What's God Got To Do With It? Estimating How Religious Affiliation Affects Individuals' Response to Extrinsic Rewards*

Michael Chaitkin, Justin Healy, Shannon Milroy

April 30, 2015

Abstract

Using data from a field experiment on incentives for private delivery of socially beneficial goods and services, we find that the effect of non-financial extrinsic rewards on condom sales performance depends dramatically on whether the sales agent is Catholic. We build on the analysis in Ashraf et al. (2014), which finds that non-financial rewards improve sales performance among 771 hair stylists in Lusaka, Zambia. We affirm that study's findings with respect to non-Catholic participants but estimate considerably different intention-to-treat effects for Catholic ones. In particular, we find that in the study setting non-financial rewards did not improve Catholics' sales performance. We also find suggestive evidence that financial rewards may have actually *suppressed* their sales performance, though these effects are imprecisely estimated. In the literature relating performance in pro-social enterprises to incentives, we could not find a study exploring how people's religion or other broad cultural attributes intermediates their responsiveness to extrinsic rewards. We have therefore identified an important consideration for policy makers and organizational leaders contemplating performance-based incentives in social sectors.

*We are deeply indebted Nava Ashraf, Oriana Bandiera, and Kelsey Jack for welcoming our effort to replicate their work, and particularly to Dr. Jack for sending us the cleanest datasets and Stata code we could have asked for. We are also grateful to Christine Choirat and Lina Song for their thoughtful comments midstream, as well as to Gary King, Stephen Pettigrew and Solé Prillaman for methodological guidance.

Introduction

There is considerable interest among policy makers in the role incentives might play in improving the provision of social goods. Programs in health and education have relied on performance-based pay as a means of improving outputs with often mixed results (Freyer 2013, Duflo 2012, Li 2014). A common concern has been that these financial incentives potentially displace the intrinsic motivations that many agents engaged in pro-social enterprises bring to their work, and therefore lead to worse outcomes than other, non-financial, incentive schemes might produce (Benabou and Tirole 2006, Besley and Ghatak 2005).

Thus far the literature on incentives has focused on characteristics of the task being incentivized or the incentive itself. Studies have examined the different way incentives impact on social or private endeavors, or on the relative merits of financial versus non-financial rewards, but far less work has been done on how the characteristics of the agents receiving the incentives affects their efficacy. This has important implications for policy makers because if different groups respond differently to incentive schemes, these initiatives may require an equivalent level of differentiation in order to be maximally effective.

We revisit a recent study that examines the mediating influence of intrinsic motivation and socioeconomic status but overlooks a key factor in shaping agents' response to various rewards: religion. We find that while rewards enhance performance of non-Catholics, Catholics respond entirely differently to incentives. Catholics receiving non-financial rewards perform no better than volunteers, while those receiving financial rewards may actually perform worse.

Methods

Study design

Our analysis builds directly atop that originally conducted on data from a field experiment that examined the effect of extrinsic rewards on condom sales by hair stylists in Lusaka, Zambia (Ashraf et al. 2014). The investigators recruited 771 stylists to sell female condoms and randomly assigned geographic clusters of them to a volunteer group and three treatment groups receiving, respectively, a large financial reward, small financial reward, and non-financial ("star") reward. All participants attended identical informational and training sessions prior to randomization. The volunteer group did not receive any direct incentive for selling condoms and so serves as a control group for the experiment. The experiment included a census survey to collect participants' baseline characteristics, a contextualized dictator game during training to measure pro-social motivation, and monthly visits to track condom sales and additional information about each salon.

Data

We analyze data furnished by the authors of Ashraf et al. (2014), inheriting that study's outcome, treatment, and independent variables. Our outcome of interest is sales performance as measured by the number of condoms sold by a stylist over the course of the experiment. Treatment refers to assignment to one of the three treatment groups or the control group. The data include several salon characteristics, including whether it caters only to male clients, whether it is located near a bar,

how many people it employs, and how many other salons operate in the same study cluster. The data also include numerous stylist traits, including whether the stylist has prior sales experience, is in the bottom quartile of asset distribution, is of low socioeconomic status, reports intrinsic work motivation, and is Catholic. The stylist's pro-social motivation, measured by donation level in the dictator game, is also included.

Identification

We model the effect of different treatments on performance by identifying the stochastic and systematic components of the underlying data generating process. Performance is measured with a count of condoms sold, which is always a non-negative integer, so we look to the family of count data distributions for the stochastic component of our model. This approach differs from that in the original study, which uses a least-squares regression model with clustered standard errors, a post-estimation technique for adjusting standard errors to account for design effects stemming from clustered sampling. Because we employ simulations to compute our quantities of interest (see Results Section below), we prefer to explicitly model variance to avoid the inefficiency and bias that can arise from clustered standard errors techniques (King and Roberts 2014). Among distributions suitable to count data, the negative binomial offers sufficient flexibility in estimating variance and foregoes stringent assumptions about the inter-temporal independence of events.

The model's systematic component simply relates the negative binomial's rate parameter to treatment and other covariates. Taken together our model can be specified as:

$$Y_i \sim f_{NegBin}(\lambda_i, \theta) \qquad \text{(stochastic component)}$$

$$\lambda_i = exp\left(\sum_{j=1}^{3} \delta_j treat_c^j + X_i \eta_i\right) \qquad \text{(systematic component)}$$

where Y_i measures total condom sales by stylist i. λ_i denotes the negative binomial's rate parameter, which is also equal to its expected value, and θ is the distribution's dispersion parameter. The three treatment groups are captured by $treat_c^j$; the intention-to-treat (ITT) effect of each treatment j on performance is δ_j. X_i is a vector of the stylist and salon characteristics mentioned above, and η_i is their effect on performance. Because λ_i must be positive, it is parameterized by exponentiating the treatment and covariate terms.

Examining Subsets of the Study Population

To determine whether the treatment effects estimated in Ashraf et al. are consistent across both Catholic and non-Catholic stylists, we divide the data into two subsets containing 178 observations of Catholics and 593 of non-Catholics, respectively. This raises concerns that although treatment is randomized in the full population[1], there may be observable differences between the treatment arms within the Catholic and non-Catholic subpopulations. Consequently, for each subset we make pairwise comparisons of covariate means between each of the treatment groups and the control group and find evidence of mild imbalance. For example, Catholics incented with the high financial

[1] Ashraf et al. provide a thorough description of their treatment assignment protocol and ample evidence of effective randomization.

reward are poorer than those in the control group, and all treated Catholics were more likely to work near a bar than controls. Similarly, non-Catholic controls have more prior sales experience and intrinsic motivation at work than their treated counterparts. We control for all covariates in our analyses since even these small imbalances cast doubt on treatment effects estimated by taking a difference of mean outcomes between treatment and control groups, . Only prior sales experience is correlated with sales performance; this might justify skepticism of inferences made about our analysis of the non-Catholic subpopulation, though as is described below, our estimated treatment effects for that group are consistent with those for the entire study population.

Quantity of Interest

To compare ITT effects between Catholic and non-Catholic participants, we use the output of our negative binomial estimation to compute expected values for condom sales using the simulation methods described in King et al. (2000). We simulate using the median values for each covariate; this "median stylist" lacks prior sales experience, earns 250,000 Kwacha (∼50 USD) per month, exhibits pro-social and intrinsic motivation, is literate in at least one language, and is not Catholic. Their salon employs two other people, serves all-female or mixed clientele, is near a bar, and shares its neighborhood with six other participating salons.[2]

Results

Expected Values

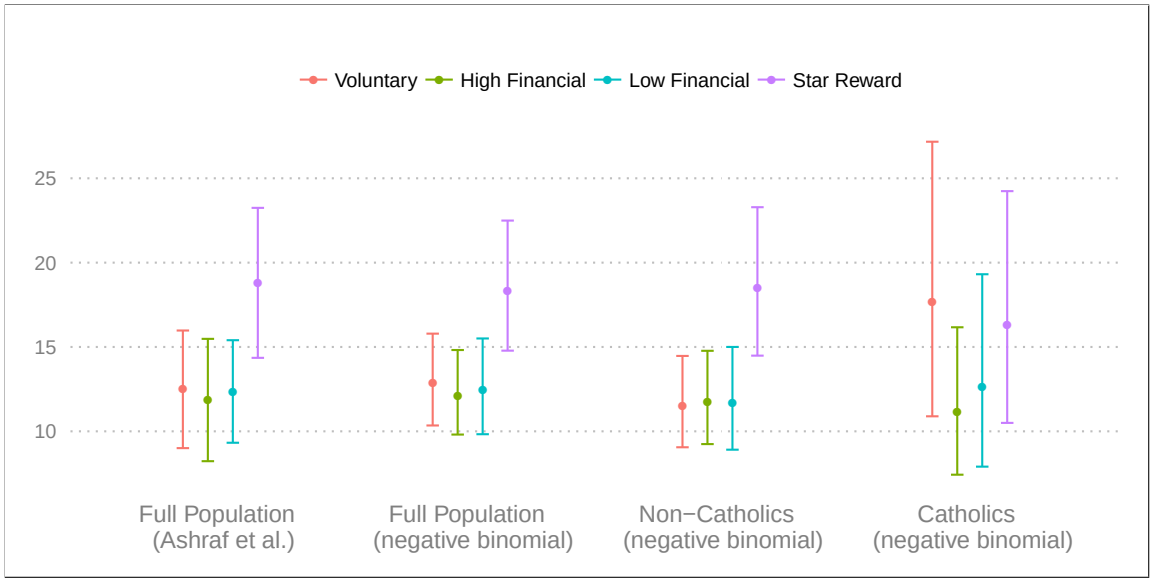
We find that although the star reward significantly improves performance among non-Catholic stylists, there is no evidence of a treatment effect on Catholics. Figure 1 depicts simulated expected values and their 95 percent confidence intervals derived from four estimations of treatment effects. The first two both pertain to the full study population: the first comes from replicating the original study[3], while the second comes from our negative binomial model specification. The third set of expected values is for the non-Catholic stylists and is consistent with the first two in that those receiving the star reward are expected to substantially outperform those in both the other treatment groups and the controls.

The fourth set of expected values, which we simulated from our estimates for Catholic stylists, reveals that there are no treatment effects among the Catholic sub-group. Within this subgroup those receiving incentives on average actually perform worse than those participating voluntarily. Remarkably, the expected values indicate that while the star reward does not bolster performance among Catholics, financial rewards may actually suppress it. That the expected sales of those receiving low financial rewards exceeds that of the high rewards recipients even hints at a dose response, though as discussed in the next section, the uncertainty of these estimates advises caution in their interpretation.

[2]Based on these characteristics we follow the same coding rules as Ashraf et al. to produce a vector of covariates to input into our simulations. When generating expected values for the Catholic and non-Catholic sub-groups, we of course omit whether the "median stylist" is Catholic. We also ran simulations on a covariates vector with all the binary variables coded to their other value. Though the expected values themselves differed, the relative treatment effects remained the same.

[3]Specifically, the expected values portrayed here derive from the estimates in column (3) of Table 2 in Ashraf et al. (p. 10).

Figure 1: Expected Condom Sales by Sample Subpopulation



Notes: Simulated expected values (points) with 95 percent confidence intervals (bars) show that although the star reward significantly improves performance among non-Catholic stylists, there is no evidence of a treatment effect on Catholics. The treatment effects on non-Catholics are consistent with findings on the full population, both as reported in Ashraf et al. and from our alternate specification. In contrast, volunteer Catholics outperform all the treatment groups, and financial rewards may even suppress performance. These expected values capture treatment effects on a "median stylist" and are robust to numerous other agent profiles.

Estimation Uncertainty

The distribution of our simulated expected values reflects estimation uncertainty (King et al. 2000). In addition to expected values, Figure 1 includes bars representing 95 percent confidence intervals. These reinforce the apparent effect of the star reward among non-Catholics while also highlighting the lack of clarity within the Catholic sub-group. Imprecision in estimates for the latter may stem at least in part from the relatively small number of Catholics in the study population.

Robustness

We perform two checks on robustness. First we reproduce our estimates with an alternate measure of sales performance. Two measures of sales performance are available in the experimental data, one based on restocking and another based on sales calculations made by field assistants. Ashraf et al. uses the restocking measure as their main performance outcome, while we prefer the calculated values because they are less prone to clustering around multiples of 12 and therefore more closely resemble a true count. Nonetheless we also run our full analysis with all four model specifications with the restock performance measure; the results are consistent with those we report here.

Second, we simulate expected values using an agent profile other than the "median stylist" de-

scribed above. The original study finds that the incentives had a greater effect when the stylist's socioeconomic status was low and that the star treatment had a greater effect amongst those with a large number of other salons in their area. To test the sensitivity of our findings we run simulations of the expected sales performance for all four models using a stylist who is of low socioeconomic status, lives in an area with 30 salons (the maximum possible number) and all other variables at their median value. Again we find that the results do not differ qualitatively from our main findings, star treatment is still the most effective in the non-Catholic population and incentives do not have a distinguishable effect in the Catholic population. We also ran the simulations of expected condoms sales using the median values for all numeric variables and the opposite value of the median for all binary variables. Again we find results consistent with our main findings of no star treatment effect among Catholics.

Conclusion

In 2010 Pope Benedict changed decades of Vatican policy by announcing that condoms could be used to stop the spread of HIV. Prior to that major Catholic figures had either impugned the efficacy of condoms or even suggested that they were contributing to the spread of disease.[4]. In such a context it is perhaps unsurprising that the original paper considered Catholicism to be a confounder in need of controlling. Yet our analysis casts doubt on that assumption and suggests that Catholicism is an important effect modifier that requires further discussion and exploration.

The authors of the original paper make a qualified case for scaling up of the non-financial reward treatment. While nothing we have shown necessarily invalidates this conclusion, we have added an important, and previously neglected, layer of complexity to the analysis. Just over one in five Zambians are Catholic (Pew Research Center 2011), and before rolling out any incentive initiative it is important to understand that this important subpopulation may react differently from rest of the population. Furthermore there are over 170 million Catholics living in Sub-Saharan Africa and similar efforts to improve condom distribution across the region may need to take this information into account.

Ultimately, this finding points the way to an important field of future research. As policy makers seek to find more sophisticated ways to leverage pro-social activities amongst various agents, it is important to better understand the ways different groups may respond to such interventions. Just as medicine has long recognized that certain populations respond differently to different drug treatments, an equivalent level of differential understanding might be required by policy makers seeking to incentivize certain behaviors. The academic literature has started to take steps in this direction. Ashraf et al. incorporates detailed analysis of individual markers for intrinsic motivation and socioeconomic status, but our findings suggest that shared religious and cultural values may also play an important role. In this sense the work of understanding pro-social incentives will move beyond traditional conceptions of classical and behavioral economics and into areas of psychology, ethnography and sociology.

[4]In 2003, for example, a senior Vatican spokesman claimed that condoms were permeable to HIV. In the same year the archbishop of Nairobi claimed, "AIDS...has grown so fast because of the availability of condoms" (BBC News 2003).

References

Ashraf N, Bandiera O, Jack BK, 2014. No margin, no mission? A field experiment on incentives for public service delivery. *Journal of Public Economics.* 120. 1-17

BBC News, 2003. Vatican in HIV condom row. http://news.bbc.co.uk/2/hi/health/3176982.stm.

Benabout R, Tirole J, 2006. Incentives and prosocial behaviour. *American Economic Review.* 96 (5). 1652-1678.

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

Duflo, Hanna R, Ryan S, 2012. Incentives work: getting teachers to come to school. *American Economic Review.* 102 (4). 1241-1278.

Freyer R, 2013. Teacher incentives and student achievement: Evidence from New York City Public Schools. *Journal of Labour Economics.* 31 (2). 373-407.

King G, Tomz M, Wittenberg J, 2000. Making the most of statistical analyses: Improving interpretation and presentation. *American Journal of Political Science.* 44 (2). 341-355.

King G, Roberts ME, 2014. How robust standard errors expose methodological problems they do not fix, and what to do about it. *Political Analysis*

Li J, Hurley J, DeCicca, Buckley, 2014. Physician response to pay-for-performance: Evidence from a natural experiment. *Health. Economics* 23 (8). 962-978.

Pew Research Center, 2011. The global Catholic population. http://www.pewforum.org/2013/02/13/the-global-catholic-population/.