partitioning the sample into terciles of psychometric traits and estimating Equation 3 for each tercile. Specifications 1, 3, and 5 use a pooled causal estimator for the entire post-intervention period, whereas Specifications 2, 4, and 6 estimate each tercile-specific treatment effect in the first half of the experiment, as well as the change in treatment effect in the second half. The estimates in the odd-numbered columns for intrinsic and prosocial motivation show that the self-tracking app has no effect on total visits in the least motivated tercile of ASHAs with respect to each trait. In contrast, the self-tracking app has positive effects in the middle terciles and even stronger effects in the top terciles of each trait. The largest treatment effect, which is observed for the top tercile of intrinsic motivation, is 46% of the mean in the generic encouragement condition. While the point estimates of the treatment effects for each tercile of extrinsic motivation have a positive slope, it is not significant, as was shown in Table 5.

The even-numbered specifications in Table 6 illustrate how the treatment effects heterogeneously evolve over time. Across all three motivational traits, the least motivated tercile exhibits a positive treatment effect in the first half of the experiment, but this effect decays to zero in the second half. A similar decay is observed in the middle tercile of extrinsic motivation, whereas the treatment effect

33 Both of these psychometric scales were administered at baseline along with the other psychometric traits. The scale items are listed in the Appendix.
in the top tercile of extrinsic motivation increases by an insignificant amount from an insignificant magnitude in the first half of the experiment. In contrast, in the middle and top terciles of intrinsic and prosocial motivation, the treatment effect is positive and precisely estimated in the first half of the experiment, and it persists through the second, half, showing no significant decay. That is, the most intrinsically/prosocially motivated ASHAs respond to the self-tracking app from the outset, and their response persists, whereas less intrinsically/prosocially motivated ASHAs respond at the outset but only transiently.

Taken together, the foregoing analysis of average and distributional effects suggest three main findings. The self-tracking app treatment leads to an average increase in client visits; consistent with theory, this effect interacts positively with intrinsic and prosocial motivation; and this complementarity is the result of the dynamic persistence of treatment response among more intrinsically and prosocially motivated workers. As to the prediction that a high-powered intrinsic incentive may dampen effort for those who are intrinsically unmotivated, the results do demonstrate a decay in client visits over time amongst the least intrinsically/prosocially motivated ASHAs. Whether this treatment effect will actually turn negative remains to be seen.

In the remainder of this sub-section, I examine the robustness of these results with respect to the validity of the data and the impact of the experimental interventions on health outcomes.

5.4 Compensating mechanisms

Table 4 examines whether the effect of the self-tracking app on reported visits is compensated by negative effects on other measures of effort. One such mechanism could be poorer targeting of visits both across space—e.g., visiting easy-to-reach clients many times, to the exclusion of other clients—and across time—e.g., visiting clients multiple times in some months but none in others. Column 1 reports estimates of a month-level panel regression in which the dependent variable is the share of pregnant clients visited by an ASHA in a given month. During the experimental period, self-tracking ASHAs visit an 8.72-percentage point greater share of their clients each month, a 23.1% improvement over the control mean. That this effect is approximately equal to the treatment effect
on total visits indicates that increased visits do not occur at the expense of coverage. Figure 7b graphically illustrates this increase in client coverage.

Columns 2-5 test for fabrication of visits. How long an ASHA spends completing a visit form is tracked by the phone directly, from the time she opens the form to when she completes it. Shorter visit duration may reflect higher productivity—e.g., greater proficiency at typing on the phone—or a socially efficient allocation of time across clients and visits. Nevertheless, it also raises concern for fabrication. Column 4 shows that self-tracking ASHAs spend 13.5% less time filling out forms; this effect is similar when considering new client visits and follow-up visits separately (Columns 2-3). Taken together, these results suggest two complementary interpretations. First, although average visit duration is 13.5% shorter, because self-tracking CHAs report 26.8% more total visits, they spend more aggregate time “inside” forms. While this does not rule out fabrication, it makes pure fabrication less plausible, since that would be expected to be a time-saving strategy manifesting in part with a higher proportion of very short visits. Second, similar effects on duration are observed for both initial and follow-up visits. This too casts doubt on the extent of fabrication, as the nature of the two types of visits is that it is easier to fabricate a follow-up visit form (which consists of checking off a list of counseling topics, which is difficult to verify) than an initial registration form (which requires typing the name of a client, her husband, her phone number, and so forth, all of which can be verified).

5.5 Health impacts

Finally, Table 4, Columns 5-12 assesses for treatment effects on four health practices and outcomes at the client level, as reported by ASHAs. These include practices such as attending antenatal care visits and receiving tetanus vaccinations, as well as pregnancy outcomes such as institutional delivery and maternal death. Columns 5-12 estimate

\[ y_{ijt} = \beta_0 + \beta_1 T_j + \beta_2 E_t + \beta_3 T_j \ast E_t + Z_i \theta + \alpha_j + u_{ijt} \]  

(5)
where \( y_{ijt} \) is the outcome of client \( i \) of ASHA \( j \) during period \( t \); \( T_j \) is ASHA \( j \)’s treatment assignment, where \( T_j = 1 \) for the self-tracking treatment; \( E_t \) denotes whether the client’s pregnancy was exposed to the experimental period; \( Z_i \) is a vector of client-level controls; and \( \alpha_j \) is a vector of ASHA dummies. The estimation sample is restricted to clients who have completed their pregnancy. \( E_t \) defines as exposed those clients who were registered by an ASHA prior to the launch of the experiment (so as to ensure no endogenous selection) but whose pregnancy concluded after the launch \((N = 1,779)\). Those who began and ended their pregnancy before the launch of the experiment are classified as non-exposed \((N = 4,820)\). The difference-in-differences estimator of interest is \( \beta_3 \), which, under the parallel trends assumption, identifies the effect of the self-tracking app on client outcomes. Standard errors are clustered at the ASHA level.

In brief, the results are equivocal. Regarding antenatal practices, the self-tracking condition leads to an increase in reported ANC visits but has no effect on ASHAs’ reports of how many tetanus vaccines the client received and whether the client has developed a birth plan. The average effect on these practices is not significantly different from zero (Column 8).

Regarding pregnancy outcomes, these too are reported by ASHAs, but using a separate form at the conclusion of the pregnancy. Whether the ASHA submits a pregnancy outcome form is itself an important outcome, as it is likely to be (negatively) correlated with whether the client has been lost to follow-up. On this margin, we observe no significant effect (Column 9). Surprisingly, Column 10 shows that, conditional on delivery, the probability of institutional delivery is 4.8 percentage points lower in the self-tracking group than in the generic encouragement group, in which 78.8% of deliveries occur at a health facility. No effect on the probability of live birth (at home or at a facility) is observed. When pooled, their average effect size is 0.06 standard deviations in favor of the generic encouragement treatment, which is on the cusp of reaching statistical significance \((p = 0.113)\).

These results on health impacts raise several questions. Assuming no misreporting of performance data, it is difficult to explain how the self-tracking treatment would reduce pregnant women’s propensity to deliver in a facility. Consistent with this expectation, ASHAs in the self-tracking condition do improve two key antenatal health practices: attending antenatal visits and
obtaining immunizations. Alternatively, the results may be biased due to selection on both registering clients and reporting their data. Finally, there may be unobserved beneficial (and detrimental) effects of the self-tracking app that are not captured in the current data.

These explanations cannot currently be disentangled, due to the limited (in both extent and quality) nature of the outcomes data that ASHAs self-report. A much more thorough survey of clients is ongoing and will enrich the analysis of effects of the self-tracking app on health.

5.6 Causal mechanisms

The experimental data offer suggestive evidence that, with the caveat that no aggregate health benefit is observed, the high-powered intrinsic incentive—the self-tracking app—is more powerful at stimulating effort than the low-powered incentive, and that it this effect is increasing in intrinsic motivation. The latter interaction effect supports the hypothesis that the treatment effect is mediated by leveraging pre-existing intrinsic motivation. Nevertheless, in this section, I briefly discuss other mechanisms that might explain the treatment effect. In particular, as the utility function (Equation 1) makes clear, an incentive can increase observed output through three channels: by increasing the marginal intrinsic utility, marginal extrinsic utility, and productivity of effort.

As regards implicit extrinsic incentive effects, two empirical results discussed above cast doubt on the possibility that the treatment effect on client visits is driven by monetary preferences: first, that ASHAs in the self-tracking condition earn no more than their counterparts in the generic encouragement condition, and second, that those who are more extrinsically motivated respond no more strongly to the self-tracking app than those who are less extrinsically motivated.

Besides a financial incentive, another embedded extrinsic incentive could arise if ASHAs interpret the self-tracking app as a monitoring tool or as a way to make good performance more visible to their employer. As suggestive evidence, nine months into the experiment, ASHAs across both treatments were asked whether their direct supervisors, the auxiliary nurse midwife (ANM) cadre, ever viewed their apps. ASHAs interact with their supervising ANM at least monthly for village health and nutrition days, and the ANMs could conceivably use either app, but especially the self-tracking app, as a supervisory tool. In fact, the proportion of ASHAs who report that their ANM
views their app is balanced between the self-tracking (20.5%) and generic encouragement (26.4%) conditions.

Finally, in addition to increasing marginal returns to effort, an incentive can also increase returns for a given level of effort by altering an agent’s production function such that the effort cost of producing a given level of output is reduced. These alterations could include a reminder effect that decreases the effort required to remember to do a task. Although plausible in principle, this reminder mechanism is inconsistent with the design of both the self-tracking and generic encouragement apps. That is, the software is “pull”-based in that the ASHA must access the software intentionally; there is no “pushing” of app content to the ASHAs. The software is therefore unlikely to function as a reminder given that the ASHA is already thinking about ASHA work when she accesses it.

6 Conclusion

It is not from the benevolence of the butcher, the brewer, or the baker that we expect our dinner, but from their regard for their own interest. We address ourselves not to their humanity, but to their self-love. . . .

—Adam Smith, The Wealth of Nations, 1776

Adam Smith’s canonical passage about the power of self-interest in creating efficient market outcomes has yet to be refuted. But as Smith (1759) himself argued, humans are also driven by motivations other than self-interest. Perhaps in no economic setting is this more true than in the prosocial sector, where agents often cite intrinsic rewards arising from enjoyable, purposeful work as a principal job benefit. Given this motivation, there may be value in addressing ourselves not only to their self-love, but also to their humaneness.34

This experiment provides evidence that a technology designed to leverage intrinsic motivation does indeed enhance intrinsic effort, in a manner that fits neatly into a standard utility maximization framework. The experiment sets forth a conceptual framework for thinking about extrinsic and

34The observation applies to the demand side context as well. For example, most people are intrinsically motivated to be healthy, but efforts to engage in healthy behaviors often fall short due to self-control problems. At a time when demand-side financial incentives such as conditional cash transfers are gaining popularity (Rawlings, 2005; Volpp et al., 2009), there may be a role for intrinsic incentives.
intrinsic motivation and incentives, and it demonstrates the feasibility of developing and testing intrinsic incentives for public service delivery in the field. I now conclude by posing two broad sets of questions for further research.

First, in what ways do intrinsic incentives interact with preferences and institutional contexts? For example, this paper has argued that the effect of an intrinsic incentive depends in part on the distribution of intrinsic motivation in the agent population. But it may also depend on other factors, especially the background extrinsic incentive regime, for which, in this experiment, there is no variation to enable identification. Are extrinsic and intrinsic rewards substitutes or complements? What is their relative cost-effectiveness? As much as the cross-effect of extrinsic incentives on intrinsic motivation has been studied, what is the effect of intrinsic incentives on extrinsic motivation? Can intrinsic incentives crowd out or crowd in extrinsic motivation?

Second, what other intrinsic incentives might be effective? The theoretical framework posits that there is a function $\psi(\cdot)$ that modulates intrinsic returns to effort. The self-tracking app is designed to increase marginal intrinsic utility by making the effort task more salient. Hence, task salience may belong in the $\psi$ function. But other incentive technologies may as well, such as an organizational culture that fosters a sense of prosocial meaning, a production system that encourages social bonding, or a gamification technology that makes mundane work take on the aspect of a game. And the effect of these technologies is likely to interact with heterogeneous preferences. Thus, although a simple device, the $\psi(\cdot)$ function points to a rich line of inquiry—dating at least as far back as the ancient Greek concepts of hedonia and eudaimonia—on what makes us happy. While this question has historically been the purview of other disciplines (including the humanities), it also lies at the heart of economics—implicating, as it does, the very notion of utility and the preferences that shape it.
References


Figure 5: Generic encouragement app: sample pages.
Table 1: Descriptive statistics.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>p25</th>
<th>p75</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. ASHA performance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average monthly case load</td>
<td>13.13</td>
<td>5.66</td>
<td>9.21</td>
<td>15.70</td>
<td>72</td>
</tr>
<tr>
<td>New cases per month</td>
<td>2.89</td>
<td>1.34</td>
<td>2.02</td>
<td>3.52</td>
<td>72</td>
</tr>
<tr>
<td>Visits per month</td>
<td>7.40</td>
<td>4.22</td>
<td>4.12</td>
<td>9.98</td>
<td>72</td>
</tr>
<tr>
<td>Average share of open cases visited each month</td>
<td>0.46</td>
<td>0.14</td>
<td>0.38</td>
<td>0.54</td>
<td>72</td>
</tr>
<tr>
<td>Average gestational week in which case opened</td>
<td>21.67</td>
<td>2.39</td>
<td>19.97</td>
<td>23.31</td>
<td>72</td>
</tr>
<tr>
<td>Average visits per case during entire pregnancy</td>
<td>2.57</td>
<td>0.96</td>
<td>1.94</td>
<td>3.06</td>
<td>72</td>
</tr>
<tr>
<td>Share of completed cases with outcome reported</td>
<td>0.81</td>
<td>0.08</td>
<td>0.76</td>
<td>0.87</td>
<td>72</td>
</tr>
<tr>
<td><strong>B. Client characteristics and pregnancy outcomes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age of client (years)</td>
<td>25.19</td>
<td>3.85</td>
<td></td>
<td></td>
<td>4913</td>
</tr>
<tr>
<td>Previous pregnancies</td>
<td>2.06</td>
<td>1.80</td>
<td></td>
<td></td>
<td>4576</td>
</tr>
<tr>
<td>Institutional delivery (all reported deliveries)</td>
<td>0.77</td>
<td>0.42</td>
<td></td>
<td></td>
<td>3275</td>
</tr>
<tr>
<td>Live birth (all reported pregnancies)</td>
<td>0.91</td>
<td>0.28</td>
<td></td>
<td></td>
<td>3835</td>
</tr>
<tr>
<td>Premature birth (all reported deliveries)</td>
<td>0.24</td>
<td>0.43</td>
<td></td>
<td></td>
<td>3501</td>
</tr>
<tr>
<td>Maternal mortality ratio (all live births, per 100,000)</td>
<td>256.70</td>
<td>5060.80</td>
<td></td>
<td></td>
<td>3506</td>
</tr>
</tbody>
</table>

*Notes: Panel A includes ASHAs in the generic encouragement condition across all time periods (15 pre-intervention and 10 post-intervention months). Panel B encompasses all pregnant clients in the generic encouragement condition. Denominators are indicated in parentheses. Premature birth is defined as birth at least three weeks before the expected due date. “p25” and “p75” denote the 25th and 75th percentiles, respectively.*
Table 2: Randomization balance.

<table>
<thead>
<tr>
<th>Treatment assignment:</th>
<th>Generic encouragement (1)</th>
<th>Self-tracking (2)</th>
<th>p-value of difference (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline performance (March 2013 - May 2014):</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visits per month</td>
<td>8.344 [4.296]</td>
<td>8.252 [4.421]</td>
<td>0.899</td>
</tr>
<tr>
<td>Monthly earnings (INR)</td>
<td>1989.5 [1459.1]</td>
<td>1974.9 [1360.1]</td>
<td>0.951</td>
</tr>
</tbody>
</table>

| **Baseline ability:** |                           |                   |                           |
| Education (years)     | 9.264 [2.556]             | 8.795 [3.140]     | 0.326                     |
| Able to read Hindi (0/1) | 0.806 [0.399]           | 0.822 [0.385]     | 0.802                     |
| Score on general health knowledge assessment (0-18) | 15.53 [1.34]           | 15.16 [1.60]      | 0.141                     |
| Score on pregnancy knowledge assessment (0-12) | 8.611 [1.369]           | 8.904 [1.556]     | 0.231                     |

| **Psychological traits:** |                           |                   |                           |
| Extrinsic motivation scale (1-5) | 4.208 [0.489]           | 4.229 [0.566]     | 0.811                     |
| Intrinsic motivation scale (1-5) | 4.289 [0.415]           | 4.279 [0.432]     | 0.894                     |
| Prosocial motivation scale (1-5) | 4.499 [0.386]           | 4.510 [0.381]     | 0.858                     |

| **Other ASHA characteristics:** |                           |                   |                           |
| Age (years) | 33.24 [6.91]           | 33.34 [7.13]      | 0.928                     |
| Belongs to a disadvantaged caste (0/1) | 0.806 [0.399]           | 0.822 [0.385]     | 0.802                     |
| Resides in work village (0/1) | 0.903 [0.298]           | 0.822 [0.385]     | 0.160                     |
| Has electricity at home (0/1) | 0.694 [0.464]           | 0.685 [0.468]     | 0.902                     |
| Has non-mud floor at home (0/1) | 0.389 [0.491]           | 0.384 [0.490]     | 0.948                     |
| Asset index (quintiles 1-5) | 2.901 [1.343]           | 3.068 [1.484]     | 0.480                     |

| **Village characteristics:** |                           |                   |                           |
| Village population (national census) | 1435.4 [851.8]           | 1321.9 [719.7]    | 0.572                     |
| Village is majority Muslim (0/1) | 0.278 [0.451]           | 0.151 [0.360]     | 0.063*                    |

| Number of ASHAs | 72 | 73 | 0.8199 |

Notes: Table reports means and standard deviations (in brackets) for the two experimental conditions, as well as a test of equality for each variable, with stars signifying *p<0.10, **p<0.05, ***p<0.01. The first two baseline performance variables (monthly new clients and visits) are obtained from the ASHAs’ CommCare form submission data during the 15 months prior the experimental intervention launch (March 2013 - May 2014). Mean monthly earnings pre-intervention is collected from the ASHAs’ payment books, in which the government payment office records payments for ASHA work. Education, Hindi literacy, and the health knowledge variables were all obtained from the baseline survey. The psychometric scales and work motivation variables were measured through the baseline survey; see the Data Appendix for more details. ASHA characteristics were measured in the baseline survey. The village population variable is based on 2001 Indian national census data. "Village is majority Muslim" was self-reported by ASHAs in the baseline survey.
Table 3: Average treatment effects on take-up and performance.

<table>
<thead>
<tr>
<th></th>
<th>Take-up</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Visual sessions</td>
<td>Audio sessions</td>
</tr>
<tr>
<td>Self-tracking app x</td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Post-intervention period</td>
<td>-0.0461*</td>
<td>-0.00884</td>
</tr>
<tr>
<td></td>
<td>(0.0258)</td>
<td>(0.0251)</td>
</tr>
<tr>
<td>Self-tracking app x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Second half of post-intervention period</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean [SD] for generic</td>
<td>0.175</td>
<td>0.147</td>
</tr>
<tr>
<td>encouragement app</td>
<td>[0.182]</td>
<td>[0.156]</td>
</tr>
</tbody>
</table>

Unit                      | daily mean | daily mean | daily mean | day | day | day | day | day | month |
| Pre-intervention units     | 0          | 0          | 0          | 455 | 455 | 455 | 455 | 455 | 15    |
| Post-intervention units    | 286        | 286        | 286        | 286 | 286 | 286 | 286 | 286 | 8     |
| Baseline controls          | Y          | Y          | Y          | N/A | N/A | N/A | N/A | N/A | N/A   |
| Month-year fixed effects   | N/A        | N/A        | N/A        | 0.016 | 0.047 | 0.042 | 0.042 | 0.474 |
| Adjusted R²                | 0.107      | 0.091      | 0.145      | 145  | 145  | 145  | 145  | 145  | 134   |
| ASHAs                      | 145        | 145        | 145        | 101184 | 101165 | 101195 | 101195 | 2326  |
| Observations               | 145        | 145        | 145        |             |             |             |             |             |

Notes: Table reports OLS estimates for cross-sectional (Columns 1-3) and fixed-effects panel regressions (Columns 4-8). Column headings are dependent variables. Standard errors (clustered at the ASHA level for panel regressions) are in parentheses. The coefficient on the post-intervention time dummy is suppressed in Columns 4-8. Baseline controls in Columns 1-3 include the stratification variables (literacy, baseline monthly visits, extrinsic motivation, intrinsic motivation, and prosocial motivation) and whether the ASHA reports that her village is majority Muslim. The mean in the generic encouragement app condition is reported for the post-intervention period (standard deviation in brackets). The differences in observations in Columns 4-7 are due to varying start dates for ASHAs at the beginning of the pre-intervention period. In Column 8, data for 11 ASHAs are missing due to these ASHAs either no longer working (3 ASHAs) or working but no longer in possession of the government ledger book in which their payments are recorded. The most recent two months of the experiment are also missing in Column 8 due to a lag in payments. Estimates are significant at the *10%, **5%, and ***1% level.

Figure 6: Impact of self-tracking app on total visits, by month.

Notes: Plot shows treatment effects of the self-tracking app on total visits in each month of the experiment, expressed as a percentage of mean visits in the generic encouragement condition in that month. Error bars are 90% confidence intervals.
Table 4: Average treatment effects on coverage, visit duration, and health outcomes.

<table>
<thead>
<tr>
<th>Share of clients visited</th>
<th>Visit duration (minutes)</th>
<th>Antenatal practices</th>
<th>Pregnancy outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>New client visits (1)</td>
<td>Follow-up visits (2)</td>
<td>All visits (3)</td>
</tr>
<tr>
<td>Self-tracking app x</td>
<td>0.0872***</td>
<td>-1.813*</td>
<td>-1.539*</td>
</tr>
<tr>
<td>Post-intervention period</td>
<td>(0.0237)</td>
<td>(1.032)</td>
<td>(0.914)</td>
</tr>
<tr>
<td>Mean [SD] for generic encouragement app</td>
<td>0.376</td>
<td>14.62</td>
<td>12.50</td>
</tr>
<tr>
<td>[0.222]</td>
<td>[15.35]</td>
<td>[12.87]</td>
<td>[13.80]</td>
</tr>
<tr>
<td>Unit</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Client controls</td>
<td>N/A</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.343</td>
<td>0.084</td>
<td>0.149</td>
</tr>
<tr>
<td>ASHA fixed effects</td>
<td>N/A</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>ASHAAs</td>
<td>145</td>
<td>9062</td>
<td>17365</td>
</tr>
<tr>
<td>Observations</td>
<td>3409</td>
<td>9062</td>
<td>17365</td>
</tr>
</tbody>
</table>

Notes: Table reports OLS estimates of fixed-effect panel regression at the month level (Column 1) and cross-sectional regressions at the level of the visit (Columns 2-4) and client (Columns 5-12). Standard errors clustered at the ASHA level are in parentheses. Column headings are dependent variables. Coefficients on the self-tracking app and post-intervention dummies are suppressed. The mean in the generic encouragement app condition is reported for the post-intervention period (standard deviation in brackets). Visit duration in Columns 2-4 is winsorized at the 1st and 99th percentiles. In Columns 5-12, client-level controls are client age, parity, mobile phone ownership, and gestational week at registration by ASHA. Columns (8) and (12) report average effect sizes (Clingingsmith et al., 2009; Kling et al., 2004) of the variables for antenatal practices and pregnancy outcomes, respectively. Estimates are significant at the *10%, **5%, and ***1% level.
Figure 7: Treatment effects on visits and client coverage.

(a) Average visits per day.

(b) Average share of pregnant clients visited each month.
Table 5: Elasticities of effort with respect to psychometric traits.

<table>
<thead>
<tr>
<th>Total visits</th>
<th>Elasticity in generic encouragement treatment</th>
<th>Marginal elasticity in self-tracking treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Extrinsic motivation</td>
<td>-0.155 (0.126)</td>
<td>0.0806 (0.169)</td>
</tr>
<tr>
<td>Intrinsic motivation</td>
<td>-0.110*** (0.0257)</td>
<td>0.102** (0.0495)</td>
</tr>
<tr>
<td>Prosocial motivation</td>
<td>-0.0885*** (0.0242)</td>
<td>0.104** (0.0499)</td>
</tr>
<tr>
<td>Social desirability</td>
<td>-0.0110 (0.0301)</td>
<td>0.0472 (0.0544)</td>
</tr>
<tr>
<td>Competitive preference</td>
<td>-0.00586 (0.0296)</td>
<td>-0.00989 (0.0501)</td>
</tr>
</tbody>
</table>

Notes: Table reports OLS estimates of fixed-effects panel regressions. Standard errors (clustered at the ASHA level) are in parentheses. The unit of analysis is the ASHA-day, and the dependent variable is log total visits. Each row is a specification that interacts the row psychometric trait with the post-intervention dummy to estimate Column 1, and with the post-intervention dummy and the self-tracking treatment to estimate Column 2. Estimates for the post-intervention dummy and the interaction of treatment dummy and post-intervention dummy are suppressed. All psychometric traits are measured on a 1-5 scale, lowest to highest. All specifications include month-year fixed effects. Estimates are significant at the *10%, **5%, and ***1% level.
Table 6: Treatment effects on total visits by terciles of psychometric traits.

<table>
<thead>
<tr>
<th></th>
<th>Tercile of row variable</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bottom</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>Extrinsic motivation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-tracking app x Post-intervention period</td>
<td>0.0384</td>
<td>0.0769**</td>
<td>0.0584*</td>
<td>0.0893**</td>
<td>0.0638*</td>
<td>0.0543</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0268)</td>
<td>(0.0292)</td>
<td>(0.0299)</td>
<td>(0.0342)</td>
<td>(0.0361)</td>
</tr>
<tr>
<td>Self-tracking app x Second half of post-intervention period</td>
<td>-0.0810***</td>
<td>-0.0662*</td>
<td>0.0141</td>
<td>0.0448</td>
<td>0.0446</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0276)</td>
<td>(0.0392)</td>
<td>(0.0232)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.0426</td>
<td>0.0429</td>
<td>0.0399</td>
<td>0.0400</td>
<td>0.0448</td>
<td>0.0446</td>
</tr>
<tr>
<td>ASHAs</td>
<td>58</td>
<td>58</td>
<td>51</td>
<td>51</td>
<td>36</td>
<td>36</td>
</tr>
<tr>
<td>Observations</td>
<td>40415</td>
<td>40415</td>
<td>35563</td>
<td>35563</td>
<td>25217</td>
<td>25217</td>
</tr>
<tr>
<td>Intrinsic motivation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-tracking app x Post-intervention period</td>
<td>0.0145</td>
<td>0.0514*</td>
<td>0.0678**</td>
<td>0.0777***</td>
<td>0.0956***</td>
<td>0.107***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0278)</td>
<td>(0.0304)</td>
<td>(0.0279)</td>
<td>(0.0277)</td>
<td>(0.0300)</td>
</tr>
<tr>
<td>Self-tracking app x Second half of post-intervention period</td>
<td>-0.0794***</td>
<td>-0.0246</td>
<td>-0.0259</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0271)</td>
<td>(0.0293)</td>
<td>(0.0357)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.0475</td>
<td>0.0477</td>
<td>0.0303</td>
<td>0.0302</td>
<td>0.0472</td>
<td>0.0471</td>
</tr>
<tr>
<td>ASHAs</td>
<td>59</td>
<td>59</td>
<td>41</td>
<td>41</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>Observations</td>
<td>40865</td>
<td>40865</td>
<td>28753</td>
<td>28753</td>
<td>31577</td>
<td>31577</td>
</tr>
<tr>
<td>Prosocial motivation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-tracking app x Post-intervention period</td>
<td>0.0236</td>
<td>0.0538*</td>
<td>0.0808***</td>
<td>0.0982***</td>
<td>0.0783**</td>
<td>0.0863**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0294)</td>
<td>(0.0305)</td>
<td>(0.0285)</td>
<td>(0.0298)</td>
<td>(0.0295)</td>
</tr>
<tr>
<td>Self-tracking app x Second half of post-intervention period</td>
<td>-0.0653***</td>
<td>-0.0352</td>
<td>-0.0242</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0217)</td>
<td>(0.0353)</td>
<td>(0.0413)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.0392</td>
<td>0.0393</td>
<td>0.0478</td>
<td>0.0478</td>
<td>0.0419</td>
<td>0.0417</td>
</tr>
<tr>
<td>ASHAs</td>
<td>65</td>
<td>65</td>
<td>42</td>
<td>42</td>
<td>38</td>
<td>38</td>
</tr>
<tr>
<td>Observations</td>
<td>45104</td>
<td>45104</td>
<td>29295</td>
<td>29295</td>
<td>26796</td>
<td>26796</td>
</tr>
</tbody>
</table>

Notes: Table reports OLS estimates of fixed-effects panel regressions at the day level. Standard errors (clustered at the ASHA level) are in parentheses. The dependent variable is total visits. Each column estimates the specified coefficient(s) in the specified tercile of the row variable. All specifications include month-year fixed effects. Estimates are significant at the *10%, **5%, and ***1% level.
Appendix

A Data sources

This experiment draws on several data sources. Data on ASHA performance come from two sources. The first is the CommCare mobile phone application that all ASHAs use, as described in Section 3.2. ASHAs self-report client visits, and visits data are at the client-day level. Since each visit is associated with a client, it is possible to track initial vs. follow-up visits to the client. In addition, ASHAs collect and report basic health and demographic data about their clients (expected delivery date, age, parity, marital status). At the conclusion of the client’s pregnancy, the ASHA fills out a case completion form reporting the maternal outcome (survival, death), the fetal outcome (miscarriage, stillbirth, live birth), and delivery location (home, health facility).

The other source of performance data is earnings. ASHA payments are recorded in a receipt book that ASHAs keep in their possession. To be paid, the ASHA visits the district hospital and submits a request for payment. Based on delivery records, the payment is transferred into the ASHA’s bank account and recorded in the ASHA’s receipt book. The receipt books only record ASHA-related payments (i.e., no personal transactions). The raw data in the receipt books are at the transaction level, but these do not differentiate between payments for deliveries vs. other activities, or for specific deliveries. The compensation data used in this paper are thus aggregated at the month level.

For data on take-up of the experimental interventions, I measure usage directly. The software has the ability to track usage by users in a highly granular manner, with a timestamp recorded for every page view across the different pages in the software. For the analysis in this paper, one unit of take-up is defined as any contiguous series of page-views with a maximum timestamp interval of ten minutes. Thus, if a user views 10 pages every two minutes over the course of 20 minutes, this is counted as one session. If she views two pages over the course of 20 minutes, this is counted as two sessions. The intention behind this rule is to capture how often ASHAs use the experimental intervention, rather than how intensively they do so. However, the results of the experiment are
robust to using individual page-views as the unit of take-up. The audio service also tracks incoming phone calls, whether the automated response call is answered, and how much content is played to the recipient before the recipient hangs up. In this paper, any instance of the system’s automated response call being received by the ASHA is counted as one session.

I collect survey data about ASHA traits, beliefs, and preferences through a baseline survey conducted the launch of the experiment and a midline survey conducted in month 9 of the experiment. The baseline survey includes modules for personal and household demographics; an assessment of health- and pregnancy-related knowledge; a series of psychometric scales; and job motivations and preferences. The psychometric scales are adapted from previously published validated scales designed to measure extrinsic, intrinsic, and prosocial preferences (Amabile et al., 1994; Grant, 2008; Wrzesniewski et al., 1997). The statements constituting each scale are listed in Appendix A.1. During the survey, the enumerator read each scale item aloud to the respondent, and the respondent was asked to state whether she strongly agreed, agreed, neither agreed nor disagreed, disagreed, or strongly disagreed with each statement.

The midline survey was administered during the ninth month of the experiment. It queried each ASHA’s work habits, motivations, time use across tasks, and comprehension and ability to use the experimental app to which she was assigned. A final module asked about each ASHA’s knowledge of the treatment assignment and performance of each of her peers within her subcenter.
The psychometric scales used in this experiment are adapted from validated scales. During the actual baseline survey, the items were interspersed with one another along with items from other scales. The extrinsic and intrinsic motivation scales draw primarily from Amabile et al. (1994), and the prosocial motivation scale draws primarily from Grant (2008). Each of the three scales also draws from items in Wrzesniewski et al. (1997). The particular items were chosen ex ante based on consultation with native Hindi speakers regarding which items could be translated into Hindi most clearly and would be most likely to be understood by the ASHA respondents. All three of these scales were pre-specified in a pre-analysis plan (Lee, 2014).

The final two scales were administered in the same exercise. The social desirability scale is adapted from Hays et al. (1989). The competitive preference scale is adapted from Bartick et al. (2002) and Amabile et al. (1994).

**Extrinsic motivation:**
- I want to be in a higher-level job in five years.
- My primary reason for working is financial—to support my family and lifestyle.
- I think a lot about how much money I have and how much I can make.
- I often think about salary and promotions.

**Intrinsic motivation:**
- I enjoy talking about health to others.
- The more difficult the problem, the more I enjoy trying to solve it.
- I enjoy doing work that is new to me.
- I want to find out how good I can be at my work.
- What matters most to me is enjoying what I do.

**Prosocial motivation:**
- ASHA work makes the world a better place.
- I want to help others through my work.
- At work, I care about improving the lives of other people.
- I do more work than is required of me to help my clients to be healthy.

**Social desirability motivation:**
- I am always respectful and considerate even to people who are rude and unfriendly.
- No matter who I am talking to, I am always a good listener.
- I sometimes feel annoyed when I don’t get my way (reverse-coded).

**Competitive motivation:**
- I often compare my ASHA work against other ASHAs’ work.
- I try to be the highest performing ASHA.
- To me, success means doing better than other ASHAs.