No margin, no mission? A Field Experiment on Incentives for Pro-Social Tasks*

Nava Ashraf (HBS and NBER); Oriana Bandiera (LSE); Kelsey Jack (Tufts)

October 24, 2011

Abstract

A substantial body of research investigates the design of incentives in firms, yet less is known about incentives in organizations that hire individuals to perform tasks with positive social spillovers. We conduct a field experiment in which agents hired by a public health organization are randomly allocated to four groups. Agents in the control group receive a standard volunteer contract often offered for this type of task, whereas agents in the three treatment groups receive small financial rewards, large financial rewards, and non-financial rewards, respectively. The analysis yields three main findings. First, non-financial rewards are more effective at eliciting effort than either financial rewards or the volunteer contract. The effect of financial rewards, both large and small, is much smaller and not significantly different from zero. Second, non-financial rewards elicit effort both by leveraging intrinsic motivation for the cause and by facilitating social comparison among agents. Third, contrary to existing laboratory evidence, financial incentives do not crowd out intrinsic motivation in this setting.

JEL codes: J33, O15, M52, D82

Keywords: incentives, non-monetary rewards, intrinsic motivation.

*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, Kelsey.Jack@tufts.edu. We thank seminar participants at Harvard Business School, London School of Economics, London Business School, NYU, University of Chicago (Harris), University College London, University of Maryland, Cornell University, the Boston Federal Reserve Bank, Tufts University, the BEAM conference, 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. The study was approved by the Harvard Business School IRB. Field implementation would not have been possible without the support of Mannaseh Phiri, Pamela Chama, Charles Kalonga, Nicholas Shilya, Richard Harrison, Marie-Helene Cloutier, Lauren Gaydosh, Arne Weiss, Mahjabeen Haji, Holly Hickling, and all of the monitors. The authors would like to especially acknowledge the dedication of Patrick Mulenga.

1 Introduction

Understanding what drives individuals to devote time and effort to different 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 with which they are affiliated. As a consequence, the design of optimal incentive contracts has been the subject of extensive economic research, both theoretical and empirical.

Empirical contributions, however, mainly focus on the effect of financial rewards in settings in which employee effort only benefits the employer (Bandiera et al. 2005, 2007, Lazear 2000). Very little attention has been paid to incentives in organizations, such as NGOs and charities, that pursue goals with positive social externalities and hire agents to perform pro-social tasks, namely tasks that create benefits enjoyed by those other than the employer and employees. Our paper begins to fill this gap by providing evidence from a field experiment designed to evaluate the effect of different incentive mechanisms on the performance of agents in a public health organization.

The theoretical literature suggests two reasons for why evidence from the private sector provides limited guidance on the design of effective incentive mechanisms for pro-social tasks, and for sectors that consist primarily of such tasks. First, mission-driven organizations benefit from matching with workers whose interests are aligned with the mission (Besley and Ghatak 2005) and these individuals might put low weight on financial gains. Second, to the extent that agents are motivated by pro-social or intrinsic motivation, financial incentives could actually reduce or "crowd out" such motivation, and reduce overall performance (Bénabou and Tirole 2003, 2006), especially when incentives are low powered (Gneezy and Rustichini 2000). These considerations raise the need for an alternative reward scheme, which can give agents a non-monetary stake in success and leverage the intrinsic motivation agents have for the pro-social task. Moreover, alternative rewards could also provide a cost-effective path for providing incentives to workers.

We implement a field experiment to test the effect of both financial and non-financial rewards on the performance of agents engaged in a task that has a pro-social component: the promotion and sale of female condoms for HIV prevention. To this end, we collaborate with a public health organization based in Lusaka, Zambia, that hires and trains hairdressers and barbers to sell condoms in their shops.[1]

The experiment randomly assigns approximately 1200 agents located in 200 distinct geograph-

[1]Using embedded community agents for health delivery is extremely common in developing countries, where community volunteers are often called upon to deliver needed services. For instance, as of June 2007, BRAC, one of the world's largest NGOs, relied on 67,000 community members to deliver basic health services to a population of 31 million in Bangladesh (BRAC 2007). In another well-documented example, community health workers in Uganda effectively distribute injectable contraceptives in rural communities (WHO 2007). School teachers are often relied on in health interventions to impart information and promote behavioral change among their students (see, as a successful example, teachers promoting safe water and hygiene, O'Reilly et al. 2007). However, the question of how to compensate these community agents remains a challenge for many non-profit employers (Bhattacharaya et al. 2001, Mathauer and Imhoff 2006).

ical areas to one of four groups. Agents in the control group receive a standard volunteer contract offered by NGOs, whereas agents in the three treatment groups receive small financial rewards, large financial rewards, and non-financial rewards, respectively. Financial incentives are provided by giving a margin on each condom sale, whereas the non-financial scheme provides social recognition in the form of stars posted on a thermometer displaying condom sales ("stars" treatment).

Our research design is informed by the theoretical insight that the effect of financial and non-financial incentives depends on the agents' intrinsic motivation and that this type of motivation can be crowded out, reducing performance. We thus design a modified altruism (dictator) game that yields a direct and quantitative measure of the agents' motivation for the cause.[2] The donation in the experimental game is a strong predictor of sale performance; agents who donate more than the median sell 51% more condoms than the average agent in the control group. We complement this measure by collecting information on a host of agents' characteristics that can proxy for their underlying motivation for selling condoms and their responsiveness to financial incentives, such as religion and self-reported profit motive.

Our design has three further features that allows us to shed light on the effect of monetary and non-monetary incentives and the underlying mechanisms. First, we offer both small and large financial rewards to test the hypothesis that motivation crowding-out is particularly detrimental to performance when rewards are small as argued in Gneezy and Rustichini (2000). Second, we measure agents' performance monthly over a one year horizon, thus allowing a longer follow up than in many employee experiments where changes in behavior may be artifacts of the experimental design, due to novelty or Hawthorne effects. Third, we exploit naturally occurring variation in the density of agents in each geographical sales area to provide evidence on the relevance of social comparisons in this setting.

We find that non-financial incentives are effective at promoting sales, while the effect of financial incentives is much smaller and not significantly different from zero. Agents in the star treatment sell twice as many condoms as agents in any other group, and the difference exists both on the extensive and intensive margins. Agents in the star treatment are 10 percentage points more likely to make any sale (a 32% increase with respect to the mean of the control group) and the difference is larger at higher quantiles of sales. We show that the differential effect of non-financial rewards is stable throughout the one-year period, ruling out that this is driven by a novelty effect.

Further analysis indicates that non-financial incentives work through two channels: they leverage intrinsic motivation for the cause and they facilitate social comparison among agents. In support of the first channel we find that non-financial incentives are more than twice as effective for agents who are motivated for the cause, as measured both by their donation in the dictator game and by personal characteristics correlated with motivation. In support of the second, we

[2] As is customary in economics we use the term "intrinsic motivation" to refer to the motivation for the task that relates to its pro-social effects. In contrast, the psychology literature distinguishes between intrinsic and pro-social motivation (both of which can be crowded-out)

find that the marginal effect of non-financial incentives is increasing in the number of neighboring salons that received the same treatment, whereas the effect of the other incentive mechanisms is zero throughout.

Finally, in contrast to existing laboratory evidence, we find no evidence that financial incentives crowd out intrinsic motivation. On the contrary, high financial rewards are more effective for agents who score higher on our motivation measure.

Our research design allows us to rule out that the star treatment increases sales by affecting demand, e.g. through advertising. Our strategy has two prongs. First, we survey a random sample of customers to probe the effectiveness of promotional materials such as posters that are given to agents in all treatments vis-a-vis the thermometer that is only given to agents in the star treatment. The survey reveals that customers are aware of the former but not of the latter. Second, we implement a "placebo thermometer" treatment, namely we randomly allocate a subsample of salons in treatments other than the star treatment to receive a thermometer that looks identical to the treatment thermometer except that it reports the average sales in the area, rather than individual sales. In contrast with the advertising hypothesis, we find that the placebo thermometer has no effect on sales.

Our paper is the first to provide evidence from a field experiment designed to compare financial and non-financial incentives for employees in a social organization. We contribute to a new and rapidly growing literature that uses field experiments to identify the causal effect of incentives on performance. The literature to date has focussed on financial incentives for private sector employees (see Bandiera et al. 2011 for a review), and for teachers, especially in developing countries (Glewwe et al. 2010, Muralidharan and Sundararaman 2011). While recent theoretical contributions make precise the role of non-financial incentives, especially status awards (Besley and Ghatak 2008, Moldovanu et al. 2007), evidence on the effectiveness of these in the workplace remain scarce. A recent exception is Kosfeld and Neckermann (2010) who design a field experiment to evaluate the effect of symbolic awards for students hired for an occasional two-hour data entry job for an NGO. In line with our findings, they also show that symbolic awards—in their case a congratulation card to the best performer—are effective at eliciting effort. In addition, however, our design allows to compare the effectiveness of financial and non-financial rewards in the same setting and to test whether such effects are sustained over an entire year, thus separating novelty effects from long-run effort responses.Evidence on the effectiveness of recognition and status rewards also comes from recent laboratory experiments (Alpizar et al. 2008, Ball et al. 2011, Eriksson and Villeval 2010) and from pro-social behavior outside the workplace, e.g. blood donations and charitable giving (Lacetera and Macis 2008, Landry et al. 2011, Soetevent 2005). Our paper also contributes to the literature on crowding out providing the first evidence from the workplace to complement existing evidence from laboratory and field experiments on charitable giving (Ariely et al. 2010, Gneezy and Rustichini 2000, Lacetera et al. 2011, Mellström and Johannesson 2008).

The rest of the paper proceeds as follows. Section 2 provides a basic theoretical framework to guide the empirical analysis, Section 3 describes the context, data sources and the research design. Section 4 discusses the identification strategy. Section 5 and 6 present the findings and Section 7 concludes with a discussion of external validity.

2 Theoretical Framework

2.1 Set-up

We develop the simplest possible theoretical framework that allows us to flesh out the difference between financial and non-financial incentives in our setting. The framework is designed to capture the fact that individuals might derive utility both from money and from the production of social value, and that financial incentives might crowd out the latter. There is one principal, "the organization" that hires one agent to produce output Y, where $Y = f(e) + \varepsilon$, e is the agents' effort, and ε is a random shock with mean 0 and variance ϱ^2, so that $E(Y) = f(e)$.

Output has positive externalities for the community. To use our empirical context as an example, Y represents condom sales that generate both revenue for the organization and positive externalities in the local community by slowing down the diffusion of HIV. Effort is non-observable, and because of the random disturbance ε, Y is not a perfect signal for it, so the organization cannot infer e by observing Y. We assume that $f' > 0$, $f'' \leq 0$, namely output is increasing in agent's effort, at a non-increasing rate. We assume both parties are risk neutral.

We assume that the agent's payoff has two components. The first is the monetary rewards he gets from the organization. The second is the non-monetary payoff he gets from contributing to the cause, that is from producing output that entails positive externalities for others in his community. This captures the fact that agents in social organizations might be motivated to exert effort "intrinsically", that is even if their performance does not affect their monetary pay-off. We analyze the case when the organization has two instruments to leverage the monetary and non-monetary components of the agent's payoff, respectively. The organization can affect the agent's monetary pay-off by offering financial bonuses, which, for simplicity, we assume to be linear in performance. In addition, the organization can affect the agent's non-monetary payoff by offering non-financial rewards, such as recognition or status goods, linked to performance.

The agent's monetary payoff is $\phi(w+mE(Y(e))$, where ϕ is the weight the agent gives to money, w is the baseline wage and m is the financial bonus. The agent's non-monetary payoff is equal to $\sigma(1+r)E(Y(e))c(m)$, where σ the weight the agent gives to non-monetary benefits whereas r measures the level of non-financial incentives offered by the organization.[3] The specification

[3]Alternatively, it could be argued that the agent's non-monetary payoff depends on the level of effort he devotes to the cause, instead of the output that effort produces. We prefer this specification because in our experiment the non-monetary incentives r are a function of output, not effort. Moreover, given the assumption of risk neutrality, the

captures the intuition that agents "care" about the organization's performance even if they are not given incentives for it.

The last term, $c(m)$,$c' \leq 0$, $c(0) = 1$, captures the possibility that non-monetary benefits are crowded out by financial rewards. The formulation is a reduced form representation of many types of social preferences, and many motives for crowding-out. For instance, it might be that the agent cares about $E(Y(e))$ because he is altruistic towards the people who benefit from the output he produces - such as those whose risk of HIV infection decreases in the example above - or because he cares about acquiring a reputation for altruism with others in his community (Benabou and Tirole 2003, 2006, Ariely et al. 2009). Finally, we assume that effort is costly for the agent and denote the disutility of effort by $d(e)$, $d' > 0$, $d'' > 0$.

2.2 The optimal response to incentives and crowding-out

The agent chooses effort to maximize:

$$max_e\{\phi(w + mE(Y(e)) + \sigma(1 + r)E(Y(e))c(m) - d(e)\} \tag{2.1}$$

s.t. $e \geq 0$. The optimal level of effort $e*$ satisfies the first order condition:

$$\phi m f'(e) + \sigma(1 + r)f'(e)c(m) = d'(e) \tag{2.2}$$

where the first and second terms represent the marginal benefit of effort on monetary and non-monetary pay-offs, respectively. The first order condition makes clear that the agent has two motives to exert effort in this setting, namely to increase earnings and to contribute to a cause he cares about. Note that if $\sigma = 0$, i.e. if the agent does not care about the cause, effort levels are solely determined by financial incentives, so that the agent exerts the minimum feasible amount of effort if his pay is not tied to performance, i.e. if $m = 0$. Note also that if $\sigma > 0$ the optimal level of effort is positive even if the agent is not offered a non-financial reward (r=0). This captures the fact that agents who are intrinsically motivated for the cause ($\sigma > 0$) exert effort even in the absence of tangible rewards.

The agent's optimal effort response to monetary incentives is:

$$\frac{de*}{dm} = -\frac{\phi f'(e*) + \sigma(1 + r)f'(e*)c'(m)}{\phi m f''(e*) + \sigma(1 + r)f''(e*)c(m)}. \tag{2.3}$$

By the second order condition the denominator is negative, so the sign of (2.3) is determined by the sign of the numerator. The first term in the numerator is positive and represents the increase

two formulations are equivalent in our theoretical setting. If the agent were risk averse, however, the level of utility from contributing to the cause would differ depending on whether we model the non-monetary payoff as a function of effort or output.

in monetary payoff, the second term is negative and represents the decrease in non-monetary payoff due to the fact that an increase in the strength of financial incentives crowds out intrinsic motivation ($c'(m) < 0$). Equation (2.3) thus illustrates that when agents are intrinsically motivated, monetary incentives can backfire and reduce effort if the crowding-out effect is stronger than the effect on the increase of the financial payoff. This is more likely to occur when the agent's weight on financial payoff ϕ is low and when his weight on non-financial payoff σ is high. The data contains proxies of ϕ and σ that can be used to test whether the effect of financial incentives is heterogeneous as predicted by the theory. Moreover, the shape of the crowding-out function $c(.)$ determines the magnitude of this effect at different levels of incentive power m. For instance, $c''(.) < 0$ would imply that motivation crowding-out is particularly detrimental to performance when incentives are low powered as argued in Gneezy and Rustichini (2000).

The agent's optimal effort response to non-monetary incentives is:

$$\frac{de*}{dr} = -\frac{\sigma f'(e*)c(m)}{\phi m f''(e*) + \sigma r f''(e*)c(m)} \tag{2.4}$$

which is always positive, as $\sigma f'(e*)c(m) > 0$, namely there is no drawback to increase the power of non-financial incentives as they do not crowd out any type of motivation.[4] The strength of non-financial incentives depends on the weight the agent puts on his non-financial pay-off σ, namely on how much he cares about the cause. We will test for this by allowing the effect of non-financial incentives to be a function of the empirical proxy of σ.

3 Context and Research Design

3.1 Context

The field experiment was run in collaboration the Society for Family Health (SFH), a public health organization based in Lusaka, Zambia. The experiment was embedded in SFH's new program for the distribution of female condoms through hair salons, and we collaborated with SFH closely at each stage of the program, including salon selection, training, incentive design and monthly sales monitoring for one year from December 2009 to December 2010.[5]

Like many NGOs in developing countries, SFH relies on community members to implement

(e.g., teachers, community leaders) and engage in pro-social public health related tasks in addition to their main income generating activities. In the program under study, hairstylists were chosen as ideal promoters of female condoms both because the familiarity between the stylist and the client creates the potential for successful targeting of female condom 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 the necessary information. Finally, hair salons are numerous and distributed throughout the city. Our census of salons, implemented as part of the research design, found just over 2500 hair salons, serving a population of about 2 million (2,198,996, according to the 2010 Census of Population and Housing for Zambia).

The first stage of the program consists in distributing invitation letters to hairstylists. The letters invite them to attend a one-day training program after which they will be given the opportunity to join the female condom distribution program. In case of multi-stylists salons, the invitation is extended to the person responsible for the management of the salon, which is either the owner or, if they are not directly involved in salon activities, the general manager. During training, stylists are provided with information on HIV/AIDS, female condom promotion, basic business skills and program details, including the compensation package.[6]

At the end of training, stylists decide whether to join the program. Those who join buy condoms from SFH to sell in their shops. The purchase and resale price is set at K500 for a pack of two condoms, the same as the male condom. SFH provides a range of promotional materials including posters and display units, and SFH representatives ("monitors") visit salons once a month to allow stylists to buy more condoms and answer queries about the program. Stylists can purchase their first condom dispenser (containing 12 packs) at training at the subsidized price of K2000. After that, dispensers or single packs can be purchased at K500 per pack either at the monthly visit or by phoning a toll-free number service dedicated to the female condom program. The number was set up to allow agents to purchase condoms if they missed the monthly visit, or if they run out in between visits.

In this context, the agents' choice variable is the level of effort to devote to the promotion and sale of female condoms. As this is a new product customers are not familiar with, the agents have to exert effort in explaining the female condom's characteristics, mode of use and benefits, to persuade customers to make a purchase. For repeat customers, the hair stylists have the opportunity to follow up on first time purchases 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

[6]Participants were offered a K40,000 show up fee. This is about eight times the average price of a hair wash service and is thus likely to cover the stylists' opportunity cost of time for a week day. In 2009, USD 1 = K5,000.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 program and materials were designed by the research team in consultation with external communication experts. The training was conducted using a variety of pedagogical approaches (lectures, exercises, games, role-play, etc.) and teaching material (individual handouts, flipcharts, videos, etc.).

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 to prevent the diffusion of HIV/AIDS, which undermines economic growth in many sub-Saharan countries. Zambia has one of the world's highest adult HIV prevalence rates at 14.3% (CSO et al. 2008). It is estimated that in 2009, 1 million people were living with HIV and 45,000 died of HIV related causes (UNAIDS 2010). Stylists are aware of the pro-social nature of the task because of extensive informational campaigns run by the Ministry of Health on the importance of condoms for HIV prevention and their benefits for society. In addition to the social benefit, condom sales might carry private benefit depending on which compensation scheme the stylists are offered, as described in the next section.

3.2 Research Design: Treatment Groups

Following the framework above, our experiment is designed to test the effect of financial and non-financial incentives on agents' and performance as indicated by the agent's first order condition (2.2). Agents are randomly assigned to one of four groups. Agents in the **control group** are hired as volunteers, namely they receive no incentives, financial or otherwise. This is a common arrangement in organizations that rely on the help of community members.

The incentive schemes are designed to match the theoretical parameters m and r as described above. In addition, we offer both small and large financial rewards to test the hypothesis that motivation crowding-out is particularly detrimental to performance when rewards are small as argued in Gneezy and Rustichini (2000).

Agents in the **small financial reward treatment group** receive K50 for each condom pack sold, a 10% margin over the retail price. K50 is the smallest bill commonly in circulation, making this the smallest payment that is easily implementable. This treatment corresponds to a small m in the theoretical framework.

Agents in the **large financial reward treatment group** receive K450 for each condom pack sold, a 90% margin over the retail price. K450 is the highest incentive compatible reward, as agents would have the incentive to buy and dispose of the condoms if the reward were larger than the purchase price. This treatment corresponds to a large m in the theoretical framework.

Agents in the **non-financial reward (stars) treatment group** receive a star for each condom pack sold. These agents are provided with a thermometer, akin to those used by charitable fundraisers, which they are instructed to post in a visible location in the salon/shop. Each sale is rewarded with a star stamped on the thermometer, which is labeled as measuring the stylist's contribution to the health of their community.[7] The thermometer is designed to create a visual link

[7]Ball et al. (2001) use stars to confer status in an experimental market and find that status is associated with obtaining a greater share of the surplus in the experimental transactions, regardless of whether the stars are earned

between packs sold and lives protected, making social impact salient (Grant 2007) and effectively rewarding stylists for marginal contributions to the cause. This corresponds to the parameter r in the model. Stars, however, have no resale value thus their effect on performance, if any, cannot be driven by financial gains. In addition, stylists who sell more than 216 packs during the experimental year are invited a ceremony at SFH headquarters together with five guests of their choice. During the ceremony the stylist is awarded a certificate by a well-known and respected figure in the health sector in Zambia.

To ease comparison across treatments, both financial and non-financial rewards are given for each pack sold and have a similar linear structure. The difference between treatments, if any, cannot therefore be ascribed to differences in incentive structure, e.g. difference in convexity or induced competition, which would arise, for instance, if one of the treatment had a tournament structure such that only the best performer(s) would receive the reward.

In all treatment groups, rewards are calculated based on restocking decisions during the previous monitoring visit and are paid by monitors at each monthly visit. To measure performance we use restocking rather than sales because restocking can be precisely measured through invoice and inventory data, whereas we cannot monitor sales directly and stylists might intentionally misreport them or report them with error. It is important to note that none of the incentive treatments makes it worthwhile for agents to buy stock if they do not plan to sell it. Indeed, even in the large financial reward treatment, the reward is less than the price of a pack. Thus restocking choices are a good proxy for sales. Finally, rewards are paid at the monthly visit after the re-stocking purchase to avoid delegating the computation of rewards to the monitors and to make sure they have the exact amount of rewards to distribute at each round.[8]

All stylists were told how many they restocked and, when applicable, how much reward they receive as a result, as well as the number of potentially protected sexual intercourses to which the number of packs corresponded, linking their effort to the social outcome of prevention of HIV/AIDS and unplanned pregnancies.[9]

based on performance in a trivia quiz or allocated randomly.

[8]All of our main results are robust to using sales based on stock observed by the monitor as our outcome measure. Incentive payments are delivered 5 weeks after the restocking purchase to avoid delegating the computation of incentives to the monitors. 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 monitoring visit. Monitors 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. We record these restocking decisions separately. Stylists in the high financial rewards 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 do significantly increase their restocking decision after receiving their incentive, relative to the volunteer control group.

[9]The following script was read to all groups, including the volunteer control, at the end of the monitoring session: "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..... sexual intercourses. You have therefore helped your clients protect themselves against STIs and unplanned pregnancies." In the reward treatments, they were also told "Because of your hard work

3.3 Research Design: Randomization

To minimize the risk of spillovers between treatment groups, randomization is carried out 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 a wealth of salon and stylist characteristics. We then imposed a grid on the electronically mapped locations of the salons, to divide the city into equal geographical areas of 650 square meters each. Within each area, we cut out a buffer of 75 meters on each side. The resulting areas, each measuring 500 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 1222 hair salons.[10]

To increase power we balance on a vector of variables that are likely to affect condom sales. These are: salon type (hairdresser, barber or mixed), 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 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 1000 random draws. Figure 1 illustrates the outcome of the randomization.

While randomization occurred before the training invitation letters were delivered, the letters themselves contained no information on the assigned treatments, to minimize the risk of spillovers as letters might have travelled across neighborhoods. This also ensures that attrition between the randomization and training stages is orthogonal to treatment, as shown in Section 4.1 below.

3.4 Research Design: Data Sources

The analysis will use six main sources of data on stylists and condom sales. First, we conducted a census survey on all stylists in Lusaka, through which we collected information on characteristics of both the salon and the respondent.[11] Eligible respondents were defined as those who could join the female condom sale program, namely those in charge of the daily management of the salon. These were typically the owner or the person managing the salon/shop on behalf of the owner for

and great sales performance in the last month, you have earned a reward of ..... (Kwacha or stars)".

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

[11] 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. Information 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 time since opening. Information on 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.

at least 4 days a week.[12]

Second, during the training program we implemented a contextualized dictator game to elicit incentive compatible measures of pro-social attitudes toward HIV causes. Participants were told that, in addition to the show-up fee (K40,000), each of them would receive K12,500 which they could keep for themselves or donate in part or in full to a well-known charity in Lusaka that provides palliative care, including antiretroviral treatment for their HIV patients.[13] The amount donated is taken as a proxy for the agents' motivation for the cause. As this is likely to be correlated with the agents' wealth, it is always used together with asset and socioeconomic status measures in the analysis that follows.

Third, we use SFH inventory records to build a precise measure of restocking, i.e. agents' purchases of condoms, with monthly frequency. This is the measure used to compute incentive payments and is our main performance measure in the analysis that follows.

Fourth, monitors recorded restocking decisions and sales, and collected information on a number of program related issues such as the visibility of promotional material at each monthly visit.[14] In addition, monitors check the logbooks stylists are asked to keep to record condom sales and customer characteristics related to their HIV risk profile. While the monthly monitoring surveys are a useful complement to the information gathered from SFH records, it is important to notice that these are available only if the salon was open and the trained stylist was present when the monitors visited their neighborhood.[15] Overall, 60% of all attempted visits were successful, and these are equally distributed across treatments.

[12] If the desired respondent was not present at the time of the monitor's visit, an appointment was scheduled if possible or the monitors returned on the following days. A maximum of three attempts were made.

[13]Specific instructions for the game were scripted and read out loud. The scripts 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 K12500. [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) treatment 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 K12500, 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 K12,500 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 5 booths where they counted the sum and separated the amount they kept from the amount they donated. The bills donated were place in an envelope sealed before leaving the booth. Each participant then deposited themselves the envelope in a box specially designed for this purpose sitting in front of the room.

[14]Five full time monitors were trained to carry out visits and they rotated between salons and treatments. Monitoring visits lasted approximately one hour, during which follow a detailed script and recorded both observational and survey data. Besides collecting data, monitors answer queries about the program, distribute promotional materials, allow the stylists to restock and hand out incentive payments.

[15]Monitors stopped visiting salons after three failed attempts in three consecutive months. These stylists, 218 in

Fifth, two months before the end of the program, we administered a customer survey to investigate the customers' familiarity with the female condom distribution program through hair salons, and their use of female and male condoms.[16]

Finally, at the end of the program we re-administered the baseline census questionnaire augmented by modules on business skills and on own health behavior.[17]

4 Identification

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

$$y_{ic} = \alpha + \sum_{j=1}^{3} \delta_{0j} treat_c^j + u_{ic} \tag{4.1}$$

where y_{ic} is a measure of condom sales by agent i located in area c over the year, and $treat_c^j$ denotes the three treatment groups. Errors are clustered at the level of the randomization unit, the geographical area c, throughout. We estimate (4.1) on the entire sample of stylists who came to training and hence were exposed to treatment. As 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 as almost all the agents who are exposed to treatment join the program. Section 5 presents detailed evidence on this issue.

The coefficients δ_{0j}measure the causal effect of the treatments on sale performance under the identifying assumption that $treat_c^j$ is orthogonal to u_{ic}. In support of this assumption, appendix table A.1 presents the means and standard deviations of agents' and salons' characteristics in each treatment, together with the largest normalized difference between treatment pairs. The table reports both the variables used to balance in the randomization procedure and additional determinant of sales used later in the analysis. Reassuringly, the randomization yields a sample that is balanced across treatments; out of 66 pairwise differences, only one is just above .25 standard deviations.[18] This notwithstanding, the identifying assumption fails if the decision to participate to

[16]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 entry, 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 to match customers with treatments. Customers were then asked a very brief set of survey questions about demographics, familiarity with the female condom, sources of information, purchase behavior and own sexual practices.

[17]At endline we re-interviewed 69 percent of the stylists from the original sample who attended training. At the time of the endline survey, stylists were reminded that the monitoring visits would not be continued but that they would be able to restock female condoms directly from SFH sales agents if they wished to continue distributing the product.

[18]The number of stylists reporting that profit is their primary motivator in their daily work is 0.26 standard deviations higher for stylists in the stars treatment than for stylists in the volunteer treatment. We control for this and all other stratification variables in the specifications below.

the training program is not orthogonal to treatment or if there are spillovers between treatments. We discuss these in turn below.

4.1 Participation decision

Of the original sample of 1222 stylists chosen to participate in the experiment and randomly allocated to one of the four groups, 771 chose to came to training and were therefore exposed to the treatments. Stylists drop out of the sample at two points between treatment assignment and treatment exposure. First, 20% of the 1222 stylists assigned to receive training invitations did not receive them as they could not be found during any of the three delivery attempts. Second, 21% of the 981 stylists who received an invitation chose not to attend the training. 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

$$p_{ic} = \alpha + \sum_{j=1}^{3} \theta_{0j} treat_c^j + X_i \eta_i + \varepsilon_{ic} \tag{4.2}$$

where $p_{ic}$is an indicator variable equal to 1 if the agent receives the invitation 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 and later with sales. Two findings are of note. First, 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. Second, the decision to attend training is correlated with some individual characteristics such as gender (barbers are more likely to attend, presumably as many already sell male condoms), and self-reported donations to HIV related causes as stylists who attend training are five percentage points more likely to report giving to HIV charities- a 23% increase over the mean. This is in line with the theoretical literature that suggests agents in mission-driven organizations share interest in the mission (a low ϕ and/or a high σ in the model) and has implications for their response to financial and non-financial incentives compared to the general population, which we will discuss in detail in the conclusions. We also note that the decision to attend is positively correlated with the number of stylists operating in the same neighborhood, suggesting peer effects might be relevant in this setting, an issue we will return to when exploring the mechanisms driving the effect of incentives.

4.2 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 buffer zone around the edge of each geographical areas where salons are located to ensure that each agent either neighbors 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 areas are assigned to the same treatment, this precaution can be undone by stylists relocating after randomization is carried out.[19] Relocated stylists were allowed to stay into the program only if they moved within the same geographical area or to a new area with the same treatment as they were originally allocated to.

Second, stylists attended the training only 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. Finally, 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 type of posters. The sole exception to this rule is that only stylists in the stars treatment receive a thermometer poster.

To mitigate information spillovers that might arise in spite of these design efforts, participants were told that the program was being rolled out to hairstylists across Lusaka, and that some training sessions may emphasize different aspects of the program. They were told not to be surprised to find that the program is slightly different in places.[20] Though this explanation may have reduced stylists surprise at learning about other incentives, we attempt to track possible spillover opportunities that may have still affected effort.

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 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 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. Reassuringly, 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.

To detect spillovers and identify the stylists who might be affected by them during the course

[19]Only 12 cases occurred where the salon moved and remained in operation, 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 involvement movement into a buffer area and the remaining 2 cases involved relocation to a different treatment. These salons were dropped from the program.

[20]Participants seemed to accept this explanation. In a debriefing after a pilot training, they mentioned that companies selling new products often have different programs in different neighborhoods.

of the experiment, we asked monitors to note all questions and complaints at every monthly visit. In over 7,000 monitor visits, only one stylist asked about different incentive schemes.

While these three pieces of evidence are reassuring, they cannot completely rule out that agents in one treatment effectively responded 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.

5 The effect of incentives on sale performance

We begin by evaluating the effect of the three experimental incentive treatments on overall sale performance during the experimental year. We measure sale performance by the number of packs restocked over the year, namely the number of packs agents purchase from SFH to sell in their salons. Restocking is precisely measured from SFH inventory data and checked against invoices signed by the agents upon purchase. Most importantly, restocking is the performance measure used to compute financial and non-financial rewards.[21] The difference between sales to customers and our measure of performance is the number of packs bought by the stylists but left unsold in the salons. To measure sales to customers we asked monitors to record the number of condom packs in the salons at every visit; sales to customers are therefore measured with error to the extent that unsold packs might be not be displayed. Despite this, the correlation between the two measures is 0.92 and similar across treatments.

During the course of the experimental year, agents sold 13,886 female condom packs, of which 5,332 were sold by agents offered non-financial rewards, and the remaining 8,554 is roughly equally divided across the other three treatments. Figure 2 reports average yearly sales by stylists in the four treatment groups. Two patterns are of note. First, there is a striking difference between stylists in the star treatment and all the others. Agents in the star treatment sell twice as many packs (14 versus 7). Second, sales levels are generally low; even in the star treatment, the average stylist sells slightly more than one pack per month. This is in line with qualitative evidence that female condoms are difficult to sell. [22]

To evaluate the effect of incentive treatments controlling for stylists' characteristics we estimate:

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

where y_{ic} is a measure of sales by agent i in area c, the variable $treat_c$ identifies the incentive treatment agent i is assigned to, and X_i is a vector of salons' and agents' characteristics that can

[21] Accordingly, we do not count the first dispenser (12 packs) bought at training because no rewards were paid for this.

[22] Stylists in our sample report that customers are afraid to try the product because of rumors about discomfort or malfunctioning. Successful sellers report the need to follow up with customers at least once or twice, since the product becomes easier to use with practice.

affect the willingness or ability to sell female condoms. Errors are clustered at the level of the randomization unit, the geographical area, throughout.

We estimate (5.1) on the entire sample of stylists who came to training and were exposed to the treatments, regardless of whether they joined the program.The coefficients δ_{0j} thus capture the effect of the incentive treatments on sales through both selection and effort. In practice, however, only 3% of the stylists who came to training did not join the program, hence incentives have no effect on selection as shown in columns 1 and 2 of appendix table A.3. Two points are of note. First, to join the program stylists had to purchase a minimum of 12 packs at the subsidized price of KW2000, which corresponds to 2/3 of the average price of a haircut in our sample. The fact that joining the program is costly allays the concern that the joining decision is moot, namely that agents might have agreed to join without ever intending to participate actively. Second, as the NGO is well known for using existing retail networks to distribute health products, and the invitation letter stated the program was an opportunity to "help the community", the stylists who selected into training were probably willing to work for little or no reward. Table A.2 indeed shows that the characteristics of the stylists who came to training differ from the general population of stylists, in particular those who chose to participate are more likely to report donating to HIV related causes. Taken together, these suggest that selection mostly took place before stylists knew about the treatment and hence it reconciles our evidence with earlier findings suggesting that incentives affect selection both in the field and lab (Bandiera et al. 2007, Dohmen and Falk 2011, Larkin and Leider 2011, Lazear 2000, Lazear et al. 2007).

Besides choosing to join the program at the beginning, stylists could also choose to quit during the course of the experimental year at no cost. Only 58 stylists (7% 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 and significantly different from zero ($p = 0.077$) only for agents in the small financial reward treatment as shown in appendix table A.3. Overall, only 10% of the 771 stylists who were exposed to treatment select out of the program either after training or later during the year, and the incentive treatments do not have a substantial impact on either selection decision. This implies that the coefficients δ_{0j} capture the effect of incentives on sales through effort rather than through selection.

To measure the impact of incentives on sales, the first two columns of table 1 estimates (5.1) using the total number of packs as the outcome variable. Four findings are of note. First, agents in the star treatments sell 7.54 more packs, that is over twice as many packs as stylists in the control group and this is robust to including stylist's, salon's and area characteristics.[23] Second,

[23]We note that agents in the large financial reward treatment face a lower marginal cost (50 instead of 500) and could, in principle, have boosted 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 maximize sales revenues. Moreover, while this does not invalidate the identification of the effect of incentives on sales performance, it might cloud the identification of the effect of incentives on effort. We do not observe agents choosing this strategy in equilibrium. Our follow up survey shows

neither financial treatment affects sales. Both coefficients are orders of magnitude smaller than the coefficient on the star treatment and 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% level. 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 sell 3.34 more packs, which is equal to 44% of the effect of star rewards and over 50% the baseline mean of 6.96 in the control group. To allay concerns that this captures differences in wealth the regression includes a measure of asset value. This is correlated with the value of donation, as expected, but not with sales. As self-reported asset value might be measured with substantial noise, we also use information on whether the agent has completed primary education and whether they speak English, which are the best proxies of socio-economic status in our setting. This measure is also correlated with donation but not with sales. Fourth, the following agent's characteristics are correlated with sales: barbers sell 3.23 more packs, possibly reflecting the fact that men are in charge of contraceptive choices in our setting, promoters with previous sale experience sell 5.21 more packs and Roman Catholics sell 3.61 fewer. The effect of the star treatment is thus larger than the effect of any personal characteristics.

Columns 3,4, and 5 estimate treatment effects on the extensive margin and at different points of the distribution of sales. The distribution exhibits bunching at 0, 12 and 24, probably due to the fact that while stylists could purchase one pack at the time from SFH, buying one dispenser (12 packs) would save on transaction costs. Overall, 62% of stylists sell no packs other than those purchased at training, 22% sell between 0 and 12, and 16% sell 24 or more.

Column 3 of table 1 shows that the likelihood of selling at least one pack is 12 percentage points higher for agents in the star treatment; this represents a 33% increase over the mean of the control group. Agents in the high and low financial reward treatments are equally likely to sell at least one pack as agents in the control group. Columns 4 and 5 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% more than stylists in the volunteer treatment, and 10 percentage points more likely to sell 24 or more, which is 80% more than stylists in the volunteer treatment. Promoters who are offered financial rewards, either large or small, do not perform differently than stylists in the control group. All coefficients are precisely estimated and very close to zero.

Taken together, the evidence in table 1 indicates that non-financial incentives are effective at promoting sales in this context, whereas financial incentives are not. Before delving into the mechanisms that underpin our findings, this section presents evidence on two issues that are key for the interpretation of the findings.

---

that only four stylists reported ever selling a pack at a price lower than KW500, and none of them was in the large financial reward 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.

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 neighbor areas in the star treatments, whereas others do not. We exploit this variation to test whether the agents who are more likely 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.

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 (5.1) in each month, using the same set of controls and clustering errors at the level of the randomization unit as above.[24] 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 fact that the torrential rains in months 3 and 4 depressed 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 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 rewards is close to zero in all months, suggesting again that the aggregate results do not hide substantial heterogeneity through time. [25]

That the effect of non-financial incentives is stable through time suggests that this is unlikely to be driven by the prospect of 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 and indeed only one managed to reach it. Had the effect of non-financial incentives been driven by the ceremony component, 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 given a

[24] In contrast to restocking data, which is available for each month of the year, customer sales data is only available for the months in which the enumerators were able to find the stylists when visiting their area. Overall, 60% of the visits were successful, and this does not vary by treatment. The most common reasons for a missed visit were that the shop was closed or that the stylist was not present.

[25] The observed pattern is also 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. 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 they sold female condoms to 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% in month 1) but moo thcus4()-0.2(n)2oo(a-0.131.oo)-3574(o)0.5(a)0.1me ien mo

number (216) they might have used to form expectations about "reasonable" sales levels. As sales were substantially below the level needed to reach 216 over one year, this effect should have also disappeared after the first months.

6 Mechanisms

The evidence in the previous section indicates that, in this setting, non-financial rewards are effective at increasing sales, whereas financial rewards are not. This section provides evidence on the mechanisms that underlie the treatment effects estimated above. As the evidence in section 5 indicates that the difference between treatments is stable throughout the duration of the experiment, the remainder of the paper will focus on aggregate performance at the year level.

We begin by making precise the distinction between the effect of incentives on the agents' effort vis a vis their effect on customers' demand, as the effect on sales can, in principle, be due to changes in either. Guided by the theoretical framework we then go beyond the average effects, and allow the treatment effects to be heterogeneous as a function of the agents' motivation for the cause (σ) or the weight they put on financial payoffs (ϕ). This will shed light on the importance of intrinsic motivation crowding-out in this setting and on whether non-financial incentives leverage intrinsic motivation as indicated in the theoretical framework.

Finally we provide evidence on the practical relevance of two further differences between financial and non-financial incentives, namely that the latter are visible to other agents and can therefore affect motivation through social comparisons, and that they can provide motivation to all employees in the salon as the star thermometer is publicly displayed.

6.1 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, if the clients are altruistic vis-a-vis the stylists and buy packs to make them earn stars or if the clients take it as a signal of the agents' type and buy packs because they share their interest in the mission.[26] 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

[26] 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 endline across treatment groups and find no significant differences.

dimensions that are correlated with sales effort as measured during the monthly visits. Our monthly surveys contain four variables that can be used for this purpose. In particular, we test whether the different incentive schemes affect the quantity of promotional materials, such as posters and "sold here" signs, displayed in the shop and the probability that the stylists fill in their logbooks as instructed. In addition monitors were asked to rank the stylists' interest in selling and promoting the female condoms and also to judge the stylists' attention level when the monitor themselves demonstrated sales techniques in the salon.

Table 2 reports the estimates of 5.1 using effort proxies as outcome variables. We find that agents in the star treatment display 0.23 more materials (10% more than the mean of the control group), are 7 percentage points more likely to fill in their logbooks (15% 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 monitors. Stylists in the two financial reward schemes do not differ from the control group in either of these three measures of effort. Finally, stylists in all treatments appear to be equally interested during the monitor's demonstration. Overall, the results in table 2 indicate that, in line with the effect on sales, non-financial incentives promote effort on three out of the four dimensions 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 customers to assess directly whether they report being affected by the thermometer. We ask customers whether they had seen promotional materials for female condoms in hair salons and, if so, to describe all that they have seen. Overall, 37% of the interviewees report having seen promotional materials. Of these, 92% had seen the promotional poster (which is the largest and most visible of the materials distributed), 36% had seen the "sold here" sign, and only 2%, or 15 people in total, report seeing the thermometer. 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 advantage in advertising.

Given the low sale 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 consist in distributing placebo thermometers to a random sample of salons in the pure volunteer and the two financial treatments. In the 8th monitoring cycle we distributed placebo thermometers to 113 randomly selected salons and standard posters to the remaining 138 that were visited during that cycle. The placebo thermometer looks identical to those given to stylists in the non-financial treatment except that the number of stars reflects total average sales by all salons, rather than the individual salon sales. We then measure the effect of the placebo thermometer on sales in the following month. The results, reported in columns 1 and 2 of table 3 show that the placebo thermometer has no effect on sales. The estimated effect of the placebo thermometer is 0.23 and not significantly different from zero, whereas agents in the star treatment sold 1.58 more packs in their first month. Columns 3 and 4 explore the possibility that the effect of the placebo

thermometer is biased downwards because stylists might have unsold stock they might sell from, and our measure of performance (restocking) fails to capture that. The results in columns 3 and 4 suggest that this is not the case, the effect of the placebo thermometer on sales is 0.09 and not significantly different from zero. Overall, table 3 indicates that the thermometer is not an effective advertising instrument, casting further doubts on the hypothesis that non-financial rewards affect sales by changing customers' behavior.

Taken together, the evidence so far suggest that the effect of non-financial rewards is due to stylists exerting more effort, rather than the treatment boosting demand. The theoretical framework makes precise that the effect of financial and non-financial rewards on effort depends on the weight stylists put on monetary and non-monetary pay-off. The next two subsections provide evidence on the empirical relevance of these mechanisms.

6.2 Motivation for the cause: crowding-out or crowding-in?

The theoretical framework makes precise that the agent's motivation for the cause, namely the weight σ they put on non-monetary payoff, determines the effectiveness of non-financial rewards and, if there is motivation crowding-out, also the effectiveness of financial rewards. The derivative of optimal effort with respect to the strength of non-financial rewards (2.4) illustrates that non-financial rewards leverage or "crowd-in" motivation, namely non-financial rewards are more effective when the agent's motivation for the cause, σ, is high. In contrast, financial incentives can crowd-out intrinsic motivation as illustrated in (2.3), and the strength of this effect $[\sigma(1+r)c'(m)]$ also depends on the agent's motivation for the cause σ. The evidence so far 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). Indeed, our findings indicate that sale performance is the same when there are no financial incentives, when financial incentives are low powered and when they are high powered. In all specifications, the differences are precisely estimated and close to zero. The findings are consistent with two interpretations. First the parameters of both our financial reward schemes could generate knife-edge cases so that the negative crowding-out effect exactly balances the positive effect on monetary payoff. Second, the average effects reported in table 1 might hide the fact that motivation crowding-out occurs only for agents who are motivated for the cause, namely those with $\sigma > 0$ in (2.3), if their contribution to the average effect is small.

To assess the empirical relevance of these mechanisms, we allow the effects of incentives to be heterogeneous as a function of the agent's motivation. As above, we proxy for the stylist's motivation by their observed willingness to donate to the cause and by their religion. The estimates in Table 1 indicate that both proxies of σ are strong predictors of sales on the intensive and extensive margins.[27]

[27]We note that to the extent that Catholic stylists have prevalently Catholic customers, the level effect estimated

To test for heterogeneous responses we estimate:

$$y_{ic} = \alpha + X_i\beta_i + \sum_{j=1}^{3} \delta_{0j} treat_c^j + \sum_{j=1}^{3} \delta_{1j} treat_c^j * \sigma_i + u_{ic} \tag{6.1}$$

where σ_i is the measure of stylist's motivation (whose level is included in the vector of stylist's characteristics X_i) and all other variables are defined above. Results in Table 4 indicate that indeed non-financial incentives leverage intrinsic 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 9.9 (se 3.2) more packs than the control group (low motivated stylists in the pure volunteer treatment), while stylists who donate less than the median amount sell 4.3 (se 2.9) more packs. The p-value of the difference is 0.101. Promoters who are not Catholic and are assigned to the star treatment sell 10.2 (se 3.0) more packs than the control group, while Catholic stylists sell 0.78 (se 2.9) fewer. The p-value of the difference is 0.06. Taken together, the findings are in line with the interpretation that non-financial incentives elicit effort by leveraging intrinsic motivation as stylists with stronger motivation respond more to the treatment.

In contrast, the findings in table 4 do not support the crowding-out hypothesis. Financial incentives are indeed ineffective both for stylists who are strongly motivated (high donors, non Catholic) and those who are not (low donors, Catholic). In contrast to the crowding-out assumption, the findings indicate that, if anything, high financial rewards actually appear to reinforce intrinsic 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 in column 1 and a p-value of 0.037 in column 2.

Our findings can be reconciled with the laboratory evidence on crowding-out (for example, Ariely et al. 2009) by noting that most experiments that find evidence of crowding out rely on the social image channel, namely on the fact that financial incentives reduce the reputational gains from pro-social activities. In our setting, however, this channel is closed as 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 500K price in all treatments. As 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

in table 1 might be capturing differences in demand rather than stylist's motivation. This does not invalidate the test, since Catholic customers should be equally unlikely to purchase condoms regardless of the treatment group to which their stylist is assigned.

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. [28]

Even if monetary incentives cannot affect the agents' social image, they can still crowd-out intrinsic motivation 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, 1991). Our findings suggest that this is not the case.[29]

6.3 Taste for money

Having ruled out that financial incentives are ineffective because they crowd-out intrinsic motivation, this subsection presents further evidence on the mechanisms that drive the response to financial incentives in this setting. The theoretical framework makes precise that the effectiveness of financial incentives depends on the agents' taste for money, namely the weight ϕ they put on their monetary payoff. Financial incentives might be ineffective in this setting because the average ϕ is low. This is fairly plausible because agents did not know which treatment they were assigned to until the end of the training program. As most NGOs rely on community members to volunteer, and indeed this is the standard SFH "contract", all agents who came to training were probably willing to join as volunteers. To the extent that agents who are willing to donate their time and effort also put a low weight on monetary payoffs, most agents in our sample will have a low ϕ.

To assess whether the effect of financial incentives is heterogeneous we use three alternative proxies for ϕ, which correspond to three underlying reasons why agents might put different weight on monetary gains. First, we exploit the fact that, under the assumption of concave utility, the same amount of money is relatively more valuable for poor stylists. To proxy for socioeconomic status we use information on the education level and English speaking ability of the stylist, and classify as "low socioeconomic status" the 19% 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.

Second, we use information on whether stylists sell other products in their shops. As most products are sold on commission, a revealed preference argument suggests that stylists who do sell other products, that is 27% of the sample, might value commissions more. At the same time, however, these agents might be at a corner solution where they devote all their effort to the

[28] In addition, qualitative evidence from focus groups in the field indicates no stigma attached to being paid for pro-social tasks, possibly because Zambia is a very poor economy, and that tasks seems more valuable if a donor, NGO or government is willing to pay for it.

[29] To minimize differences across treatments other than those arising from the compensation schemes, agents in all groups were reminded about their contribution to social value whenever they made a sale. This may have mitigated the chance of agents re-assessing their motivation for the task.

product that yield the highest margin, and therefore do not respond to variation in margins of other products.

Third, we use information on the stylists' main motive driving their occupational choice, and distinguish between those who list profit as their main motive, 34% of the sample, and those who do not.[30] Agents who are mostly motivated by profit for their main line of business might put a larger weight on the monetary payoff from all activities. We estimate:

$$y_{ic} = \alpha + X_i\beta_i + \sum_{j=1}^{3} \delta_{0j} treat_c^j + \sum_{j=1}^{3} \delta_{1j} treat_c^j * \phi_i + u_{ic} \tag{6.2}$$

where ϕ_i is the measure of stylists' motivation (whose level is included in the vector of stylists' characteristics X_i) and all other variables are defined above.

Table 5 estimates heterogeneous treatment effects along the three dimensions of ϕ_i. We find 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 treatment), low socio-economic status stylists sell 3.6 more packs when offered large financial rewards and 4.7 more packs when offered small financial rewards. While only the effect of large rewards is precisely estimated at conventional levels ($p = 0.046$), we cannot reject the null that small rewards have the same effect. Finally, the results in columns (2) and (3) show that the effect of financial incentives does not depend on the stylists' profit motives or sale experience with other products.[31]

Taken together, the evidence in table 5 indicates that financial incentives are only effective for a minority of stylists, namely the poorest in our sample.

6.4 Social comparison

A crucial difference between the treatments is that only the non-financial treatment enables stylists to make their sale performance visible to third parties. Stylists can do so either by displaying the thermometer in their salon or privately to the relevant parties. Enumerators' records from monthly visits indicate that, on average, the thermometer was publicly displayed in 43% of the salons. This provides a lower bound to the share of agents who choose to make their performance known to others as we do not observe whether they show it to selected individuals, or post it at other times when the enumerators are not in the salon.

While the evidence in section (6.1) casts doubt on the hypothesis that the thermometer affects customer demand, an implication of the difference in visibility is that stylists in the star treatment

[30] Other choices (%) were "selling products" (0%), "making people look nice" (44%), "being connected to the community" (14%), "being one's own boss" (7.5%)

[31] We also test for heterogeneous treatment effects by stylist gender and find no significant differences in sales outcomes.

can compare their performance to the performance of their peers in the same neighborhood, while stylists in the other treatments cannot. This might elicit effort if stylists are motivated by wanting to outperform their peers, or if they are encouraged by the effort of others dedicated to the same cause. Referring to our theoretical framework, a larger peer group might increase the power of the non-monetary incentive r.[32]

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 stylist's salons, that is the number of trained stylists in the same geographical area. By design, the randomization procedures ensures that the number of salons in each geographical area is balanced across treatments (see Table A2). 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 average area has 4.5 salons 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 none of the pairwise Kolmogorov-Smirnov tests rejects the null of equality.

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

$$y_{ic} = \alpha + X_i\beta_i + \gamma N_c + \sum_{j=1}^{3} \delta_{0j} treat_c^j + \sum_{j=1}^{3} \delta_{1j} treat_c^j * N_c + u_{ic} \tag{6.3}$$

where N_c is the number of trained salons in neighborhood c, where the neighborhood is the unit of randomization and covers an area of 500 square meters. The specification thus controls for area specific characteristics that 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 γcaptures these effects.

Three findings are of note. First, the interaction coefficient between the number of peers and the star treatment ($\delta_{13} = 1.06$;se=.38) is statistically and (38))-3010.90911(i)10910010 imf(.)Tj1(h)0.4(e)BT0ETE

or the ability to observe others' performances helps the stylists assess what is expected of them. This finding is robust to alternative sample restrictions, such as trimming at the 95th percentile, and alternative functional form specifications, such as replacing N_c with an indicator for whether N_c is above its median value.[33] [34]

Second, the number of peers is not correlated with sales in the control group ($\gamma = . - .052$, se=.16). This allays the concern that density captures other area specific characteristics that are correlated with sales. Third, the interaction coefficients between the number of peers and the two financial treatments are small and not significantly different than zero ($\delta_{11} = -.18$; se=.17; $\delta_{12} = .15$; se=.17). This allays the concern that density captures area specific features that make incentives more effective, such as the differential selection of stylists types mentioned above.

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 the star treatment 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% 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 posters or the number of other promotional materials is small and not statistically different from zero.

6.5 Stars as public goods

Another key difference between the financial and non-financial incentive schemes is that money is divisible and can be given to specific individuals, while stars are not divisible and attributed to the

5% 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. One possible interpretation is that even multi-employee salons are still quite small, and that the non-divisibility of stars would have more bite at larger salon size, but in our context we can rule out that the effectiveness of non-financial incentives is due to their non-divisibility.

7 Conclusions

We run a field experiment to provide evidence on the effectiveness of financial and non-financial rewards for pro-social tasks. We find that agents who are offered non-financial rewards ("stars" in this setting) exert more effort than either those offered financial rewards or those offered volunteer contracts. Non-financial rewards elicit effort by facilitating social comparisons among agents and by leveraging the agents' pro-social motivation. The magnitude of the effects are such that, as non-financial rewards are considerably cheaper than financial incentives, they dominate all other contracts on a cost-benefit comparison.[35]

As is customary 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. First, 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 they do (Bandiera et al. 2007, Dohmen and Falk 2011, Larkin and Leider 2011, Lazear 2000, Lazear et al. 2007). In general, we expect incentives to affect selection, as 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 fact that all agents in our experiment accepted the invitation to "help the community" without knowing whether and how much they would be paid also sheds light on why agents who were offered financial rewards did not perform better than those who were asked to volunteer, as financial gains from this task might have a low utility weight for both sets of agents. Our results are consistent with the possibility of financial rewards being effective for agents with different preferences -e.g. those who would have joined had they known about financial compensation- or even for other, perhaps less socially useful, tasks performed by the same agents.

The second key feature of our setting is that the task at hand is not the agents' main occupation. This has two implications for the relative effectiveness of non-financial vis-a-vis financial rewards.

[35]Beside the cost of the actual rewards (the stars), the cost of non-financial rewards include both the upfront expense necessary to figure out an effective design and the cost of administering the rewards. In our setting the latter was the same in all treatments as monitors attempted to visit all salons each month.

First, these are agents who have selected as their main occupation small entrepreneurship in the private sector. Non-financial rewards might be more effective for them as 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.

Second, even with the most generous financial reward scheme, earnings from condom sales are a small fraction of overall earnings because both demand for the product and earnings from each sales are low. As 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, although not necessarily more profitable as financial rewards are more costly. In line with this, we do find that financial incentives are effective for poorer agents for whom the marginal utility of money is higher. Likewise, the power of non-financial incentives depends on their relative scarcity. In our setting no other tasks was compensated with non-financial rewards. If non-financial rewards given for different tasks are substitutes, they might be less effective when they are used more widely.

While we implemented a specific type of non-financial rewards, 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.

An obvious limit to the use of non-financial rewards is that they cannot replace money as the main medium of compensation, and are thus of limited use in jobs where, due to the nature of the agency problem, performance pay accounts for a large share of total pay. Our findings however suggest that they can be a cost-effective means to motivate agents in the many settings where the fraction of variable pay over total pay is small. Ultimately, to assess whether 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.

References

[1] Alpizar, F., F. Carlsson and O. Johansson-Stenman. 2008. "Anonymity, reciprocity, and conformity: Evidence from voluntary contributions to a national park in Costa Rica." Journal of Public Economics, 92: 1047-1060.

[2] Ariely, D., Bracha, A and Meier, S. 2009. "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially." *American Economic Review,* 99 (1).

[3] Ball, S., C. Eckel, P.J. Grossman and W. Zame. 2011. "Status in Markets." *Quarterly Journal of Economics*, 116(1): 161-188.

[4] Bandiera, O., Rasul, I. and Barankay, I. 2005. "Social Preferences and the Response to Incentives: Evidence from Personnel Data"." *Quarterly Journal of Economics,* 120(3): 917-962.

[5] Bandiera, O., Rasul, I. and Barankay, I. 2007. "Incentives for Managers and Inequality Among Workers: Evidence from a Firm Level Experiment." *Quarterly Journal of Economics*, 122: 729-75.

[6] Bandiera, O., Rasul, I. and Barankay, I. 2011. "Field Experiments with Firms." *Journal of Economic Perspectives vol. 25(3), pages 63-82, Summer*

[7] Bénabou, R. and Tirole, J. 2003. "Intrinsic and Extrinsic Motivation." *Review of Economic Studies* 70(3): 489-520.

[8] Bénabou, R. and Tirole, J. 2006. "Incentives and Prosocial Behavior." *American Economic Review* 96(5): 1652-1678.

[9] Timothy Besley & Maitreesh Ghatak, 2005. "Competition and Incentives with Motivated Agents," American Economic Review, American Economic Association, vol. 95(3), pages 616-636, June

[10] Besley, T. and Ghatak, M. 2008. "Status Incentives." *American Economic Review*, 98(2): 206-211.

[11] Bhattacharyya, K., Winch, P., LeBan, K., and Tien, M. 2001. "Community Health Worker Incentives and Disincentives: How They Affect Motivation, Retention, and Sustainability." Published by the Basic Support for Institutionalizing Child Survival Project (BASICS II) for the United States Agency for International Development. Arlington, Virginia.

[12] Cameron, J., & Pierce, W.D. 1994. "Reinforcement, Reward, and Intrinsic Motivation: A Meta-analysis." *Review of Educational Research*, 64, 363-423.

[13] Cohn, A., E. Fehr, B. Herrmann and F. Schneider. 2011. "Social Comparison in the Workplace: Evidence from a Field Experiment," IZA Discussion Papers 5550, Institute for the Study of Labor (IZA).

[14] Deci, E. L., Koestner, R. and Ryan, R.M. 1999. "A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation." *Psychological Bulletin* 125(6): 627-68.

[15] Dohmen, T. and A. Falk. 2011. "Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender," *American Economic Review*, 101: 556-90.

[16] Eisenberger, R., & Cameron, J. 1996. "Detrimental Effects of Reward: Reality or Myth?" *American Psychologist*, 51, 1153-1166.

[17] Eriksson, T., Villeval, M.C. 2008. "Other-Regarding Preferences and Performance Pay. An Experiment on Incentives and Sorting." *Journal of Economic Behavior and Organization*, 68, 412-421.

[18] Eriksson, T. and Villeval, M. 2010. "Respect as an Incentive." *GATE Working Paper No. 1023.*

[19] Glewwe, P., N. Ilias & Michael Kremer. 2010. "Teacher Incentives," *American Economic Journal: Applied* Economics, 2: 205-27.

[20] Gneezy, U., and A. Rustichini. 2000. "Pay Enough or Don't Pay At All." *Quarterly Journal of Economics*, 791-810.

[21] Grant, A. M. 2007. "Relational job design and the motivation to make a prosocial difference." *Academy of Management Review*, 32: 393-417.

[22] Kosfeld, M. and Neckermann, S. 2010. "Getting More Work for Nothing? Symbolic Awards and Worker Performance." *American Economic Journal: Microeconomics (forthcoming).*

[23] Lacetera, Nicola & Macis, Mario. 2008. "Social Image Concerns and Pro-Social Behavior: Field Evidence from a Nonlinear Incentive Scheme" *Journal of Economic Behavior and Organization* 76: 225–237

[24] Lacetera, N., M. Macis and R. Slonim. 2011. "Will There Be Blood? Incentives and Displacement Effects in Pro-Social Behavior" *American Economic Journal: Economic Policy* (forthcoming)

[25] Landry, C., A. Lange, J.A. List, M.K. Price, N.G. Rupp. 2011. “The Hidden Benefits of Control: Evidence from a Natural Field Experiment.” NBER Working Paper No. 17473.

[26] Larkin, I., and S. Leider. 2011. "Incentive Schemes, Sorting and Behavioral Biases of Employees: Experimental Evidence." *American Economic Journal: Microeconomics* (forthcoming).

[27] Lazear, E. 2000. “Performance Pay and Productivity.” *The American Economic Review,* 90 (5): 1346-1361.

[28] Lazear, E., Malmendier, U. & Weber, R. 2006. "Sorting, Prices, and Social Preferences." NBER Working Papers 12041, National Bureau of Economic Research, Inc.

[29] Ministry of Health, Zambia, Central Statistic Office, Zambia, and ORC Macro, Calverton, Marlyand. July 2006. “Zambia HIV/AIDS Service Provision Assessment Survey 2005.”

[30] Mathauer, I. and Imhoff, I. 2006. “Health Worker Motivation in Africa: the Role of Non-financial Incentives and Human Resource Management Tools.” *Human Resources for Health,* 4(24).

[31] Mellström, C. and M. Johannesson, 2008. "Crowding Out in Blood Donation: Was Titmuss Right?," *Journal of the European Economic Association*, 6: 845-863

[32] Moldovanu, B., A. Sela and X. Shi, 2007. "Contests for Status," *Journal of Political Economy*, 115: 338-363

[33] Muralidharan K. and V. Sundararaman. 2011 "Teacher Performance Pay: Experimental Evidence from India", *Journal of Political Economy*, 119: 39-77.

[34] O’Reilly et al. 2008. “The Impact of a School-based Water and Hygiene Programme on Knowledge and Practices of Students and Their Parents: Nyanza Province, Western Kenya, 2006.” *Epidemiology and Infection,* 136, 80-91.

[35] PATH, UNFPA. 2006. “Female Condom: A Powerful Tool for Protection.” Seattle: UNFPA, PATH.

[36] Soetevent, A.R. 2005. “Anonymity in giving in a natural context: An economic field experiment in thirty churches.” *Journal of Public Economics,* 89: 2301-2323.

[37] Stanback, J., Mbonye, A. and Bekita, M. 2007. “Contraceptive Injections by Community Health Workers in Uganda: A Nonrandomized Community Trial.” *Bulletin of the World Health Organization,* 85: 768-773.

[38] Tirole, J. and Bénabou, R. 2003. "Intrinsic and Extrinsic Motivation." *Review of Economic Studies*, 70: 489–520.

[39] Tirole, J. and Bénabou, R. 2006. "Incentives and Prosocial Behavior." *American Economic Review*, 96(5): 1652-1678.

[40] Titmuss, R. 1970. The Gift Relationship: From Human Blood to Social Policy. Reprinted by the New Press.

Table 1: Average treatment effects on sales

Dependent variable	Number of packs sold		=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*	*.368*	*.341*	*.128*
	(1)	(2)	(3)	(4)	(5)
Large financial reward	0.769	1.217	0.002	0.014	0.033
	[1.618]	[1.666]	[0.063]	[0.059]	[0.040]
Small financial reward	0.378	0.847	-0.022	-0.015	0.012
	[1.528]	[1.560]	[0.067]	[0.060]	[0.040]
Star reward	7.482***	7.542***	0.119*	0.134**	0.103**
	[2.448]	[2.540]	[0.067]	[0.066]	[0.050]
Barbershop		3.230**	0.093**	0.093**	0.031
		[1.608]	[0.040]	[0.041]	[0.031]
Mixed (barber and hair salon)		3.857	-0.052	-0.036	0.003
		[3.935]	[0.070]	[0.071]	[0.053]
Near a bar (=1 if yes)		0.576	-0.047	-0.029	-0.000
		[2.127]	[0.076]	[0.064]	[0.049]
Low assets index (bottom quartile)		1.111	0.006	-0.000	0.018
		[1.720]	[0.051]	[0.052]	[0.035]
Number of employees (log)		1.647	-0.071	-0.063	0.037
		[2.784]	[0.066]	[0.066]	[0.049]
Sells other products in salon (=1 if yes)		5.211***	0.084**	0.084**	0.073**
		[1.740]	[0.040]	[0.041]	[0.036]
Number of salons in the same area (log)		-0.257	-0.008	-0.012	-0.024
		[0.911]	[0.031]	[0.027]	[0.020]
Dictator game donation above median		3.335***	0.149***	0.140***	0.013
		[1.131]	[0.031]	[0.032]	[0.028]
Low socioeconomic status (=1 if yes)		-1.063	-0.010	-0.012	-0.042
		[1.409]	[0.046]	[0.047]	[0.029]
Roman Catholic (=1 if yes)		-3.606***	-0.083**	-0.073*	-0.034
		[1.379]	[0.041]	[0.040]	[0.034]
Motivated by profit		0.796	0.023	0.016	0.018
		[1.375]	[0.038]	[0.037]	[0.032]
Constant	6.929***	0.736	0.352***	0.315***	0.113
	[1.123]	[4.056]	[0.107]	[0.103]	[0.086]
R-squared	0.0285	0.0661	0.0498	0.0479	0.0282
Observations	771	765	765	765	765

Notes: OLS estimates. Standard errors clustered at cell level. * p<0.10 ** p<0.05 *** p<0.01. Low socioeconomic status =1 if the agent does not speak English or has not completed primary education. Motivated by profit =1 if the agent reports "making money" was a reason to choose their occupation.

Table 2: Average treatment effects on effort measures

Mean in control group	*2.23*	*0.45*	*2.53*	*2.13*
Standard deviation in control group	*1.6*	*0.5*	*0.64*	*0.69*
	(1)	(2)	(3)	(4)
	0.100	0.02	-0.005	0.01
	0.0 3	0.024	0.030	0.040
	-0.0	0.01	0.02	0.05
	0.0	0.025	0.031	0.042
	0.22 **	0.0 0***	-0.04	0.103**
	0.0 3	0.025	0.030	0.041
	2.2 ***	0.4 3***	2.545***	2.1 ***
	0.22	0.0 1	0.0 0	0.0
-	0.03	0.0053	0.00 5	0.01 5
	4 0	44	45 3	4034

( ) .

. * 0.10 ** 0.05 *** 0.01.

( ) .

1 .

' "

( ) (0-3)". '

" ".

3, 1.

Table 3: Placebo thermometer

Dependent variable	Sales (restocking)		Sales (calculated)	
Mean in control group	*0.71*	*0.67*	*0.68*	*0.68*
	(1)	(2)	(3)	(4)
Placebo thermometer	0.238	0.248	0.097	0.056
	[0.334]	[0.341]	[0.281]	[0.277]
Controls		yes		yes
R-squared	0.0018	0.0623	0.0005	0.0451
Observations	296	295	296	295

Notes: Standard errors clustered at cell level. * p<0.10 ** p<0.05 *** p<0.01. The dependent variable is measured as the number of packs restocked based on invoices (Columns 1 and 2) or sold based on monitor calculations (Columns 3 and 4) in the round following distribution of the placebo thermometer. Placebo thermometer =1 if stylist received a thermometer poster reporting average sales of condoms across stars treatment (12 packs). All regressions include the same vector of controls as in Column 3, Table 1.

Table 4: Heterogeneous treatment effects, by pro-social motivation

Dependent variable is number of packs sold

Motivation variable	High donation (above median)	Non-Catholic
Mean in control group = 6.96		
	(1)	(2)
Motivation variable	0.730	-1.159
	[1.537]	[1.311]
Large financial reward	-2.295	-2.224
	[1.589]	[2.049]
Small financial reward	1.067	-2.126
	[1.944]	[2.674]
Star reward	4.294	-0.783
	[2.884]	[2.956]
High financial X Motivation variable	5.904**	4.410**
	[2.621]	[2.098]
Low financial X Motivation variable	-0.587	3.848
	[2.337]	[3.089]
Stars X Motivation variable	5.568	10.959***
	[3.374]	[3.906]
Controls	yes	yes
R-squared	0.0727	0.0748
Observations	765	765
Effect of large reward when motivation variable =1	3.609	2.187
	[2.397]	[1.792]
Effect of small reward when motivation variable =1	0.480	1.722
	[1.953]	[1.769]
Effect of stars reward when motivation variable =1	9.862***	10.18***
	[3.225]	[2.999]

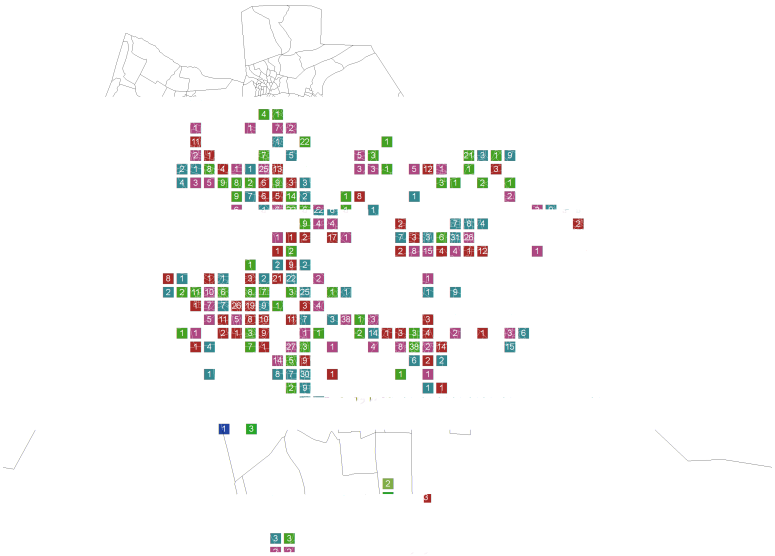
Notes: Standard errors clustered at cell level. * p<0.10 ** p<0.05 *** p<0.01. All regressions include the same vector of controls as in Column 3, Table 1.

Table 5: Heterogeneous treatment effects, by financial motivation

De e de a ab e be ac d			
M a a ab e	L c ec c a	Se e F - d c e	
Mean in Volunteer control group = 6.96			
	(1)	(2)	(3)
M a a ab e	-4.129**	3.509	2.703
	[1.614]	[2.167]	[2.142]
La e a c a e a d	0.840	1.645	2.176
	[1.983]	[1.697]	[1.955]
S a a c a e a d	0.004	0.845	0.619
	[1.734]	[1.656]	[1.761]
S a e a d	6.883**	4.946**	9.095***
	[2.886]	[2.093]	[2.608]
H a c a X M a a ab e	2.801	-2.538	-3.177
	[2.733]	[3.116]	[2.978]
L a c a X M a a ab e	4.752	-0.229	0.315
	[3.302]	[3.482]	[3.573]
S a X M a a ab e	4.142	9.923*	-4.29
	[3.788]	[5.127]	[4.412]
C	e	e	e
R- a ed	0.0676	0.0787	0.0687
Ob e a	765	765	765
E ec a e e a d e a a ab e =1	3.641**	-0.894	-1.001
	[1.809]	[2.996]	[2.669]
E ec a e a d e a a ab e =1	4.756	0.616	0.934
	[2.903]	[3.212]	[3.199]
E ec a e a d e a a ab e =1	11.03***	14.87**	4.805
	[3.137]	[5.827]	[4.271]

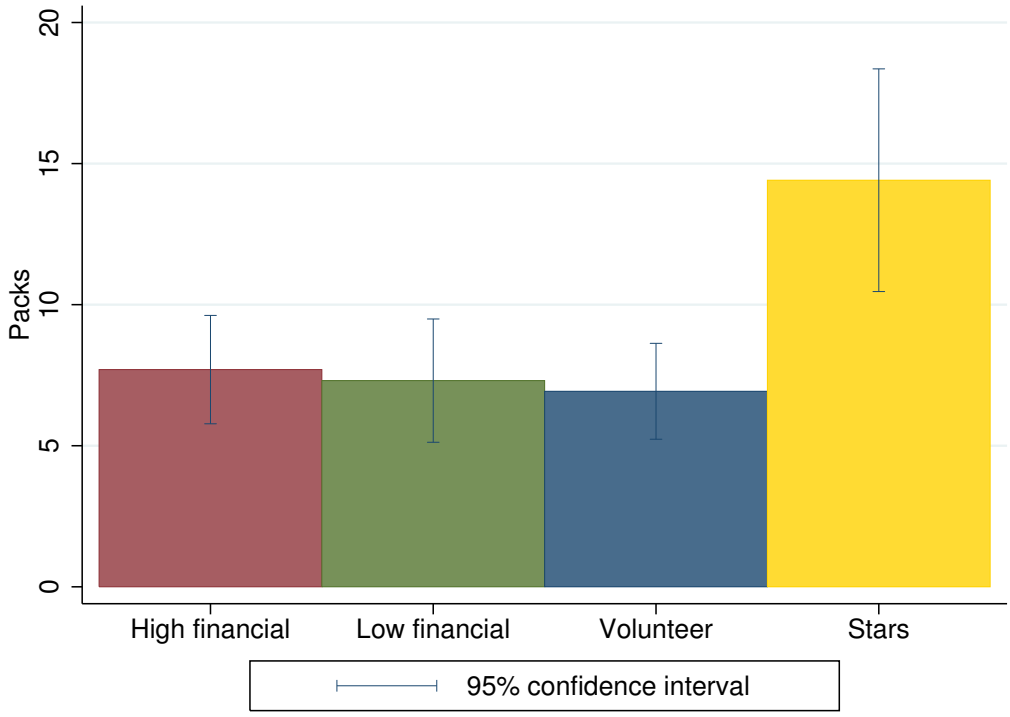
N e : S a da d e c e ed a ce e e . * <0.10 ** <0.05 *** <0.01. A e e c de e a e ec c a C 3, Tab e 1.

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 invited salons 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% confidence intervals.

Figure 3: Total sales and share of agents with positive sales, by month

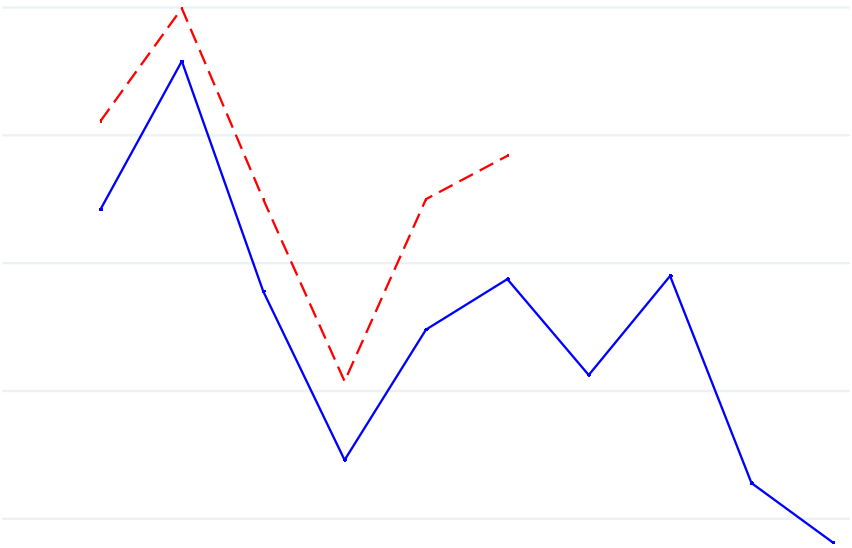
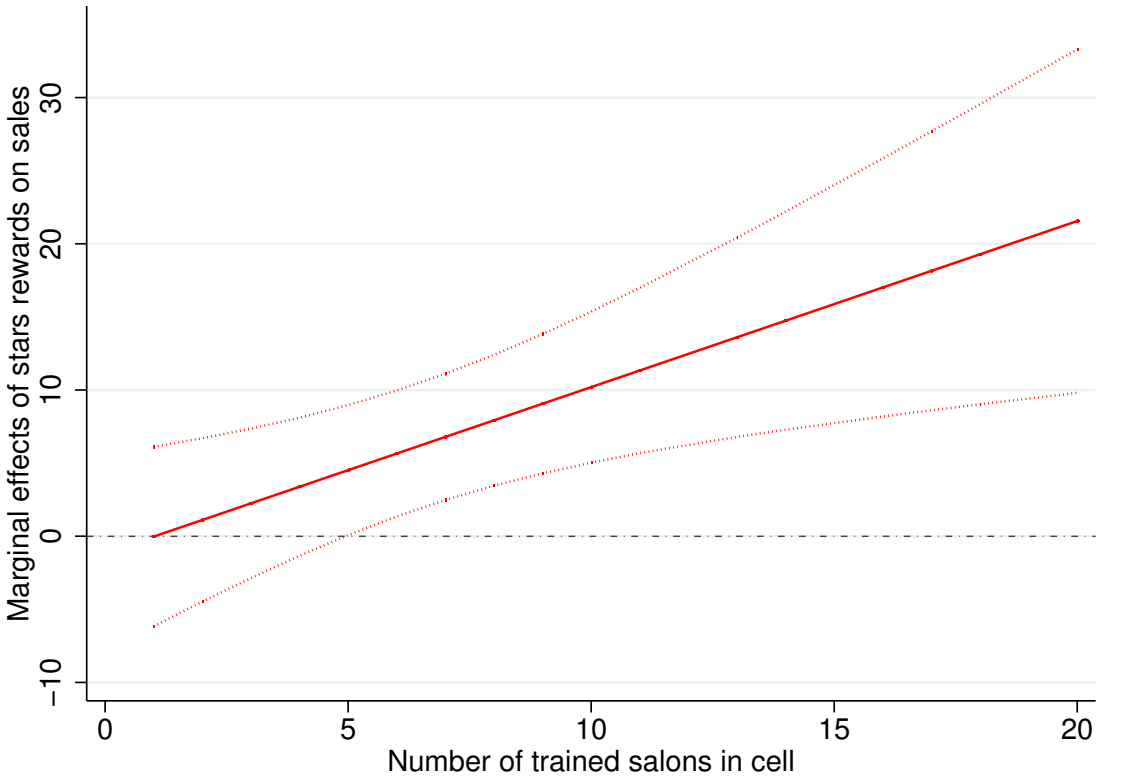


Figure 5: Imputed marginal effect of star rewards at different values of cell density



Notes: The solid line plots the imputed marginal effect of the star treatment at each value of cell density. This is computed as the sum of the coefficient of stars plus the coefficient of the interaction of stars and cell density multiplied by the respective value of cell density estimated in a regression of sales on of sales on the three treatments, the three treatments interacted with cell density, and controls listed in the footnote to Figure 3. The dotted lines represent the 95% confidence interval based on standard errors clustered at the cell level.

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

	0.51	0.52	0.41	0.4	0.15
	0.501	0.501	0.4 4	0.501	
	0.42	0.412	0.42	0.4 1	0.0 2
	0.4	0.4 4	0.4	0.501	
1	0. 21	0. 3	0.	0. 54	0.152
	0.2 0	0.345	0.304	0.354	
	0.21	0.220	0.1 4	0.203	0.0 33
	0.414	0.415	0.3	0.403	
	1. 2	1. 13	1. 1	1. 03	0.114
	0. 2	0. 20	1.2 1	0. 04	
	0.240	0.2 0	0.25	0.311	0.114
	0.42	0.450	0.440	0.4 4	
	.5	.500	.22	. 0	0.0
	.	.343	.430	5. 5	
	0.5	0.552	0.5	0.5	0.0 3
	0.4 2	0.4	0.4 4	0.4 2	
1	0.203	0.23	0.200	0.142	0.1 2
	0.403	0.42	0.401	0.34	
1	0.234	0.214	0.254	0.22	0.0 3
	0.425	0.411	0.43	0.420	
	0.3 0	0.324	0.43	0.2 4	0.2 1
	0.4 4	0.4	0.4	0.442	

, , .

. 1

. 1 " "

. ,

, .

Table A.2: Participation decision

D a ab	R c a	A a
	C a a a C a c a	
Mean in control group = 0.80		
	(1)	(2)	(3)	(4)
La a c a a	-0.005	-0.008	0.02	0.018
	[0.033]	[0.029]	[0.042]	[0.041]
S a a c a a	0.029	0.028	-0.023	-0.015
	[0.034]	[0.031]	[0.042]	[0.040]
S a a	-0.006	0.002	-0.042	-0.03
	[0.031]	[0.031]	[0.046]	[0.046]
Ba b		0.063**		0.059*
		[0.028]		[0.033]
M (ba b a a a)		0.041		0.036
		[0.041]		[0.052]
N a a ba (=1)		0.019		0.065
		[0.037]		[0.048]
L a (b a)		-0.062*		-0.005
		[0.033]		[0.036]
N b ( )		-0.003		-0.018
		[0.016]		[0.018]
S c a		0.016		-0.005
		[0.025]		[0.032]
N b a a a a		0.003***		0.002*
		[0.001]		[0.001]
G HIV ca (=1)		0.028		0.055**
		[0.025]		[0.026]
L c c c a (=1)		0.012		-0.069*
		[0.025]		[0.036]
R a Ca c (=1)		0.012		0.022
		[0.025]		[0.026]
M a b		-0.046*		-0.026
		[0.026]		[0.029]
C a	0.799***	0.724***	0.797***	0.684***
	[0.021]	[0.042]	[0.032]	[0.061]
R- a	0.0012	0.0164	0.0032	0.0226
Ob a	1222	1216	981	977

N : L a bab a a c a c . * <0.10 ** <0.05 *** <0.01. Va ab a a c b Tab A.1 c G HIV ca c b a - a HIV/AIDS.

Table A.3: Treatment effects on selection

Dependent variable	Selection at training		Selection during monitoring	
	Stylist did not join program		Stylist was dropped from program	
Mean in Volunteer control group	*0.042*		*0.052*	
	(1)	(2)	(3)	(4)
Large financial reward	-0.009	-0.011	0	0
	[0.016]	[0.009]	[0.025]	[0.024]
Small financial reward	-0.017	-0.009	0.059**	0.052*
	[0.016]	[0.011]	[0.034]	[0.033]
Star reward	-0.017	-0.011	0.051*	0.047
	[0.015]	[0.009]	[0.034]	[0.034]
Barbershop		0.017		-0.003
		[0.012]		[0.022]
Mixed (barber and hair salon)		-0.003		-0.052*
		[0.018]		[0.018]
Near a bar (=1 if yes)				0.007
				[0.025]
Low assets index (bottom quartile)		-0.005		0.001
		[0.007]		[0.023]
Number of employees (log)		-0.002		0.035***
		[0.006]		[0.010]
Sells other products in salon		0.014*		-0.004
		[0.010]		[0.021]
Number of salons in the same area		0.001*		0
		[0.000]		[0.001]
Dictator game donation above median		0.028***		-0.003
		[0.009]		[0.015]
Low socioeconomic status (=1 if yes)		-0.01		-0.006
		[0.009]		[0.021]
Roman Catholic (=1 if yes)		0.006		-0.016
		[0.011]		[0.019]
Motivated by profit		0.014		0.009
		[0.010]		[0.017]
Pseudo R-squared	0.0094	0.1046	0.0163	0.0447
Observations	771	766	771	765

Notes: Linear probability model with standard errors clustered at the cell level. * p<0.10 ** p<0.05 *** p<0.01. Variables are as described in Table A.1. The variable describing proximity to a bar is dropped in column (2) since it is perfectly collinear with the dependent variable.

Table A.4: Missed monitoring visits, by treatment group

			(10)	
	0	1 3	4 6	7 9	10
	0.21	0.35	0.13	0.24	0.07
	0.20	0.30	0.11	0.34	0.05
	0.21	0.33	0.20	0.20	0.07
	0.18	0.37	0.12	0.28	0.05
	0.20	0.34	0.14	0.26	0.06

: D , .