

Estimating Social Preferences and Gift Exchange at Work

Addendum to Pre-Registration

Feb 2021

Stefano DellaVigna, John List, Ulrike Malmendier and Gautam Rao

This document presents the pre-registered design for Experiment 3 for the revision of our paper “*Estimating Social Preferences and Gift Exchange at Work*.” This pre-registration is being done before running the experiment, but after obtaining approval from U Chicago IRB and conducting a small pilot (n=200) of the task on MTurk to ensure feasibility.

Goals. The experiment is designed to complement the two existing experiments (Exp 1: Productivity Experiment and Exp 2: Labor Supply Experiment) while achieving the following eight goals: (1) an extensive-margin design (i.e. how much work you do in total, rather than how fast you work in a given fixed amount of time), (2) a similar task to the previous experiments, (3) keeping gift and piece-rate variation, while also (4) investigating altruism versus warm glow (i.e. whether participants respond to the return to the employer). The experiment will (5) focus on outputs (units of work done) rather than inputs (minutes spent), (6) be pre-registered as the previous two versions and (7) replicate the core of Experiment 2. We plan to recruit workers on MTurk, which allows us to (8) provide a large sample at moderate cost, which means the design is more scalable and can be easily used by others. The task will be externally valid while also providing a fairly natural way to measure (and vary) the return to the employer: the workers will check a set of addresses for the Alumni Office at the University of Chicago and flag errors. Since envelopes with an incorrect address are often returned to the sender or lost, flagging such errors is important so the mailer will not be wasted. Since the data set of nearly 6,000 addresses has approximately 10% of addresses with errors or missing fields (based on random checks), we will truthfully inform workers that on average they save us 10% times the cost of a mailer per address checked.

Design Details. We plan to recruit 2,000 workers on MTurk for a one-time 8-12 minute job paying \$1.60 (in line with MTurk typical pay).¹ The recruitment will be done by the University of Chicago, which is the employer. Since the recruitment is for an actual work task, the workers signing up will not have to sign a consent form and will be taken directly to the task. (This is IRB approved.) The experiment thus can plausibly count as a natural field experiment. All workers then see the description of the task. *“In this task, you will check if addresses from alumni of the University of Chicago we have on file are accurate. It is important for our institution to reach out to the alumni to keep them engaged and informed about our activities. Based on our experience with similar data, about 1 out of 10 addresses in our database are incorrect and in the past these mistakes have caused the letters to be returned to us. Using Google Maps, you will help us ensure that we are not sending envelopes to incorrect or nonexistent addresses. You will*

¹ We will exclude from analysis workers who categorize 20 or more addresses out of 40 incorrectly. The “correct” classification will be assumed to be that made by 70% or more of participants who evaluated a particular address.

be given 40 addresses to check, and we expect this to take you between 5 and 10 minutes. As mentioned in the advertisement, the Becker Center will pay you \$1.60 for your work. After you finish, we will ask you if you have time for any extra work today.”

This description of the task is truthful, makes the case for the value of the work, and is parallel to Experiment 2 as much as possible. In the next page, we provide more detail on how to check addresses, including an example. After this, the subjects move on to the list of addresses, which are displayed in groups of 5 on a page (to avoid confusion).

Only after the conclusion of this required work period does the experimental variation start. All workers are asked whether they would be willing to do some more work, coding up to 20 additional addresses. We have five main experimental arms: **Control (25%), Low Piece Rate (25%), Medium Piece Rate (12.5%), High Piece Rate (12.5%), and Gift Exchange (25%)**, where the shares in parentheses indicate the share of the sample assigned to the arm. This largely parallels experiment 2, except for the addition of a medium piece rate which will make it easier to identify parameters non-parametrically (more on this below), as well as the fact that there is only one gift exchange arm. We are replicating the monetary gift from Experiment 1, but not the non-monetary gift of a mug, or the early gift (which took place before the end of the work period). In the MTurk context, a non-monetary gift is both harder to implement and seemingly out of place as well, so we dispensed with that. The early gift is not possible for a different reason: in online platforms, attrition is a first-order issue and a gift at the start of the experiment would likely have affected the share of workers that complete the task compared to the other experimental arms; in turn, this would bias the estimate of the impact on the extra work margin. As the design stands, we ensure identical attrition across tasks because all randomization takes place after the required work period.

Importantly, we cross the 5 arms above with a manipulation of **normal-return to employer (50%) versus higher-return to employer (50%)**. This is achieved by communicating to the workers the value of their work in terms of the cost of a typical mailer (50 cents), versus the cost for a larger mailer (1 dollar). Note that, since about 10% of addresses are incomplete, these imply a return to employer of 5 cents and 10 cents per address check, respectively. Since the Alumni office sends off both types of mailers (and other mailers), this is truthful variation that, we found, is relatively easy to explain. This variation allows us to test the altruism versus warm-glow model.

Here we reproduce the instructions for this critical part of the experiment, which are as parallel to Experiment 2 as possible:

“Thank you for your work today. You have completed the work we hired you for and will be receiving the \$1.60, as advertised.

*[Only if **GIFT**: In addition, as a token of appreciation, the Becker Center is giving you an additional 40 cents for helping today. Therefore, we are paying you a total of \$2.00.]*

*Thanks to your work, we estimate that for every address you check we are saving roughly 5 cents [10 cents in **HI RETURN**] on average (1 out of 10 addresses has a mistake on average. Each mistake you identify saves us approximately 50 cents [1 dollar in **HI RETURN**] in costs of mailing our next newsletter)*

If you happen to have some time and are willing to do some extra work, that would be appreciated.

*[**CONTROL or GIFT**: Unfortunately, we cannot compensate you for this extra work.] [**PIECE RATES**: We*

will pay you 1/2/4 extra cents [**LOW/MID/HIGH PIECE RATE**] for every address you check, up to 20 addresses. For example, if you check 10 additional addresses, we will pay you 10/20/40 extra cents. [**LOW/MID/HIGH PIECE RATE**] Even checking a few addresses would help, but we totally understand if you are not available to do more.

Would you be willing to help us by checking up to 20 additional addresses? () Yes () No”

If the worker agrees to do more, we show 20 additional addresses to check in batches of 5. The worker can end the task at any time and get to a page with the code to redeem the promised pay (inclusive of the gift and/or any additional earnings if relevant).

Design Features. The design is as parallel to Experiment 2 as possible given the setting with, we believe, some small improvements. We discuss first how we aimed to make it parallel.

1. (*Pay*) In Experiment 2 we pay workers \$60 to come for a 2-hour in-person job. In the proposed Experiment 3 we pay MTurk workers \$1.60 for a task of checking 40 envelopes. While we had to pay \$30 per hour to recruit enough workers for Experiment 2, \$1.60 is an appropriate pay for a 10-minute task on MTurk (we recruited the workers for our pilot within hours, confirming this) and in line with pay rates used in recent papers using MTurk.
2. (*Measure of Effort*) In Experiment 2 we measure the extent of extra work in terms of extra minutes. In the proposed Experiment 3 we instead measure the extra work in terms of extra output (addresses coded), thus measuring “outputs” rather than “inputs”. This switch is required by the online setting, in which we just could not monitor minutes of work (workers can just leave the screen open and multi-task) and it would thus be atypical on MTurk to ask for minutes of work instead of units of output; hence, the switch to units of output produced as our metric.
3. (*Magnitudes of Piece Rate and Gift*) We carefully calibrated magnitudes to keep the various treatment interventions in Experiment 3 comparable to Experiment 2. So the gift is 40 cents, 25% of the total pay of \$1.60, the same percent as in Experiment 2, a gift of \$15 to a total pay of \$60. In Experiment 2 the piece rates for doing extra work were set at 1X the rate for the initial period (high piece rate) and at 0.5X (low piece rate). In Experiment 3 we similarly have the 1X and 0.5X piece rates (corresponding to 4 cents and 2 cents per address, respectively), but in addition we also added a lower, 0.25X piece rate (1 cent per extra address). The comparability, proportionally speaking, of the gift and of the piece rates make it easier to compare the structural estimates, as we discuss below.

Experiment 3 has two main advantages in design compared to Experiment 2.

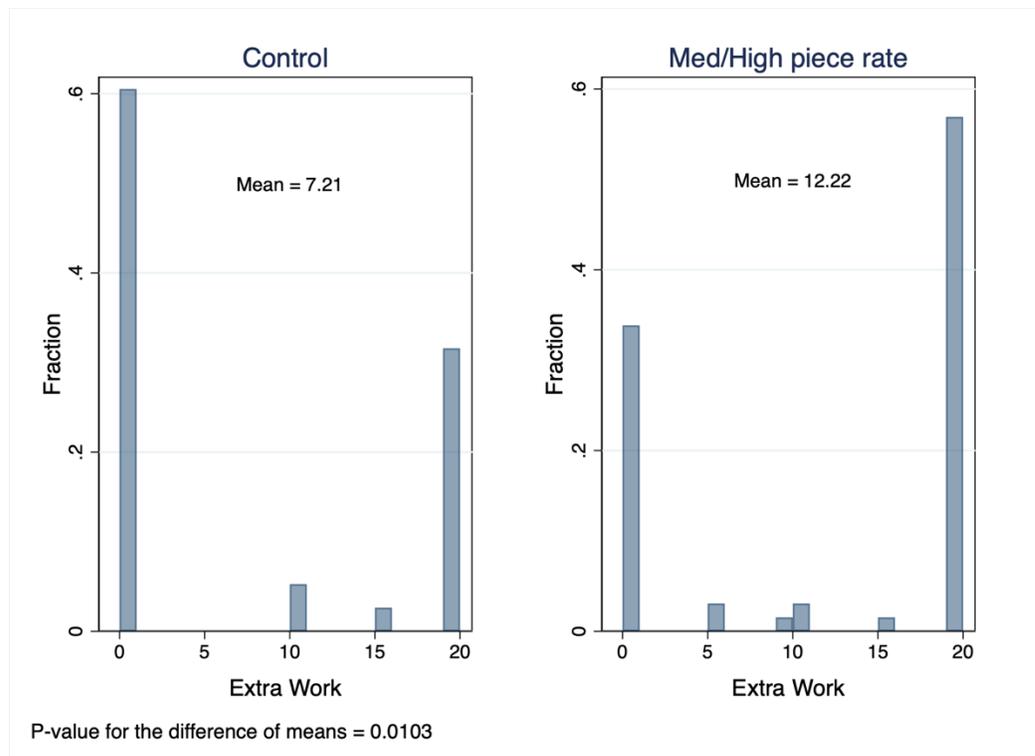
- a. (*Value of Work*) First, the value of work has a natural interpretation: for each address which the worker points out is incorrect, the Alumni office saves the cost of a mailing. We estimate a rate of incorrect addresses of about 10 percent, that is, cases with incorrect zip code, wrong city, address is non-existent, etc. In all these cases the mailing is not likely to reach destination. By not sending it to such wrong addresses, the center saves the cost of a pointless mailer. Since some mailers cost 50 cents (the more basic ones) and others around \$1, we have a natural variation of the cost as well that we can convey to the workers. We considered a number of

alternatives to vary the value, including having a sample of addresses with a different percent of incorrect addresses (but this also alters the cost of effort), or giving the saving for 4 mailings in a year, as opposed to the savings for one mailer. We converged ultimately on the number which we provide in the current design for its simplicity and validity.

- b. (*Non-parametric identification*) While we provide structural estimates of the social preference parameters in the paper, it would be desirable if possible to provide non-parametric identification of the parameters. Indeed, in Experiment 2 we present a result akin to this, showing that the effect of the gift is equivalent to about one half the effect of the lower piece rate. While this allows some non-parametric identification, ideally we would have a piece rate with an effect that equals the one of the gift, since then the identification of the gift effect on social preferences is trivial. As it is in Experiment 2, the fact that the gift effect is one half the effect of the lower piece rate implies that we need to extrapolate the effect, which makes the estimation of curvature of the cost of effort more relevant. This is the motivation to add a lower piece rate, at $\frac{1}{4}$ the rate for the initial batch, so that if the gift exchange effect is small but still sizable, it may be identified by the low piece rate.

The main limitation of the design of Experiment 3 is that we are implementing only one (monetary) gift arm for the reasons stated above.

Pilot. To test the design, we ran two 100-worker pilots, for a total of 200 workers, with encouraging results. We were able to recruit the target sample within hours at the set wage, did not receive any substantial negative feedback from workers in the comment box offered at the end, with mostly positive feedback to the task. Further, the task proceeded well, with no hiccups. While any results were not the focus of the pilot, we do find encouraging evidence of a responsiveness of extra work to the piece rate. Below we plot the distribution of extra effort comparing the control group to the pooled medium- and high-piece rate. In both cases, most workers either do zero extra work, or they do all the extra work. In the piece rate group, there is a sizably larger share of workers doing the latter. This suggests a sizable elasticity, which is a keystone of Experiment 2. We stress that we did not use the pilot to make design changes (indeed, the pilot has the same design), but to insure that there were no hiccups of implementation.



Structural Estimates. We aim to analyze the results in light of the same model which we used for Experiments 1 and 2, so as to enable comparison of the structural estimates across experiments. We focus on two main parameters which we hope to identify and compare with Experiments 1 and 2.

First, we will aim to estimate, as in Experiment 1, a model with both warm glow a and altruism α . The key difference between the two is that the altruism term takes into account variation in the return to the employer, while the warm glow term evaluates an average employer return. We will aim to compare the value of the altruism parameter α in Experiment 1. While it is of course too early to tell from the pilot data, the estimates on the observations from the pilot suggest that we should be able to get quite precise estimates on α , similarly precise to what we find in Experiment 1. We should thus be able to draw a meaningful comparison.

Second, we will aim to compare the change in social preference due to a 25 percent gift. Assuming that the social preferences take the warm glow form, as opposed to pure altruism, we will estimate the increase in the warm glow a , or Δa , due to the gift.² The estimates on the pilot data suggests that we should reach about the same level of precision for this term as in Experiment 2, and we will then compare the estimates to those in Experiment 1 and Experiment 2.

² To estimate the warm glow model, as in Experiments 1 and 2, we will normalize the warm glow term as $A = a \cdot \bar{p}_E$, where \bar{p}_E denotes the average return to the employer. To keep the estimates consistent with Experiment 2, we will use $\bar{p}_E = \$0.04$ per address checked, equal to the replacement cost of hiring another worker to do the work instead (since workers are paid \$1.60 for 40 addresses in the main task, corresponding to \$0.04 per address checked). This normalization is also quite close to the lower return to the employer of \$0.05 per address.

Both comparisons would be, in our view, quite meaningful and would illustrate the value of the model and estimation in allowing one to draw comparisons across quite different experimental designs. I believe this point is in line with the editorial question and intent.

We should point out that one parameter which we intend to be context-specific and which we do not intend to be comparable across tasks is the baseline warm glow a . Indeed, this baseline warm glow is not even point identified in Experiment 2 as Table 4 reports. This is because in the warm glow model it is hard to exactly back out separately the level of baseline motivation from curvature when there is no response to the return to the employer. Nonetheless, the change in warm glow is clearly identified, which is the key reciprocity parameter we want to capture.

As an auxiliary analysis, we may estimate a warm-glow model (that is, assuming the altruism parameter $\alpha=0$) allowing for a change in social preference Δ_{high} when the return to the employer is high. This may capture in a reduced-form way whether the worker thinks a high return to employer is unfair (holding fixed the return to the worker), say due to inequity aversion. This specification will be relevant if increasing the return to the employer *reduces* worker effort.