Estimating Social Preferences and Gift Exchange at Work

Stefano DellaVigna, John List, Ulrike Malmendier, Gautam Rao

Analysis Plan Update, September 2018

Summary

In this update to the analysis plan, we propose an additional set of field experiments on social preferences and gift exchange in a one-time work task. The new set of experiments will be based on the same model of social preferences as the previous set of experiments. However, unlike the previous experiment which was partly a *within-subject* design, the new design will be entirely a *between-subject* design. In order to do so, and as an independent contribution, we will change the measure of effort from effort within a given unit of time (i.e. the speed of work) to willingness to work extra time. The task performed will also change, from preparing mailers for a charity to data-entry (coding up data as a research-assistance task). This pre-registration is completed in advance of any data collection for the new experiment (other than collection of pilot data to ensure the setting and design are appropriate).

Motivation and Design Summary

In our previous experiment, we used a within-subject design to increase statistical power. That is, each worker was hired to fold envelopes for a one-time job. The work schedule consisted of stuffing envelopes for 20 minutes, a short break, then the next group of envelopes to be stuffed for 20 minutes, and so on 10 times. The 20-minute work periods had two key manipulations -- different piece rates for the worker, and different returns to the employer -- so as to allow us to estimate the elasticity of effort to variation in the private piece rate and in the return to the firm. Having such variation within-person allows us to control for person fixed effects in productivity, increasing substantially the statistical power. In the last 2 work rounds in the experiment, we had gift treatments, which were between subjects. In each 20-minute work period, we measured the effort as the number of envelopes stuffed within 20 minutes.

A possible concern with this design is that the series of 10 manipulations make the task less natural, and could have an impact on the findings. An alternative is an experiment where all the variation, not just the gift variation, but also the variation in piece rate and in return to the employer, is between subjects. The difficulty is doing so without losing considerable statistical power, given that one would not be able to control for the person fixed effects in such a design, and further one would observe the results of a particular treatment only in the relevant between-subject sub-group of subjects, as opposed to for each subject. This leads to a further loss in statistical power for a given total number of subjects.

To achieve the goal of a between-subject design without compromising statistical power, our new planned experiment will change the measure of effort. The gift exchange literature has measured effort by how fast people work *within a fixed amount of time*, within a 6-hour period

for Gneezy and List (EMA) and Kube et al. (AER, JEEA) and within a 20-minute period in our first experiment. This is more generally true also for real-effort experiments focused on other topics, like Gneezy, Niederle, and Rustichini (QJE), and Gill and Prowse (AER).

The key problem with this type of design is that effort within a fixed amount of time is inelastic to motivation, which drastically reduces power. How many more envelopes, or library books, can one code within a fixed time period, even as motivation goes up? There is a clear ceiling. Thus, it is not surprising that the elasticity of effort to incentives that we estimate is low, in the order of 0.1. Other studies that estimate elasticities in real-effort tasks, like DellaVigna and Pope (RES) or De Araujo et al (JESA), estimate similar or lower elasticities. At low elasticities, it is hard to be well-powered, and more so when there is substantial between-subjects heterogeneity in ability. Thus our previous requirement of using a within-subject design to increase power, with a large number of subjects to boot.

The main idea of the new experiment is to change the margin of effort to one that is *more elastic to motivation*. Abeler et al (AER) and Gneezy et al (JEEA) are inspirations for this: subjects work for a fixed amount of time, and then the experimenter asks subjects if they would stay for an extra period. *How long people stay extra to work* is the measure of effort. In this design, people are quite sensitive in their staying decision: based on our pilot data, the elasticity is close to 1, as opposed to 0.1 or 0.2 in the older design, providing significantly more power for a given sample size.

Here is the proposed design. We hire people for a one-time two-hour data coding job at the Becker Center for $60. (We determined that this was the required market rate, see below. It is also a natural task.) This part is identical in all treatments, except for treatment '*early gift*'. After we pay the workers, we ask them if they would be able to stay a bit longer to code some extra data, for up to one extra hour. The key outcome variable is *how many extra minutes they stay*, from 0 (minimum) to 60 (maximum). These are the six between-subject treatments:

(a) (*control*) The control group is requested to stay longer, but informed that we are unable to pay for the extra work;
(b) (*monetary gift*) The monetary gift treatment group is paid $15 above the pre-announced pay as a token of gratitude, and then asked whether they can stay longer, for no extra pay;
(c) (*non-monetary gift*) The non-monetary gift treatment is given a gift-wrapped thermos of $15 value as a token of gratitude, and then asked whether they can stay longer, for no extra pay;
(d) (*medium piece rate*) The medium-piece rate is asked if they can stay longer, at the rate of 25 cents per minute, that is, $15 per hour (half the rate of the first two hours); if subjects stay for the full hour, they earn $15, which is the same as the gift;
(e) (*high piece rate*) The subjects are offered 50 cents per minute, that is, $30 per hour, if they can stay longer; if subjects stay for the full hour, they earn $30, which is the same as the initial wage;
(f) (*early gift*) The non-monetary gift treatment is given a gift-wrapped thermos of $15 value as a token of gratitude, at the beginning of the experiment. At the end of required two hours of work, the subjects are asked whether they can stay longer, like control group (a).

Further, within each treatment we cross-randomize a simple and, we think, externally valid measure of the value of work for the employer: we tell a high-value group that "*getting the extra data entered today is really valuable to us.*" Thus, the treatment has a total of 6*2= 12 between-subject cells.

The precise scripts, including the addition of a short pause to complete paperwork between making the payments / gifts and requesting the extra work, are in the next section.

This between-subject experiment aims to identify the model of social preference and gift exchange taking advantage of the same variation as in the previous within-subject design. In particular, like in the previous experiment we take advantage of variation in the piece rate to identify the cost of effort function (which is now interpreted as effort of doing the extra work). This was always the key methodological component to the design, which we now emphasize even more, calling it a "piece-rate-metric design" for real-effort experiments.

There are two respects in which the new design does not replicate the previous design. First, we will not have negative gift treatments. In the previous design, the negative gift treatment was possible (without deception) because we could compare work under an unexpectedly low pay of $3 (per 20 minutes) to work under the previous pay of $7 (per 20 minutes) for the same employer. The previous pay worked as a reference point, against which the lower pay plausibly served as a loss (which we verified later in debriefing surveys). With a between-subject design, and in particular a design that employs length of extra work as measure of effort, we could not find a plausible way to do this. Still, we replicate both the monetary gift and the non-monetary gift treatments, which is the vast majority of the gift exchange treatments in the literature. Second, with the more natural data coding task, we do not explicitly state an exact value of work to the employer, which was a component of our old design. Nevertheless, in our new experiment the value of work to the employer can plausibly be inferred to be the cost of future subjects to code the same work, $30/hour. Thus, the value of work has a more natural interpretation.

Instructions, Pilot, Model, and Power Calculations

Recruitment and Sample Size. We will recruit subjects for the $60 2-hour data coding task both online, posting ads on Craigslist, as we did for our first experiment, and on the University of Chicago campus, posting ads on Marketplace and using student mailing lists to advertise. For physical on-campus recruitment we will post flyers in locations where students usually find ads for jobs.

It is possible that the subject participation from both of these sources may not be enough to get to our envisioned sample size, which is 300 subjects, that is, about 50 for each treatment conditioned. We thus ensured the collaboration of the University of Illinois at Chicago and Chicago State University, where we will be able to post on their campus as well to speed up recruitment. Overall, the experiment is expected to take between three and six months to recruit the envisioned sample of 300 workers across the six treatments. We will stop sooner if we run out of subjects, and may run longer if recruitment is slower than anticipated but steady.

Randomization. We will randomize subjects into the 12 treatments by strata of 12 consecutive subjects. In every strata, each treatment will be randomly assigned to one subject.

Instructions. This is the exact wording for the experiment, with the different treatments

"*[Sit each participant in an individual work space]* Hello, today you will be working in a one-day-only job to enter data for the Becker Center at the University of Chicago. This data is going to be used in an academic paper. Please look in front of you. You should see several sheets of data that need to be entered. You will be entering the data into a Google Spreadsheet. The spreadsheet has two tabs. The first tab is a legend that describes how you should enter the data. The second tab is a blank spreadsheet for you to enter the data into. Please look over the legend and let me know if you have any questions.
As mentioned in the advertisement, the Becker Center will pay you $60 to enter the data for two hours. [**Early Gift:** In addition, as a token of appreciation, the Becker Center is giving you this thermos with a retail value of $15 for helping today.] Payment will occur at the end of the two hours. After that, we will ask you if you have time for any extra work today. I will be in and out of this room today, but please email me if you have any questions. Do you have any questions before I leave? *[Pause for questions]* Great. Please begin working.
[Start Time: _____:_____(AM/PM)] *[End Time:_____:_____(AM/PM)]*

[At the end of the 2-hour session] Thank you for your work today. You have completed the work we hired you for, so here is the $60, as advertised. [**Monetary Gift/Non Monetary Gift:** In addition, as a token of appreciation, the Becker Center is giving you (**Monetary Gift:** an additional $15 for helping today. Therefore, we are paying you a total of $75./ **Non Monetary Gift:** this thermos with a retail value of $15 for helping today.)]
Could you please show me the last sheet of data that you have entered? *[The experimenter will take a moment to look at the sheet]* Please allow me a minute to look through the data you have entered *[The experimenter will take a minute to go through the data]* Thank you, this looks good*.…..[do receipt]...*

If you happen to have some time available, even a few minutes, and are willing to do some extra work, that would be appreciated. [**High Return:** Getting the extra data entered today is really valuable to us.] Would you be willing to help us enter some more of the data for up to one hour? [**Control/Monetary Gift/Non-Monetary Gift/Early Gift**: Unfortunately, we cannot compensate you for this extra time.] [**M-H Piece Rate:** We will pay you [¢25/ ¢50] for every minute of work that you do, up to one hour. For example, if you do an extra 20 minutes of work, we will pay you $5 [$10] extra.] Even a few extra minutes of work would help, but we totally understand if you have to leave soon. As soon as you are ready to leave, I will be right outside and will immediately [**M-H Piece Rate:** pay you the extra amount and] escort you from the building. *[If the subject does not want to stay and enter data, escort them from the building.] [If the subject is willing to stay, tell them to e-mail you when they are done for the day and you will show them out of the building.]* "

Pilot. We collected pilot data to ensure that the design was feasible, and to enable power calculations (below). We recruited subjects for a 2-hour data coding task both on Craigslist and on a website posting to University of Chicago students, initially at the rate of $15 per hour, approximately what we paid for the first experiment. But it became clear that $30 for two hours of work was not competitive pay and was attracting very few applicants. This is both because there is a fixed cost aspect in the job, and a 2-hour job needs to pay a higher hourly rate than a 5-hour job; but also the stronger economy and a significantly higher minimum wage in Chicago than in 2015. Over the course of 4 weeks, we ran a small-scale experiment randomizing pay for a 2-hour job across postings at $30, $50, $70 and $90 total pay. There was a strong increase in

response at $50 and $70; we thus settled for $60 per hour, which gives us a nice round number as a per-minute (and hourly) rate, and also ensures that we stay below $100 as total compensation for the experiment, even including payment for the extra work (at $100 the participants would have to have taxes withheld).

As a complementary approach, we also benefited from pilot data that one of us (Stefano) collected together with Devin Pope on an MTurk sample. Stefano and Devin are going to run a follow up to their REStud 2018 paper with an online real effort task in which they vary in several ways the experimental design, including using a design where workers code 40 historical cards (taking about 8 minutes) and then are asked if they would stay longer. Stefano and Devin examine the impact on working longer of varying the piece rate for the extra work, just as we plan to do in the new design. The pilot sample of 400 workers was highly responsive to the incentives in their decision to stay longer, with an elasticity between 0.3 and 1, depending on the specification. This is much higher than the elasticity of about 0.1 (or lower) in the version with a fixed amount of time. This higher envisioned elasticity is a key motivation for the design and it allows us to do the power calculations below. (We note that the DellaVigna-Pope evidence is clearly separate from this project given that it is not focused on gift exchange and it is not in the context of a natural short-term job.)

Model. The model remains the same as in the previous draft. We can identify the social preference weight a under the (plausible) assumption that the value for the employer of the extra work, p_E, equals the amount that the employer would have to pay to have more workers do the coding, that is, $30/hour (which is the rate advertised for the job, and thus plausibly the rate that we would have to recruit subjects for). We will also have a treatment that qualitatively affects the perceived return to the charity (with the wording "*Getting the extra data entered today is really valuable to us.*"), which we can identify as being associated to a higher a. We do not have, though, a treatment that varies in a quantitative way the return to the employer, and thus we cannot use this treatment to compare the "warm glow" specification versus the "pure altruism" specification. Given the strong evidence in the first experiment against the pure altruism specification, we will proceed with a "warm glow"-type specification, $a*p_E$., assuming p_E=30. All other components of the model, especially the cost of effort function, remain the same. In the high-return-to-the-employer condition, we assume that the social preference weight a can be higher.

Outcome Variables and Estimation. We pre-register the main outcome variable to be the number of minutes of extra work that subjects in different conditions do. As an alternative dependent variable, we will use the share of subjects that stay on for at least some time. We do not use the productivity in the initial two hours as outcome variable since five of the six treatments take place *after* this initial work period. For the early gift treatment, the gift could potentially have an impact on productivity in the first 120 minutes, but this prediction is underpowered, as the pilot calculations below make clear. Thus, we will report it, but it is not a main outcome of interest. For the early gift treatment as well, we will focus on the impact on the extra work, testing whether there is a decay effect, compared to the effect of gift given shortly before the request for extra work.

For our main comparison, we plan to pool across the value of work, and compare the results across the six groups. For most power, we will also compare the pooled three gift treatments to the control group and the group with the low piece rate. We will do these comparisons unconditional, and controlling for any control variable that we have (such as demographics and pool we recruited from).

We will use these same variables for the structural estimation. The model is the same as in the previous version, except that the choice variable *e* is now in units of minutes, as opposed to envelopes. The estimation of the model will be done still assuming both an exponential cost of effort function and a power cost of effort function. The estimation now needs to take into account the censoring at 0 minutes (not staying longer) and at 60 minutes (staying all the 60 minutes). We thus estimate the model with maximum likelihood, allowing for censoring at these two thresholds.

Power. An important advantage of the structural estimation, combined with the pilot data collection, is that we can do model-based power calculations to inform the design. This is different than the more standard reduced-form power calculations.

In the table below we compare simulations with power calculations for two alternative *between-subject* designs: (i) a design a la Gneezy-List and Marechal et al. with measures of how many units of output the workers get done in 2 hours comparing a control group; and (ii) a design like our new one, with the unit of measure being how many minutes the workers stay beyond the required 120 minutes. Both designs, unlike our previous one, are between subjects; the difference is that (i) is the between-subjects version of our old design, and (ii) the new design. For all three treatments we assume that the N subjects are divided 5 ways into a control group, two gift groups, a low-piece rate group, and a high-piece rate group. (For these calculations, we do not count the *early gift* treatment.) For design (ii) (the new one) we use the envisioned piece rates, \$0.25 per minute (low piece rate) and \$0.50 per minute (high piece rate). For the old design, we set the piece rates such that the per-minute pay is about the same, that is, \$0.12 per envelope (low piece rate) and \$0.25 per envelope (high piece rate). These piece rates are also comparable, if slightly higher, to the ones used in our old experiment. (Notice that at about 100 envelopes per hour, the high-piece-rate treatment gives an average hourly pay of \$25, comparable to the \$30 per hour paid in the new design). For the gift group, we simulate the case in which the gift has a relatively low impact, creating additional motivation Δs equal to 2/5 of the low piece rate, and the case in which it has a medium impact, creating additional motivation Δs equal to 4/5 of the low piece rate. (If one thought that the gifts have a larger impact, say, equal to the low piece rate, the power would equal the one for the low piece rate treatment and thus readable in Column 2)

For the simulation of the Gneezy-List design, we use the structural parameters estimated in our main study assuming a power cost of effort function (Table 2, Column 2). Two important parameters are the curvature of the cost of effort function, which is approximately 10 (thus implying an elasticity of 1/10th) and the baseline social preference, which we take to be $s=0.1$, consistent with the estimates in Table 2, Column 2, since the warm glow $a=0.4$ was multiplied by an average return to the employer of *0.3*, implying an $s=0.12$. The third key parameter is the standard deviation of effort across treatments, which is now the between-subjects s.d., 0.3, while in the old experiment it was the within-subject standard deviation, 0.1 (since the person fixed

effects are controlled for). The final parameter, k, determines fix the level of productivity; the level of k is irrelevant to the power results, as it is just a level shifter. We set it to match the average productivity for envelope stuffing.

As the simulations show, under the Gneezy-List design with fixed time and varying speed of work, the piece rate increase leads to a modest rise in effort, from 195 to 210 with the low piece rate and to 220 with the high piece rate. Even assuming a per-treatment sample size of 50 (and thus total N=250), both piece rate treatments are underpowered, with a probability of detecting a statistically significant difference relative to the control group (at the 5% level) of p=0.41 (high piece rate) and p=0.23 (low piece rate). The gift effects, which we assume to be smaller than the low piece rate, are very poorly powered. This power calculation illustrates nicely our motivation in moving away from a between-subject version of this design to a within-study version. We should note that this power calculation is a very useful application of the structural estimates, as it is done easily using the benchmark estimates, just replacing the standard deviation.

For comparison, Panel C shows a corresponding power calculation holding everything the same, except using the within-subject standard deviation of effort. This simulation shows that a within-power experiment, as implemented in our older design, has significantly more statistical power

In panel B, we report instead the results of simulations for the new design. We assume that in the new design the model is the same, with power cost of effort, but the unit of effort is the number of (extra) minutes in addition to the required 120 minutes, not the units of work done as in (i). For this new design, we assume an elasticity of effort of 1, building on the pilot results on the MTurk sample reported above. Two other parameters are important: the standard deviation of k and the baseline motivation s. We have suggestive estimates from the online MTurk task suggesting a rather large standard deviation equal to 1 (implying that many workers will be at the boundary, working 0 or 60 minutes) and a baseline motivation of 0.1, as in the simulations above. We then set the cost of effort at a level to match the observed participation in extra work in the pilots.

Panel B in the table shows that the piece-rate treatments are now highly powered, moving the number of minutes of extra work from 5 in the control to an average of 27 in the low piece rate group and 39 in the high piece rate group. Assuming 50 subjects per treatment, a small-size gift effect is decently powered (p=0.65), and a medium-size gift effect is very well powered (p=0.97). Even assuming a smaller sample of 30 subjects per treatment, the piece rate treatments are still very well-powered, and the medium-size gift effects is still reasonably powered (p=0.80). These results are a striking difference from the ones in Panel A, motivating the new design. These power results are in fact slightly higher than the ones from the within-subject version of the fixed-time design, mirroring our old experiment (Panel C), motivating the new design.

We should stress that we cannot know for sure the exact values until we run the experiment, so we point out some comparative statics. A parameter that lowers statistical power is the standard deviation, which we assumed to be rather large at 1. Assuming a standard deviation that is twice as large, the high-piece rate is still powered (p=0.92), but the other treatments are not as powered

any more. Conversely, though, if the standard deviation is lower the power is even higher. The power is also decreasing in the baseline motivation and is increasing in the elasticity.

Table 1. Power Calculations, Fixed-Time Design vs. Duration-of-Work Design

Dependent Variable:	Effort				
Dependent Variable:	Control Group	Low Piece Rate	High Piece Rate	Gift (Low Impact)	Gift (Mid Impact)
	(1)	(2)	(3)	(4)	(5)
Panel A. Between-Subjects Fixed-Time Design, Effort is No. Units Produced in 2 Hours (Traditional Gift Exch. Design)					
Assumptions: elasticity 0.1 (gamma=10), baseline social preferences s=0.1, standard deviation of effort cost=0.3					
Predicted Average Effort	195	210	220	201	208
Power Calculation (N=50 per treatment), compared to Control, 5% significance		p=0.23	p=0.41	p=0.09	p=0.19
Power Calculation (N=30 per treatment), compared to Control, 5% significance		p=0.13	p=0.29	p=0.06	p=0.15
Panel B. Between-Subjects Duration-of-Work Design, Effort is Extra Work Minutes (0-60) (Proposed New Design)					
Assumptions: elasticity 1 (gamma=1), baseline social preferences s=0.1, standard deviation of effort cost=1					
Predicted Average Effort	5 min	27 min	39 min	14 min	22 min
Power Calculation (N=50 per treatment), compared to Control, 5% significance		p=0.99	p=1.00	p=0.65	p=0.97
Power Calculation (N=30 per treatment), compared to Control, 5% significance		p=0.91	p=1.00	p=0.38	p=0.80
Panel C. Within-Subjects Fixed-Time Design, Effort is No. Units Produced in 2 Hours (Similar to Old Design)					
Assumptions: elasticity 0.1 (gamma=10), baseline social preferences s=0.1, standard deviation of effort cost=0.1					
Predicted Average Effort	187	203	212	195	201
Power Calculation (N=50 per treatment), compared to Control, 5% significance		p=0.94	p=1	p=0.4	p=0.87
Power Calculation (N=30 per treatment), compared to Control, 5% significance		p=0.77	p=0.98	p=0.28	p=0.67

Notes: All panels assume power cost of effort, corresponding to a contant elasticity of effort to motivation. In the Fixed-Time Design, effort is measured as the number of units of output within a fixed time frame, 2 hours in this case. This is the traditional design for the field experiments on gift exchange. To enhance comparability with the previous envelope-stuffing experiments, the effort cost parameter k is set to match the average productivity in terms of envelopes stuffed per minute; importantly, the level of k is immaterial to the power calculations. In the Duration-of-Work Design that we are proposing, subjects first work for 2 hours (as required), then they can choose to work for up to extra 60 minutes, or they can stop working immediately. The unit of effort is the number of extra minutes of work, from 0 to 60. Panels A and C adopt the Fixed-Time Design with parameters matching our structural estimates in Table 2, Column 2, except that Panel A uses the relevant s.d. of effort for a between-subject experiment, while Panel B uses the relevant s.d. for a within-subject experiment, as in our old experiment. In Panel B we present simulation for the Duration-of-Work using parameters consistent with pilot data for the new design from MTurk runs by DellaVigna and Pope (work in progress). Column (1), (2), (3), (4), (5) are the five treatments we are planning to run: (1) Control, i.e. no piece rate incentive; (2) Low piece rate (3) High piece rate; (4) Gift assumed to increase instrinsic motivation s=0.1 by an 2/5 of the low piece rate; (5) Gift assumed to increase intrinsic motivation s=0.1 by 4/5 of the low piece rate. For Panel A and C, the Low piece rate has a piece rate of 0.125 and the High piece has a piece rate of 0.2; for Panel B, the Low piece has a piece rate of 0.25 and the High piece rate has a piece rate of 0.5. The predicted average level of effort is simulated assuming a large number of observations (N>30,000). For the power calculations, we use 200 simulations and compute the statistical power, that is, the fraction of the simulations in which the mean of a given treatment is significantly different from the mean in the control at the 5% level.