No margin, no mission? A field experiment on incentives for public service delivery

No margin, no mission? A field experiment on incentives for public service delivery*

Nava Ashraf, Oriana Bandiera and B. Kelsey Jack

February 13, 2014

Abstract

We conduct a field experiment to evaluate the effect of extrinsic rewards, both financial and non-financial, on the performance of agents recruited by a public health organization to promote HIV prevention and sell condoms. In this setting: (i) non-financial rewards are effective at improving performance; (ii) the effect of both rewards is stronger for pro-socially motivated agents; (iii) the effect of both rewards is stronger when their relative value is higher. The findings illustrate that extrinsic rewards can improve the performance of agents engaged in public service delivery, and that non-financial rewards can be effective in settings where the power of financial incentives is limited.

JEL codes: J33, O15, M52, D82

Keywords: financial incentives, non-monetary rewards, pro-social motivation, public service delivery

*Ashraf: HBS and NBER, nashraf@hbs.edu; Bandiera: Department of Economics and STICERD, LSE, o.bandiera@lse.ac.uk; Jack: Department of Economics, Tufts University and NBER, kelsey.jack@tufts.edu.

1 Introduction

Understanding what motivates individuals to devote time and effort to work endeavors is a question that lies at the core of the social sciences. The answer is crucial both to understanding observed behavior and to designing incentive mechanisms that align the individuals' interests with the interests of the organization for which they work. As a consequence, the design of optimal incentive contracts has been the subject of extensive theoretical and empirical research.

Empirical contributions, however, mainly focus on the effect of financial rewards in settings in which employee effort only benefits the employer (Bandiera et al. 2011; Oyer and Schaefer 2011). Much less attention has been paid to incentives in organizations, such as governmental and non-governmental organizations, which hire agents to perform pro-social tasks, namely, tasks that create benefits enjoyed by those other than the employer and employees. A notable exception is the literature on the effect of monetary incentives on teachers' performance, which finds markedly mixed results. (Duflo et al. 2012; Fryer 2011; Lavy 2002; Glewwe et al. 2010; Muralidharan and Sundararaman 2011).

The theoretical literature suggests reasons why the effect of extrinsic rewards on performance in private and pro-social tasks might differ. Mission-driven organizations benefit from matching with workers whose interests are aligned with the mission, and these individuals might respond less to incentives or even deliver a weaker performance if incentives displace other sources of motivation. In particular, to the extent that agents are motivated by the externalities generated through pro-social tasks, this motivation may interact positively or negatively with extrinsic incentives (Benabou and Tirole 2003, 2006; Besley and Ghatak 2005; Dixit 2002).

Informed by these insights, we design a field experiment to evaluate the effect of extrinsic rewards on the performance of agents in a public health organization. The experiment is designed to compare the effects of monetary and non-monetary incentives, as both are commonly used in practice,[1] but their relative effectiveness for public service delivery is understudied.[2] The experiment is designed to measure the interaction between extrinsic rewards and the pro-social motivation of the agents, and to test whether this interaction differs between financial and non-financial rewards. We collaborate with a public health organization based in Lusaka, Zambia, which recruits and trains hairdressers and barbers to provide information about HIV prevention and sell condoms in their shops.

The experiment randomly assigns 205 distinct geographical clusters containing 1,222 agents to one of four groups that receive different rewards based on condom sales. Agents in the control

[1] Many organizations, ranging from large corporations to NGOs, use a range of non-financial performance rewards to motivate their employees. For example, Larkin (2011) uses observational data to study a non-linear incentive scheme that provides employees of a software firm with a "gold star" and company-wide recognition if they meet an annual performance threshold. His evidence suggests that employees forgo 27,000 USD worth of revenue to obtain the non-financial reward.

[2] Kube et al. (2012) compare the effect of monetary and non-monetary rewards on the performance of agents engaged in a task (book sorting) that has no pro-social elements. They find that the non-monetary reward, a water bottle, is more effective than the equivalent cash amount.

group receive no rewards, while agents in the three treatment groups receive financial margins at the bottom and the top of the feasible range, and non-financial rewards, respectively. The smaller and larger financial-margin treatments pay a 10% and 90% margin on each condom sale, respectively, whereas the non-financial scheme ("star" treatment) gives agents a "thermometer" display, showing condom sales and stamps on it, one star for each sale.

The first part of our empirical analysis shows that non-financial rewards are effective at promoting sales: agents in the star treatment sell over twice as many condoms as agents in any other group, on average. We track agents' performance over one year and thus can separate responses due to the novelty of the program from long-run responses. The estimates are stable throughout the one-year period, thus ruling out novelty effects. The magnitude of the estimated treatment effects is such that, had all agents been offered non-financial incentives, they would have sold 22,496 condoms, compared to 10,686 / 11,938 / 12,504 had they all been offered the volunteer contract, small financial margins and large financial margins, respectively.

That financial incentives are ineffective might be due to earnings from condom sales being a small fraction of overall earnings, because both demand for the product and earnings from each sale are low. Since demand for the product and the cost of effort are orthogonal to treatment, our results imply that the agents' marginal utility of stars is higher than their marginal utility of money, given their initial endowments of money and stars. In general, we expect there to be a threshold level of financial rewards such that all rewards above that threshold would be more effective at eliciting effort than non-financial incentives, and indeed, below, we show that financial rewards are effective for the poorest agents in the sample, for whom their relative value is higher.

The second part of the analysis explores mechanisms driving the estimated treatment effects. We begin by assessing whether treatments differ because they make the agents exert different levels of effort or because they affect demand directly. We provide three pieces of evidence on this matter. First, we show that agents in the star treatment behave differently on dimensions correlated with sales effort, such as displaying promotional materials and filling in sales records.This rules out that the star treatment increases sales exclusively by increasing demand. Second, we survey a random sample of customers to probe the effectiveness of different promotional materials; most surveyed customers recall and correctly describe the promotional posters given to agents in all treatments, but only a negligible minority mentions the thermometer that is only given to agents in the star treatment. Third, we implement a "placebo star-reward" treatment, namely we randomly provide a subsample of salons in the control and financial reward treatments with a thermometer that, to a third party, looks identical to the treatment thermometer, and hence is an equally effective advertising tool, but carries no reward for the agent, as the stars stamped on it represent the average sales in the area. I.e., the placebo star treatment has no effect on sales.

The next step of our analysis provides evidence on the interaction between extrinsic incentives and intrinsic motivation for the cause. To this purpose, we measure motivation through an

adapted dictator game where agents can make a donation to an existing charity that provides care to HIV/AIDS patients. We find that the donation is a strong predictor of sales performance; agents who donate more than the median sell 51 percent more condoms than the average agent in the control group. We find that agents who are motivated by the cause respond more strongly to both financial and non-financial rewards, suggesting that extrinsic incentives are complementary to pro-social motivation in this context.

The final step of our analysis shows that the responses to both financial and non-financial incentives are stronger when the utility associated with financial and non-financial rewards, respectively, is higher. In particular, we show that financial incentives increase sales for the poorest agents in our sample, for whom the relative value of rewards is higher. To measure the relative value of non-financial incentives, we exploit the intuition that these might be more valuable when they are visible to a larger peer group. To implement this test, we exploit the naturally occurring variation in the number of salons in each neighborhood. We find that the marginal effect of non-financial incentives is increasing in the number of neighboring salons that also received non-financial incentives, whereas the response to the other incentive treatments is not affected by the number of neighboring salons that receive the same treatment.

Our findings contribute to the broad literature evaluating the effect of incentives in for-profit firms and to the nascent literature studying how to motivate agents engaged in pro-social activities (see, for example, Gneezy and Rustichini 2000; Lacetera et al. 2011; Meier 2007; Mellström and Johannesson 2008). Most of the related literature on public services delivery focuses on performance incentives for teachers (Duflo et al. 2012; Fryer 2011; Lavy 2002; Glewwe et al. 2010; Muralidharan and Sundararaman 2011) with two recent exceptions, both of which analyze the delivery of health services. Miller et al. (2012) evaluate the effect of providing financial incentives to school principals to reduce anemia among students in rural China and find a modest effect. Olken et al. (2012) study an intervention that links the disbursement of aid to the performance of health services at the village level in Indonesia and find that linking aid to performance improves health indicators.[3] In the context of this literature, our paper provides the first field comparison of monetary and non-monetary incentives and how these interact with motivation for public services delivery.

The rest of the paper proceeds as follows. Section 2 describes the context, data and research design. Section 3 discusses the identification strategy. Sections 4 and 5 present the findings, and section 6 concludes with a discussion of costs and benefits of the different schemes and the external validity of our findings.

[3]Related research examines the effect of salary levels on selection into the health sector and performance (Propper and Van Reenen 2010; Dal Bó et al. 2011).

2 Context, Data and Research Design

2.1 Context

The field experiment was run in collaboration with the Society for Family Health (SFH), a public health organization based in Lusaka, Zambia. The experiment was embedded in SFH's new program for HIV prevention through the distribution of female condoms by hair salons. Our setting is representative of many health delivery programs in developing countries, where embedded community agents are called upon to deliver services and products, and where effective incentive design remains a significant challenge (Bhattacharyya and Winch 2001; Mathauer and Imhoff 2006). We collaborated with SFH closely at each stage of the program, including salon selection, training, incentive design and monthly restocking visits, for one year, from December 2009 to December 2010.[4] To ensure behavior was not affected by experimenter effects, we designed the experiment to fit within SFH standard procedures and agents were not aware that they were part of an experiment.[5]

In the program under study, hairstylists were identified as ideal agents for the delivery of this health service, both because the familiarity between the stylist and the client creates the potential for successful targeting of female condoms to "at risk" customers, and because during the period that a client is in the salon, he or she is a captive audience, allowing the stylist to provide information about HIV prevention generally, and specifically about the female condom. Finally, hair salons are numerous and distributed throughout Lusaka. Our census of salons, implemented as part of the research design, found just over 2,500 hair salons, serving a population of about 2 million (2,198,996, according to the 2010 Census of Population and Housing for Zambia).

In this context, the agents' choice variable is the level of effort to devote to the diffusion of information about HIV and to the sale of female condoms. Since this product is new and unfamiliar to customers, the agents must exert effort in explaining the female condom's proper use and benefits in order to persuade customers to make a purchase. For repeat customers, the hair stylists have the opportunity to follow up in order to encourage repeat use and troubleshoot any barriers to future purchase. Effort is costly in terms of forgone time spent discussing other topics that might be either more enjoyable or lead to the sale of other products available in the salon, such as clothes or hair products. Promoting female condoms has a strong pro-social component, since the use of condoms creates positive externalities for society at large. Condoms are an effective means of preventing the spread of HIV/AIDS and Zambia has one of the world's highest adult HIV-prevalence rates at 14.3 percent (Government Republic of Zambia 2010). Stylists are aware of the pro-social nature of the

[4]Female condoms are embraced by many in the public health community as the only female-controlled tool for HIV/AIDS and other STI prevention (PATH UNFPA 2006). Young, married women are the fastest growing demographic infected with HIV (UNAIDS et al. 2004). Adoption rates for female condoms are higher in Sub-Saharan Africa than in most parts of the world and earlier work in Zambia indicates that both men and women have expressed interest in the female condom (HLSP 2007).

[5]The experiment is a "natural field experiment" in Harrison and List's (2004) taxonomy, in that all research activities were embedded in SFH's normal activities.

task because of extensive informational campaigns run by the Ministry of Health on the importance of condoms for HIV prevention.

The program has four stages: (i) SFH attempts to distribute invitation letters to a one-day training program for the sale of female condoms to 1,222 stylists; (ii) of these, 981 can be reached and receive the letter; (ii) of these, 771 accept, undergo training, find out which type of reward they can earn (if any), and choose whether to purchase condoms from SFH to sell in their salons; (iii) of these, 747 join, are required to purchase 12 packs at the subsidized price of 2000 ZMK (166 ZMK per pack) and are given a range of promotional materials, including posters and display units. Thereafter, dispensers or single packs can be purchased at 500 ZMK per pack, either during a monthly restocking visit by SFH representatives or by calling a toll-free number dedicated to the female condom program. These are standard SFH practices for the distribution of health products.[6] The retail price is set at 500 ZMK for a pack of two condoms, which is the same price as the male condom.

2.2 Data

Our sample consists of the 771 stylists who participated in the training program and were exposed to treatment. Our main outcome variable is sales performance. Our preferred measure of sales is the number of packs each stylist restocks from SFH over the study period. Restocking is precisely measured from SFH inventory data and checked against invoices signed by the agents upon purchase. Restocking is mechanically correlated with customer sales, as there is no reason for agents to buy stock if they do not plan to sell it. Most importantly, restocking is the performance measure used to compute financial and non-financial rewards. Since the latter are not paid on the 12 packs agents were required to purchase at training, these 12 packs are excluded from our restocking measure. Table 1 shows that, on average, agents restock 9 packs, and the median is 0; namely, more than half of the agents do not purchase condoms from SFH other than at training. The standard deviation is 18 packs, indicating a fair amount of variation in performance. The sales data illustrate that the demand for female condoms is low, but that some agents manage to overcome this.

Our alternative measure of performance is calculated by SFH sales representatives, by subtracting the hairdresser's stock at month t from the sales representative's record of stock at t-1. Sales representatives measure stock each time they visit the salon by counting the number of packs on display and confirming with the stylists that no other packs are stored elsewhere. This variable suffers from measurement error due to the fact that unsold packs might not be visible to the SFH representative and/or hidden intentionally. Despite this potential for errors, the correlation between

[6]SFH representatives were instructed to stop attempting to visit stylists who could not be found for three consecutive visits, i.e., three consecutive months. By the end of the experimental year, 218 salons fell in this category. These stylists, however, were still formally enrolled in the program, and they could have called the toll-free number to resume the visits or restock condoms and are included in the sample throughout with sales of zero for each restocking visit.

the two measures is 0.92. Table 1 shows that the average calculated sales are 13.9 packs. The discrepancy between the two measures is due to the fact that calculated sales includes the 12 packs purchased at training and that it is likely to be biased upward, as every pack the sales representative cannot see in the salon is counted as sold.

In addition to sales performance, table 1 reports four variables collected by SFH sales representatives to proxy for the agents' sales effort: (i) the quantity of promotional materials displayed in the shop, such as posters and "sold here" signs (mean 2.26, sd .9); (ii) the probability that the stylists fill in their logbooks as instructed (47% of them do); (iii) the sales representatives' subjective evaluation of the stylists' interest in selling and promoting the female condoms (mean of 2.15 out of 3); and (iv) their judgement of the stylists' attention level at the time of the visit (mean of 2.52 out of 3).

Panel B, table 1 reports descriptive statistics of the agents' and salons' traits. These are collected via a census survey administered to all stylists in Lusaka before the start of the program.[7] Panel B illustrates that the salons in our sample are evenly split between hairdressers (only female clients) and barbers (only male clients) and a small minority (8%) caters to both men and women. 88% of the sample salons is located near a bar, which is a strong determinant of demand for condoms. 23% of the sample agents are Catholic, which might be a hindrance to selling condoms, as these were not approved by the Catholic Church at the time of the experiment. Salon size is typically small: the average salon has 1.75 employees and the median is 2. The average number of trained salons in the same area, a proxy for competition in the market for female condoms, is 4.5. A substantial share of agents (27%) sell products in their salons, suggesting that at least some of these agents already have some experience acting as sale agents. Panel C, however, shows that none of the agents has experience selling health products and male condoms, which could be substituted for by female condoms.

Panel B also shows that 19% of agents in the sample have "low socio-economic status", which encompasses those who do not speak English or have not completed primary education. In the absence of a reliable measure of wealth, these are the best proxies of socio-economic status in our setting. The next variable aims to measure the main source of motivation for the agents' day-to-day job. We asked stylists to identify what they enjoy most about their job among: "making money", "being own boss", "making people look nice", "being connected to the community", and "other".

[7] To minimize interference with the normal management of the condom distribution program, the survey was presented as a research activity and not linked in any way to the condom distribution program. The census was carried out from July to September 2009; the survey lasted for an average of 35 minutes. Two data collection teams worked concurrently. The first team consisted of scouts responsible for locating all salons and collecting GPS data. The second team then visited the shop and carried out the interview. Questions regarding the business included the type and quantity of equipment owned (mirrors, chairs, roller trays, dryers, etc.), the number of employees, the number and type of clients, the nature and prices of offered services and products, the monthly revenues and profit, and the time since opening. Questions about the manager included demographics, the stylist's peer network, employee status in the salon, monthly earnings, length of employment/ownership, other-regarding preferences/attitude, and living conditions.

The share of stylists choosing each of these are 35 percent, 6 percent, 44 percent, 14 percent and 1 percent, respectively. To measure work motivation we generate a dichotomous variable, coding the first two options as extrinsic and the second two as intrinsic.[8] 58% of agents report being intrinsically motivated, according to this definition.

To elicit an incentive-compatible measure of pro-social motivation toward HIV causes, we designed a contextualized dictator game, which was implemented by SFH personnel during the training program.[9] Agents were told that, in addition to the training show-up fee (40,000 ZMK), each of them would receive 12,500 ZMK, which they could keep for themselves or donate, in part or in full, to a well-known charity in Lusaka that provides care to HIV/AIDS patients, including antiretroviral treatment.[10] The amount donated is taken as a proxy for the agents' motivation for the cause. Since this is likely to be correlated with the agents' wealth, it is always used together with asset- and socio-economic-status measures in the analysis that follows. It is important to acknowledge that while donations may be higher than the individuals' truly-preferred amounts, because of social pressure (DellaVigna et al. 2012), the measure is still valid for our purposes as long as ranks are preserved so that more motivated agents donate more than agents who are less motivated, even if they all donate more than they would in the absence of social pressure. The average donation in the dictator game is 5,728 ZMK; that is 45% of the total endowment. More importantly for our purpose, the measure exhibits considerable variation, as its standard deviation is 3,744 ZMK.

Finally, panel C reports other agents' traits that, while not used directly in the analysis, can be helpful to put the experiment and the results in context. The weekly income of the average salon is 332,569 ZMK; 94% agents are literate and can thus easily read SFH's invitation letter and further communications, and 85% are literate in English.

[8] Results are robust to alternative ways of coding the baseline survey responses, for instance by coding "being own boss" and "other" as separate motivations.

[9] Previous work by Lagarde and Blaauw (2013) on South African nurses shows that the dictator game is effective in identifying socially-motivated workers. They found that student nurses who were more generous in an adapted dictator game to anonymous patients were more likely to choose a rural hardship post, where both social impacts and personal sacrifices are larger.

[10] Specific instructions for the game were scripted and read aloud. The script read: "We have recently received additional money for today's training. As a consequence we have sufficient funds to give each of you an additional 12500 ZMK. [This was in addition to the 40,000 show up fee]. You can choose how much of this sum to keep for yourselves and how much to donate to Our Lady's Hospice, a local charity that provides palliative care that includes offering ART (antiretroviral therapy) for their HIV patients. If you wish to donate, please put your donation in the envelope provided with this form [form has pre-printed ID number on it] and drop it in the collection box. Note that the amount you donate is totally up to you: you can give nothing, part of the 12,500 ZMK, or the entire thing. The amount you contribute will be kept completely confidential. We will give you a few minutes to think about it. When you've taken a decision, please drop your envelope in the box at the front." While instructions were being read, the helpers distributed identical pre-arranged packets of 12,500 AMK in small bills to each participant. While the need to collect individual measures of altruism obviously prevents us from guaranteeing full anonymity, the design ensured that individual choices were not observable by other participants or by the training personnel. After receiving the money, stylists were guided one at a time to one of five booths where they counted the sum and separated the amount they kept from the amount they donated. The bills donated were placed in an envelope and sealed before leaving the booth. Each participant then deposited the envelope in a box sitting in front of the room, specially designed for this purpose.

2.3 Research Design: Treatment Groups

Agents are randomly assigned to one of four groups. Agents in the **control group** are recruited as volunteers and receive no incentives, financial or otherwise.

Agents in the **large financial-margin treatment** receive 450 ZMK for each condom pack sold, a 90 percent margin over the retail price. 450 ZMK is the highest incentive-compatible margin, since agents would have the incentive to buy and dispose of the condoms if the reward were larger than the purchase price. To put these numbers in context, consider that the average stylist in our sample charges 3000 ZMK for a haircut. Whether devoting time to condom sales is more profitable than devoting time to cutting hair therefore depends on whether stylists can sell a pack in less than 1/6th of the time it takes them to do a haircut, other things equal. Thus the power of the rewards depends on other things that affect sales, including demand.[11]

Agents in the **small financial-margin treatment** receive 50 ZMK for each condom pack sold, a 10 percent margin over the retail price. 50 ZMK is the smallest bill commonly in circulation, making this the smallest payment that is easily implementable. The comparison of the two financial incentive treatments allows us to test whether financial incentives crowd out motivation when they are low-powered, as found in Gneezy and Rustichini (2000).

Agents in the **non-financial reward (star) treatment group** receive a star for each condom pack sold. These agents are provided with a thermometer display, akin to those used in charitable fundraisers. Each sale is rewarded with a star stamped on the thermometer, which is labelled as measuring the stylist's contribution to the health of their community. The thermometer display is designed to create a visual link between packs sold and health outcomes, making social impact salient (Grant 2007) and effectively rewarding stylists for marginal contributions to the cause. In addition, stylists were told that all those who sell more than 216 packs over a year would be awarded a certificate at a ceremony.

Three points are of note. First, rewards are a function of the number of condoms each agent restocks every month from SFH's invoice data. Rewards are not paid for the packs purchased at a subsidized price during the training program.

Second, SFH representatives visit each salon once per month[12] and pay rewards (financial and non-financial) based on the number of packs the agent bought during the previous restocking visit. This ensures that rewards can be computed at headquarters and that the sales representatives carry

[11] We note that agents in the large financial-margin treatment face a lower marginal cost (50 instead of 500) and can, in principle, boost sales by reducing the price. This incentive is common to all sales-based bonuses and quota schemes, i.e., sales people can increase sales by passing some of their reward to customers. This practice is not detrimental to the principal as long as they want to maximize sales revenues. We collect data on prices to test whether agents implemented this strategy.

[12] Five full-time sales representatives were trained to carry out visits and they rotated between salons and treatments. Restocking visits lasted approximately one hour, during which sales representatives followed a detailed script and recorded both observational and survey data. Besides collecting data, representatives answered queries about the program, distributed promotional materials, allowed the stylists to restock and handed out incentive payments.

the exact quantity of rewards to give at each visit.[13] [14]

Third, the design of the non-financial reward scheme was driven by the need to balance two equally important considerations: realism and comparability with the financial incentives schemes. We thus included a commonly observed feature of non-financial rewards (the certificate to top performers) while ensuring that agents in all treatments earn a reward for each pack sold. Therefore, at low sale levels, financial and non-financial incentives have the same linear structure, at high sale levels the non-financial scheme has an additional lump sum benefit past a given threshold. Whether this difference can drive differences in performance is a matter for empirical analysis.

2.4 Research Design: Randomization

Assignment to treatment is randomized at the neighborhood level with buffer zones between neighborhoods, so that all agents in the same neighborhood are assigned to the same treatment and salons' neighbors are either in the same treatment or not part of the program. To implement the design, we first conducted a census of all hair salons in Lusaka, collecting GPS coordinates and numerous salon and stylist characteristics. We then imposed a grid on the GPS-mapped locations of the salons, to divide the city into equal geographical areas of 650 by 650 meters each. We excluded a buffer of 75 meters on all sides of the grid cell, resulting in at least 150 meters between salons in adjacent areas. The resulting areas, each measuring 250,000 square meters, served as the unit of randomization. Salons located in buffer areas were not invited to join the program. The final sample for randomization consists of 205 distinct neighborhoods, containing 1,222 hair salons.[15]

To increase power, we balance on a vector of variables that are likely to affect condom sales. These are: salon type (hairdresser or barber); salon size (proxied by the number of employees); whether the salon is located near a bar (a proxy for condom demand); the number of salons in the

[13] Delaying the delivery of rewards by five weeks may lower the value of the reward if stylists have high discount rates, though stylists making regular sales receive rewards each restocking visit. Though restocking decisions are offset by five weeks from incentive delivery, the different incentive treatments do have the potential to influence the impact of liquidity constraints on restocking. Specifically, stylists in either of the financial incentives may have more cash on hand after the delivery of incentives from restocking during the previous visit. Sales agents elicited restocking decisions before incentives were handed out to mitigate this problem. However, if stylists changed their mind about restocking after receiving incentives, they were allowed to purchase more. We record these restocking decisions separately. Stylists in the high financial-margins treatment do not change their decision significantly more than stylists in the volunteer control, which suggests that liquidity constraint differences do not have a meaningful effect on restocking. Stylists in the star reward treatment, on the other hand, do significantly increase their restocking decision after receiving their reward, relative to the volunteer control group.

[14] At the end of the restocking session, all agents in the control group were told: "Now, I have good news for you today. Because of your hard work and great sales performance in the last month, you have potentially protected... [# of packs x 2]... sexual intercourses. You have therefore helped your clients protect themselves against STIs and unplanned pregnancies." Agents in the reward treatments were told "Now, I have good news for you today. Because of your hard work and great sales performance in the last month, you have earned a reward of ..... (Kwacha or stars). In addition to that, you have potentially protected ..... [# of packs x 2] sexual intercourses. You have therefore helped your clients protect themselves against STIs and unplanned pregnancies."

[15] Salons/shops that reported planning to close/move in the next six months were excluded from the sample, as were neighborhoods that contained only one salon.

same cell; the agents' total assets; and whether the agent sells other products in their salon. Randomization is implemented via the minmax t-stat method for the vector of balance variables across 1,000 random draws. Figure 1 illustrates the outcome of the randomization. Table A.1 presents the means and standard deviations of agents' and salons' characteristics in each treatment, together with the p-value corresponding to the F-statistic from a test of significance for each treatment pairs and the largest normalized difference across treatment pairs. All normalized differences are small and out of the 66 tests, 7 have $p < 0.05$, corresponding to three variables that differ between two or more treatments. These are: (i) the share of mixed salons, i.e. salons that serve both men and women, which is higher in the star treatment; (ii) the share of stylists with low socio-economic status, which is lower in the control group, and (iii) the share of stylists who report "making people look nice" and "being connected to the community" as the factors they enjoy the most about their job over "making money" and "being my own boss", which are lower in the star treatment and higher in the control group. All of our specifications include these variables and we test whether the response to incentives differs according to these variables in section 5.

3 Identification

To evaluate the effect of different incentive schemes on sale performance we estimate:

$$y_{ic} = \alpha + \sum_{j=1}^{3} \delta_{0j} treat_c^j + X_i \eta_i + u_{ic} \tag{3.1}$$

where y_{ic} measures condom sales by agent i located in area c over the year. Our main measure of sales performance is the same as that used to compute rewards, that is, the number of packs each stylist buys from SFH over the study period, excluding the 12 packs purchased at training. For robustness, we also estimate (3.1) using sales as calculated by SFH sales representatives by subtracting the hairdresser's stock at month t from the sales representative's record of stock at t-1. Sales representatives measure stock each time they visit the salon by counting the number of packs on display and confirming with the stylists that no other packs are stored elsewhere. This measure includes the 12 packs purchased at training.

$treat_c^j$ denotes the three treatment groups and X_i is a vector of agents' characteristics that can be correlated with sales. These include: salon type (barber or hairdressers) and size, stylist's sale experience, religion, socio-economic status and wealth, and motivation for the cause. The outcome variables and stylist- and salon-level characteristics are reported in Table 1.

Errors are clustered at the level of the randomization unit, the geographical grid-cell area c, throughout. We estimate equation (3.1) on the entire sample of stylists who came to training and hence were exposed to treatment. Since agents choose whether to participate in the program after

learning about incentives, the coefficients δ_{0j} capture the effect of incentives on sales performance through both the margins of selection and effort. In this setting, however, the role of selection is limited since almost all the agents who were exposed to treatment joined the program. Section 4 presents detailed evidence on this issue.

The coefficients δ_{0j} measure the causal effect of the treatments on sales performance under the identifying assumption that $treat_c^j$ is orthogonal to u_{ic}. This notwithstanding, the identifying assumption fails if the decision to participate in the training program is not orthogonal to treatment, or if there are spillovers between treatments. We discuss these in turn below.

3.1 Participation decision

The randomization algorithm yields a sample of 1,222 hairstylists to be invited to the one day training program and subsequently, to sell condoms. SFH representatives managed to deliver the invitation letter to 981 stylists. The letter, reproduced in appendix figure A.1, stressed both private and public benefits of the program. In particular, the letter suggested that joining the program might attract new customers to the salons and might help the community by facilitating HIV prevention. In the case of multi-stylist salons, the invitation is extended to the person responsible for the management of the salon, who is either the owner or the general manager. To attract the largest possible number of agents and ensure a representative sample, stylists are offered 40,000 ZMK (USD 8) to attend the one-day training. This is over 13 times the average price of a haircut and is therefore likely to exceed the stylists' expected earnings for a weekday. Using information on self-reported earnings, 40,000 ZMK corresponds to 69 percent of weekly earnings for the median salon.

Of the 981 stylists who received the invitation letter, 771 attended the training, perhaps as a result of the generous show-up fee and/or the financial and social benefits from joining the program, as stated in the letter. During training, stylists are provided with information on HIV/AIDS, female condom promotion, basic business skills, and program details, including the randomly assigned compensation package.[16]

Regardless of the high participation rate, the identifying assumption fails if the treatments affect selection at either stage. However, since stylists were not informed about treatments until the end of training, selection ought to be orthogonal to treatment. Appendix table A.2 reports the estimates of

$$p_{ic} = \alpha + \sum_{j=1}^{3} \theta_{0j} treat_c^j + X_i \eta_i + \varepsilon_{ic} \tag{3.2}$$

[16]The training took place between October and December 2009 and lasted for 40 days, running from Monday through Thursday for 10 weeks, with a maximum of 50 stylists attending in a single day. Training sessions were staggered and balanced across treatment groups, so that the timing of the training did not vary systematically between treatments.

where p_{ic} is an indicator variable equal to 1 if the agent receives the invitation letter in columns 1 and 2, and an indicator variable equal to 1 if the agent chooses to attend training in columns 3 and 4. X_i is a vector of agents' characteristics that can be correlated with the participation decision. Reassuringly, the estimates in table A.2 clearly show that the participation decision is orthogonal to treatment: all coefficients θ_{0j} are small and not significantly different from zero.[17]

3.2 Spillovers

The identifying assumption fails if, because of spillovers, the control group is not a proper counterfactual for how agents in the treatment groups would have behaved in the absence of treatment. This might be the case if, for instance, agents in the control group change their behavior as a result of knowing that other agents have been offered rewards. Four design features were employed to minimize the risk of spillovers across treatment groups.

First, we created a 150-meter buffer zone between each geographical area in which salons are located to ensure that each agent neighbors either other stylists in the same treatment group or stylists who are not part of the program. While the research design ensures that all stylists in the same geographical area are assigned to the same treatment, this precaution can be undone by stylists relocating after randomization is carried out. Relocated stylists were allowed to stay in the program only if they moved within the same geographical area or to a new area with the same treatment as their original assignment. [18]

Second, stylists attended the training with other stylists belonging to the same treatment group. Third, the enumerators who delivered the invitation letters were, themselves, unaware of which training day pertained to which treatment. Fourth, the program was designed to appear similar across treatment groups to an outside observer. Most importantly, the sale price was identical across treatments and all stylists received the same promotional materials, which included aprons, "sold here" signs, t-shirts and different types of posters. The sole exception to this rule is the thermometer poster, which was given only to stylists in the star treatment.

To assess the potential for spillovers through the stylists' social network, our baseline survey asked respondents about their relationships with other stylists in Lusaka. Reassuringly, the median stylist reported only one connection, whether a relative, friend or acquaintance, with another stylist in the city. To monitor the evolution of this variable over the course of the program, we collected

[17] The decision to attend training is correlated with some individual characteristics, such as gender (barbers are more likely to attend), and self-reported donations to HIV-related causes. Stylists who attend training are five percentage points more likely to report giving to HIV charities. While this is in line with the theoretical literature that suggests that agents in mission-driven organizations share an interest in the mission, the magnitude of the difference between participants and non-participants is small, as the vast majority of invited stylists chooses to participate.

[18] Only 12 cases occurred in which the salon moved and remained in operation and staffed by the stylist involved in the research project. In 7 of these cases, the salon relocated within the same treatment cell. Three of the cases involved movement into a buffer area and the remaining 2 cases involved relocation to a different treatment. These salons were dropped from the study and all subsequent restocking observations are recorded as zeros.

information on new connections with other stylists during each monthly visit. During the first four months of the program, 60 to 80 percent of stylists reported at least one new connection with another stylist in the city. After the fourth month, very few new connections were reported. Over 90 percent of the new acquaintances reported during the first four months met during the training and are therefore in the same treatment group. Finally, to detect spillovers and identify the stylists who might be affected by them during the course of the experiment, we asked sales representatives to note all questions and complaints at every monthly visit. In over 7,000 restocking visits, only one stylist asked about different incentive schemes.[19]

While these three pieces of evidence are reassuring, they cannot completely rule out that agents in one treatment effectively responded to not being assigned to another. In the next section, we will exploit variation in treatments of neighboring areas to assess the empirical relevance of this concern.

4 The Effect of Incentives on Sale Performance

We begin by estimating equation (3.1) to evaluate the effect of the three experimental reward treatments on overall sales performance. Throughout, we report estimates of δ_{0j}, with and without a vector of salons' and agents' characteristics that can affect the willingness or ability to sell female condoms, and therefore explain some of the variation in sales.

As discussed above, we estimate equation (3.1) on the entire sample of stylists who came to training and hence were exposed to treatment. This implies that the coefficients δ_{0j} capture the effect of incentives on sales performance through both the margins of selection and effort. The next two sections provide evidence on the relative importance of these margins.

4.1 Selection

Of the 771 stylists attending the training, 747 (97%) joined. Not surprisingly, columns 1 and 2 of appendix table A.3 show that incentives had no impact on the decision to join. Stylists could also choose to quit during the course of the experimental year at no cost. Only 58 stylists (7 percent of those exposed to treatment) did so; of these, 53 never made a sale. The effect of the incentive treatments on the choice to select out is small for all treatments and significantly different from zero ($p = 0.077$) only for agents in the small financial-margin treatment when accounting for individual and salon characteristics, as shown in column 4 of appendix table A.3.

Overall, only 10 percent of the 771 stylists who were exposed to treatment selected out of the program either at training or later during the year, and the incentive treatments did not affect

[19] Most questions regarded queries that originated from customers on the characteristics of the product. The most common complaint was that the condoms were difficult to sell.

either selection decision. This implies that the coefficients δ_{0j} capture the effect of incentives on sales through effort rather than through selection.

4.2 Sales

Figures 2 and 3, and table 2 show the effect of incentives on average sales and at different points of the sale distribution.

Beginning with average sales, figure 2 shows that there is a striking difference between stylists in the star treatment and all others. Agents in the star treatment sell twice as many packs over the year. This is confirmed by the estimates in columns 1 and 2 of table 2. Four findings are of note. First, agents in the star treatment sell 7.66 more packs, which is over twice as many packs as stylists in the control group. This result is robust to the inclusion of stylist-, salon- and area characteristics and is not driven by outliers in the star treatment group.[20] Second, neither financial-incentive treatment affects sales.[21] Both coefficients are substantially smaller than the coefficient on the star treatment and are not significantly different from zero. The null hypothesis that the effect of either financial treatment is equal to the effect of the star treatment can also be rejected at the 1 percent level or lower. The evidence thus casts doubt on the relevance of a specific form of crowding-out effect, namely that crowding-out only dominates when financial rewards are low-powered, so that small rewards reduce performance while large rewards increase it (Gneezy and Rustichini 2000).[22]

While the sales level of the average stylist is low, the difference between treatments is large in aggregate. The estimates in column 2, Table 2 imply that if all of the 771 agents had been offered non-financial incentives, they would have sold 22,496 condoms; that is 11,810 more condoms than the counterfactual scenario, in which they were all hired as volunteers. Had they all been offered

[20] We obtain similar results if the top 1 percent of sellers are dropped from the sample.

[21] We note that agents in the large financial-margin treatment face a lower marginal cost (50 instead of 500) and could, in principle, have boosted sales by reducing the price. While this does not invalidate the identification of the effect of incentives on sales performance, it changes the interpretation of the effect of incentives on effort. We do not observe agents choosing this strategy in equilibrium. Our end-line survey shows that only four stylists reported ever selling a pack at a price lower than 500 ZMK, and none of them were in the large financial-margin treatment. This, of course, does not rule out that the agents tried lowering the price, but this had no effect on sales, which is consistent with demand for this product being inelastic. The stylists' ability to take advantage of the low elasticity to increase price was limited by the fact that the same product was available from other outlets, e.g., chemists and drugstores, at 500 ZMK. Unbranded versions were available free of charge from health clinics. In our focus groups, both stylists and customers report some sales at 1000 ZMK. We note that at this price, stylists in the control group and star treatment also get a margin per pack sold (500 ZMK), but this is considerably lower than the margin received by stylists in HPFT (950 ZMK), so that the ranking of treatments in terms of incentive power is unchanged as long as stylists in all treatments are able to sell at the same price.

[22] To be specific, our 95-percent confidence interval on low financial incentives relative to the volunteer control group allows us to rule out negative coefficients greater than 2.2 packs, or 0.12 standard deviations. As a comparison, Gneezy and Rustichini (2000) find a crowding-out effect of 0.54 standard deviations associated with paying a low financial incentive, and Ariely et al. (2009) find a 0.16 standard deviation decrease associated with public payment of financial incentives. We are therefore able to rule out crowding-out effects of the magnitudes found in these two papers at $\alpha < 0.025$.

small or large financial margins, they would have sold 11,938 and 12,504 condoms, respectively.[23]

Third, we find that our experimental measure of motivation is correlated with sales and the effect is large: agents who donate more than the median amount to the HIV charity sell 3.36 more packs, which is equal to 44 percent of the effect of star rewards and almost 50 percent of the baseline mean of 6.96 in the control group. The fact that the donation in the experimental game predicts sales reassures us that social pressure to donate, if any, did not mask actual differences in motivation. To allay concerns that the donation measure captures differences in wealth, the regression includes a measure of the stylist's own assets. This is correlated with the value of donation, as expected, but not with sales. Since self-reported assets might be measured with substantial noise, we also use information on whether the agent has completed primary school and whether they speak English, which are good proxies of socio-economic status in our setting. This measure is also correlated with donation but not with sales. Fourth, the following agent characteristics are correlated with sales: barbers sell 3.32 more packs, possibly reflecting the fact that men are in charge of contraceptive choices in our setting, promoters with previous sales experience sell 5.18 more packs and Roman Catholics sell 3.65 fewer packs. The effect of the star treatment is thus larger than the effect of any personal characteristic.

Fourth, column 3 shows that all results are robust to using sales calculated by SFH representatives as the outcome variable. Recall that our main outcome variable does not include the 12 packs the agents purchased at training, as all agents were required to do so and these are not counted for the computation of rewards. In contrast, the calculated sales measure includes these 12 packs and its mean is correspondingly higher. The qualitative results are unchanged, as agents in the star treatment sell more than agents in any other treatment group. Consistently with the fact that the calculated sales variable is measured with error, both the estimated star-treatment effect and the effect of other agents' traits (pro-social motivation, type of salon, religion, sales experience) are somewhat smaller but precisely estimated throughout. Table A.4 shows that results are also robust to winsorizing (at 90% and 95%) alternative samples and SFH representatives fixed effects.

Figure 3 illustrates the distribution of sales in the four groups. The distribution exhibits bunching at 0, 12 and 24 packs, probably due to the fact that while stylists could purchase one pack at a time from SFH, buying one dispenser (12 packs) saves on transaction costs. Overall, 62 percent of stylists sell no packs other than those purchased at training, 22 percent sell between 0 and 12,

[23]To express these differences in a more relevant metric for comparing public health outcomes, our estimates imply that offering non-financial incentives to all agents would have saved 112 disability-adjusted life years (DALYs), compared to 53 DALYs in the counterfactual volunteer scenario, 60 DALYs with small financial margins, and 62 DALYs with large financial margins. This calculation is based on a model calibrated for Zambia by Population Service International (PSI 2012). The cost per DALY saved by enrolling all 771 agents in a single contract type, including both fixed and variable costs, is USD 2,078 in the volunteer contract group, USD 1,861 in the low financial scheme, USD 1,785 in the high financial scheme and USD 1,003 in the star reward group. To put this cost in context, Garber and Phelps (1997) estimate the value of a DALY at approximately twice annual income. The per-capita annual income in Zambia in 2010 was USD 1,020, so the cost of the star reward treatment compares favorably to the value of the health benefits it generates.

and 16 percent sell 24 or more.[24] Conditional on selling any, stylists sell an average of 24 packs in addition to the 12 purchased at training. Figure 3 shows that the treatment effects differ on the extensive- and intensive margins. In particular, 47 percent of agents in the star treatment sell at least one pack besides those purchased at training, compared to 35 percent in the other groups. The figure also illustrates that the average difference between the star treatment and the other three groups is driven by agents who sell more than 12 packs. This is confirmed by the estimates in columns 3, 4, and 5 of table 2.

Column 4 of table 2 shows that the likelihood of selling at least one pack in addition to those purchased at training is 12 percentage points higher for agents in the star treatment; this represents a 33-percent increase over the mean of the control group. Agents in the high- and low financial-margin treatments are equally likely to sell at least one pack as agents in the control group. Columns 5 and 6 show that the difference across treatments is stable at different points of the distribution in absolute value, but it increases in proportion to the mean level in the control group. Promoters in the star treatment are 13 percentage points more likely to sell 12 or more packs, which is 39 percent more than stylists in the volunteer treatment, and 10 percentage points more likely to sell 24 or more, which is 80 percent more than stylists in the volunteer treatment. Promoters who are offered financial margins, either large or small, do not perform differently than stylists in the volunteer treatment. All coefficients are precisely estimated and very close to zero.

A possible explanation for why financial rewards are not effective at promoting sales is that even in the high margin treatment, earning potential was low because of low demand for the product. Even stylists in the top quintile of sales and in the high margin treatment made only 3.5 percent of self-reported annual earnings from condom sales. While field experiments on performance typically analyze the effect of incentives that account for a substantial share of earnings, two other experiments use comparably small incentives and still find large positive effects. Chetty et al. (2012) show that offering academic referees a USD 100 gift card to complete their reports on time decreases median survival times by 43 percent. The value of the incentive is equivalent to 0.1 percent of the average assistant professor salary in the US.[25] Goette and Stutzer (2008) show that offering a lottery ticket worth USD 4.3 increases blood donations by 12 percent among a large sample of potential donors in Switzerland.

More importantly for the interpretation of our results, low demand cannot explain why non-financial rewards are effective, unless the non-financial reward treatment affects demand directly. Our research design allows us to rule out that the star treatment increases sales by affecting demand, an issue we will return to in section 6.

[24] Results are robust to estimating Tobit models.

[25] Data from Scott and Sigfried (2011) refer to the mean salary of assistant professors in PhD-granting institutions. Mean salaries for associate- and full professors are USD 117,231 and USD 159,816, respectively.

4.3 Spillovers and timing

Before delving into the mechanisms that underpin our findings, this section presents evidence on two key issues for interpretation. First, we provide evidence that allays the concern that the estimated effect of the non-financial treatment might be contaminated by spillovers, namely by agents in other treatments reacting to not having been given stars. As illustrated in figure 1, some non-star areas border neighbor areas in the star treatments, whereas others do not. We exploit this variation to test whether the agents who are more likely to be affected by spillovers have higher or lower sales. Reassuringly, we find that being close to agents in the star treatment does not affect sales for agents in other groups, which casts doubt on the relevance of spillovers in our setting. Of the 586 salons not in the star treatment, 41 percent are located in areas adjacent to star treatment areas, specifically in one of the 8 areas bordering a star treatment area. The estimated treatment effect for being adjacent to a star treatment area is 1.30 (s.e. 1.39).[26]

Second, we provide evidence that the treatment effects are stable through time, thus ruling out that the aggregate effect of non-financial rewards on sales is due to the novelty of being offered star rewards, or similar forms of Hawthorne effects. To do so, we exploit the fact that the SFH inventory files contain the exact dates of restocking and estimate equation (3.1) in each month, using the same set of controls and clustering errors at the same level of the randomization unit as above. Figure 4 reports month-specific treatment effects. Two patterns are of note. First, the effect of the star treatment is positive and of similar magnitude in all months except the fifth, when it is close to zero. This might be due to the torrential rains in months 3 and 4 depressing sales, so that agents could not sell the stock purchased in those months and did not need to restock in month 5. The magnitude of the star treatment effect is somewhat higher in the first two months and above the mean of the control group in most months, implying that agents in the star treatment sell at least twice as many packs as agents in the control group at any given point in time. Not surprisingly, however, the effect on monthly sales is less precisely estimated than on yearly sales. Second, the effect of both large and small financial margins is close to zero in all months, suggesting again that the aggregate results do not hide substantial heterogeneity through time.[27]

The stability over time of the effect of non-financial incentives suggests that the effectiveness of

[26] Although the concern for spillovers might be stronger from the star treatment, given the visibility of the thermometer, we also check for spillovers from the financial margin treatments. Being in a cell adjacent to any financial margin treatment has no statistically significant effect on sales.

[27] We note that the observed pattern is consistent with agents in the star treatment exerting effort only at the beginning to establish a regular customer base, and sell to the same customers throughout the year. While this is not the only interpretation of the patterns, from the principal's point of view this is not less desirable than reaching new customers, but the interpretation of the effect of stars through time differs if this is the case. To shed light on this issue, we use the agents' reports on whether the customers to whom they sold female condoms had used them before. The share of sales made to customers who had never used a female condom is naturally higher in earlier months (80 percent in month 1) but remains substantial in later months (44 percent by month 10), suggesting that agents were reaching out to new customers throughout the program. More importantly for the interpretation of the treatment effects, the share of new customers does not decline faster for agents in the star treatment.

the star reward is unlikely to be driven by the prospect of qualifying for the ceremony. This can be inferred from the fact that, given the volume of sales, the threshold for being entitled to the ceremony (216 packs sold in one year) was unattainable for most agents. Indeed, stylists who sold at least one pack and who were assigned to star treatment sold on average 3.1 packs per month, and only one stylist managed to sell enough to qualify for the ceremony. Had the effect of non-financial incentives been driven by the ceremony component alone, it should have disappeared after a few months, as most agents realized the threshold was far beyond reach. The same logic suggests that the effect of the star treatment is not driven by the fact that agents in that treatment were motivated by career concerns, in the form of networking with high-ranking SFH officers at the ceremony, to gain employment with the organization.[28]

5 Mechanisms

The evidence in the previous section indicates that, in this setting, non-financial incentives are effective at increasing sales, whereas financial incentives are not. This section provides evidence on the mechanisms that underlie the treatment effects estimated above. Since the evidence in section 4 shows that the difference between treatments is stable throughout the duration of the experiment, the remainder of the paper will focus on aggregate year-long performance.

5.1 Agents effort vs. customer demand

While all stylists are given the same posters and other promotional materials, a key difference between the star treatment and all others is that only agents in the star treatment are given the thermometer, which provides a visible measure of the stylists' performance and their contribution to the program. Visibility could, in principle, lead to higher sales for a given level of effort through an advertising effect, or if the clients are altruistic vis-à-vis the stylists and buy packs to make them earn stars, or still, if the clients take it as a signal of the agents' type and buy packs because they share an interest in the mission.[29] Assessing whether stars result in higher sales because they encourage effort or increase demand is key for a correct interpretation of the findings and to derive implications for incentive design.

To this purpose, we first test whether agents in the star treatment behave differently along dimensions that are correlated with sales effort, as measured during the monthly visits. Table 3

[28] Stylists who participated in focus groups mentioned they were quick to realize the ceremony threshold was not attainable, but that nevertheless having a target and seeing how they progressed towards it through the stars motivated them to work in its own right.

[29] A related consideration is that the star treatment could have attracted more customers to the salon. We compare the change in the number of salon customers between the baseline and the end-line across treatment groups and find no significant differences. Specifically, the change in the reported number of regular customers between baseline and endline is small and imprecisely estimated in all treatment groups.

reports the estimates of equation 3.1 using effort proxies as outcome variables. We find that agents in the star treatment display 0.25 more materials (11 percent more than the mean of the control group), are 7 percentage points more likely to fill in their logbooks (15 percent more than the mean in the control group), and score 0.10 more points, or 1/7th of a standard deviation more, on the "interest" variable recorded by the sales representatives. Stylists in the two financial margin schemes do not differ from the control group for any of these three measures of effort. Finally, stylists in all treatments appear to be equally interested during the sales representative's demonstration. Overall, the results in table 3 indicate that, in line with the effect on sales, non-financial incentives promote effort on three out of the four dimensions that we can measure, while financial incentives do not.

Next, we test whether the star treatment changes customers' behavior, leading to higher sales. First, we survey 2,000 customers to assess directly whether they report being affected by the thermometer.[30] We ask customers whether they had seen promotional materials for female condoms in hair salons and, if so, to describe what that they had seen. Overall, 37 percent of the interviewees report having seen promotional materials. Of these, 92 percent had seen the promotional poster (which is the largest and most visible of the materials distributed), 36 percent had seen the "sold here" sign, and only 2 percent, or 15 people in total, report seeing the thermometer. Of these 15, 5 had previously used a female condom, but none had bought them at a hair salon. This casts doubt on the interpretation that the thermometer attracts more attention than the standard promotional materials, giving stylists in the non-financial treatment an advertising advantage.

Given the low sales volume, however, the customer survey might fail to capture the responses of the small subset of customers who are indeed affected by the thermometer. The second step of our strategy consists of distributing a placebo star reward treatment to a random sample of salons in the volunteer control group and the two financial treatments. In the 8th restocking cycle, we distributed placebo thermometers to a randomly-selected half of the salons not in the star treatment and standard promotional posters to the remaining half. The placebo thermometer looks identical to those given to stylists in the star treatment, except that the number of stars reflects average sales by all salons, rather than the individual salon sales. The effect of the placebo thermometer on sales gives us a measure of the effect of the star treatment through advertising, as salons in the placebo treatment look the same to an outside observer as salons in the star treatment.[31] Table 4 compares the effect of the placebo thermometer to that of the star treatment. Two comparisons are

[30] To interview customers we selected 16 dense Lusaka markets, four for each experimental treatment. Surveyors conducted random-intercept surveys with individuals in the markets by approaching every fifth individual entering through the main market entrance, and asked if they would be willing to answer a few questions. Once consent was obtained, we asked whether the respondent frequented a hair salon in the market where the survey took place and a very brief set of survey questions about demographics, familiarity with the female condom, sources of information, purchase behavior and own sexual practices.

[31] Note that the placebo thermometer does not allow us to rule out whether clients buy packs to help the hairstylist accumulate stars, because, by design, the number of stars in the placebo thermometer does not reflect individual sales. Our earlier finding that none of the clients who report having seen the thermometer buy condoms from the stylists casts doubt on the relevance of this mechanism.

of interest. Column 1 estimates treatment effects for all agents at the same point in time, that is in the visit round that follows the distribution of the placebo thermometer. The comparison is thus clear of time-varying factors that might affect sales in all treatment groups. Column 2 estimates treatment effects in the first period after the treatment was implemented. This is period 1 for the star treatment and period 9 for agents who received the placebo thermometer in round 8. This comparison is thus clear of factors, such as novelty effects, that might affect sales right after the treatment is implemented.

Table 4 shows that the placebo star reward has no effect on sales and its effect is significantly different from that of the star treatment. Columns 3 and 4 explore the possibility that the effect of the placebo star reward is biased downward because stylists might have unsold stock from which they might sell, and our measure of performance (restocking) fails to capture that. The results in columns 3 and 4 suggest that this is not the case. Overall, table 4 indicates that the thermometer is not an effective advertising instrument, casting further doubts on the hypothesis that non-financial rewards affect sales by changing customer behavior.

5.2 Pro-social motivation and the response to incentives

Results in table 2 make clear that both rewards and pro-social motivation affect sales performance. We now provide evidence on their interaction, namely on whether they reinforce or crowd each other out. To assess this, we allow the effects of incentives to be heterogeneous as a function of the agent's pro-social motivation and we estimate:

$$y_{ic} = \alpha + X_i\beta_i + \sum_{j=1}^{3} \delta_{0j} treat_c^j + \sum_{j=1}^{3} \delta_{1j} treat_c^j * \sigma_i + u_{ic} \tag{5.1}$$

where σ_i is the agent's donation in the adapted dictator game (whose level is included in the vector of stylist's characteristics X_i) and all other variables are defined above.

The results in column 1, table 5 indicate that both financial and non-financial incentives leverage pro-social motivation. The effect of non-financial incentives is large and precisely estimated only for motivated stylists. In particular, stylists who donate more than the median amount in the experimental dictator game and are assigned to the star treatment sell 10.0 (s.e. 3.2) more packs than the control group (low-motivated stylists in the volunteer group), while stylists assigned to star treatment who donate less than the median amount sell 4.3 (s.e. 2.9) more packs than do low-motivated stylists in the volunteer group. The p-value of the difference is 0.096. This implies that non-financial incentives *crowd in* pro-social motivation in our experiment.

Perhaps more surprisingly, the findings in table 5 indicate also that high financial margins appear to reinforce pro-social motivation; namely, the difference between the effect of high financial incentives on high- and low-motivated stylists is positive with a p-value of 0.026.

These findings contribute to a body of laboratory and field experiments on charitable giving (Ariely et al. 2009; Gneezy and Rustichini 2000; Lacetera et al. 2011; Mellström and Johannesson 2008) that test whether financial rewards crowd-out other sources of motivation. Most of these studies focus on social reputation, namely the possibility that financial incentives reduce the reputational gains from pro-social activities. In our setting, however, this channel is closed since the two financial schemes and the control group were designed to be observationally identical to an outside observer to minimize the risk of contamination via information spillovers. In particular, customers could not observe whether agents were receiving rewards for condom sales, and all condoms were sold at the same 500 ZMK price in all treatments. Since it is common practice for retail agents to receive a margin on the price of the goods they sell, the most likely inference from the customer's perspective is that all hairstylists were paid monetary margins, but we cannot pin down customers' beliefs in our setting (or, more germane for our analysis, hairstylists' beliefs about customers' beliefs about their motivation). More importantly, we would not expect differential inference about incentives across the volunteer and financial treatments, particularly since stylists in the volunteer control group have no way to credibly signal that they were not getting paid.[32]

Since it is unlikely that monetary incentives affected the agents' social image differentially across treatments in our setting, we are able to assess the degree to which they might crowd out an agent's internal sense of motivation. This could happen through a self-signaling mechanism by which the agents receive less "warm glow" because financial incentives make them re-assess their own motives for devoting effort to the task (Deci 1971; Andreoni 1990). Our findings suggest that this is not the case.

To provide further evidence on the interaction between treatments and other sources of agents' motivation, we use a self-reported measure of work motivation. Our measure, described in section 2.2 above, proxies for the importance of intrinsic motivation relative to extrinsic motivation. It is important to note that, in contrast to the donation in the experimental dictator game, this variable measures agents' motivation for their main job, which has no pro-social component. Results in column 2 reveal that the response to incentives is affected by the agents' motivation for their main job. Those who rank intrinsically-oriented motivations above other motivations and are assigned to the star treatment sell 10.48 (s.e. 2.99) more packs than those who indicate individual-oriented motivations and are assigned to the pure volunteer treatment. The interaction of intrinsic motivation and the star treatment (large financial rewards treatment) is positive but not precisely estimated, with $p = 0.134$ ($p = 0.144$).

[32] In addition, qualitative evidence from focus groups indicates no stigma attached to being paid for pro-social tasks, possibly because Zambia is a very poor economy, and that tasks seem more valuable if a donor, NGO or government is willing to pay for it. Customers reported that the price at which condoms were sold ruled out that stylists were being paid extremely well for performing the task, and that knowing that they were paid a margin similar to that paid for other products did not tarnish their reputation.

5.3 Heterogeneous responses by the value of financial rewards

To provide evidence on the mechanisms that drive the response to financial incentives, we test whether the effectiveness of financial incentives depends on their value for different agents. We exploit the fact that, under the assumption of concave utility, the same amount of money is more valuable for poor stylists. To proxy for socio-economic status we use information on the education level and English-speaking ability of the stylist, and classify as "low socio-economic status" the 19 percent of stylists in our sample who either do not speak English or have not completed primary education. In the absence of a reliable measure of wealth, these are the best proxies of socio-economic status in our setting. We estimate:

$$y_{ic} = \alpha + X_i\beta_i + \sum_{j=1}^{3} \delta_{0j} treat_c^j + \sum_{j=1}^{3} \delta_{1j} treat_c^j * \phi_i + u_{ic} \tag{5.2}$$

where ϕ_i measures socio-economic status (whose level is included in the vector of stylists' characteristics X_i) and all other variables are defined above.

Column 3 of table 5 shows evidence in favor of the hypothesis that financial incentives are effective when their relative value is higher, i.e. for low-socio-economic-status stylists. Compared to stylists in the control group (high socio-economic status in the volunteer group), low-socio-economic-status stylists sell 3.7 more packs when offered large financial margins and 4.9 more packs when offered small financial margins. Both effects are precisely estimated at conventional levels. This notwithstanding, non-financial incentives are more effective than financial incentives for all agents.

5.4 Heterogeneous responses by the value of non-financial rewards

In line with the previous test, we now test whether the effectiveness of non-financial incentives depends on their relative value. To do so, we exploit the fact that treatments were randomized at the neighborhood level and hence agents in different neighborhoods have a different number of peers; that is, agents in the same treatment group, in their vicinity. As the non-financial treatment enables stylists to make their sale performance visible to third parties, its effectiveness might depend on the number of peers who can see it. For instance, social prestige associated with stars or reputational gains from contribution to society might be higher when they can be shown-off to a larger number of people, or stylists might be motivated by wanting to outperform their peers, or encouraged by the effort of others dedicated to the same cause.[33] To shed light on the practical relevance of this mechanism, we allow the effect of treatments to vary with the number of potential peers in the vicinity of the stylists' salons; that is, the number of trained stylists in the same geographical area.

[33]SFH representatives' records from monthly visits indicate that, on average, the thermometer was publicly displayed in 43 percent of the star treatment salons and the literature on charitable giving provides evidence that donations are larger when they are visible to others (Soetevent 2005; Karlan and McConnell 2012).

By design, the randomization procedures ensure that the number of salons in each geographical area is balanced across treatments (see appendix table A.1). This, together with the fact that selection into training is orthogonal to treatment, implies that the average number of *trained* salons is balanced as well. The median (mean) number of trained salons in an area is 3 (4.5) with a standard deviation of 5, and none of the tests of equality of means between treatment pairs rejects the null. Reassuringly, the distribution of the variable is also similar across treatments, and no pairwise Kolmogorov-Smirnov test rejects the null of equality.

To evaluate whether the star treatment is more effective when the peer group is larger, we estimate:

$$y_{ic} = \alpha + X_i\beta_i + \gamma N_c + \sum_{j=1}^{3} \delta_{0j} treat_c^j + \sum_{j=1}^{3} \delta_{1j} treat_c^j * N_c + u_{ic} \tag{5.3}$$

where N_c is the number of trained salons in area c (or a dummy that equals 1 if the number of trained salons in area c is larger than the median), where the area is the unit of randomization and covers 250,000 square meters. The specification thus controls for salon density, which itself may affect sales, regardless of treatment. For instance, customer demand for condoms might be higher in areas with more salons because more customers transit through these areas, or lower if there are more alternative outlets. Also, stylists in denser areas might be more effective sellers because they face stronger competitive pressure. The coefficient γ captures these effects.

We report the findings in column 4 of table 5 and in figure 5. Column 4 of table 5 reports the effect of all three treatments separately, for salons located in areas with fewer than the median number of peers and those located in areas with a larger number of peers. We find that the coefficient of the star treatment is large and precisely estimated in the large peer group, as agents in that group sell 9.14 more packs than agents in the omitted category. The size of the peer group itself is uncorrelated with sales; this allays the concern that density captures other area-specific characteristics that are correlated with sales. Moreover, financial incentives are ineffective in both cases and this allays the concern that density captures area-specific features that make any form of incentives more effective.

We repeat the analysis using a continuous measure of the number of peers and we find similar results. The interaction coefficient between the number of peers and the star treatment ($\delta_{13} = 1.06$; $s.e. = 0.38$) is statistically and economically significant. The magnitude of the coefficient is such that the effect of stars increases by 5.3 packs (70 percent of the average effect estimated in table 2) for one standard deviation increase of the number of peers. Figure 5 reports the marginal effect of the non-financial treatment ($\delta_{03} + \delta_{13}$) evaluated at different values of N_c with 95 percent confidence bands. This shows that the effect of stars is positive throughout and precisely estimated when the number of peers is 5 or larger.

The findings support the idea that the non-financial treatment partly works by allowing social

comparisons; non-financial incentives are more effective when the number of potential peers is higher. It is important to note that this finding does not necessarily imply that stylists compete to collect stars; rather, stylists might be encouraged by the effort of others, or the ability to observe others' performances helps the stylists assess what is expected of them.[34] Indeed, stylists who participated in focus groups reported being motivated by showing off their own sales levels and viewing the sales levels of their peers, and also using the sales information on the thermometer to identify successful sellers to ask for sales tips. The finding that the star treatment was significantly more effective, the more dense the peer group, is robust to alternative sample restrictions, such as trimming at the 95th percentile.[35]

To corroborate our interpretation that the interaction between the number of peers and the star treatment captures the incentive effect of social comparison, we note that agents in areas with more trained salons are significantly more likely to display the thermometer in their salons. One standard deviation increase in N_c is associated with a 14 percentage-point higher likelihood of displaying the thermometer, a 23-percent increase from its mean value, and the correlation is precisely estimated. Crucially, for the interpretation of our findings, this is not driven by agents choosing to advertise more in denser areas; indeed the correlation between N_c and the likelihood of displaying other promotional posters or the number of other promotional materials is small and not statistically different from zero.[36]

6 Conclusions

We conduct a field experiment to provide evidence on the effectiveness of financial and non-financial rewards within health services delivery. We find that agents who are offered non-financial rewards ("stars" in this setting) exert more effort than either those offered financial margins (10% and 90%

[34]Further analysis, not reported, allows the effect of non-financial incentives to be heterogeneous, according to the stylists' motivation for the cause, the number of possible peers and the interaction of the two. The evidence favors the interpretation that the two mechanisms act independently; both high and low donors sell more when surrounded by more peers, but high donors sell more for any given number of peers.

[35]Further analysis, not shown, indicates that the distance between salons within the same neighborhood does not affect the effectiveness of the star treatment, presumably because neighborhoods are sufficiently small (500 meters by 500 meters).

[36]A second source of variation that might be associated with the utility weight of non-financial rewards is the variation in the number of salon employees. In contrast to money, stars are not divisible and cannot be attributed to the employee who made the sale, and the thermometer does not bear the name of any particular stylist working in the salon. A priori, a larger number of employees might be associated with a lower value of non-financial rewards if stylists free-ride on the effort of their colleagues or with a higher value, if group dynamics lead to encouragement and higher effort. To provide evidence on whether this mechanism is relevant in our context, we allow the effect of non-financial incentives to be heterogeneous as a function of the number of employees. In our sample, 49 percent of salons are operated by a single person, 34 percent have two employees, 12 percent have three and the remaining 5 percent have four or more. We find that the difference between financial and non-financial incentives is constant at different salon sizes, thus ruling out possible differences due to differences in divisibility. The power of this test is limited by the observed variation in salon size, as most multi-employee salons are quite small, but, in our context, we can rule out that the effectiveness of non-financial incentives is due to their non-divisibility.

commission on the suggested retail price) or those offered volunteer contracts, and generate higher sales of packs of condoms per year. Non-financial rewards elicit effort by leveraging the agents' pro-social motivation and by facilitating social comparisons among agents. While we implemented a specific type of non-financial reward, the general design principles are easily replicable and adaptable to other settings. Our rewards were a linear function of sales, which minimized discouragement or gaming effects typically associated with non-linear schemes. Moreover, rewards were made clearly visible to third parties, thus allowing social comparisons between different agents engaged in the same task, which proved effective at eliciting effort. Finally, they were awarded by a reputable and well-known organization, which might have contributed to their value.

We designed the incentive treatments to reward sales performance rather than usage, since sales performance can be precisely measured while usage cannot. It is nevertheless important to discuss the link between sales and usage, since the health impact of the treatments depends on the latter. We can provide two pieces of evidence indicating that customers indeed used the condoms. First, the stylists' logbooks, in which they are asked to record customer characteristics for every sale, reveal that by the end of the experiment 56 percent of buyers had purchased female condoms at least once before. This suggests the repeat customers used their previous purchases. Second, in line with this, 13 percent of respondents to our customer survey report using the condom.

The customer survey data also reveals that the effect of incentives on sales might actually underestimate the effect on usage. Indeed, while 16 percent of the respondents report receiving information on female condoms from their stylists, only 0.5 percent report buying from them because (unbranded) female condoms were available at the same price through other outlets such as chemists and bars, and available free of charge from health clinics. However, the share of respondents who ever used a female condom is more than double among those who report receiving information from their stylists (27 percent) vs. those who do not (12 percent), suggesting that the effect of the agents' effort in promoting the condoms on usage is larger than the effect on sales through hair salons.

Two considerations are important to inform the scaling-up of the non-financial reward treatment to include all eligible stylists in Lusaka. First, the fact that stars are more effective when stylists are surrounded by other stylists in the same treatment suggests that the effect estimated from a share of treated stylists might be a lower bound for the effect of stars, when these are offered to all stylists, as the number of potential peers would be larger in the latter case. Second, the fact that the effect of stars is stable throughout the experimental year provides reasons for cautious optimism that this scheme might be effective at motivating agents in the long run. While we cannot measure the effect past the experimental year, the absence of a clear trend reassures us that the effect is unlikely to discontinuously disappear as the treatment is extended past the year.

As is often the case in field experiments, the interpretation of the findings and their wider applicability depends on the key features of the specific setting. In our case, two features are of note. The first key feature is that, to minimize the possibility of information spillovers among agents

in different treatment groups, agents were not informed of the existence or type of rewards when they were first invited to participate in the training for condom distribution. This reconciles our finding that incentives do not affect the selection of agents into the job with earlier evidence from the private sector and from the laboratory that suggests substantial selection effects (Bandiera et al. 2007; Dohmen and Falk 2011; Larkin and Leider 2012; Lazear 2000; Lazear et al. 2012). In general, we expect incentives to affect selection, since different schemes might attract different numbers and types of agents. This is likely to be particularly relevant in the social sector to the extent that organizations are better off by hiring agents who are attracted by the mission as opposed to a generous incentive scheme.

The second key feature of our setting is that the task at hand is not the agents' main occupation and the agents we study have selected entrepreneurship in the private sector as their main occupation. Non-financial rewards might be more effective for them because they reward the only pro-social component of their jobs. On the other hand, if non-financial rewards interact with the agents' pro-social motivation, they might be even more effective for agents who self-select into the social sector as their main occupation. Ultimately, to assess whether and how non-financial rewards can be effective in other settings, future research will need to provide evidence on how the nature of the reward interacts with the nature of the task to attract, motivate and retain employees.

Acknowledgements

We thank Jim Andreoni, Anat Bracha, Tim Besley, Raj Chetty, Maitreesh Ghatak, Matt Gentzkow, Uri Gneezy, Emir Kamenica, Ian Larkin, Imran Rasul and seminar participants at Harvard Business School, London School of Economics, London Business School, NYU, University of Chicago (Harris and Booth), University College London, University of Maryland, Cornell University, Stanford University, the Boston Federal Reserve Bank, Tufts University, the BEAM conference, Toulouse School of Economics, as well as from the University of Zurich and UCSD for helpful comments at different stages of this project. Financial support was provided by the National Science Foundation, 3ie, STICERD, WAPPP, HBS DFRD, and the Sustainability Science Program. Field implementation would not have been possible without the support of Mannaseh Phiri, Pamela Chama, Charles Kalonga, Nicholas Shilya, Richard Harrison, Debbie Goldschmidt, Marie-Helene Cloutier, Lauren Gaydosh, Arne Weiss, Mahjabeen Haji, Holly Hickling, and all of the SFH sales representatives. The authors would like to especially acknowledge the dedication of Patrick Mulenga.

References

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal 100*(401), 464–477.

Ariely, D., A. Bracha, and S. Meier (2009). Doing good or doing well? image motivation and monetary incentives in behaving prosocially. *The American Economic Review 99*(1), 544–555.

Bandiera, O., I. Barankay, and I. Rasul (2007). Incentives for managers and inequality among workers: Evidence from a firm-level experiment. *The Quarterly Journal of Economics 122*(2), 729–773.

Bandiera, O., I. Barankay, and I. Rasul (2011). Field experiments with firms. *The Journal of Economic Perspectives 25*(3), 63–82.

Benabou, R. and J. Tirole (2003). Intrinsic and extrinsic motivation. *Review of Economic Studies 70*(3), 489–520.

Benabou, R. and J. Tirole (2006). Incentives and prosocial behavior. *The American Economic Review 96*(5), 1652–1678.

Besley, T. and M. Ghatak (2005). Competition and incentives with motivated agents. *American Economic Review 95*(3), 616–636.

Bhattacharyya, K. and P. Winch (2001). Community health worker incentives and disincentives. Technical report, BASICS II.

Chetty, R., E. Saez, and L. Sandor (2012). What policies motivate pro-social behavior? an experiment with referees at the journal of public economics. Mimeo, Harvard University.

Dal Bó, E., F. Finan, and M. Rossi (2011). Strengthening state capabilities: The role of financial incentives in the call to public service. *UC Berkeley*.

Deci, E. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology 18*(1), 105–115.

DellaVigna, S., J. List, and U. Malmendier (2012). Testing for altruism and social pressure in charitable giving. *The Quarterly Journal of Economics 127*(1), 1–56.

Dixit, A. (2002). Incentives and organizations in the public sector: An interpretative review. *Journal of Human Resources*, 696–727.

Dohmen, T. and A. Falk (2011). Performance pay and multidimensional sorting: Productivity, preferences, and gender. *The American Economic Review 101*(2), 556–590.

Duflo, E., R. Hanna, and S. Ryan (2012). Incentives work: Getting teachers to come to school. *The American Economic Review 102*(4), 1241–1278.

Fryer, R. (2011). Teacher incentives and student achievement: Evidence from new york city public schools. National Bureau of Economic Research Working Paper No. 16850.

Garber, A. and C. Phelps (1997). Economic foundations of cost-effectiveness analysis. *Journal of Health Economics 16*(1), 1–31.

Glewwe, P., N. Ilias, and M. Kremer (2010). Teacher incentives. *American Economic Journal: Applied Economics 2*(3), 205–227.

Gneezy, U. and A. Rustichini (2000). Pay enough or don't pay at all. *The Quarterly Journal of Economics 115*(3), 791–810.

Goette, L. and A. Stutzer (2008). Blood donations and incentives: Evidence from a field experiment. IZA Working Paper No. 3580.

Government Republic of Zambia (2010). Zambia country report: Monitoring the declaration of commitment on hiv and aids and the universal access biennial report. Technical report, Lusaka, Zambia.

Grant, A. (2007). Relational job design and the motivation to make a prosocial difference. *The Academy of Management Review 32*(2), 393–417.

Harrison, G. and J. List (2004). Field experiments. *Journal of Economic Literature 42*(4), 1009–1055.

HLSP (2007). Materials development for female condom marketing: A report on key informant and focus group findings, Zambia. Technical report, HLSP Limited.

Karlan, D. and M. McConnell (2012). Hey look at me: The effect of giving circles on giving. National Bureau of Economic Research Working Paper No. 17737.

Kube, S., M. Marechal, and C. Puppe (2012). The currency of reciprocity: Gift exchange in the workplace. *The American Economic Review 102*(4), 1644–1662.

Lacetera, N., M. Macis, and R. Slonim (2011). Rewarding altruism? a natural field experiment. National Bureau of Economic Research Working Paper No. 17636.

Lagarde, M. and D. Blaauw (2013). Pro-social preferences and self-selection in to rural jobs: Evidence from south african nurses. Mimeo, London School of Hygiene and Tropical Medicine.

Larkin, I. (2011). Paying $30,000 for a gold star: An empirical investigation into the value of peer recognition to software salespeople. Mimeo, Harvard Business School.

Larkin, I. and S. Leider (2012). Incentive schemes, sorting, and behavioral biases of employees: Experimental evidence. *American Economic Journal: Microeconomics 4*(2), 184–214.

Lavy, V. (2002). Evaluating the effect of teachers? group performance incentives on pupil achievement. *Journal of Political Economy 110*(6), 1286–1317.

Lazear, E. (2000). Performance pay and productivity. *The American Economic Review 90*(5), 1346–1361.

Lazear, E., U. Malmendier, and R. Weber (2012). Sorting in experiments with application to social preferences. *American Economic Journal: Applied Economics 4*(1), 136–163.

Mathauer, I. and I. Imhoff (2006). Health worker motivation in africa: The role of non-financial incentives and human resource management tools. *Human Resources for Health 4*(1), 24–41.

Meier, S. (2007). Do subsidies increase charitable giving in the long run? matching donations in a field experiment. *Journal of the European Economic Association 5*(6), 1203–1222.

Mellström, C. and M. Johannesson (2008). Crowding out in blood donation: Was titmuss right? *Journal of the European Economic Association 6*(4), 845–863.

Miller, G., R. Luo, L. Zhang, S. Sylvia, Y. Shi, P. Foo, Q. Zhao, R. Martorell, A. Medina, and S. Rozelle (2012). Effectiveness of provider incentives for anaemia reduction in rural china: A cluster randomised trial. *BMJ: British Medical Journal 345*, 1–10.

Muralidharan, K. and V. Sundararaman (2011). Teacher performance pay: Experimental evidence from india. *Journal of Political Economy 119*(1), 39–77.

Olken, B., J. Onishi, and S. Wong (2012). Should aid reward performance? evidence from a field experiment on health and education in indonesia.

Oyer, P. and S. Schaefer (2011). *Personnel economics: hiring and incentives*, Volume 4, pp. 1769–1823. Elsevier.

PATH UNFPA (2006). Female condom: A powerful tool for protection. Technical report, UNFPA.

Propper, C. and J. Van Reenen (2010). Can pay regulation kill? panel data evidence on the effect of labor markets on hospital performance. *Journal of Political Economy 118*(2), 222–273.

PSI (2012). Family planning interventions impact estimation model. Technical report, Population Services International.

Scott, C. E. and J. J. Siegfried (2011). American economic association universal academic questionnaire summary statistics. *The American Economic Review 101*(3), 664–667.

Soetevent, A. (2005). Anonymity in giving in a natural context: A field experiment in 30 churches. *Journal of Public Economics 89*(11), 2301–2323.

UNAIDS, UNFPA and UNIFEM (2004). Women and HIV/AIDS: Confronting the crisis. Technical report, A joint report by UNAIDS, UNFPA and UNIFEM.

Table 1: Summary statistics

Panel A: Outcome variables	mean	median	min	max	sd	N
Packs sold (restocked)	9.01	0.00	0.00	216.00	18.08	771
Packs sold (calculated)	13.90	12.00	0.00	148.00	15.77	771
Promoter attention	2.52	2.56	0.00	3.00	0.30	725
Promoter interest	2.15	2.12	0.00	3.00	0.38	697
Logbook filled	0.47	0.50	0.00	1.00	0.23	725
Total displays (promotional material)	2.26	2.20	0.00	8.00	0.90	726
Panel B: Control variables						
Salon is a hair salon (0-1)	0.48	0.00	0.00	1.00	0.50	771
Salon is a barbershop (0-1)	0.44	0.00	0.00	1.00	0.50	771
Salon is both a barbershop and hair salon (0-1)	0.08	0.00	0.00	1.00	0.27	771
Salon is near a bar (0-1)	0.88	1.00	0.00	1.00	0.32	770
Salon size (number of employees)	1.75	2.00	1.00	9.00	0.99	770
Number of trained salons in the same area	4.46	3.00	1.00	30.00	5.06	173
Stylist sells other products in salon (0-1)	0.27	0.00	0.00	1.00	0.45	771
Stylist is in the bottom quartile of the asset distribution (0-1)	0.21	0.00	0.00	1.00	0.40	771
Stylist's socio-economic status is low (0-1)	0.19	0.00	0.00	1.00	0.40	771
Stylist's dictator-game donation (Kwacha)	5,728.94	5,000.00	0.00	40,000.00	3,744.67	767
Stylist's reported work motivation is intrinsic (0-1)	0.58	1.00	0.00	1.00	0.49	771
Stylist's religion is Catholic (0-1)	0.23	0.00	0.00	1.00	0.42	771
Panel C: Other Descriptors						
Weekly income of the salon (Kwacha)	332,569	250,000	0	10,000,000	572,050	700
Stylist can read and write in at least one language (0-1)	0.94	1.00	0.00	1.00	0.23	771
Stylist can read and write in English (0-1)	0.85	1.00	0.00	1.00	0.35	770
Total number of products sold	0.47	0.00	0.00	6.00	0.94	771
Stylist sells hair products (0-1)	0.70	1.00	0.00	1.00	0.46	212
Stylist sells cosmetics (0-1)	0.33	0.00	0.00	1.00	0.47	212
Stylist sells clothing (0-1)	0.14	0.00	0.00	1.00	0.34	212
Stylist sells jewelry (0-1)	0.15	0.00	0.00	1.00	0.35	212
Stylist sells talktime (0-1)	0.11	0.00	0.00	1.00	0.32	212

Notes: Sample includes all salons that attended training (N=771). *Packs sold (restocked)* is the number of packs (excluding the initial dispenser sold at training) that the stylist chooses to buy and restock over a 10-month period, based on invoices. *Packs sold (calculated)* is the number of packs sold, including the initial dispenser sold at training, based on sales agents' calculations. *Promoter attention* is a measure of stylist's level of attention, on average, across all sales agent visits during the "interpersonal communication" session on a 0-3 scale (with 0 being "not interested", and 3 being "very interested"). Similarly, *Promoter interest* is the sales agent's subjective rating, on average, of the stylist's level of interest in promoting female condoms on a scale of 0 to 3. *Logbook filled* is an indicator variable that takes a value of 1 if filled-in logbook sheets were collected by the sales agent, averaged across non-missing visits. *Total displays* is the average number of posters, brochures, "sold here" signs, flipcharts, condom dispensers, and certificates visible in the shop during non-missing restocking visits. A salon was considered *near a bar* if there was a bar within a 5-minute walk. The *number of trained salons in the same area* average across the 173 neighborhoods used as the unit of randomization along with training attendance outcomes. Stylists were classified as having *low socio-economic status* if they could not read and write in English, or if they had not completed primary school. The *asset quartile* was determined based on ownership of durables by the stylist. We define self-reported *work motivation to be intrinsic* if the agent reports "being connected to the community" or "making people look nice" as their preferred aspect of the job, in contrast to "making money" and "being own boss". Refer to Data Appendix for more details on variables.

Table 2: Average treatment effects on sales

Dependent variable	Packs sold (restocked)		Packs sold (calculated)	=1 if sells at least one pack	=1 if sells 12 or more packs	=1 if sells 24 or more packs
Mean in control group	*6.93*	*6.96*	13.30	*.368*	*.341*	*.128*
	(1)	(2)	(3)	(4)	(5)	(6)
Large financial reward	0.769	1.179	-0.647	-0.002	0.01	0.031
	[1.618]	[1.763]	[1.851]	[0.067]	[0.063]	[0.042]
Small financial reward	0.378	0.812	-0.142	-0.025	-0.018	0.011
	[1.528]	[1.547]	[1.620]	[0.066]	[0.060]	[0.040]
Star reward	7.482***	7.660***	5.996**	0.118*	0.131**	0.101**
	[2.448]	[2.554]	[2.427]	[0.066]	[0.066]	[0.049]
Salon is a barbershop (0-1)		3.316**	3.624**	0.094**	0.093**	0.032
		[1.611]	[1.490]	[0.041]	[0.042]	[0.031]
Salon is both a barbershop and hair salon (0-1)		3.94	3.009	-0.05	-0.035	0.004
		[3.944]	[3.136]	[0.071]	[0.071]	[0.053]
Salon is near a bar (0-1)		0.545	0.796	-0.048	-0.031	-0.005
		[2.143]	[2.004]	[0.074]	[0.063]	[0.050]
Salon size (log number of employees)		1.557	0.575	-0.071	-0.062	0.036
		[2.776]	[2.678]	[0.066]	[0.067]	[0.049]
Number of trained salons in the same area		0.027	0.074	0.001	0.000	-0.001
		[0.087]	[0.093]	[0.003]	[0.003]	[0.002]
Stylist sells other products in salon (0-1)		5.183***	2.794*	0.084**	0.084**	0.073**
		[1.718]	[1.548]	[0.039]	[0.040]	[0.036]
Stylist in the bottom quartile of asset distribution (0-1)		1.159	0.322	0.007	0.000	0.018
		[1.724]	[1.630]	[0.051]	[0.052]	[0.035]
Stylist's socio-economic status is low (0-1)		-0.998	-0.926	-0.009	-0.012	-0.042
		[1.410]	[1.207]	[0.046]	[0.047]	[0.029]
Stylist's dictator-game donation above the median (0-1)		3.364***	2.234**	0.152***	0.143***	0.016
		[1.137]	[1.123]	[0.031]	[0.032]	[0.028]
Stylist's reported work motivation is intrinsic (0-1)		-0.512	-0.424	-0.035	-0.034	-0.03
		[1.328]	[1.191]	[0.036]	[0.035]	[0.032]
Stylist's religion is Catholic (0-1)		-3.652***	-3.215***	-0.084**	-0.073*	-0.035
		[1.387]	[1.198]	[0.042]	[0.040]	[0.033]
Constant	6.929***	0.431	8.334**	0.351***	0.311***	0.086
	[1.123]	[3.851]	[3.851]	[0.098]	[0.093]	[0.073]
R-squared	0.0285	0.0659	0.0547	0.0505	0.0485	0.0267
Observations	771	765	743	765	765	765
Large financial = Small financial (p-value)	0.803	0.819	0.753	0.697	0.58	0.584
Large financial = Stars (p-value)	0.00719	0.0122	0.006	0.0467	0.048	0.152
Small financial = Stars (p-value)	0.00365	0.00622	0.009	0.0177	0.0124	0.0546

Notes: OLS estimates. Standard errors are clustered at cell level. * p<0.10 ** p<0.05 *** p<0.01. The dependent variable in columns 1-5, *Packs sold (restocked)* is the total number of packs (excluding the initial dispenser sold at training) that the stylist chooses to buy and restock during the study period, based on invoices. The dependent variable in column 6, *Packs sold (calculated)* is the total number of packs sold (including the initial dispenser sold at training), based on representatives' calculations. The sample size varies across columns because of missing values in some covariates. Variables are as described in Table 1. P-values in the bottom three rows are from a Wald test for equality of coefficients between treatments.

Table 3: Average treatment effects on effort measures

Dependent variable	Total displays	Logbook filled	Promoter attention	Promoter interest	Average standardized effect
Mean in control group	2.285	0.479	2.498	2.111	
Standard deviation in control group	*1.19*	*0.28*	*0.41*	*0.42*	
	(1)	(2)	(3)	(4)	(5)
Large financial reward	0.072	0.028	-0.004	0.024	0.03
	[0.102]	[0.029]	[0.034]	[0.035]	[0.036]
Small financial reward	-0.099	0.008	0.022	0.049	-0.005
	[0.127]	[0.028]	[0.044]	[0.049]	[0.050]
Star reward	0.245**	0.065**	-0.044	0.096**	0.090**
	[0.120]	[0.031]	[0.034]	[0.044]	[0.041]
Controls	yes	yes	yes	yes	yes
R-squared	0.101	0.0234	0.035	0.0605	
Observations	722	722	721	694	726
Large financial = Small financial (p-value)	0.152	0.502	0.516	0.605	0.049
Large financial = Stars (p-value)	0.123	0.219	0.237	0.116	0.133
Small financial = Stars (p-value)	0.0137	0.074	0.12	0.417	0.087

Notes: OLS estimates weighted by the number of observations for each salon. All outcomes are averages are at the salon level across all restocking visits. Standard errors are clustered at the cell level. * p<0.10 ** p<0.05 *** p<0.01. *Total displays* is the average number of posters, brochures, "sold here" signs, flipcharts, condom dispensers, and certificates visible in the shop during non-missing restocking visits. *Logbook filled* is an indicator variable that takes a value of 1 if filled-in logbook sheets were collected by the sales agent, averaged across non-missing visits. *Promoter attention* is a measure of stylist's level of attention, on average, across all sales agent visits during the "interpersonal communication" session on a 0-3 scale (with 0 being "not interested", and 3 being "very interested"). Similarly, *Promoter interest* is the sales agent's subjective rating, on average, of the stylist's level of interest in promoting female condoms, on a scale of 0 to 3). Column 5 reports the average standardized effect for the four effort variables. All regressions include the same vector of controls as in Table 2. P-values in the bottom three rows are from a Wald test for equality of coefficients between treatments.

Table 4: Placebo star reward

Dependent variable	Packs sold (restocked)		Packs sold (calculated)	
	placebo rounds	first round	placebo rounds	first round
Mean in control group	0.469	0.469	1.156	1.156
	(1)	(2)	(3)	(4)
Placebo thermometer	0.415	0.01	-0.05	0.01
	[0.386]	[0.398]	[0.375]	[0.398]
Star reward	1.629***	1.736**	1.535***	1.736**
	[0.598]	[0.712]	[0.480]	[0.712]
Controls	yes	yes	yes	yes
R-squared	0.0656	0.0948	0.117	0.0948
Observations	319	318	319	318
Placebo thermometer = Stars (p-value)	0.0536	0.0105	0.00107	0.0105

Notes: Standard errors are clustered at cell level. * p<0.10 ** p<0.05 *** p<0.01. The sample is restricted to salons that completed a restocking visit in round 9, who were either in the star reward treatment group or who received either a placebo thermometer or an additional promotional poster in round 8. *Placebo thermometer* = 1 if stylist received a thermometer poster reporting average sales of condoms across stars treatment (12 packs) during the previous restocking visit. The dependent variable in Columns (1) and (2), *Packs sold (restocked)*, is the number of packs (excluding the initial dispenser sold at training) that the stylist chooses to buy and restock in the month following the placebo intervention or the first round the treatment (either placebo or star) took effect, based on invoices. The dependent variable in Columns (3) and (4), *Packs sold (calculated)*, is the number of packs sold, including the initial dispenser sold at training, based on representatives' calculations. Columns (1) and (3) report sales for the first round in which the placebo thermometer could affect sales (round 9). Columns (2) and (4) report sales for the first round after the treatment was implemented (round one for the star reward treatment and round 9 for the placebo thermometer and promotional material control). One star reward treatment salon did not complete the first round restocking visit so is dropped from columsn 2 and 4. All regressions include the same vector of controls as in Table 2. P-values in the bottom row are from a Wald test for equality of coefficients.

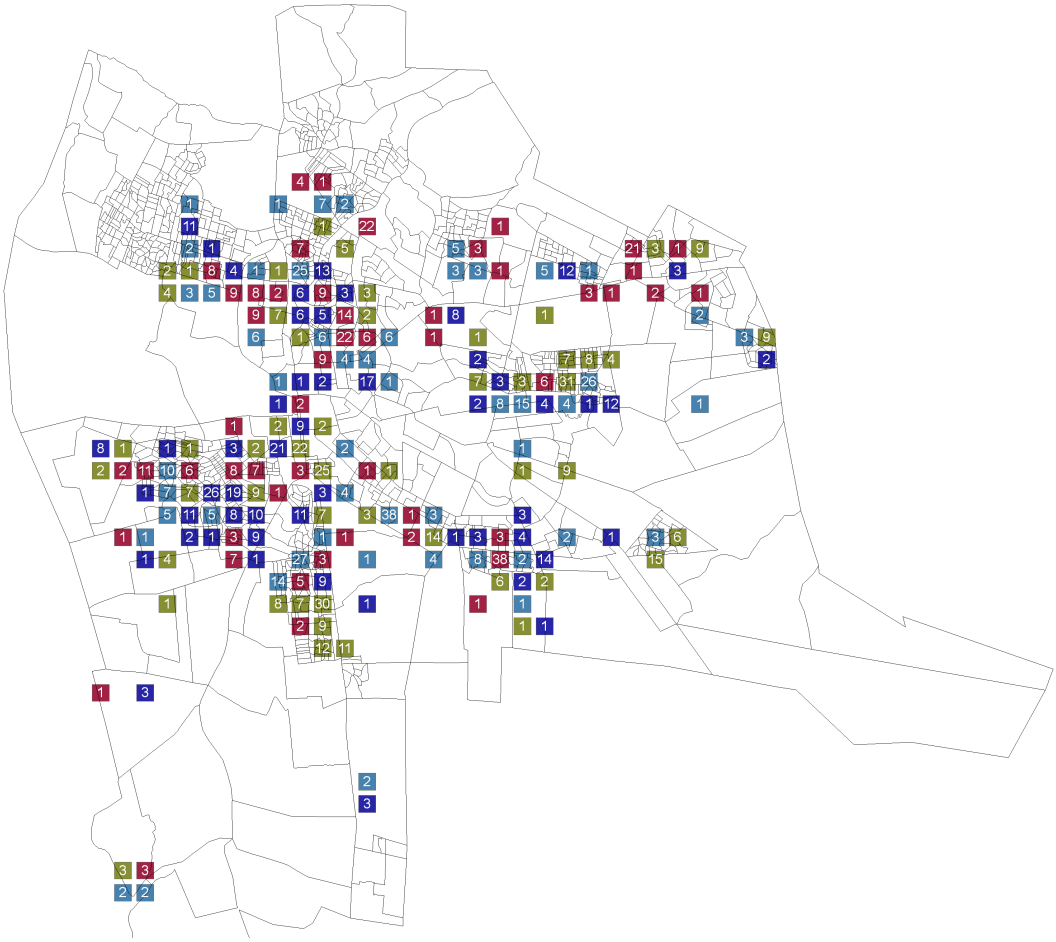
Table 5: Heterogeneous treatment effects, by stylist motivation

Dependent variable is Packs sold (restocked)

Interaction variable	Stylist's dictator game donation is above the median	Stylist's reported work motivation is intrinsic	Stylist's socio-economic status is low	Number of trained salons in the same area is above median
Mean in control group = 6.96				
	(1)	(2)	(3)	(4)
Motivation variable	0.771	-3.631*	-4.126**	-0.983
	[1.531]	[1.958]	[1.610]	[2.302]
Effect of large financial when interaction variable =0	-2.364	-1.66	0.775	2.584
	[1.642]	[2.447]	[2.091]	[2.939]
Effect of small financial when interaction variable =0	1.068	-0.321	-0.077	-0.201
	[1.936]	[2.841]	[1.719]	[2.803]
Effect of stars when interaction variable =0	4.341	3.858	7.016**	2.427
	[2.897]	[3.816]	[2.906]	[3.660]
Effect of large financial when interaction variable =1	3.546	2.63	3.682**	0.223
	[2.490]	[2.228]	[1.839]	[1.741]
Effect of small financial when interaction variable =1	0.383	0.999	4.869*	1.326
	[1.933]	[1.768]	[2.910]	[1.705]
Effect of stars when interaction variable =1	10.010***	10.480***	11.080***	9.144***
	[3.238]	[2.986]	[3.108]	[2.966]
Controls	yes	yes	yes	yes
R-squared	0.073	0.071	0.067	0.073
Observations	765	765	765	765
Large financial: P-value on the interaction term	0.026	0.144	0.301	0.484
Small financial: P-value on the interaction term	0.769	0.686	0.139	0.511
Stars: P-value on the interaction term	0.096	0.134	0.281	0.127

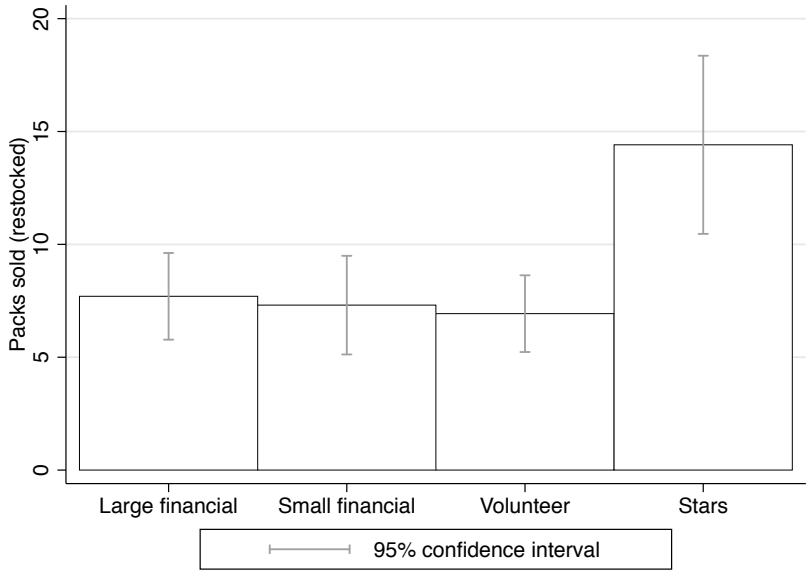
Notes: Standard errors are clustered at cell level. * p<0.10 ** p<0.05 *** p<0.01. The dependent variable, *Packs sold (restocked)* is the total number of packs (excluding the initial dispenser sold at training) that the stylist chooses to buy and restock over the study period, based on invoices. All regressions include the same vector of controls as in Table 2. Variables are described in Table 1. The median number of trained salons per area, across areas, is 3.

Figure 1: Randomization of map cells into treatment groups



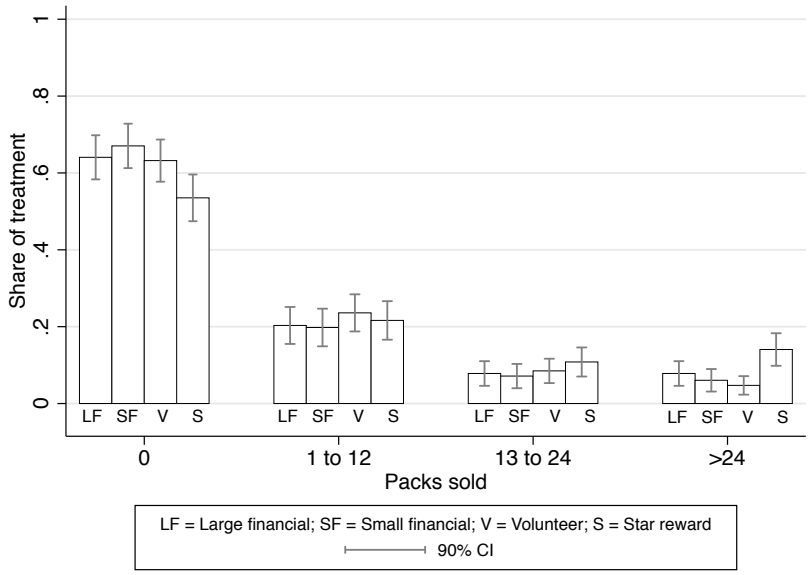
Notes: Treatment groups and volunteer control group are shown by the cell colors. The number of salons attending the training are written in each cell.

Figure 2: Average yearly sales by treatment group



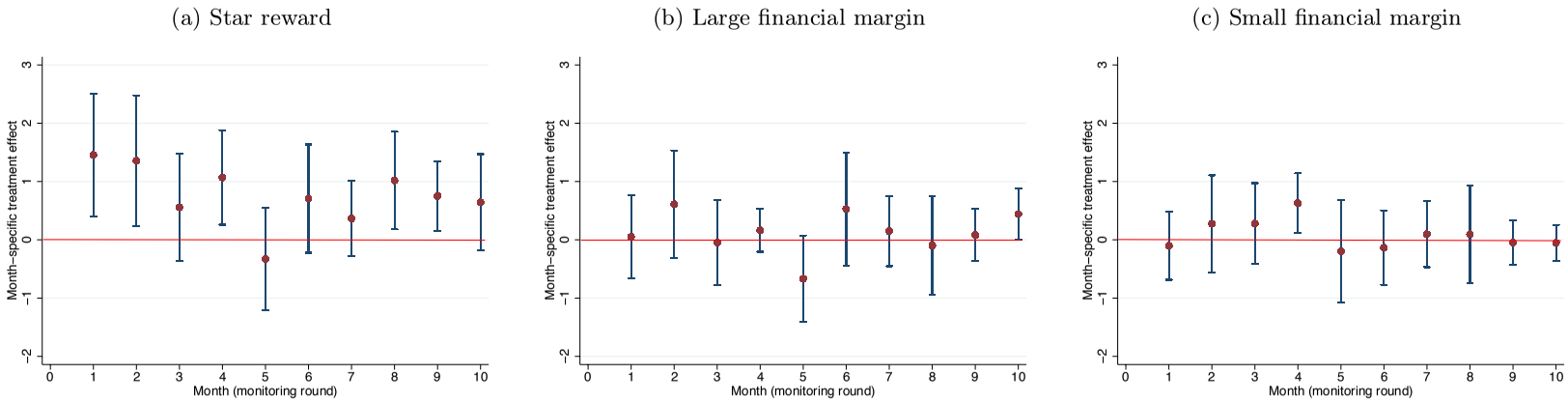
Notes: Each bar measures the average number of packs sold over the year by agents in each of the four groups with 95 percent confidence intervals.

Figure 3: Distribution of packs sold by treatment



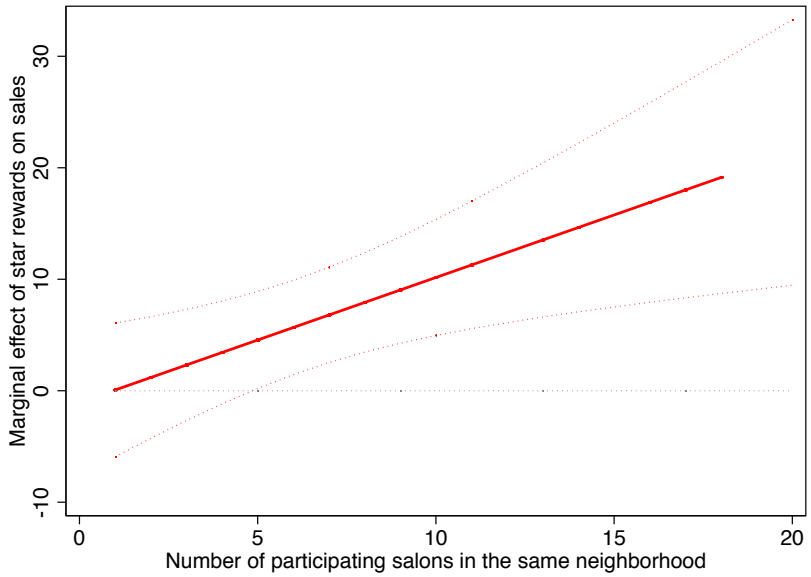
Notes: For each treatment group, packs sold are binned into the four categories displayed on the x-axis. The height of the bars shows the share of the treatment in each bin, which sum to one in each treatment. The error bars correspond to the 90 percent confidence interval.

Figure 4: Month-specific treatment effects



Notes: Each dot represents the estimated effect of the star treatment (Panel A), large financial margin treatment (Panel B) and small financial margin treatment (Panel C) in a regressions of sales on the three treatments, and controls listed in table 2. The vertical lines represent 95 percent confidence intervals based on standard errors clustered at the cell level.

Figure 5: Effect of star rewards as function of the number of salons



Notes: The solid line plots the imputed marginal effect of the star treatment at each total number of salons in the same neighborhood. This is computed as the sum of the coefficient of stars plus the coefficient of the interaction of stars and number of salons in the same neighborhood, multiplied by the respective value of neighborhood density estimated in a regression of sales on the three treatments, the three treatments interacted with neighborhood density, and controls. The dotted lines represent the 95 percent confidence interval is based on standard errors clustered at the cell level.

APPENDIX

Table A.1: Agents' and salons' characteristics at training, by treatment group

	Large financial	Small financial	Stars	Volunteer	Largest pairwise normalized difference	Large financial vs. volunteer	Small financial vs. volunteer	Stars vs. volunteer	Large vs. small financial	Large financial vs. stars	Small financial vs. stars
Randomization balance variables											
Salon is a barbershop (0-1)	0.427	0.412	0.427	0.481	0.0982	0.193	0.296	0.254	0.826	0.999	0.836
	[0.496]	[0.494]	[0.496]	[0.501]							
Salon is both a barbershop and hair salon (0-1)	0.0573	0.0604	0.157	0.0425	0.274	0.644	0.601	0.002	0.931	0.010	0.020
	[0.233]	[0.239]	[0.365]	[0.202]							
Salon is near a bar (0-1)	0.921	0.863	0.897	0.854	0.152	0.301	0.908	0.521	0.266	0.538	0.535
	[0.270]	[0.345]	[0.304]	[0.354]							
Salon size (log number of employees)	0.958	0.958	0.989	0.948	0.0885	0.828	0.850	0.351	0.100	0.502	0.552
	[0.296]	[0.278]	[0.349]	[0.292]							
Number of trained salons in the same area	4.364	4.333	4.302	4.818	0.080	0.695	0.671	0.597	0.980	0.953	0.976
	[5.723]	[5.467]	[3.907]	[5.100]							
Stylist sells other products in salon (0-1)	0.240	0.280	0.259	0.311	0.114	0.278	0.651	0.417	0.511	0.724	0.727
	[0.428]	[0.450]	[0.440]	[0.464]							
Stylist is in bottom quartile of asset distribution (0-1)	0.219	0.220	0.184	0.203	0.0633	0.778	0.769	0.709	0.987	0.527	0.524
	[0.414]	[0.415]	[0.388]	[0.403]							
Other controls											
Stylist's socio-economic status is low (0-1)	0.203	0.236	0.200	0.142	0.172	0.254	0.017	0.228	0.575	0.962	0.502
	[0.403]	[0.426]	[0.401]	[0.349]							
Stylist's dictator-game donation is above median (0-1)	0.597	0.552	0.587	0.597	0.0638	0.994	0.316	0.861	0.192	0.845	0.513
	[0.492]	[0.499]	[0.494]	[0.492]							
Stylist's reported work motivation is intrinsic (0-1)	0.552	0.588	0.476	0.689	0.312	0.018	0.070	0.000	0.560	0.169	0.037
	[0.499]	[0.494]	[0.501]	[0.464]							
Stylist's religion is Catholic (0-1)	0.234	0.214	0.254	0.226	0.0663	0.837	0.792	0.523	0.606	0.582	0.362
	[0.425]	[0.411]	[0.437]	[0.420]							
Number of observations	192	182	185	212							
Number of geographic cells	44	42	43	44							

Notes: The first four columns report means, with standard deviations in brackets, for the four groups. P-values are reported for F-tests for equal coefficients from a regression with standard errors clustered at the cell level. Variables are described in Table 1. All means are taken at the individual level, with the exception of *Number of salons in the same area*, which is measured at the randomization cell level. The *Number of geographic cells* refers to the unit of randomization.

Table A.2: Participation decision

Dependent variable	Received invitation (Conditional on assigned to program)		Attended training (Conditional on received invitation)	
Mean in control group = 0.80				
	(1)	(2)	(3)	(4)
Large financial reward	-0.005	-0.008	0.02	0.015
	[0.033]	[0.029]	[0.042]	[0.042]
Small financial reward	0.029	0.029	-0.023	-0.016
	[0.034]	[0.031]	[0.042]	[0.041]
Star reward	-0.006	0.000	-0.042	-0.034
	[0.031]	[0.031]	[0.046]	[0.047]
Salon is a barbershop (0-1)		0.060**		0.056*
		[0.028]		[0.033]
Salon is both a barbershop and hair salon (0-1)		0.023		0.028
		[0.040]		[0.053]
Salon is near a bar (0-1)		0.023		0.067
		[0.037]		[0.050]
Salon size (log number of employees)		0.044		-0.033
		[0.039]		[0.045]
Total number of salons in the same area		0.003***		0.002*
		[0.001]		[0.001]
Stylist sells other products in salon (0-1)		0.013		-0.006
		[0.026]		[0.032]
Stylist is in bottom quartile of asset distribution (0-1)		-0.057*		-0.004
		[0.033]		[0.036]
Stylist's socio-economic status is low (0-1)		0.014		-0.069*
		[0.025]		[0.036]
Stylist gives to HIV causes (0-1)		0.025		0.055**
		[0.025]		[0.026]
Stylist's reported work motivation is intrinsic (0-1)		0.035		0.003
		[0.023]		[0.028]
Stylist's religion is Catholic (0-1)		0.011		0.021
		[0.025]		[0.026]
Constant	0.799***	0.648***	0.767***	0.707***
	[0.021]	[0.062]	[0.032]	[0.075]
R-squared	0.0012	0.0164	0.0032	0.0218
Observations	1222	1216	981	977
Large financial = Small financial (p-value)	0.351	0.215	0.259	0.392
Large financial = Stars (p-value)	0.975	0.806	0.147	0.238
Small financial = Stars (p-value)	0.316	0.363	0.65	0.646

Notes: Coefficients are marginal effects from a probit model. Errors clustered at the cell level. * p<0.10 ** p<0.05 *** p<0.01. Variables are described in Table 1 with the exception of *Stylist gives to HIV causes*, which is a binary, self-reported measure of donating funds to people living with HIV/AIDS, and *Total number of salons in the same area* which represents the neighborhood population of salons assigned to receive an invitation. P-values in the bottom three rows are from a Wald test for equality of coefficients between treatments.

Table A.3: Treatment effects on selection

Dependent variable	Stylist did not join program		Stylist quit after joining		Visits stopped after 3 consecutive misses	
Mean in Volunteer control group	*0.042*		*0.052*		*0.259*	
	(1)	(2)	(3)	(4)	(5)	(6)
Large financial reward	-0.009	-0.011	0.000	-0.001	0.054	0.069
	[0.016]	[0.009]	[0.025]	[0.024]	[0.050]	[0.055]
Small financial reward	-0.017	-0.01	0.059**	0.051*	0.06	0.059
	[0.016]	[0.011]	[0.034]	[0.033]	[0.056]	[0.055]
Star reward	-0.017	-0.011	0.051*	0.049	-0.017	-0.023
	[0.015]	[0.009]	[0.034]	[0.034]	[0.052]	[0.052]
Salon is a barbershop (0-1)		0.017		-0.003		-0.071*
		[0.012]		[0.022]		[0.036]
Salon is both a barbershop and hair salon (0-1)		-0.002		-0.049*		0.031
		[0.018]		[0.018]		[0.066]
Salon is near a bar (0-1)		-0.006		0.091***		0.03
		[0.017]		[0.029]		[0.055]
Salon size (log number of employees)		0.001*		0.000		-0.004**
		[0.000]		[0.001]		[0.002]
Number of trained salons in the same area		0.014		-0.002		-0.004
		[0.011]		[0.021]		[0.040]
Stylist sells other products in salon (0-1)		-0.006		0.001		0.067
		[0.007]		[0.023]		[0.044]
Stylist is in bottom quartile of asset distribution (0-1)		-0.009		-0.005		-0.052
		[0.009]		[0.021]		[0.040]
Stylist's socio-economic status is low (0-1)		0.028***		-0.002		-0.099***
		[0.009]		[0.014]		[0.034]
Stylist's dictator-game donation is above median (0-1)		-0.011		-0.009		0.055
		[0.009]		[0.016]		[0.035]
Stylist's reported work motivation is intrinsic (0-1)		0.006		-0.015		0.029
		[0.011]		[0.019]		[0.040]
Stylist's religion is Catholic (0-1)				0.007		-0.002
				[0.025]		[0.047]
Pseudo R-squared	0.0094	0.1017	0.0163	0.0418	0.0044	0.0264
Observations	771	766	771	765	771	765
Large financial = Small financial (p-value)	0.627	0.905	0.0237	0.0346	0.902	0.841
Large financial = Stars (p-value)	0.56	0.96	0.0556	0.0567	0.137	0.0612
Small financial = Stars (p-value)	0.983	0.887	0.823	0.953	0.148	0.0899

Notes: Coefficients are marginal effects from probit model. The outcome in Columns (1) and (2) equals 1 if the stylist did not show up to training or did not join the program after training. The outcome in Columns (3) and (4) equals 1 if the stylist requested that sales agents stop visiting. The outcome in Columns (5) and (6) equal 1 if the salon was closed for three consecutive restocking visits and sales agents removed the salon from the rotation. Errors are clustered at the cell level. * p<0.10 ** p<0.05 *** p<0.01. Variables are described in Table1. The variable describing stylist religion is dropped in Column (2) because it is perfectly collinear with the dependent variable. Sample size varies across columns because of missing values in some covariates. P-values in the bottom three rows are from a Wald test for equality of coefficients between treatments.

Table A.4: Robustness checks: Average treatment effects on sales

Dependent variable	Packs sold (restocked)	Winsorized at		In sample through final round	Ln[Packs sold (restocked)]	Monthly sales, with agent FE
		95%	90%			
Mean in control group	*6.962*	*5.769*	*5.769*	*9.800*	*1.035*	*0.823*
	(1)	(2)	(3)	(4)	(5)	(6)
Large financial reward	1.179	1.426	0.386	2.92	0.045	0.166
	[1.763]	[1.396]	[1.224]	[2.146]	[0.199]	[0.160]
Small financial reward	0.812	0.652	-0.165	2.762	-0.032	0.211
	[1.547]	[1.219]	[1.143]	[2.397]	[0.190]	[0.171]
Star reward	7.660***	7.096***	4.472***	10.675***	0.483**	0.896***
	[2.554]	[2.025]	[1.543]	[3.651]	[0.211]	[0.229]
Salon is a barbershop (0-1)	3.316**	2.477**	1.734*	2.897	0.297**	0.427**
	[1.611]	[1.233]	[0.881]	[2.128]	[0.130]	[0.181]
Salon is both a barbershop and hair salon (0-1)	3.94	1.667	0.04	5.403	-0.063	0.509
	[3.944]	[2.353]	[1.585]	[6.210]	[0.229]	[0.484]
Salon is near a bar (0-1)	0.545	0.012	-0.038	1.046	-0.094	0.076
	[2.143]	[1.726]	[1.366]	[3.372]	[0.218]	[0.198]
Salon size (log number of employees)	1.557	0.503	-0.237	7.249*	-0.119	0.343
	[2.776]	[1.961]	[1.383]	[4.127]	[0.211]	[0.308]
Number of trained salons in the same area	0.027	0.02	0.022	-0.086	0.002	0.000
	[0.087]	[0.070]	[0.059]	[0.103]	[0.010]	[0.001]
Stylist sells other products in salon (0-1)	5.183***	3.254***	2.225**	7.244***	0.324**	0.630***
	[1.718]	[1.141]	[0.869]	[2.501]	[0.126]	[0.218]
Stylist in the bottom quartile of asset distribution (0-1)	1.159	0.506	0.532	1.761	0.056	0.113
	[1.724]	[1.290]	[1.022]	[2.701]	[0.157]	[0.193]
Stylist's socio-economic status is low (0-1)	-0.998	-0.511	-0.492	-2.839	-0.058	-0.162
	[1.410]	[1.131]	[0.861]	[2.112]	[0.137]	[0.156]
Stylist's dictator-game donation above the median (0-1)	3.364***	2.671***	2.166***	3.217**	0.430***	0.426***
	[1.137]	[0.889]	[0.698]	[1.625]	[0.098]	[0.134]
Stylist's reported work motivation is intrinsic (0-1)	-0.512	-0.668	-0.856	-0.079	-0.112	-0.039
	[1.328]	[1.040]	[0.826]	[1.848]	[0.112]	[0.150]
Stylist's religion is Catholic (0-1)	-3.652***	-2.509**	-1.536*	-5.126***	-0.272**	-0.413***
	[1.387]	[1.154]	[0.893]	[1.818]	[0.128]	[0.150]
Constant	0.431	2.231	3.904*	-1.787	0.850***	-0.26
	[3.851]	[2.726]	[2.000]	[5.942]	[0.303]	[0.483]
R-squared	0.0659	0.0748	0.0622	0.0988	0.0555	0.0217
Observations	765	765	765	491	765	6106
Large financial = Small financial (p-value)	0.819	0.532	0.587	0.947	0.654	0.803
Large financial = Stars (p-value)	0.012	0.00743	0.00829	0.029	0.0311	0.00152
Small financial = Stars (p-value)	0.006	0.00137	0.0018	0.037	0.00892	0.00479

Notes: OLS estimates. Standard errors are clustered at the cell level in Columns (1) through (5) and at the salon level in Column 6. * p<0.10 ** p<0.05 *** p<0.01. The dependent variable, *Packs sold (restocked)* is the total number of packs (excluding the initial dispenser sold at training) that the stylist chooses to buy and restock over the study period, based on invoices. Column (2) winsorizes *Packs sold (restocked)* at the 95th percentile for each treatment, and Column (3) does the same at the 90th percentile. Column (4) excludes salons that were dropped from or exited the program at any point. Column (5) reports the *Packs sold (restocked)* in natural logs. Column (6) is run at the salon*month level to facilitate the inclusion of sales-agent fixed effects. If no restocking visit was attempted, the observation is not included in the regression. Variables are described in Table 1. P-values in the bottom three rows are from a Wald test for equality of coefficients between treatments.

Figure A.1: Invitation letter

Become a CARE Promoter!

A great opportunity to help the fight against HIV/AIDS and promote your business!

_____ 2009

Dear Sir/Madam _____ of _____

Society for Family Health (SFH) wishes to invite you to enroll your salon in a CARE female condom promotion program. Your salon would become an official distribution point of the CARE female condom. This represents a great opportunity to improve your business performance through increased visibility and to contribute to the fight against HIV/AIDS in Zambia. **What's SFH?**

SFH is a non-governmental organization whose mission is to improve the health status of Zambians using social marketing techniques, increasing demands and supply of essential health products. Our programs include the promotion of CARE female condoms by hairdressers and barbers.

What's the advantage of joining the program?

As of now, numerous hair salons and barber shops in Lusaka, Chipata, Livingstone, and Kitwe have successfully joined the program. Hairdressers and barbers from these salons and shops tell us that *participating in the program has provided them with the immense satisfaction of helping their community and has attracted additional clients to come to the salon for other services.*

How do I join?

If you are interested in getting involved, we ask you to attend training on HIV/AIDS prevention, adequate use of the female condom and promoting and selling strategies. The training will be held

on _____

at **LUSAKA HOTEL (ON CAIRO ROAD, NEAR KATONDO STREET)** in Lusaka.

What happens at the training?

- SFH staff will teach you about the female condom (the product itself, how to use it adequately to prevent HIV/AIDS and pregnancy and how to promote it), prevention of HIV/AIDS transmission in general, and all-around promoting and selling strategies.
- SFH will provide lunch, tea break, an attendance fee of K40,000.
- SFH will have CARE available for everyone to purchase at a subsidized rate if they wish to sell CARE to their clients.

What we ask of you:

- To arrive promptly on time at the training: it will start at **8:30hrs**. Hairdressers reporting late for the training will be turned away.
- To bring the **invitation card** (see below). Hairdressers reporting without their invitation card will be turned away and will not receive the K40,000 attendance fee. The invitation is **exclusive** and **non-interchangeable**. You have been randomly selected to participate in this training because SFH does not have the resources at this point to train everyone. So, it is very important that if anyone attends this training, it is you. At registration, we will check if the invitation was addressed personally to you.
- To be committed to HIV, STIs, and unplanned pregnancy prevention and to want to teach your clients more about these issues.
- To be willing to sell the CARE female condom to your clients.

What happens after the training?

A SFH CARE monitor will visit your salon/barbershop every 5 weeks to:

- Record sales;
- Provide new stock of CARE female condoms;
- Provide continuous support and advice;

If you own this salon/barbershop as well as other salon(s)/barbershop(s) and one of your employees has already been invited to this training, please ignore this invitation. If not, your participation will be highly appreciated.

We thank you in advance for your usual cooperation.

Yours faithfully,

Miriam Mukamba, HIV Program Manager

INVITATION CARD #

CARE promoter training

_____ , Lusaka Hotel

This card needs to be presented to SFH staff on the day of the training to be allowed to attend.

FOR ONLINE PUBLICATION

Table B.1: Robustness check: Average treatment effects on calculated sales

Dependent variable	Packs sold (calculated)		=1 if sells at least one pack	=1 if sells 24 or more packs	=1 if sells 34 or more packs
Mean in control group	13.29	13.30	0.89	0.17	0.06
	(1)	(2)	(3)	(4)	(5)
Large financial reward	-0.9	-0.647	-0.019	-0.026	0.002
	[1.590]	[1.851]	[0.048]	[0.046]	[0.027]
Small financial reward	-0.538	-0.142	-0.04	0.04	-0.005
	[1.606]	[1.620]	[0.048]	[0.042]	[0.025]
Star reward	5.678**	5.996**	0.004	0.103*	0.083**
	[2.370]	[2.427]	[0.042]	[0.054]	[0.040]
Salon is a barbershop (0-1)		3.624**	-0.014	0.094**	0.034
		[1.490]	[0.029]	[0.036]	[0.028]
Salon is both a barbershop and hair salon (0-1)		3.009	0.028	-0.024	0.001
		[3.136]	[0.047]	[0.052]	[0.037]
Salon is near a bar (0-1)		0.796	0.025	-0.037	-0.005
		[2.004]	[0.046]	[0.050]	[0.036]
Salon size (log number of employees)		0.575	-0.017	0.006	0.054
		[2.678]	[0.048]	[0.054]	[0.036]
Number of trained salons in the same area		0.074	0.001	0.000	0.001
		[0.093]	[0.002]	[0.002]	[0.001]
Stylist sells other products in salon (0-1)		2.794*	-0.029	0.048	0.038
		[1.548]	[0.028]	[0.040]	[0.024]
Stylist in the bottom quartile of asset distribution (0-1)		0.322	-0.01	-0.011	0.015
		[1.630]	[0.035]	[0.040]	[0.029]
Stylist's socio-economic status is low (0-1)		-0.926	0.017	-0.027	-0.004
		[1.207]	[0.027]	[0.034]	[0.023]
Stylist's dictator-game donation above median (0-1)		2.234**	-0.005	0.02	0.036
		[1.123]	[0.028]	[0.028]	[0.022]
Stylist's reported work motivation is intrinsic (0-1)		-0.424	0.014	0.002	-0.025
		[1.191]	[0.021]	[0.029]	[0.025]
Stylist's religion is Catholic (0-1)		-3.215***	-0.004	-0.038	-0.045**
		[1.198]	[0.031]	[0.034]	[0.020]
Constant	13.295***	8.334**	0.884***	0.137	-0.014
	[1.182]	[3.851]	[0.075]	[0.085]	[0.055]
R-squared	0.0279	0.0547	0.0073	0.0289	0.0364
Observations	749	743	743	743	743
Large financial = Small financial (p-value)	0.812	0.753	0.534	0.0757	0.788
Large financial = Stars (p-value)	0.005	0.006	0.499	0.0127	0.0613
Small financial = Stars (p-value)	0.008	0.009	0.210	0.201	0.0339

Notes: OLS estimates. Standard errors are clustered at cell level. * p<0.10 ** p<0.05 *** p<0.01. The dependent variable, *Packs sold (calculated)* is the total number of packs sold over the study period (including the initial dispenser sold at training), based on representatives' calculations. The sample size varies across columns because of missing values in some covariates. 22 salons never completed a restocking visit (7 in the *Large financial reward* condition, and 5 in each of other three treatments) and hence are omitted from the analysis. Variables are described in Table 1. P-values in the bottom three rows are from a Wald test for equality of coefficients between treatments.

Table B.2: Customer survey

	Mean [SD]	Seen female condom ad in salon	Stylist talked about female condom	Ever used female condom
Mean in control group		*0.419*	*0.188*	*0.141*
	(1)	(2)	(3)	(4)
Market assigned to large financial reward	0.200	-0.061*	0.004	0.008
	[0.400]	[0.034]	[0.027]	[0.038]
Market assigned to small financial reward	0.324	-0.103***	0.018	0.019
	[0.468]	[0.021]	[0.020]	[0.035]
Market assigned to star reward	0.258	0.017	-0.016	0.042
	[0.438]	[0.019]	[0.017]	[0.040]
Female	0.386	-0.026	0.076***	0.044
	[0.487]	[0.021]	[0.017]	[0.026]
Age	29.31	-0.003**	0.000	0.003**
	[8.364]	[0.001]	[0.001]	[0.001]
Ever used female condom	0.132	0.070**	0.084***	
	[0.339]	[0.028]	[0.023]	
Seen female condom ad in salon	0.374		0.284***	0.087***
	[0.484]		[0.030]	[0.025]
Stylist talked about female condom	0.170	0.476***		0.783***
	[0.376]	[0.038]		[0.021]
Bought female condom in salon	0.0048	0.235***	0.377***	0.017
	[0.0690]	[0.077]	[0.089]	[0.014]
Has a committed partner	0.651	-0.005	0.021	0.031
	[0.477]	[0.027]	[0.019]	[0.026]
Has a casual partner	0.209	-0.006	-0.002	0.070***
	[0.407]	[0.036]	[0.026]	[0.019]
Uses male condoms	0.320	0.058**	0.02	-0.001
	[0.466]	[0.026]	[0.022]	[0.020]
Uses salon in market	0.608	0.037	0.034**	0.044**
	[0.488]	[0.022]	[0.012]	[0.017]
R-squared		0.172	0.179	0.0721
Observations	2089	1686	1686	1686
Large financial = Small financial (p-value)		0.266	0.578	0.631
Large financial = Stars (p-value)		0.0317	0.45	0.315
Small financial = Stars (p-value)		0.0000425	0.0585	0.424

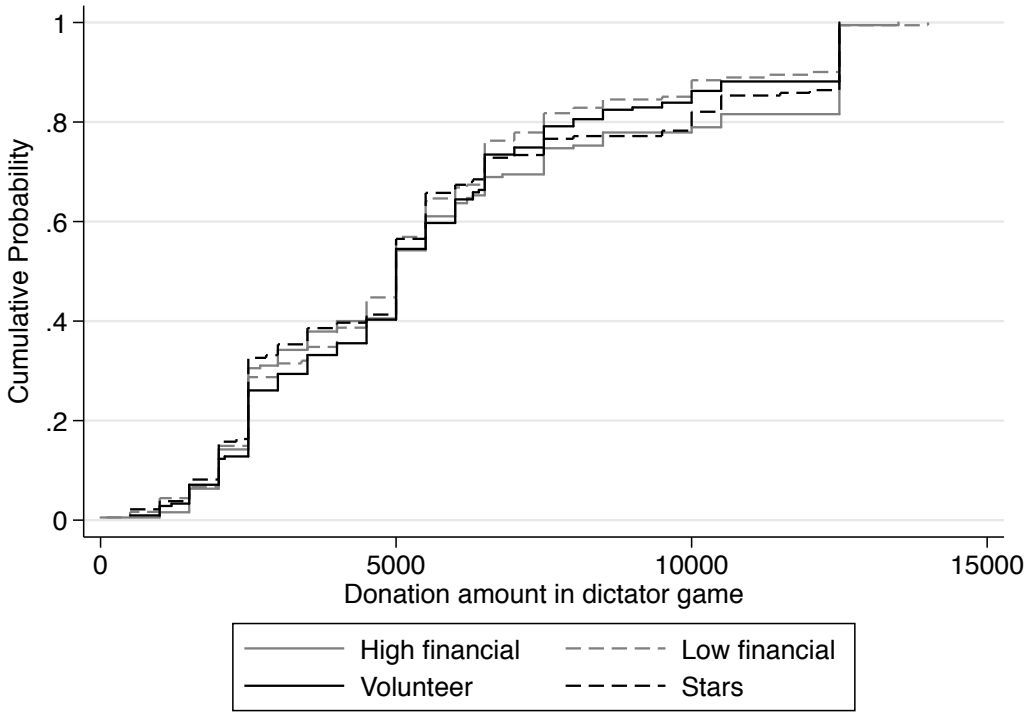
Notes: Results from OLS estimates. Standard errors are clustered at cell level. * p<0.10 ** p<0.05 *** p<0.01. The interaction terms between the treatment groups and the dummy variables about female condoms are not reported in the table for brevity. Only 10 customers in the survey reported having bought female condoms at the salons, so the results using this measure as the dependent variable is not reported. Sample size varies across columns because of missing values in some covariates. P-values in the bottom three rows are from a Wald test for equality of coefficients between treatments.

Table B.3: Robustness check: Average treatment effects on logbook sales

Dependent variable	Packs sold (logbook data)		=1 if logbook reports 24 or more packs	=1 if logbook reports 36 or more packs
Mean in control group	*13.74*	*13.75*	*0.146*	*0.0449*
	(1)	(2)	(4)	(5)
Large financial reward	0.406	0.535	-0.017	0.014
	[2.153]	[2.367]	[0.044]	[0.028]
Small financial reward	3.153	3.637	0.048	0.004
	[2.629]	[2.545]	[0.042]	[0.024]
Star reward	12.851***	11.785***	0.141**	0.118***
	[3.819]	[3.826]	[0.056]	[0.044]
Salon is a barbershop (0-1)		3.899*	0.065	-0.009
		[2.279]	[0.040]	[0.027]
Salon is both a barbershop and hair salon (0-1)		7.155	-0.042	-0.004
		[5.897]	[0.056]	[0.042]
Salon is near a bar (0-1)		2.139	-0.015	0.009
		[3.122]	[0.054]	[0.038]
Salon size (log number of employees)		5.917	0.062	0.011
		[5.372]	[0.056]	[0.041]
Number of trained salons in the same area		0.001	0	0.001
		[0.129]	[0.002]	[0.002]
Stylist sells other products in salon (0-1)		5.649**	0.013	0.036
		[2.423]	[0.037]	[0.025]
Stylist in the bottom quartile of asset distribution (0-1)		2.73	-0.014	0.056*
		[3.083]	[0.045]	[0.030]
Stylist's socio-economic status is low (0-1)		-3.588	-0.019	-0.016
		[2.372]	[0.038]	[0.023]
Stylist's dictator-game donation above the median (0-1)		4.191**	0.001	0.041*
		[1.735]	[0.029]	[0.021]
Stylist's reported work motivation is intrinsic (0-1)		-2.13	-0.01	-0.038
		[1.958]	[0.032]	[0.023]
Stylist's religion is Catholic (0-1)		-5.150***	-0.061*	-0.034
		[1.946]	[0.036]	[0.024]
Constant	18.025***	6.637	0.089	0.01
	[1.485]	[6.965]	[0.085]	[0.067]
R-squared	0.0649	0.13	0.0368	0.0528
Observations	651	649	649	649
Large financial = Small financial (p-value)	0.305	0.243	0.105	0.716
Large financial = Stars (p-value)	0.00148	0.00351	0.00555	0.0394
Small financial = Stars (p-value)	0.0202	0.0419	0.0989	0.0183

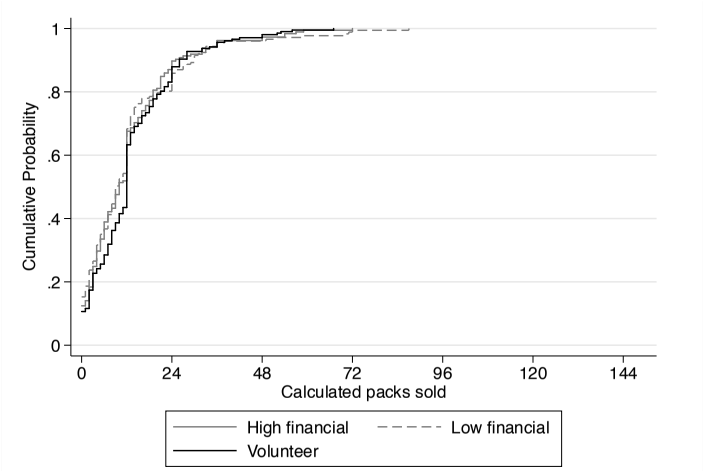
Notes: OLS estimates are weighted by the number of observations for each salon. * p<0.10 ** p<0.05 *** p<0.01. The dependent variable, *Packs sold (logbook data)* is measured as the total number of packs sold over the study period, according to logbooks kept by the stylists. Within the sample of stylists with available logbook data, all sell at least one pack according to the logbook data. We therefore omit the outcome *=1 if logbook reports 0 or more packs*. Variables are described in Table 1. The sample size varies across columns because of missing values in some covariates. P-values in the bottom three rows are from a Wald test for equality of coefficients between treatments.

Figure B.1: CDF of dictator game donations

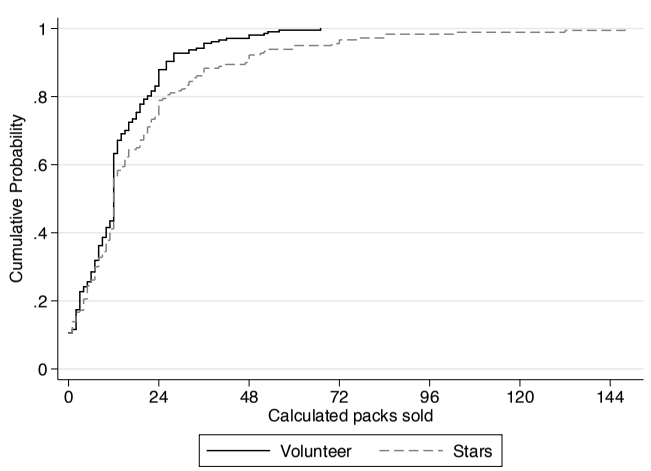


Notes: Cumulative distribution function of dictator game donations at training, by treatment group. Figure omits a single high outlier (=40,000 K) in the high financial reward treatment.

Figure B.2: CDF of calculated sales



(a) Volunteer, high financial and low financial reward treatments



(b) Volunteer and star reward treatments

Notes: Cumulative distribution function of packs sold (calculated), which measures the number of packs sold over the 10-month period (including the initial dispenser sold at training), based on sales representatives' calculations.