Pre-Analysis Plan

Outline

The goal of this experiment is to investigate how signaling motives for prosocial behavior interact with notched incentives. Subjects can raise money for charity by completing effort tasks. In some treatments, subjects are also offered a private bonus (paid to them) for completing 15 tasks or more. In some treatments a subject's incentive scheme, total effort and resulting donation and earnings are shown to other subjects together with a picture of their face.

The key predictions are

1. Making effort visible to others reduces the amount of bunching at 0 and at 15 completed tasks (the bonus threshold).
2. Making effort visible to others moves subjects from bunching at 15 completed tasks to completing strictly more than 15 tasks.
3. If effort is visible to others, then an increase in the bonus amount moves subjects from completing less than 15 tasks to completing strictly more than 15 tasks.
4. If effort is visible to others, then an increase in the bonus amount moves subjects from completing strictly more than 15 tasks to completing even more tasks.

Treatments

The experiment features a 2×4 treatment design, cross-randomizing visibility and bonus incentive, using the following treatment levels.

Visibility (2 levels)

Badge Tasks completed, total donation raised, total personal gain and the bonus scheme are recorded on a digital badge that looks similar to an employee badge. The badge also displays tasks completed in a progress bar going up to the maximum of 38 tasks, and a picture of the subject's face that they take using their webcam. The badge is displayed during the task, and subjects are informed that their badge will be seen by other subjects.

No badge Subjects do not receive a badge, and all information about their contribution remains private.

Bonus (4 levels)

No bonus Subjects raise 8c per completed task for a charity of their choice. They do not receive any personal gain from completing tasks.

40c bonus Like *No bonus*, but additionally subjects receive 40c, if they complete 15 tasks or more.

80c gift + 40c bonus Like *40c bonus*, but in addition to the 40c bonus at 15 tasks, subjects also receive 80c for any effort level. The 80c is given "as a gift" and displayed in the progress bar like a bonus that is achieved for completing 0 tasks.

\$1.20 bonus Like *40c bonus*, but the bonus amount is \$1.20.

Sample

My goal is to obtain 2,400 participants using the Amazon MTurk platform. The study inclusion criteria are

- Location is US
- Number of HITs Approved greater than or equal to 50
- HIT Approval Rate for all Requesters' HITs greater than or equal to 95%

The target treatment cell sizes are $N = 400$ for the four cells $(\text{badge}, \text{no badge}) \times (\text{no bonus}, \$1.20 \text{ bonus})$ and $N = 200$ for the four cells $(\text{badge}, \text{no badge}) \times (\text{40c bonus}, \text{80c gift + 40c bonus})$.

Incentive adjustment

Because the experiment requires a webcam, I expect recruitment to be slower than usual. In order to achieve a sufficient sample size within a short timeline, I will adjust the participation reward as follows.

If the completed sample size (as reported by MTurk)

- is less than 200 at 48 hours after launch, or
- is less than 800 at 7 days after launch, or
- is less than 1,600 at 14 days after launch, or
- is less than 2,000 at 21 days after launch

then I will cancel the current HIT (keeping all existing data), and repost it with an increase in the participation reward by an amount between 10c and 50c. This accumulates, so if I miss the 800 target at 7 days, and later the 1,600 target at 14 days, I will then do a second increase.

Sample exclusion

Subjects will be dropped from the main analysis if they

- do not consent to study participation
- do not consent to creating and sharing their badge in part 2
- do not complete the 5 required tasks in part 1
- do not submit a picture of their face
- show any behavior suspicious of manipulating their submission or use of computer aids, for example correctly completing a transcription in less than 20 seconds

I have implemented technical hurdles to prevent subjects from participating multiple times. For subjects that appear multiple times in my sample despite this, I only consider their oldest completed submission.

Analysis of Effort

I define effort as the number of tasks completed for charity. To maintain sufficient statistical power my analysis of effort focuses on the following three outcomes.

1. Average effort
2. Share of subjects in different parts of effort distribution
3. Distribution of effort above the bonus threshold

I analyze these outcomes as follows.

1. Average effort

I estimate the average treatment effect on effort, by estimating the model

$$\text{Effort}_i = \beta_0 + \beta_T \text{Treatment}_i + \beta_X X_i + \beta_{T\times X} \text{Treatment}_i \times X_i$$

using OLS, where Treatment_i is a vector of treatment group dummies and X_i a vector of pre-treatment covariates. The treatment group dummies are the full interaction of the visibility treatment with the bonus treatments (excluding an omitted group). The covariates are *education*, *sex*, *age*, *race/ethnicity*, *Instagram usage*, *hours worked on MTurk in past 7 days*, *hours worked on MTurk so far today*, and without any interaction among each other. The covariates are measured as categorical group dummies just as they are elicited in the survey. *race/ethnicity* allows for multiple answers, and I treat any combination as a distinct group. I either center the covariates or use weighted effect coding.

I use Eicker-Huber-White standard errors to construct confidence intervals and test whether the average treatment effects of (i) the badge (1 test), (ii) the bonus levels (3 tests), and (iii) the interactions (3 tests) are different from zero.

Robustness

I repeat the analysis using (i) no covariates and (ii) fully interacted covariates.

2. Share of subjects in different parts of the distribution

Effort is a discrete variable with support $\{0, \ldots, 38\}$. To simplify the analysis of the treatment effects on the distribution of effort I partition the support as follows:

1. **Bunching at zero:** Effort $\in \{0, 1\}$
2. **Anti-bunching at zero:** Effort $\in \{2, \ldots, 14\}$
3. **Bunching at bonus:** Effort $\in \{15, 16\}$
4. **Anti-bunching at bonus:** Effort $\in \{17, \ldots, 38\}$

I define bunching to include more than one effort level, in order to take care of noise and unknown alternative explanations that predict a one-step increase as a treatment response.

I estimate the full probability mass function of effort within each of the $2 \times 4 = 8$ treatment groups, and then do inference on the parts of the distribution that I am interested in. For some tests I combine data from *40c bonus* and *80c gift + 40c bonus* into one treatment group, *40c pooled*, in order to increase statistical power.

To estimate the probability mass functions I use a seemingly unrelated regression framework stacking the linear equations

$$1\{\text{Effort}_i = e\} = \beta_T^e \text{Treatment}_i$$

for $e \in \{0, \ldots, 38\}$. I then estimate the heteroscedasticity robust covariance matrix and conduct the following tests.

(a) The badge reduces bunching at zero (4 tests)

$$H_0 : \beta^0_{\text{badge,b}} + \beta^1_{\text{badge,b}} = \beta^0_{\text{no badge,b}} + \beta^1_{\text{no badge,b}}$$

for $b \in \{\text{no bonus}, \text{40c bonus}, \text{80c gift + 40c bonus}, \text{\$1.20 bonus}\}$

(b) With no badge, a bonus induces bunching at the bonus threshold (2 tests)

$$H_0 : \beta^{15}_{\text{no badge,b}} + \beta^{16}_{\text{no badge,b}} = \beta^{15}_{\text{no badge,no bonus}} + \beta^{16}_{\text{no badge,no bonus}}$$

for $b \in \{\text{40c pooled}, \$1.20 \text{ bonus}\}$

(c) With a badge, a bonus induces less bunching at the bonus threshold than without (2 tests)

$$\begin{aligned} H_0 : &\beta^{15}_{\text{badge,b}} + \beta^{16}_{\text{badge,b}} - (\beta^{15}_{\text{badge,no bonus}} + \beta^{16}_{\text{badge,no bonus}}) \\ &= \beta^{15}_{\text{no badge,b}} + \beta^{16}_{\text{no badge,b}} - (\beta^{15}_{\text{no badge,no bonus}} + \beta^{16}_{\text{no badge,no bonus}}) \end{aligned}$$

for $b \in \{\text{40c pooled}, \$1.20 \text{ bonus}\}$

(d) With a badge, a bonus induces anti-bunching at the bonus (2 tests)

$$H_0 : \sum_{e=17}^{38} \beta^{e}_{\text{badge,b}} = \sum_{e=17}^{38} \beta^{e}_{\text{badge,no bonus}}$$

for $b \in \{\text{40c pooled}, \$1.20 \text{ bonus}\}$

(e) With a badge, a bonus induces more anti-bunching at the bonus than without a badge (2 tests)

$$H_0 : \sum_{e=17}^{38} \beta^{e}_{\text{badge,b}} - \beta^{e}_{\text{badge,no bonus}} = \sum_{e=17}^{38} \beta^{e}_{\text{no badge,b}} - \beta^{e}_{\text{no badge,no bonus}}$$

for $b \in \{\text{40c pooled}, \$1.20 \text{ bonus}\}$

(f) With a badge, an increase in the bonus amount increases anti-bunching at the bonus (2 tests)

$$H_0 : \sum_{e=17}^{38} \beta^{e}_{\text{badge,\$1.20 bonus}} = \sum_{e=17}^{38} \beta^{e}_{\text{badge,b}}$$

for $b \in \{\text{40c bonus}, \text{80c gift + 40c bonus}\}$

(g) With a badge, an increase in the bonus amount increases anti-bunching at the bonus, more than without a badge (2 tests)

$$\begin{aligned} H_0 : &\sum_{e=17}^{38} \beta^{e}_{\text{badge,\$1.20 bonus}} - \sum_{e=17}^{38} \beta^{e}_{\text{badge,b}} \\ &= \sum_{e=17}^{38} \beta^{e}_{\text{no badge,\$1.20 bonus}} - \sum_{e=17}^{38} \beta^{e}_{\text{no badge,b}} \end{aligned}$$

for $b \in \{\text{40c bonus}, \text{80c gift + 40c bonus}\}$

Exploratory analysis of transition probabilities

I use the approach of Kline and Tartari (2016) to estimate bounds on the transition probabilities between the four elements of the effort partition. I restrict the transition probabilities by assuming that an increase in visibility or the bonus amount cannot decrease effort.

For each pairwise treatment comparison the main parameters of interest are the three transition probabilities from the lower parts of the distribution into anti-bunching at the bonus. Since this technique can fail to yield informative bounds and because I am concerned to be underpowered I do not consider it part of my main analysis.

Robustness

I repeat the analysis using the alternative partition

1. **Bunching at zero:** Effort $= 0$
2. **Anti-bunching at zero:** Effort $\in \{1, \ldots, 14\}$
3. **Bunching at bonus:** Effort $= 15$
4. **Anti-bunching at bonus:** Effort $\in \{16, \ldots, 38\}$

3. Distribution of effort above the bonus threshold

I hypothesize that an increase in the bonus increases the effort of subjects with effort strictly above the bonus threshold. To test this hypothesis I compare the top quantiles of the effort distribution in the low bonus level group (*40c pooled*) with the same quantiles in the high bonus level group (*$1.20 bonus*). I ignore the *No bonus* and the *No badge* groups, because I expect substantially fewer subjects to go beyond the bonus threshold in these groups, which limits the statistical power of any distributional tests.

I focus my analysis on the conditional effort distributions $(e_i | F_e(e_i) \geq p_{\text{badge}}, \text{visibility}_i = \text{badge}, \text{bonus}_i)$ with CDFs denoted as $\tilde{F}_{e|\text{badge,bonus}}$ and $p_{\text{badge}} \equiv \Pr(e_i \geq 17 | \text{bonus}_i = \text{40c pooled}, \text{visibility}_i = \text{badge})$. I use the respective sample analogues to conduct three kinds of tests.

1. A test against $H_0 : \tilde{F}_{e|\text{badge,\$1.20 bonus}} >_{\text{FOSD}} \tilde{F}_{e|\text{badge,40c pooled}}$
2. A test against $H_0 : \tilde{F}_{e|\text{badge,\$1.20 bonus}} \not>_{\text{FOSD}} \tilde{F}_{e|\text{badge,40c pooled}}$
3. Taking $\tilde{q} = \max\{17, \tilde{F}^{-1}_{e|\text{badge,\$1.20 bonus}}(p_{\text{badge}})\}$ a series of tests on the

top part of the effort CDF, using the estimates from section 2:

$$\begin{aligned} H_0(q) : \max\{0, (\sum_{e=0}^{q} \beta^{e}_{\text{badge,\$1.20 bonus}}) - (1 - p_{\text{badge}})\} \\ = \max\{0, (\sum_{e=0}^{q} \beta^{e}_{\text{badge,40c pooled}}) - (1 - p_{\text{badge}})\} \end{aligned}$$

for $q \in \{\tilde{q}, \tilde{q}+1, \ldots, 38\}$

For 1. I use a one-sided Kolmogorov-Smirnov test. For 2. I use an empirical likelihood ratio test with Bootstrapping inference as outlined in Davidson and Duclos (2013). I pre-specify the interval of interest for restricted dominance to be $[18, 37]$. In addition, I report the largest interval that is still rejected at $\alpha \in \{0.01, 0.05, 0.1\}$. For 3. I use the same testing procedure as in section 2.

Robustness

- I repeat the analysis using
  - (i) $p_{\text{badge}} \equiv \Pr(e \geq 16 | b = \text{40c pooled, visibility} = \text{badge})$, $\tilde{q} = \max\{16, \tilde{F}^{-1}_{e|\text{badge,\$1.20 bonus}}(p_{\text{badge}})\}$, and $[17, 37]$ as interval of interest for restricted dominance
  - (ii) *40c bonus* as low bonus level group
  - (iii) *80c gift* + *40c bonus* as low bonus level group

Analysis of Beliefs

At the end of the experiment I show each subject 5 pairs of badges and ask for each pair (i) "Who is more generous?" (ii) "Who do most other MTurk workers say is more generous?" The badges represent real data, but are shown randomly with or without a picture, with most badges being shown without a picture.

Sample exclusion

- I only consider answers to badges that are shown without a picture, in order to remove any effect of seeing a picture on beliefs.
- I only consider answers to question (ii) in line with the hypothesis that own beliefs about what others' see as generous are the main channel behind image motivated behavior.

- I drop all answers given within 3 seconds or less, in order to reduce the noise induced by subjects answering without considering the information displayed on the badges. I also drop the first of the five answers, since I do not observe how fast it was given.

Graphical analysis

A pair consists of two badges, badge 1 and badge 2, displayed in random order either left and right, or right and left. For each pair I calculate the share of subjects answering "badge 2" to question (ii) pooling different display orders. If the same subject answers multiple times on the same pair, I compute their average answer for that pair before averaging answers across subjects. I also compute confidence intervals for these averages, clustering on the subject level.

The resulting averages make up social image functions, which I summarize graphically. Denote the share of subjects answering "badge 2" to question (ii) as $\gamma_{b_1,b_2}^{e_1,e_2}$, where b_i and e_i are the bonus level and effort level displayed in badge $i \in \{1,2\}$. In a graph I fix e_1 and b_1, and then plot $\gamma_{b_1,b_2}^{e_1,e_2}$ varying e_2 and b_2. I draw these plots for $e_1 \in \{1,15,16,17\}$ and $b_1 = 40$c bonus. If two badges of a pair are identical, I do not collect data on them and set $\gamma b_1, {b_1}^{e_1,e_1} = 0.5$.

Statistical analysis

To maintain statistical power I focus on anti-bunching at the bonus, using *40c bonus* as control group, and cases in which a higher bonus amount makes a higher level of effort appear less generous. The main prediction is that the social image function at and above the bonus threshold changes with treatments. I test for all $e_2 \geq e_1$ with $e_1 \in \{16,17\}$ against

(a) $H_0(e_2) : \gamma_{\text{40c bonus,\$1.20 bonus}}^{e_1,e_2} \geq \gamma_{\text{40c bonus,40c bonus}}^{e_1,e_2}$

(b) $H_0(e_2) : \gamma_{\text{40c bonus,\$1.20 bonus}}^{e_1,e_2} \geq \gamma_{\text{40c bonus,80c gift + 40c bonus}}^{e_1,e_2}$

(c) $H_0(e_2) : \gamma_{\text{40c bonus,80c gift + 40c bonus}}^{e_1,e_2} \geq \gamma_{\text{40c bonus,40c bonus}}^{e_1,e_2}$

Any rejection in (a) or (b) indicates a downward shift in the image function as predicted by my model. I also predict to reject $H_0(e_2)$ from (a) at more and higher levels of e_2 than $H_0(e_2)$ from (c).

Robustness

I repeat the analysis including answers made within 3 seconds or less.

Exploratory Analysis of Heterogeneity

I also conduct an analysis on heterogeneity of treatment effects based on beliefs and gender. As I expect to lack statistical power for making meaningful statements on heterogeneity, I report it as exploratory analysis that might be helpful for future research.

Beliefs

I test whether beliefs in a shift in the social image function are predictive of behavior. I select subjects that are shown a pair $P = (e_1, b_1, e_2, b_2)$ with $e_1 \in \{16, 17\}$, $b_1 \in \{40\text{c bonus}, 80\text{c gift} + 40\text{c bonus}\}$, $e_2 \in [e_1 + 1, e_1 + x]$ $b_2 = \$1.20$ bonus, and another pair $P' = (e'_1, b'_1, e'_2, b'_2)$ with $(e'_1, b'_1, e'_2) = (e_1, b_1, e_2)$ and $b'_2 \neq b_2$.

For the selected sample, I create a "reversal" dummy variable indicating whether a subject has simultaneously answered "badge 1" in P and "badge 2" in P'. I then calculate heterogeneity of the treatment effects from section 2, hypothesizing that those with a reversal show a stronger pattern than those without it.

The benefit of choosing a smaller x is that answers to pairs with e_2 closer to e_1 are possibly more predictive of behavior. The benefit of choosing a larger x is a bigger sample size. I choose x as follows: For a given x denote the number of reversals as $n_r(x)$. I choose the smallest integer $x \geq 1$ that satisfies $n_r(x)/n_r(38 - e_1) \geq 0.8$.

Gender

I split the sample by *sex*, and calculate heterogeneity in treatment effects.

Other Degrees of Freedom

Differential attrition due to consent form

Subjects see a consent form at the beginning of the study. Subjects also see a second consent form before creating their badge and completing effort tasks for charity. In *Badge* treatments the second form asks for consent to show the subject's picture to other workers. In *No badge* treatments the second form only asks for consent to continue the study. This difference in consent forms might lead to differential attrition, possibly biasing any comparison of effort between the two visibility treatments.

To check the robustness of my results to differential attrition, I assume that attrition is monotone and compute Lee (2009) bounds. These bounds can be tightened by computing them within cells of pre-treatment covariates that are thought to be predictive of attrition. However, covariates need to be chosen carefully chosen, as a larger number of covariates makes it more likely that monotonicity is violated in the sample due to sampling error.

I pre-specify to use *education*, *sex*, *age* and *Instagram usage* as covariates. As an additional protection against a sample violation of monotonicity, I combine the least frequent levels of each of these categorical variables into an "other" category until "other" comprises at least 20% of the population.

Adjustment for multiple hypotheses testing

I conduct $1+3+3=7$ tests in the analysis of average effort and $4+2+2+2+2+2+2=16$ tests in the analysis of the partitioned effort. Given the large number of tests, I expect that controlling for the family-wise error rate would make it too difficult to reject hypotheses.

Instead of controlling the family-wise error rate, I choose to control the false discovery proportion (FDP), which is defined as the share of false rejections among all rejections and set to be zero if there are no rejections. I use the FDP-StepM method described in Romano, Shaikh and Wolf (2008), which controls $\Pr(\text{FDP} > \gamma) \leq \alpha$. I pre-specify $\gamma = 0.1$ and $\alpha = 0.05$.

References

Davidson, Russell, and Jean-Yves Duclos. 2013. "Testing for Restricted Stochastic Dominance." *Econometric Reviews*, 32(1): 84–125.

Kline, Patrick, and Melissa Tartari. 2016. "Bounding the Labor Supply Responses to a Randomized Welfare Experiment: A Revealed Preference Approach." *American Economic Review*, 106(4): 972–1014.

Lee, David S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Review of Economic Studies*, 32.

Romano, Joseph P., Azeem M. Shaikh, and Michael Wolf. 2008. "Formalized Data Snooping Based On Generalized Error Rates." *Econometric Theory*, 24(02).