Estimating Social Preferences and Gift Exchange at Work

Stefano DellaVigna John A. List Ulrike Malmendier Gautam Rao *

This version: August 2021

Abstract

We design three field experiments to estimate how workers' social preferences towards their employer motivates their work effort. We vary the pay rates offered to workers, the return to the employer, and employer generosity demonstrated via unexpected gifts. Workers exert effort even without private incentives, but their effort is insensitive to the return to the employer. This is consistent with 'warm glow' but not pure altruism. The gifts have no effect on productivity, but engender extra work. This difference is explained partly by the finding that extra work is much more responsive to incentives than is productivity.

*Stefano DellaVigna, UC Berkeley and NBER, sdellavi@berkeley.edu; John A. List, University of Chicago and NBER, jlist@uchicago.edu; Ulrike Malmendier, UC Berkeley and NBER, ulrike@berkeley.edu; Gautam Rao, Harvard University and NBER, grao@fas.harvard.edu. This paper replaces an earlier draft titled "Estimating Social Preferences and Gift Exchange with a Piece-Rate Design". We thank James Andreoni, Roland Benabou, Patrick Dejarnette, Florian Englmaier, Constanca Esteves-Sorenson, Robert Gibbons, David Gill, Johannes Hermle, Mitch Hoffman, Supreet Kaur, Rosario Macera, Klaus Schmidt, Dmitry Taubinsky, Vera te Velde, and the audiences at Bonn, Munich, Zurich, the ASSA 2016 Conference, the SITE Conference on Experimental Economics, the NBER Conference on Organizational Economics, and especially our discussants James Andreoni and Florian Englmaier for useful comments. We also thank Debbie Blair, Alec Brandon, Natalie Cox, Michael Guy Cuna, Johannes Hermle, Justin Holz, Lukas Kiessling, Xinyue Lin, Maxim Massenkoff, David Novgorodsky, Waldo Ojeda, Lina Ramirez, Avner Shlain, Jessica Shui, Ferenc Szucs, Jonas Tungodden, and Brian Wheaton for excellent research assistance. Our experiment was approved under University of Chicago IRB 140661 and pre-registered as trial AEARCTR-0000502 (DellaVigna, List, Malmendier, and Rao, 2021a)

1 Introduction

What motivates workers to exert effort? Many jobs do not involve piece rates or other outcome-contingent pay, perhaps because of difficulties in measuring output or providing incentives to multi-tasking workers (Holmstrom and Milgrom, 1991). In such cases, workers may still work hard due to repeated-game incentives, but also because they care about their contribution to the firm or its mission. For this reason, Akerlof and Kranton (2005) and Besley and Ghatak (2005) make the case that organizations should select workers with pro-social preferences. But what is the strength and nature of these pro-social preferences?

The literature points to examples of the role of workers' social preferences towards their employers, such as Kahneman, Knetsch, and Thaler (1986)'s survey evidence on fairness in wage setting or Mas (2006)'s evidence of negative reciprocity after a disappointing pay increase. Overall, however, there is little systematic field evidence about the nature of workers' social preferences towards employers. Do workers take into account their employer's payoffs, as in pure altruism models (Becker, 1974)? In this case, employees would work harder when their effort is of higher value to the employer. Or is the right model one akin to warm glow (Andreoni, 1989) or social norms regarding effort? In that case, workers may value contributing to the employer, but could be insensitive to the actual employer payoff. Also, does extra employer generosity matter, as in gift-exchange models pioneered by Akerlof (1982)? Is there a negative response to ungenerous employers, as suggested by Mas (2006) and Krueger and Mas (2004)?

Our paper showcases three field experiments designed to estimate the nature and shape of social preferences at work. As in prior gift-exchange field experiments, workers are hired for a one-time task, to shut down repeated-game incentives and thus isolate social preferences. They are then exposed to different employer actions, such as surprise pay raises (Gneezy and List, 2006), pay cuts (Kube, Marechal, and Puppe, 2013), or in-kind gifts (Kube, Marechal, and Puppe, 2012). We also create variation in the return to the employer from workers' effort as in Englmeier and Leider (2012a). The differences in worker effort across the treatments provide evidence on workers' baseline levels of social preferences as well as reciprocity to the employer's generosity.[1]

We link the design to a simple model of social preferences to identify the underlying social preferences toward employers. A general first insight of our work is that previous field experiments, while providing very valuable qualitative evidence on social preferences, did not allow for estimation of the social preferences. In particular, two elements were missing from the design.

First, a key unobservable is the cost of worker 'effort'.[2] Assume for example that an unexpected

[1] We focus on workers' warm glow or pure altruism towards their employer / its mission, and how this varies with employer generosity. Other forms of social preferences could also be important in work settings, such as pay-inequity aversion (Breza, Kaur, and Shamdasani, 2018) or concerns about a fair division of the surplus between the employer and workers. We do not study such preferences in this paper. Nor do we capture social preferences which emerge as a result of repeated interactions and social relationships between worker and employer.

[2] We use the term 'effort' to refer to the costly actions that workers must take to produce output that the employer values. In the case of the productivity experiment, effort involves working faster to produce more in a fixed amount of time. In the case of the extra-work experiments, effort corresponds to the additional time spent working or the additional units of work taken on. In both cases, we assume that effort maps one-to-one to the empirically-observed

pay increase leads to 20 percent higher effort, as is the case initially in Gneezy and List (2006). The increased effort could reflect 20 percent higher altruism towards the firm, under a cost function with unit elasticity, or a 100 percent higher altruism under an inelastic cost function with elasticity 0.2. Without information on the cost function, it is impossible to tell. Yet, the two estimates imply a very different role for reciprocity in the workplace.

Second, for the tasks used in these experiments, such as data entry, the return the worker's effort generates for the employer is unclear, even though it is a key variable to distinguish different types of social preferences. Under pure altruism towards the employer, employee effort increases in the return to the employer. Under 'warm glow' (broadly construed to include positive feelings from doing meaningful work, adhering to a social norm of working hard, or signaling prosociality), effort may instead not depend on the employer's profit. This distinction could not be tested in prior experiments since the return to the employer was typically unobserved and held constant.

We design three field experiments which address both issues.[3] Similar to previous experiments, we set up a one-time job opportunity for a clerical task, and include gift treatments. Our experiments feature what we call a *pay-rate design*, which employs variation in pay rates for workers in addition to the gift treatments. Observing effort under different pay rates—financial rewards proportional to effort—allows us to identify the cost of effort function, thus addressing the first issue noted above.[4] In addition, we address the second issue by informing workers about the value of their work to the employer, or by ensuring that there is a natural measure for the value.

In Experiment 1, we measure productivity in a six-hour clerical job, where workers prepare mailers for multiple charities and a grocery store. We adopt a hybrid design that includes both within- and between-subject variation, with the aim to increase statistical power (as in Cohn, Fehr, and Goette, 2014). Also with power in mind, we hired a large sample of 446 workers.

The within-subject dimension is that, every 20 minutes, workers stuff a batch of envelopes for a different employer (charity). Each employer pays a different fixed and/or piece rate, and we also inform the workers about the different average per-envelope returns to the charities. To control for order effects, we randomize workers into two different orders.

The between-subject part of the experiment takes place in the final 2 (out of 10) batches. All participants work again for a charity that previously paid \$7 per batch. In the control group, the charity again pays \$7. In the *positive monetary gift* group (as in Gneezy and List, 2006), the charity now pays \$14. In the *in-kind gift* group (as in Kube, Marechal, and Puppe, 2012) the charity pays \$7, but in addition provides a gift-wrapped thermos of the value of \$14. In the *negative monetary gift* group (as in Kube, Marechal, and Puppe, 2013), the charity pays only \$3.[5]

outcome variables (productivity or units of extra work), with different cost-of-effort functions.

[3]The data for the three experiments are available at DellaVigna, List, Malmendier, and Rao (2021b).

[4]When effort is measured as output produced, pay rates are simply piece rates. When effort instead corresponds to time spent working, pay rates correspond to wage rates.

[5]All the gifts—and more generally all payments to workers—are paid for by the employers, and the returns to worker effort (the raised donations) also go to the employer. As we tell the subjects (truthfully), the Becker Center at the University of Chicago is collaborating with the charities and facilitating the employment, while not paying for the work. The one exception is the two paid training sessions, which by design are paid by the Becker Center.

We establish four key results. First, moving from a 0c piece rate to a 20c piece rate increases output by 12 percent. While this difference is highly statistically significant, it implies a modest elasticity of productivity of 0.1.[6] Second, productivity increases by 10 percent when the envelopes are utilized by the employer compared to paid training rounds where they are not, holding constant the piece rate. This suggests *some* form of social preferences: workers value that the work counts for the employer. Third, a doubling of the employer return in the form of a donor match leads to a negligible (and statistically insignificant) productivity increase of 1.6 percent. These findings are more consistent with a warm glow model than with pure altruism. Fourth, we estimate no statistically significant impact for any of the gift treatments.

This experiment allows us to consider, within one design, variation in incentives for the workers, return to the employer, and response to the gifts. Yet, it also has limitations. First, the within-subject structure is less natural, with the potential to confuse subjects. Second, the precision of the social preference estimates is limited by the fact that the outcome (envelopes stuffed) is quite inelastic. Even sizable shifts in worker altruism or warm glow towards the employer, say due to a gift, would result in fairly small impacts on productivity, which are hard to detect. Indeed, it is only due to the large sample and within-subject structure that we can still extract reasonably precise inferences. This small elasticity appears to be the rule, rather than the exception, for real-effort productivity tasks: previous papers with piece-rate variation estimate elasticities of 0.03 (Araujo et al., 2016 and Goerg, Kube, and Radbruch, 2019) and 0.04 (DellaVigna and Pope, 2018).

To address these issues required us to identify a simpler, between-subject design that is still well-powered, i.e., where the outcome is significantly more elastic to incentives. We are aware of no work tasks with high elasticity of productivity. Instead, building on the design of Abeler et al. (2011), we move from measured output over a fixed unit of time—productivity— to how much *extra work* workers are willing to do—a form of labor supply.[7]

In Experiment 2, we hire 300 workers for a one-time 2-hour data coding job for $60. After the two hours are completed, and workers were paid, we ask "*Would you be willing to help us enter some more of the data for up to one hour?*" In a control group, we state that "*unfortunately, we cannot compensate you for this extra time.*" For the two wage-rate treatments, we inform them that "*we will pay you [¢25/ ¢50] for every minute of work that you do, up to one hour.*"

In Experiment 3, we hire 2,000 workers online for $1.60 to perform a one-time task of checking the accuracy of 40 mailing addresses of University of Chicago alumni. Upon completion, we ask workers if they would consider checking up to 20 more entries. For the piece-rate treatments, we state that "*we will pay you [1/2/4] extra cents for every address you check, up to 20 addresses*".

In both these experiments, which target extra work, workers are highly responsive to financial incentives (whether wage rates or piece rates). In Experiment 2, the average number of extra

[6]Note that this is the elasticity parameter estimated by the structural model, as described in Section 5.1, which also accounts for motivation due to social preferences.

[7]Consistent with our findings below, DellaVigna and Pope (forthcoming) show that this extra-work margin is much more responsive than productivity to financial incentives in an online sample, even when holding the work task (data entry) constant. In an alternative approach, Goerg, Kube, and Radbruch (2019) estimates an elasticity of 0.35 for a real-effort slider task if subjects are allowed to browse the internet during an experiment.

minutes worked increases from 2.5 in the control group to 15.6 in the medium- wage-rate group and to 29.9 in the high- wage-rate group. In Experiment 3, the number of extra addresses checked increases from 3.8 in the control group to 8.4 in the low-piece-rate group, 9.7 in the medium-piece-rate group, and 12.6 in the high-piece-rate group.

The other treatment arms are geared towards identifying social preferences. In Experiment 2, we focus on different gift treatments. In a *monetary gift* group, we give the workers, after completion of the 2 hours of work and before asking about the extra work, an additional $15 as a token of appreciation. In an *in-kind gift* group, we give them a thermos of $15 value. Further, to study how quickly any gift exchange decays, in an *early in-kind gift* group, we gift workers a thermos before they start the 2 hours of work. In all of these gift treatments, there is no wage rate offered for extra work. Finally, within each arm we inform half of the workers that their work has especially high value to the employer.

In Experiment 3, we have one *monetary gift* arm, giving workers an additional 40 cents before asking for the extra work. (The online setting made it implausible to give in-kind gifts, and an early gift would likely have affected attrition.) We also vary quantitatively the return to the employer, communicated as the cost saving to finding wrong addresses, which depends on the cost of mailers.

Turning to the gift treatments, in Experiment 2, they raise the average extra work by 5.6 minutes (monetary gift, s.e.=2.1), 4.2 minutes (in-kind gift, s.e.=2.3), and 6.6 minutes (in-kind early gift, s.e.=2.4). This constant effect across the gift treatments differs from the findings of a larger effect of non-monetary gifts in Kube, Marechal, and Puppe (2012) and the suggestive evidence of decay of the effect of the gift over time in Gneezy and List (2006). In Experiment 3, the monetary gift raises the number of additional addresses checked by 1.9 (s.e.=0.5).

Turning to the variation of the return to the employer of worker effort, in Experiment 2, the qualitative statement on the high value of work to the employer has only a small and statistically insignificant impact of 2.2 minutes (s.e.=2.1). In Experiment 3, the quantitative higher value to the employer leads, if anything, to slightly less extra work (-0.7 addresses checked, s.e. 0.4).

Recapping, the three experiments share a common pay-rate design, but differ in many ways. They span an office-type setting (Experiment 1 and 2) versus online work (Experiment 3); they involve a pay of $60-$70 (Experiments 1 and 2) versus $2 (Experiment 3); they have a within-subject design (Experiment 1) versus a between-subject design (Experiments 2 and 3); they focus on inputs (time spent, Experiment 2) versus outputs (Experiments 1 and 3); and they present the value of work to the employer qualitatively (Experiment 2) versus quantitatively (Experiments 1 and 3). Despite these differences, two key findings emerge fairly consistently.

First, experimental variation in the perceived return to the employer, whether communicated quantitatively or qualitatively, does not seem to affect work effort sizably, provided it is positive. This evidence fits most obviously with the model of motivation which we labeled 'warm glow' and is consistent with the findings of Englmeier and Leider (2012a).

Second, gifts have a statistically significant impact in Experiments 2 and 3, though of a magnitude smaller than even the lowest level of financial incentives we test. How do we reconcile these

findings with the null effect of gifts in Experiment 1? The difference can be at least partially reconciled with the vast difference in elasticities. When we translate the gift results into reciprocity parameters, we cannot reject that the impact of monetary gifts on social preferences is the same in Experiment 1 ($a_{Gift} = 0.15$, s.e. 0.13), Experiment 2 ($a_{Gift} = 0.30$, s.e. 0.14), and Experiment 3 ($a_{Gift} = 0.07$, s.e. 0.03). The impact of the in-kind gift is statistically smaller in Experiment 1 ($a_{Gift} = -0.09$, s.e. 0.10) than in Experiment 2 ($a_{Gift} = 0.18$, s.e. 0.14 and $a_{EarlyGift} = 0.36$, s.e. 0.15), but the difference is nowhere near as large as in the reduced-form differences. We view this ability to compare across designs, and at least partially reconcile some of the differences, as a major advantage of the pay-rate design, and the structural estimation that it enables.

This paper relates to the literature on social preferences at work, providing evidence on workers' *vertical* social preferences towards their employer and its mission (e.g., Besley and Ghatak, 2005;Cullen and Perez-Truglia, 2018), complementing a larger literature about *horizontal* social preferences between co-workers (e.g. Bandiera, Barankay, and Rasul, 2005; Charness and Kuhn, 2007; Cohn, Hermann, and Schneider, 2014; Hjort, 2014; Breza, Kaur, and Shamdasani, 2018). Our first contribution to this literature is providing evidence that worker effort *is* partly motivated by social preferences towards employers, but that these preferences are not consistent with pure altruism. Instead, they can be characterized by a form of warm glow (broadly construed).

Our second contribution is to the literature on gift exchange in the field, which is reviewed in Online Appendix Table 1. We contribute evidence for moderate levels of gift exchange, which can be detected only using the higher-elasticity extra-work design in Experiments 2 and 3. This field-experimental literature builds on laboratory experiments on gift exchange, starting from Fehr, Kirchsteiger, and Riedl (1998). These experiments endow 'workers' with a 'cost of effort' function (a monetary transfer) and also inform subjects of how their 'effort' affects the payoffs of the other player (the 'firm'). Our experiments build a bridge towards this lab design by estimating the cost of effort and specifying the impact on the employer payoff in the field.

Methodologically, our paper makes the case for a *pay-rate design*. None of the gift exchange papers we are aware of and few effort or productivity experiments more broadly use pay-rate variation to price out behavioral treatments (with Augenblick, Niederle, and Sprenger, 2015 and DellaVigna and Pope, 2018 being recent exceptions). A pay-rate design enables the researcher to estimate the underlying behavioral parameters and thus allow comparison across experiments with widely different tasks and setups.[8] We also point to the value of estimating and documenting the elasticity of outcomes using standard financial incentives, which can help researchers identify tasks which allow for highly-powered and cost-effective experiments.

[8]This is therefore related to the literature on *structural behavioral economics* (e.g., Laibson, Maxted, Repetto, and Tobacman, 2015; Conlin, O'Donoghue, and Vogelsang, 2007; Augenblick, Niederle, and Sprenger, 2015; DellaVigna, List and Malmendier, 2012; DellaVigna, List, Malmendier, and Rao, 2017), reviewed in DellaVigna (2018). Our design also relates to Al-Ubaydli et al. (2015), which contains comparisons of pay rates and positive reciprocity to show how the choice of incentive scheme conveys information to workers about the nature of monitoring.

2 A Simple Model

To provide a framework for our contribution, we begin by presenting a simple theoretical model of worker effort. The model formalizes the predictions regarding productivity (as in Experiment 1) and regarding extra work (as in Experiments 2 and 3). We abstract from specific features, such as the different batches of work in Experiment 1, which we detail in the structural estimation.

Worker i chooses optimal effort e_i as a function of pay incentives, social preferences and cost of effort. For simplicity, we assume risk neutrality. This allows us to write the worker's problem as

$$\max_{e_i \geq 0} u(e_i) \quad = \quad L + p_W e_i - C_i(e_i) + A\left(Gift, p_E, p_W\right) e_i \tag{1}$$

The first component of the utility function captures the monetary payoff from exerting effort e_i: a lump-sum payment $L \geq 0$ and a pay rate $p_W \geq 0$ to the worker. The second component is the cost of effort $C_i\left(e_i\right)$, which can differ across individuals i (to capture differences in individual ability). For any i, we assume the regularity conditions $C'\left(\right) > 0$, $C''\left(\right) > 0$, and $\lim_{e\to\infty} C'\left(e\right) = \infty$, guaranteeing the existence of a unique solution. Note that we assume that effort maps one-to-one to the observable work outcomes we study: productivity in Experiment 1 and extra work time or output in Experiments 2 and 3.

The third component captures a type of social preference: the worker cares about the payoff from his efforts to the employer with a social preference coefficient A, which may depend upon unexpected gifts $Gift$ from the employer. We discuss the altruism and warm-glow interpretations of this coefficient A below. This set-up is similar to Bellemare and Shearer (2011) and Englmeier and Leider (2012c), among others. Maximization problem (1) yields the first-order condition

$$p_W + A\left(Gift, p_E, p_W\right) - C'\left(e_i^*\right) \quad = \quad 0 \text{ or} \tag{2}$$

$$e_i^*\left(Gift, p_E, p_W\right) \quad = \quad C'^{-1}\left(p_W + A\left(Gift, p_E, p_W\right)\right). \tag{3}$$

where $C'^{-1}\left(\right)$ is the inverse function of $C'\left(\right)$, which exists and is monotonically increasing by the assumptions above. The second-order conditions are satisfied since $-C''\left(e^*\right) < 0$. In the following we will assume an interior solution.[9] The optimal effort e^* is increasing in the social preference parameter A and in the pay rate p_W (provided A does not decrease enough in p_W). Thus, in our model, financial incentives and prosocial preferences provide the motivation to exert effort.[10]

Pure Altruism. Under this first interpretation, the worker takes into account the employer's actual *net* return from effort, $A = \alpha(p_E - p_W)$. An altruistic worker values each dollar the employer makes (through their effort) the same as α dollars in their own pocket. Capturing reciprocity models, the altruism parameter α towards the employer may depend on the receipt of a gift from

[9] A sufficient condition to ensure an interior solution is $C'\left(0\right) = 0$ and $A > 0$. While one of the assumed cost of effort functions will not satisfy this assumption, in practice zero effort is not observed in our experiment.

[10] We assume that financial incentives affect motivation directly without crowding out social preferences or intrinsic motivation. In our model, how much the return to the employer motivates workers depends on their social preferences. Similarly, whether gifts affect motivation depends on whether they change social preferences. How motivation in turn translates into effort (and thus output) depends on the cost of effort.

the employer. Thus, the social preference term is $A = (\alpha + 1_{Gift}\alpha_{Gift})(p_E - p_W)$.

Warm Glow. Models of warm glow, broadly construed, imply that the worker simply derives utility from doing their part by exerting effort for their employer, regardless of how the effort translates into payoffs for the employer; thus, $A = a\overline{p_E}$, where $\overline{p_E}$ is the average return to the employer and is included to normalize the a term. This case is inspired by the idea of warm glow in Andreoni (1989), where donors derive utility from giving to a charitable cause, but not necessarily from the public good provided itself.[11] This specification also captures, in reduced form, mission preferences as in Besley and Ghatak (2005), intrinsic motivation (Benabou and Tirole, 2003), a social norm to exert effort for an employer, signaling (Benabou and Tirole, 2006), or a utility from exerting effort doing meaningful work (Ariely, Kamenica, and Prelec, 2008).[12] As before, we allow for warm glow to change as a result of receiving an unanticipated gift, so $A = (a + a_{Gift}1_{Gift})\,\overline{p_E}$.

Standard Gift-Exchange Experiment. Consider a gift exchange experiment without a pay rate. For the combined altruism and warm glow case, the optimal efforts are:

$$\begin{aligned} e^*_{noGift} &= C'^{-1}\left(\alpha p_E + a\overline{p_E}\right) \quad \text{and} \\ e^*_{Gift} &= C'^{-1}\left(\left(\alpha + \alpha_{Gift}\right)p_E + \left(a + a_{Gift}\right)\overline{p_E}\right). \end{aligned} \tag{4}$$

Can one back out the social-preference parameters from the observed effort e_{noGift} and e_{Gift}? Two crucial pieces of information are missing. First, we do not know what workers assume the return to the employer p_E is, since they are not informed about it. Second, we do not know the cost function $C(e)$. Hence, it is impossible to identify the social preferences.

For illustrative purposes, consider a power cost function $c(e) = ke^{1+\gamma}/(1+\gamma)$, with a constant elasticity $1/\gamma$ with respect to the return to effort.[13] The solutions under pure altruism are:

$$e^*_{noGift} = \left(\frac{\alpha p_E}{k}\right)^{1/\gamma} \quad \text{and} \quad e^*_{Gift} = \left(\frac{\left(\alpha + \alpha_{Gift}\right)p_E}{k}\right)^{1/\gamma}.$$

By dividing through and inverting, we obtain

$$\frac{\alpha + \alpha_{Gift}}{\alpha} = \left(\frac{e^*_{Gift}}{e^*_{noGift}}\right)^{1/\gamma}. \tag{5}$$

While we cannot back out the altruism (and reciprocity) parameters α and α_{Gift} without knowledge of the return p_E, we can infer the increase in altruism $(\alpha + \alpha_{Gift})/\alpha$, *provided* we know the curvature γ. In the quadratic cost-of-effort case ($\gamma = 1$), for example, an x percent increase in

[11] The warm glow could also depend on the return to the firm, in which case it would be indistinguishable from pure altruism in our setting.

[12] Explicitly modeling these motives would lead to additional predictions. For example, signaling (Benabou and Tirole, 2006) would dampen the response to a pay-rate increase because exerting effort is less diagnostic of intrinsic motivation under a higher pay rate. We recognize this limitation of our model and hope that future work will further disentangle these alternative social motives for worker effort.

[13] The first-order condition is $k\left(e^*\right)^{\gamma} = v$ (in this case equal to αp_E and $\left(\alpha + \alpha_{Gift}\right)p_E$ for the altruism case). Thus, $\partial e^*/\partial v = (1/k\gamma) * (v/k)^{1/\gamma - 1}$ and the elasticity is $\eta_{e,v} = (1/k\gamma) * (v/k)^{1/\gamma - 1}\, v\,(v/k)^{-1/\gamma} = 1/\gamma$.

effort due to a gift implies an x percent increase in altruism. But for higher curvature ($\gamma > 1$), the underlying increase in altruism is higher than x percent. Thus the elasticity $1/\gamma$ plays a key role in estimating the underlying preferences.

The power cost function has the special feature of constant elasticity. A plausible alternative is that the elasticity decreases as effort increases, such as in the case of the exponential cost function, $C(e) = k \exp(\gamma e)/\gamma$.[14] In this case, the solutions are

$$e^*_{noGift} = \frac{1}{\gamma} \log\left(\frac{\alpha p_E}{k}\right) \quad \text{and} \quad e^*_{Gift} = \frac{1}{\gamma} \log\left(\frac{(\alpha + \alpha_{Gift}) p_E}{k}\right).$$

We can transform the solution and divide through to obtain

$$\exp\left[\gamma\left(e^*_{Gift} - e^*_{noGift}\right)\right] = \frac{(\alpha + \alpha_{Gift})}{\alpha}. \tag{6}$$

Expression (6) highlights another implication. Consider an experiment with a positive gift treatment, which increases output by x units, and a negative gift treatment, which decreases output by x units. Would these equal-sized impacts of the gifts on effort imply that positive reciprocity has the same magnitude as negative reciprocity? Expression (6) shows that is not the case: the x unit increase for the positive gift requires a larger proportional change in altruism (positive reciprocity) compared to the corresponding change in altruism (negative reciprocity) for the negative gift. Intuitively, it is harder to increase effort at the margin than to reduce it.

Generalized Gift Exchange Experiment. What design would then allow for estimation of social preferences? As outlined above, one needs to measure the return to the employer, p_E, and to identify the cost of effort parameters k and γ.

To accomplish the first part, one can inform subjects about the return p_E, provided the task permits. To identify the cost of effort, however, one needs further identifying variation, such as varying experimentally the pay rate p_W. From (3) notice that

$$\frac{\partial e^*}{\partial p_W} = \frac{\partial C'^{-1}(p_W + A)}{\partial p}\left(1 + \frac{\partial A}{\partial p_W}\right). \tag{7}$$

Expression (7) shows that variation in pay rate p_W helps pin down the cost of effort function. However, the cost of effort will be identified jointly with the social preferences, given that A features in (7) and, in addition, for the altruism case $\partial A/\partial p_W = -\alpha$ (in the warm glow case $\partial A/\partial p_W = 0$). Thus, it is useful to also observe workers in a "training period," in which the work does not benefit the firm and in which the incentive p_W is paid by a third party. In this case, we assume $A = 0$, and the effort of the worker is driven solely by pay-rate incentives.[15]

[14] The first order condition is $k \exp(\gamma e^*) = v$ where v is the return per unit of effort. Thus, $e^* = (1/\gamma) \log(v/k)$. Then $\partial e^*/\partial v = (1/\gamma) * (k/v)/k$ and the elasticity is $\eta_{e,v} = (1/\gamma v) * v/((1/\gamma) \log(v/k)) = 1/\log(v/k)$.

[15] An implicit and common assumption is that the social preferences do not extend to the experimenter. In a robustness check we allow for some social preference also during the training period.

Next, consider how worker effort responds to changes in the return to the employer p_E:

$$\frac{\partial e^*}{\partial p_E} = \frac{\partial C'^{-1}(p_W + A)}{\partial p} \frac{\partial A}{\partial p_E}. \tag{8}$$

Under warm glow, $\partial e^*/\partial p_E = 0$: the workers do not respond to changes in the value of effort since $\partial A/\partial p_E = 0$. Under altruism, instead, the workers are sensitive to the return, since $\partial A/\partial p_E = \alpha$. Indeed, for the altruism case, combining (7) and (8) shows that the ratio of the response to the employer's return p_E and the response to the pay rate p_W identifies the altruism α :

$$\frac{\partial e^*}{\partial p_E} / \frac{\partial e^*}{\partial p_W} = \frac{\alpha}{1-\alpha}. \tag{9}$$

3 Productivity Experiment (Experiment 1)

3.1 Design and Data

Design. Our first design follows previous gift-exchange field experiments in that: (i) we measure productivity in an office task, (ii) workers are recruited for a one-time task to avoid repeated-game confounds, and (iii) workers are assigned (in a between-subject design) randomly into a gift treatment or a control group to causally estimate the effect of a gift on productivity.

It differs in that (i) we implement a *pay-rate design* by varying the piece rate p_W offered to the workers, and (ii) we make explicit, and vary, the value of the task to the employer, p_E.

To do so, we combine between-subject and within-subject variation. Specifically, subjects work through several batches of a task. Across the batches, we vary *within subject* the piece rate and the return to the employer. To control for order effects, we randomize the order of the batches. In the final batches, we introduce *between-subject* unexpected 'gifts' from the employer.

The model above informs the choice of task and employer. Coding of library books, for example, does not lend itself readily to quantifying the return to the employer. We partnered with three charities to prepare envelopes for fund-raising campaigns. Since similar campaigns have been done before, we could convey the average employer return p_E of each envelope prepared, and the return could plausibly be higher for envelopes for which a donor (truthfully) pledged to match the raised funds. Furthermore, we set different piece rates p_W for each charity for a 20-minute batch: (\$7 fixed pay, no piece rate), (\$3.5, 10¢ piece rate), and (no fixed pay, 20¢ piece rate). The lump-sum pay is set such that the earnings are constant for a person of average productivity (about 35 envelopes per 20 minutes) and, hence minimize the gift effects from piece-rate variation.

We hire temporary workers with a Craigslist ad for a one-time six-hour job (in one day) on Saturdays and Sundays on the University of Chicago campus.[16] The participants are taken to a room where a research assistant explains the work following a script. The Becker Center at the

[16] A typical ad reads: '*The Becker Friedman Institute is seeking individuals to help prepare letters for fundraising and advertising campaigns. No experience necessary. Employment is for six hours over a single day THIS weekend. [...] Employees can expect to earn around $60 for the day.*' We exclude anyone attempting to sign up a second time.

University of Chicago is truthfully presented as partnering with the employers – the charities – to facilitate the work. The participants receive a sheet indicating the pay for the ten batches of work, except for batches 9 and 10 which are reported as TBD (to be determined).

The workers fold and place materials in envelopes, working through a mailing list, for 20 minutes. They then take a 10-minute break, and move on to the next batch, and so on for ten batches. During the break, the research assistants count the envelopes produced by each participant and check the accuracy of five envelopes per worker. The envelopes include fund-raising material for three charities for 8 batches and an advertising campaign for a local grocery store for 2 batches.

We randomize the order as follows: Under Order A, shown in Figure 1, the participants fold envelopes at the 10-cent piece rate (and $3.50 flat pay) for the first four batches, albeit with different treatments. The first batch is a training period. We tell the participants that they '*will earn a fixed amount of $3.50 plus $0.10 per envelope completed during this training. [...] The training is paid for by the Becker Center. We will be discarding all of the envelopes prepared in this training session*'. Thus, the employer – the charity – does not directly benefit from the marginal productivity.[17]

In batches 2 and 3, workers stuff envelopes for charity 1: '*[Charity Name] will be paying for your work. The pay is $3.50 plus $0.10 per envelope completed, as noted on your schedule.*' In batch 2, there is a higher return to the employer due to a donor match: '*[Charity Name] has received a matching grant that will match every dollar raised by these letters 1 to 1 up to $2,000 total. [...] Historically, charities like [Charity Name] have yielded roughly $0.60 per mailer with such campaigns, including the match. Given that [Charity Name] is offering a $0.10 per-envelope payment today, it expects to get roughly $0.50 for each additional envelope that you prepare during this session.*' Notice that we emphasize the net return. In batch 3, there is the same piece rate, but no match. In batch 4, workers stuff envelopes for a grocery store at the same 10-cent piece rate and with a similar stated 30-cent return to the employer.[18]

After a 50-minute lunch break, the workers restart with a new training period (batch 5) on the material for Charity 2, which pays 20 cents per envelopes.[19] Next, they engage in consequential work for Charity 2 in batch 6 at the same 20-cent piece rate. In batches 7 and 8, they then stuff envelopes for Charity 3 at the 0-cent piece rate, with a charity match in batch 8. In rounds 9 and 10, the gift exchange randomization takes place, as we discuss below.

The treatments in Order B are mirror images of Order A, other than for the training rounds: batch 8 in order A becomes batch 2 in order B, and so on. That is, while the training sessions remain in rounds 1 and 5, we switch the pay schemes in the training period between order A and B. The order randomization aims to disentangle the treatment effects from learning and tiredness.[20]

[17] There is no deception in the experiment and the envelopes, which have fictional names and addresses, are discarded as announced. These training batches are presented as necessary to ensure the accuracy of the later batches. We assume that the Becker Center, which pays for the training batch, does not enter the social preferences of the workers.

[18] This allows us to check whether there was anything special about having a charity as employer. The grocery store is not our main employer because we could not find a compelling way to vary the return to the employer.

[19] This second training session is justified by the (slight) difference in materials for the different employer.

[20] We did not do a full randomization of the order of the treatments for logistical reasons, but also because the two orders allowed us to maximize power by placing next to each other treatments we intended to compare, such as match and no-match (for the same piece rate), thus minimizing the confound of productivity changes over time.

The arrows in Figure 1 illustrate four planned comparisons:[21] (i) batches 6 vs. 7 in order A (3 vs. 4 in order B) with the piece rate change from 0 cent to 20 cents, with the additional comparison with the 10-cent piece rate in batch 3 in order A (batch 7 in order B), though the treatments are not contiguous; (ii) batches 5 vs. 6 comparing when the effort counts at all for the employer, versus not (the training); (iii) batches 7 vs. 8 for the return to the employer; (iv) batches 3 vs. 4 in order A (6 vs. 7 in order B) for having a grocery store versus a charity as employer.

After batch 8, we randomize workers into the gift treatments, and therefore send the workers into separate rooms. In the Control treatment, we say that in this and the next session "*[Charity name] will pay $7 just as it paid in a previous session.*" In the Positive Monetary Gift, we say that "*[Charity name] will pay $14 instead of the standard $7 that it paid in a previous session.*" In the Negative Monetary Gift, we say that "*[Charity name] will pay $3 instead of the standard $7 that it paid in a previous session.*" We follow Kube, Marechal, and Puppe (2013) in providing no explanation for the wage change. In the Positive In-Kind Gift, we say that "*[Charity name] will pay $7 as it paid in a previous session. As a token of appreciation, the charity is also giving you this thermos with a retail value of $14.*" We then offer a gift-wrapped thermos (as in Kube, Marechal, and Puppe, 2012) with the name of the charity to make clear the gift is coming from the employer. After the announcement of the gift (if any), the workers fold envelopes for batches 9 and 10. One of the two batches has a donor match raising the return to 60 cents.

At the end, we conduct a short debriefing survey, thank the subjects, pay them according to their accumulated earnings, and walk them to an exit.

Randomization. We randomize (i) into order A or B; (ii) the charities (into three orders) to the role of Charity 1, 2, and 3; (iii) whether the charity match is in batch 9 or 10. The order of the 2x3x2=12 types of sessions was randomly drawn, and we then looped through the 12 sessions. The final randomization, to the gift treatments, is made at the individual level within a session, stratifying on the pre-lunch performance to maximize statistical power.

Data. We ran 24 sessions with 131 workers from October 2013 to January 2014.[22] We then paused to ensure that the design that we had settled on based on simulations worked appropriately. In November 2014, we registered the design, including the model and structural estimation, and the envisioned number of sessions (72 in total). The only design change was the addition of the in-kind gift treatment. Between November 2014 and May 2015, we ran 49 sessions with 319 subjects.[23] After excluding 4 subjects who left early, the final sample includes 446 workers, the largest sample size that we are aware of among in-person gift exchange field experiments with workers.

The sample (Column 1 of Online Appendix Table 2) is 52 percent female, covers a wide age range, and overrepresents unemployed individuals. Column 2 shows that productivity is higher

[21] In the pre-registration we emphasize equally comparisons taking place in batches 1-4 and in batches 5-10. However, the steep learning in batches 1-4 confounds these comparisons. Thus, we focus on batches 5-10 for the reduced form results. The structural estimates use all of the variation in the data, including the early periods.

[22] In September and October 2013, we ran 4 sessions with a pilot design. We used the data from the 17 subjects in this pilot to set the pay rate, since we aimed to equate on average earnings across the three different piece rates.

[23] The sessions add up to 73 because, in one of the sessions, one of the letters was shown incorrectly, and the research assistant opted to repeat the session. In the spirit of intent-to-treat, we also retain this session.

for employed individuals and females, as well as for 25-34 years olds relative to both younger and older participants. Using this specification, we form an index of predicted productivity. In Columns 3-6, we examine the randomization with respect to the covariates (Panel A), as well as with respect to the index of predicted effort (Panel B). Order A is somewhat overrepresented in males (who have lower productivity) and workers 55 years and older (who have somewhat higher productivity), with no evidence of selection of higher predicted-effort individuals (Panel B) into Order A. The randomization into the various gift treatments or into the different charity orders reveals no systematic patterns. Thus, the treatments are overall balanced on the covariates.

3.2 Experimental Findings

Piece-Rate Response. As Figure 2a shows, increasing the piece rate from $0¢$ to $20¢$ leads to a statistically significant, if quantitatively small, increase of 4 envelopes (12 percent). Figure 2a also provides a comparison to the $10¢$ piece rate, which shows an increase from 0c to 10c, with about the same response to 10c and 20c, suggestive of an impact to a positive pay rate, rather than the exact amount. However, the batches are not contiguous in this case, and the simple comparison of $10¢$ with either $0¢$ or $20¢$ (without accounting for learning or tiredness effects) should be treated with caution. Online Appendix Figure 1 plots the "raw" results by displaying average output by batch for the two orders. Consistent with the reduced-form analysis, one can clearly detect a response to the piece rate: the only two instances in which productivity decreases substantially from one batch to the next are cases of piece-rate decreases: batches 6 to 7 in order A ($20¢$ to $0¢$) and 8 to 9 in order B ($10¢$ to $0¢$).

Baseline Social Preferences. As Figure 2b shows, the batches with higher return to the employer (due to a match on the donations) yield (statistically insignificant) 0.6 additional envelopes (1.7 percent), an effect size much smaller than the piece-rate response. This finding speaks against a pure-altruism interpretation of workers' social preferences towards their employer.

Figure 2c presents complementary evidence on the impact of consequences to the employer, comparing the training round in batch 5, when the envelopes are discarded, to batch 6, when the envelopes are sent; notice that the piece rate is held constant. Productivity is 3.5 envelopes (10 percent) higher when the letters are used, a sizable and statistically significant difference.

Turning to additional experimental variation, worker effort does not differ between the charities (Online Appendix Figure 2a), which we thus pool. Employee effort is somewhat *higher* when the employer is a firm (a grocery store) versus a charity (Online Appendix Figure 2b), holding constant the return to the employer and the piece rate. This suggests that the social preferences we identify towards charities as employers may not over-estimate the social preferences towards employers.

Gift Exchange. Figure 2d presents empirical results for batches 9 and 10. Compared to the control group, the positive gift treatment produces 0.4 additional envelopes, a difference that is not significant. The negative monetary gift treatment has an even smaller effect, keeping productivity essentially constant compared to the control group. The in-kind gift actually leads to a *decrease* of productivity of 1.1 envelopes, although the difference is not significant at conventional levels.

To enhance the statistical power in the estimate of the gift effects, in Table 1 we control for the average productivity of a worker in batches 1-8 (Column 2) and in batches 5-8 (Column 3). Since the specification in Column 3 has higher explanatory power, we use it for the subsequent analysis with controls. The addition of controls lowers the standard errors by a quarter, essentially without changing the point estimates. We can reject that the monetary gift increases productivity by more than 5.6 percent, much smaller than the 20 to 25 percent increases in Gneezy and List (2006) and Gilchrist, Luca, and Malhotra (2016). We cannot reject the 4 percent increase in output found in the highly-powered experiment of Cohn, Fehr, and Goette (2014), which combines within- and between-subject variation and plausibly involves an extra-work dimension as well. We can reject that a negative gift lowers effort by more than 1.6 envelopes, a 4.4 percent decrease, a much smaller effect than the 20 percent decrease in Kube, Marechal, and Puppe (2013). We can also reject that the in-kind gift increases productivity by more than 0.7 envelopes, a 2 percent increase, again much smaller than the 25 percent increase in Kube, Marechal, and Puppe (2012). The results are similar using log output as a measure (Columns 4-6), the specification implied by a power cost function.

In Online Appendix Figure 3 and Online Appendix Table 3, we also consider the decay of any gift effect, comparing batches 9 and 10, and find little evidence of systematic patterns. We also do not find evidence that the gift interacts with the return to the employer, comparing a batch with matching donation to a batch without. With regards to the heterogeneity in the gift effect, there are hints of an output increase with the positive gift at lower quantiles. Further, all the key results are parallel using log output (Online Appendix Figure 4), as implied by a power cost of effort, as well as restricting to participants who report being employed (Online Appendix Figure 5).

4 Extra Work Experiments (Experiments 2 and 3)

4.1 Design Idea

One major insight of the results so far is that productivity experiments—the most common design in the literature—feature low elasticity, which makes it difficult to identify the shape of worker preferences. In Experiments 2 and 3, instead of focusing on productivity within a given amount of time as in Experiment 1, we focus on the provision of extra work after a job task is completed, measured as extra time (Experiment 2) or extra units produced (Experiment 3). As we will document, this margin is more sensitive to incentives than the productivity margin and thus provides higher statistical power to detect even modest gift exchange effects.

Both experiments use a between-subject design. The treatment groups are (i) a control group with no compensation for extra work, (ii) two or three pay-rate groups with incentives to provide extra work, and (iii) one or more groups with a gift, but no monetary incentive to complete extra work. Further, within each of these groups, half of the workers are randomized to a high-value-of-work treatment, communicated either as a qualitative statement (Experiment 2), or through a quantitative measure of value to the employer (Experiment 3).

4.2 Design and Data, Experiment 2

Design. Figure 3a summarizes the design of Experiment 2. We hire workers for a one-time 2-hour data coding job at the Becker Center for $60, advertising the position both on a campus bulletin as well as on the website Craigslist. As the workers come in, we escort each to individual offices and ask them to code sheets of data from a research project into a Google spreadsheet for 2 hours. Except for one arm (which we return to later), there is no experimental manipulation at this stage.

After the two hours of work are completed, the research assistant checks back into the offices, pays workers the promised $60, and briefly reviews and accepts the completed work. At this point, the research assistant asks the worker: "*Would you be willing to help us enter some more of the data for up to one hour?*" Subjects are randomized into one of five conditions. In the control arm, we add: "*Unfortunately, we cannot compensate you for this extra time.*" For the two wage-rate treatments, we inform workers that "*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.*" Finally, three arms feature gift treatments. For the *monetary gift* and the *in-kind gift* group, the script parallels the script in the control group up until the workers have finished the two hours of pay. At this point, in the *monetary gift* group, we tell workers "*In addition, as a token of appreciation, the Becker Center is giving you an additional $15 for helping today. Therefore, we are paying you a total of $75.*" We then pay $75 and otherwise proceed as in the Control group. In *the in-kind gift* group, we tell workers "*In addition, as a token of appreciation, the Becker Center is giving you this thermos with a retail value of $15 for helping today*" and we give workers a thermos with the University of Chicago name on it. The experiment then proceeds as in the Control group. The additional time that subjects choose to work in each treatment is the key outcome and is the notion of effort in this experiment.

The explicit "ask" for extra work arguably provides a clearer channel for workers to reciprocate compared to the productivity experiment. One may however be concerned that asking subjects for extra work immediately after they were given a gift may set up a quid-pro-quo or create social pressure to do extra work. Alternatively, the gift may be seen as manipulative and thus backfire. We therefore introduced a gap between the gift and the request for extra work: the research assistant, after the pay and after the gift distribution but before asking for extra work, goes over and briefly checks some of the data coding. As a further robustness check, and to study the persistence of any gift effect, a final group, the *early in-kind gift* group, received a thermos before starting the 2 hours of coding, with the same exact wording as for the *in-kind gift* group.[24]

In sum, we have six treatment arms, and we randomize the workers into each arm with equal probability. Further, we cross-randomize (also between subjects) a statement on the value of work to the employer. Directly before asking for extra work, we inform the *high-return* group (but not the *normal-return* group) "*Getting the extra data entered today is really valuable to us.*"[25]

[24]The early gift treatment was intended to delink the gift from the "ask" for extra work. However, it could also be perceived as less strategic and thus be *more* effective.

[25]While we interpret this treatment literally as increasing the perceived value of the work to the employer, it may also operate through increased social pressure.

Data. We pre-registered our design in September 2018, including a power calculation using the structural estimates for the productivity experiment and assumptions for the elasticity of extra work. Shortly after the registration, we started collecting data and stopped in May 2019 once we reached the pre-specified sample target of 300 workers. Our sample (Online Appendix Table 4) is 50 percent female, with 37 percent recruited via Craigslist, and the rest being students. The Craigslist participants, especially older ones, stay for a greater number of extra minutes (Column 2). The treatments are balanced on the observables (Columns 3-8).

4.3 Design and Data, Experiment 3

Design. Figure 3b presents the design of Experiment 3. We hire 2,000 workers online on the MTurk platform for a one-time task that lasts about 10 minutes for pay of $1.60. The task consists of checking the accuracy of 40 mailing addresses of University of Chicago alumni. The 40 addresses are randomly drawn for each worker from a sample of 5,612 addresses. As we truthfully inform the workers, "*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.*" The workers are given instructions on how to search the addresses (without sharing the recipient's name) and then label each address as correct or in error.

Upon completion of the task, we thank the workers and then provide them with (randomized) information on the value of the work that they have done: "*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)*". This variation, which is cross-randomized, is again truthful as some mailings cost more than others, such that identifying a wrong address saves the employer more money.

A control group is then asked "*If you happen to have some time and are willing to do some extra work, that would be appreciated. Unfortunately, we cannot compensate you for this extra work."* The piece-rate groups see the same first sentence, but then see *"[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]"* As in Experiment 2, the piece rate for the high-piece-rate group corresponds to the average pay in the initial task ($1.60/40=4 cents per address), and half of that for the medium-piece-rate group. Unlike in Experiment 2, the larger sample of the online, low-cost setting allows us to add a low-piece-rate group, with a piece rate one quarter the initial compensation.

The monetary gift arm is similar to the control group, except that we add, before the information about the value of work, "*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.*" The gift equals 25 percent of the initial pay, as in Experiment 2 and similar to Experiment 1. In this online experiment, we could not plausibly, and naturally, offer an in-kind gift. Similarly, we did not add an early-gift

group, which would likely have affected attrition, a well-known issue for online tasks. (As it is, an advantage of the extra-work design is that the randomization takes place only *after* completion of the initial task, with no room for differential attrition.)

To summarize, we have five experimental arms, with the share of the sample in parentheses: Control (25%), Low Piece-Rate (25%), Medium Piece-Rate (12.5%), High Piece-Rate (12.5%), and Gift (25%). Each arm is crossed with the High/Normal return to work manipulation. The additional number of addresses checked in each treatment is the key outcome for this experiment.

Data. In February 2021 we pre-registered our design, and then collected data for 9 days until we reached the pre-specified sample of 2,000 workers. Following the pre-analysis plan, we eliminate workers with a high coding error rate, resulting in a final sample of 1,954 workers. Given that the task is a natural field experiment we did not collect demographic information.

4.4 Experimental Findings, Experiments 2 and 3

We present our empirical results for the amount of extra work in Experiments 2 and 3 in Figures 4 and 5 and in Table 2. As main outcome variables, we pre-specified the minutes of extra work in Experiment 2 and extra addresses checked in Experiment 3.

Pay-Rate Response. Figure 4a documents very sharp responses to varying wage rates in Experiment 2. The average number of extra minutes worked increases from 2.5 (control) to 15.6 (medium wage-rate) and 29.9 (high wage-rate). These differences are statistically significant and correspond to a high elasticity: a doubling of the wage rate (25¢ to 50¢) is associated with more than double the extra minutes worked, indicating an elasticity over 1. Figure 5a presents parallel evidence for Experiment 3. The average number of extra envelopes increases from 3.8 (control) to 8.4 (low piece-rate), to 9.7 (medium piece-rate) and to 12.6 (high piece-rate), indicating a similarly high elasticity. We are not aware of any real-effort experiment on productivity that even approaches such elastic responses. In our view, this is a major advantage of our chosen design.

Baseline Social Preferences. In Figure 4b we consider the response in Experiment 2 to the qualitative manipulation of the return to the employer, phrased as: "*Getting the extra data entered today is really valuable to us.*" Since this manipulation is cross-randomized, we pool across all six gift and wage-rate arms. We estimate a small and statistically insignificant increase of 2.5 minutes from receiving the high-return statement. In Figure 5b we consider the parallel manipulation for Experiment 3, which involved doubling the gross return to the employer of the work completed. The response to the higher-return condition is noisy, and if anything, negative.

Gift Exchange. Figure 4c compares the number of minutes worked in the control group to the minutes worked in the three gift groups in Experiment 2. We find clear evidence of gift exchange. The gift treatments increase the average labor supply from 2.5 minutes in the control group to 6-9 minutes, about halfway to the medium-wage-rate treatment. The effect is similar for the monetary gift and the early in-kind gift, and slightly smaller for the in-kind gift. Thus, not only is there a gift effect on labor supply, but the effect does not decay quickly over our time period used.[26]

[26]It is also possible to compare the productivity in the initial 120 minutes in the Early-Gift condition to the

Figure 5c shows the parallel finding for Experiment 3 for the monetary gift treatment. The average number of extra addresses increases from 4.1 in the control group to 6 in the gift group.

Robustness. Figure 6a presents the c.d.f. of the number of extra minutes worked in Experiment 2 in the 6 main arms. The medium wage-rate treatment leads to an increase in the share staying extra time (the extensive margin) relative to the control group, but even more in the number of extra minutes among those staying longer (the intensive margin). The high wage-rate treatment is associated with a large increase along both margins. The gift treatments are associated with a substantial increase in the share of workers staying extra minutes compared to the control group; however, most of the workers that stay extra do not stay long. Thus, the gift triggers only a moderate amount of extra effort. The patterns are similar in Experiment 3, in which the effect of the gift is smaller than the effect of any of the piece rates (Figure 6b).

Online Appendix Table 5 reports estimates with controls: gender, age groups, and an indicator for the Craigslist sample in Experiment 2, and fixed effects for the day of participation and the hour blocks in Experiment 3. The point estimates with controls are similar to the ones without. In Columns 3 and 4 of Table 2 we obtain similar estimates from tobit regressions modeling the censoring (at 0 and 60 minutes in Experiment 2 and at 0 and 20 envelopes in Experiment 3). In Columns 5 and 6 we present marginal effects from probit regressions on an indicator for doing extra work, yielding similar, but noisier, estimates. In Columns 2 and 4 of Online Appendix Table 5 we consider whether the effect of gifts in Experiments 2 and 3 is higher when workers are informed of a higher return for the employer, but we do not find any systematic relationship.

Online Appendix Figure 6 shows that the results for Experiment 2 are parallel in terms of units of extra work as a fraction of the work completed in the first two hours, as opposed to merely extra time spent, addressing concerns that workers were not productive in the extra time. (The fraction is 0 for workers who do not stay extra.) Conversely, Online Appendix Figure 7 shows that the results for Experiment 3 are parallel if we measure the extra work in terms of units of time (minutes) as opposed to addresses coded. Online Appendix Figures 8 and 9 split the key findings for Experiment 2 by the recruitment method, Craigslist versus on campus. In both samples, very few subjects stay extra time in the control group. The Craigslist sample, though, is more sensitive to both the wage-rate treatments and the gift treatments than the student sample. Methodologically, this might hold import for those interested in comparing results across subject populations.

5 Structural Estimates

We now move from a reduced-form analysis to our structural estimation. We are able to utilize data from all three experiments, and indeed find them complementary in a similar way in which they provided useful insights when combined in reduced-form.

productivity in the other treatments to examine whether the gift has an impact on productivity. In the pre-analysis plan we did not focus on this comparison as we expected it to be statistically under-powered given the general inelasticity of productivity tasks and the relatively small sample receiving the (early) gift (50 workers). As Online Appendix Table 6 shows, there is no statistically significant evidence of a difference.

5.1 Estimation Approach

Set-up. To estimate the model, we build on Shearer (2004) and Andreoni and Sprenger (2012) and assume that workers maximize (1) with a cost of effort:

$$C_i(e_i) = c(e_i) * \eta_i \tag{10}$$

For the first term in (10), $c(e)$, we consider two families: a power function, as used in some previous literature, and an exponential function. The power function is $c(e) = e^{1+\gamma}/(1+\gamma)$, with $\gamma > 0$ denoting the inverse of the elasticity of effort to the return to effort. In the exponential specification, $c(e) = \exp(\gamma e_{i,t})/\gamma$ with $\gamma > 0$. Both functions satisfy the desired properties $C'(e) > 0$, $C''(e) > 0$, and $\lim_{e\to\infty} C'(e) = \infty$.[27]

The second term, η_i, introduces heterogeneity as a multiplicative factor. It η_i captures the impact of both the observables, X_i, and unobservables (a noise term). Since η_i ought to be positive (as a negative draw implies a negative cost of effort), we assume a log-normal distribution where the mean is potentially a function of the observables, $ln(\eta_i) \sim N(k(X_i), \gamma^2\sigma^2)$.

Under the power specification, the first-order condition (2) implies

$$p_W + A(Gift, p_E, p_W) - (e_i)^\gamma * \exp[k(X_i) - \gamma * \epsilon_i] = 0,$$

with $\epsilon_i \sim N(0, \sigma^2)$. Taking the second term to the right-hand side and taking natural logs, and solving for $\log(e_i)$, we obtain the estimating equation

$$\log(e_i) = \frac{1}{\gamma}[\log(p_W + A(Gift, p_E, p_W))] - \frac{1}{\gamma}k(X_i) + \epsilon_i. \tag{11}$$

The first term is the motivation term, which incorporates private incentives (the pay rate p_W) and the pro-social component A. This term also shows that $1/\gamma$ is the elasticity of effort to motivation. The second term is a level shifter due to differences in the level of the cost function; we return to this term below. The final term is the error term. We take a random-coefficients approach and assume that the source of the error term is unobserved differences in the (log-normal) cost of effort.

Similarly, under the exponential cost-of-effort specification, we obtain

$$e_i = \frac{1}{\gamma}[\log(p_W + A(Gift, p_E, p_W))] - \frac{1}{\gamma}k(X_i) + \epsilon_i. \tag{12}$$

The exponential function thus leads to the same specification, except with effort, rather than log effort as the outcome of interest. This model thus micro-founds reduced-form specifications with the outcome variables effort (under exponential cost) or log effort (under power cost).

Productivity Experiment. While this set-up is common to our three experiments, there

[27] The exponential cost function does not satisfy the property $C'(0) = 0$, allowing for the possibility of optimal effort at the zero corner. In practice, this does not matter given that the lowest effort observed in any round in Experiment 1 is 7 envelopes and we include in the effort the required initial 120 minutes in Experiment 2.

are some differences, especially in the controls $k(X_i)$. In Experiment 1, we measure effort as the number of letters prepared in each of ten batches $t = 1, ...10$. We also make the additional, identifying assumption that the cost function is additively separable across batches.[28] We can then write the utility maximization batch-by-batch, with a choice of effort $e_{i,t}$ for worker i in batch t.

Given that we observe an individual multiple times, we can control for individual fixed-effects k_i. In addition, we control for the evolution of the cost of effort over batches with a function $f(t)$, which captures learning by doing or fatigue. We allow for indicators d_2, d_3, and d_4 for batches 2, 3 and 4, d_{5-8} for batches 5-8 and d_{9-10} for batches 9-10. This specification is motivated by the overall flatness of the output function from batch 5 on. The indicator d_{9-10} ensures that the estimated gift effects are not biased by a change in the cost of effort in the last rounds.[29] As an auxiliary approach, we allow for a quadratic function $f(t) = \eta_1 t + \eta_2 t^2$ and similarly for a cubic.

Thus, the estimating equation for the power cost case is

$$\log(e_{i,t}) = \frac{1}{\gamma}\left[\log\left(p_{W,t} + A\left(Gift, p_{E,t}, p_{W,t}\right)\right)\right] - \frac{1}{\gamma}k_i - \frac{1}{\gamma}f(t) + \epsilon_{i,t}. \tag{13}$$

and similarly for the exponential cost function, except with $e_{i,t}$ as outcome variable. We estimate equation (13) with a non-linear least-squares regression, clustering the standard errors by session.

Extra-Work Experiments. In Experiment 2, we measure effort as the total minutes worked inclusive of the required 120 minutes, since those initial minutes presumably contribute to the tiredness. Hence, an individual who stays 30 minutes is coded as exerting effort $e_i = 150$. Similarly, in Experiment 3, the total number of envelopes completed includes the 40 required envelopes. We assume that the value to the employer p_E of each additional minute in Experiment 2 is the per-minute cost of an additional worker, that is, 50 cents (since we pay \$60 for 2 hours of work).[30] In Experiment 3, it is the value we communicated, 5 cents or 10 cents per address checked.

We model the censoring at the corner solution— $e_i = 120$ (workers who opt not to stay) and $e_i = 180$ (workers who stay for the full 60 extra minutes) in Experiment 2, and 40 envelopes versus 60 envelopes in Experiment 3—and estimate the model by maximum likelihood.

5.2 Estimates in the Productivity Experiment

Baseline Social Preferences. We first estimate the model using batches 1-8 in Experiment 1. We allow for both pure altruism towards the charity employer with social-preference weight A equal to $\alpha(p_E - p_W)$ and warm glow with social-preference weight A equal to $a\overline{p_E}$, where we take the

[28] DeJarnette (2015) finds that effort in a real-effort task across rounds is mostly habit-forming, with a moderate effect size which mostly decays after 15 minutes. Given our 10-minute breaks between the batches and the use of two orders, violations in time separability, while certainly possible, are unlikely to have major effects on our estimates.

[29] We cannot dummy out every batch, since that would take out the comparison to the training batches which are always in batches 1 and 5. The function $f(t)$ in the registration differs in two ways. First, we assumed $d_2 = 0.5 * d_3$, since we thought that we could not estimate d_2 separately from d_3. Since we can in fact do so (though we do not reject the one-half restriction), we allow for a more general specification. Second, we assumed $d_{9-10} = d_{5-8}$, a restriction which we relax. We show that adopting the pre-registered specification leads to similar results.

[30] We acknowledge that the value of work in Experiment 2 is not stated to the workers and is thus potentially ambiguous; we follow the assumption that it is 50 cents, as laid out in the pre-registration.

average return to the employer $\overline{p_E}$ to be 0.3.[31] We first use the power-cost specification. Empirical estimates in Column 1 of Table 3 imply that social preferences are entirely of the warm-glow type, with significant warm glow ($\hat{a} = 0.46$, s.e. 0.07), putting weight on the average employer return equal to about half the weight put on private payoffs, and a precise zero for the pure altruism coefficient ($\hat{\alpha} = -0.01$, s.e. 0.04). We also estimate a substantial warm glow towards the grocer as employer, $\hat{a}$ =0.73 (s.e. 0.08). (We cannot tease apart altruism and warm glow given that the return to the grocer does not vary.) Finally, the estimated curvature of the cost function ($\hat{\gamma} = 9.5$) implies a low elasticity of effort of $1/9.5 = .10$.

To highlight the identification for α, note from (9) that $(\partial e^*/\partial p_E)/(\partial e^*/\partial p_W) = \alpha/(1-\alpha)$. Output increases by 0.6 envelopes for a 30-cent increase in the employer return (that is, $\partial e^*/\partial p_E \simeq 2$) and by 4 envelopes for a 20-cent increase in the piece rate (that is, $\partial e^*/\partial p_W \simeq 20$), implying $\alpha \simeq 0.11$. The estimated value, which uses all the variation in the data, is even closer to zero.

To visualize the fit of the model, we compare the observed patterns with the fit of the models in Figures 2a-c assuming only altruism or only warm-glow social preferences (see Online Appendix Table 7). While both models fit the response to incentives well (Figure 2a), the altruism model has difficulty fitting the combination of a small response to the match rate (Figure 2b) and a large response to training (Figure 2c). The warm-glow model, instead, predicts no response to the match rate (as a core insight) and matches the training effect with a higher level of warm glow a.

We present a number of alternative specifications. In Column 3 of Table 3, we use the exponential cost-of-effort function, and thus the number of envelopes prepared, as opposed to the log, as the dependent variable. The results are nearly identical (except for the parameter γ which has a different scaling). In Online Appendix Table 8, we also show that the estimates are similar if (i) we model the learning-by-doing with quadratic or cubic polynomials rather than with a set of indicators; (ii) the altruism model includes only the return to the employer, $A = \alpha p_E$, i. e., does not take into account that the piece rate detracts from the employer return; (iii) we allow for some warm glow during the training rounds, restricted to be half the size of the other periods, $A = (1/2) * .3$. In the latter case, there is some (limited) evidence of pure altruism.

Gift Treatments. In Columns 2 and 4 of Table 3, we use the full data set from Experiment 1 (batches 1 to 10) to estimate the sensitivity of the social preference parameters to the gift treatments. Whether reciprocity is intention-based or action-based (Fehr and Gaechter, 2000; Charness and Rabin, 2002; Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006), a reciprocal worker who receives a surprisingly generous treatment from the employer is likely to display more positive social preferences towards the employer.

Given the results so far, we assume warm-glow social preferences. The estimates of baseline social preferences parallel, not surprisingly, the estimates using only the first eight batches. Importantly, we do not find statistically significant evidence of reciprocity for any of the gift treatments. We estimate a decrease of social preferences of -0.04 (s.e. 0.12) for the negative gift treatment, as

[31] If we did not rescale and assumed $A = a$, all results would be identical, with the estimated warm glow coefficients equal to those in the table multiplied by 0.3 (and with the same statistical significance).

well as for the in-kind gift treatment, -0.09 (s.e. 0.10), and a more positive effect of 0.15 (s.e. 0.13) for the positive monetary treatments. We return to these magnitudes below.

In Online Appendix Table 9, we consider a series of robustness checks. First, we consider alternative specifications for the decay parameter $f(t)$, a quadratic or cubic polynomial in the batches, and a registered specification for the round indicators $f(t)$, yielding similar results.[32] Next, we allow for the warm glow effect of a gift, a_{Gift}, to decay to δa_{Gift} in round 10. We also present estimates assuming pure altruism, rather than warm glow. The model has trouble converging, but at least for the exponential cost function yields similar conclusions.

5.3 Estimates in the Extra-Work Experiments

Experiment 2. Regarding the baseline social preferences, we cannot separately estimate altruism and warm glow in Experiment 2 given that we do not have quantitative variation in the return to the employer. Thus, given the evidence in Experiment 1, we assume warm-glow preferences. In principle, the baseline warm glow parameter a is identified with the wage-rate variation. Yet, Columns 1 and 2 of Table 4 (corresponding to, respectively, exponential and power cost of effort) show a very wide confidence interval for a. Figure 4a shows the reason for such imprecise identification. The model predicts that effort should be a concave function of the wage rate, as equations (12) and (11) show. Yet, the increase in extra work is about linear in incentives: the increase from the medium wage-rate (25¢/minute) to the high wage-rate (50¢/minute) is about the same size as from the control group (0¢/minute) to the low wage-rate (25¢/minute). Since higher values of a contribute to a less concave predicted pattern, we cannot reject the upper bound for a.

The imprecise identification for the baseline warm-glow parameter a, though, does *not* compromise the identification of the other social-preference parameters, which are the focus of Experiment 2. The qualitative statement about high return to the employer raises social preferences by a modest 0.109 (s.e. 0.093). Importantly, the gift treatments increase the social preferences moderately, by 0.303 (s.e. 0.144) for the monetary gift, 0.181 (s.e. 0.139) for the non-monetary gift, and 0.360 (s.e. 0.146) for the non-monetary early gift. These estimates are nearly identical whether we assume an exponential cost function or a power cost function (Columns 1 and 2).

How is it possible to estimate the social-preference shifts if the level of the social preferences is not estimated with any degree of precision? This is a virtue of the pay-rate design: it "prices out" the parameter connected to a particular treatment comparison. As Figures 4a and 4c show, the impact on extra work of the gift treatments, compared to the control, is about half the size of the impact of the medium wage-rate treatment (25¢/minute). This prices out the reciprocal social preference, $a_{Gift} * \bar{p_E}$, to be less than 0.25. Given $\bar{p_E} = 0.5$, this implies $a_{Gift} \approx 0.2 - 0.4$. Similarly, the impact of the employer high-return treatment (Figure 4b) is clearly smaller than the effect of the medium wage rate, again identifying the underlying social preference.

[32] Online Appendix Figure 10 presents the estimated time path under a quadratic or cubic polynomial, comparing to the benchmark specification. In Online Appendix Figure 11, we report the fit of the model across all ten rounds, for the benchmark indicator function $f(t)$ and for the exponential cost specifications, averaging across the different gift treatments in rounds 9-10. Overall, the model does well in fitting the data.

In Figures 4a-c we display the fit of the estimated model in Column 1 of Table 4. The estimates fit the observed patterns quite well, except for overfitting somewhat the number of minutes worked in the control group and in the gift groups. Relatedly, we compare the distribution of extra work in Figure 6a to the model prediction in Online Appendix Figure 12a. In the control and gift groups the model slightly underestimates the share of people staying over time, but at the same time it overpredicts the share staying for more than 20 minutes. The assumption of a log-normal error term, which we pre-specified, does not perfectly fit the distribution of types in the data.

In Online Appendix Table 10 we test further the robustness of the estimates. First, we estimate the model with maximum likelihood, but do not include the initial 120 minutes in the cost-of-effort calculation. Second, we employ a minimum-distance estimator, instead of maximum likelihood, using as moments the share of subjects who stay in each 5-minute bin. Third, we use another minimum-distance estimates with a sparser set of moments. For each of these three estimators, the estimates for the reciprocity parameters are similar to the benchmark.

Experiment 3. For Experiment 3, given the quantitative variation in the return to the employer, we estimate a model with both altruism towards the employer and warm glow in Columns 3 and 5 of Table 4. As in Experiment 1, we estimate a very precise zero pure-altruism coefficient, $\alpha = -0.006$ (s.e. 0.006). We estimate a sizable, if not very precisely estimated, degree of baseline warm glow at $a = 0.104$ (s.e. 0.071). Given the lack of evidence for pure altruism, in Columns 4 and 6 we estimate a warm-glow model in which we allow the information on the high return to the employer to affect the warm-glow coefficient. We estimate that higher returns to the employer, if anything, affect baseline social preferences *negatively*, possibly because a higher stated return to the employer leads the workers to be less satisfied with a given level of pay.

Turning to the reciprocity parameters, we estimate an impact of the monetary gift on the warm-glow coefficient of $a_{Gift} = 0.076$ (s.e. 0.028). The effect size can be inferred from the reduced-form estimates. The gift treatment leads to an increase in extra work that is less than half the size of the low-piece rate treatment, thus implying $a_{Gift} * \bar{p_E}$ is less than half of 0.01 (the 1-cent-per-address incentive in the low-piece-rate). Given $\bar{p_E} = 0.04$, this implies $a_{Gift} \approx 0.1$. Relative to the baseline warm glow of $a = 0.104$, the increase is large, although it is more modest in absolute terms.

The fit of the model for Experiment 3 is very good (displayed by the crosses in Figures 5a-c), including the fit of the distribution of effort (Online Appendix Figure 12b, compared to Figure 6b), save for the heaping at round numbers which the model does not reproduce. Online Appendix Table 11 shows that a similar set of robustness checks as for Experiment 2 yield very similar results for all the social preference parameters.

5.4 Comparison of Estimates in the Three Experiments

Across the three experiments, we estimate that baseline social preferences are substantial (even though in Experiment 2 they are estimated imprecisely), and do not respond to information, either quantitative or qualitative, on the return to the employer (provided this return is positive). Indeed, the estimates for the pure altruism parameter α for Experiment 1 and 3 are very close to zero and

nearly identical (Figure 7a). These results are consistent with a "warm glow" model, as opposed to pure altruism.

At the same time, the gift treatments appear to lead to divergent findings, as Figure 7b shows, when comparing effect sizes of the gift across the three experiments. Relative to the control group, the monetary gift increases productivity by only 1 percent, while it increases the extra work by over 200 percent in Experiment 2 and by 46 percent in Experiment 3. Similarly, there is a wide difference in the effect of the in-kind gift treatments.

In Figure 7c, we make the same comparison, but in terms of the underlying reciprocity parameters a_{Gift}, the shifts in the warm-glow baseline social-preferences, which become $a + a_{Gift}$ in response to the gift. The differences in the reciprocity parameters are nowhere near as large. In response to the monetary gift, we estimate $a_{Gift} = 0.15$ (s.e. 0.13) in Experiment 1, $a_{Gift} = 0.30$ (s.e. 0.14) in Experiment 2, and $a_{Gift} = 0.07$ (s.e. 0.03) in Experiment 3. We cannot reject that the three structural parameters are the same, as the p-values in Figure 7c show. For the in-kind gift, we estimate $a_{Gift} = -0.09$ (s.e. 0.10) in Experiment 1 and $a_{Gift} = 0.18$ (s.e. 0.14) and $a_{EarlyGift} = 0.36$ (s.e. 0.15) in Experiment 2. The difference is statistically significant compared to the early-gift results ($p = 0.008$), but the quantitative difference is nowhere near as large.

Thus, the comparison in terms of the social preference parameters attenuates what, *prima facie,* would appear to be a large and irreconcilable divergence between the two experiments in the impact of the gifts. It is accounting for the different elasticities of the outcomes, using the pay-rate design, that permits such partial reconciliation: the elasticity of productivity with respect to motivation is at least an order of magnitude smaller than the elasticity of extra work to motivation. Thus, for a given reciprocity shift due to a gift, the observed impact on productivity will tend to be at least an order of magnitude smaller than on labor supply.

5.5 Implications of Pay-Rate Design and Structural Estimates

The structural estimates allow us to compare results across our experiments, and thus to partially reconcile apparently conflicting findings. We highlight here three additional implications.

First, the structural estimates can be used for out-of-sample predictions. For example, in the pre-analysis plan for Experiment 2, we used the structural estimates from the productivity experiment, together with a conjecture about the elasticity of labor supply, to do a model-based power calculation for the new design, compared to alternative designs.

Second, the structural model allows us to revisit the estimates in previous gift-exchange field-experiments. As we discussed in Section 2, under a power cost of effort function, we can derive a measure of reciprocity, the proportional increase in warm glow $(a + a_{Gift})/a$ due to the gift, provided we know the elasticity $1/\gamma$ of productivity to motivation (equation 5). While we do not observe the elasticity from the previous experiments, a starting assumption is that it is comparable to the estimate from our productivity task. Under this assumption, as Online Appendix Table 12 shows, some previous papers imply very large reciprocity effects, such as a 400 or 700 percent increase in social preferences with a positive gift, and an 88 percent decrease with a negative gift.

Third, we can use the estimates for counterfactual exercises. While the social-preference estimates are context specific, as a calibration exercise in Online Appendix Figure 13a we consider how employee effort responds to pay rate variation (on the x axis) under the assumption of no social preferences ($a = 0$) versus for the baseline social preferences from the productivity experiment ($\hat{a} = .4$). Under no social preferences, introducing a pay rate has dramatic impacts on output; in the presence of warm glow, instead, effort is quite high even with no pay rate and it increases slowly with the pay rate increases. Given this, the optimal pay rate for firm profits (Online Appendix Figure 13b) is positive (3 cents) without social preferences, but is instead zero under our estimates for warm glow. As Englmeier and Leider (2012c) highlight, incentives and social preferences are largely substitutes in motivating workers.

6 Discussion and Conclusion

In this Section, we take stock of our results and discuss them in the context of the literature.

Baseline Social Preferences. In all three experiments, worker effort does *not* respond to variation in the stated return to the employer. Thus, workers' social preferences towards the employer is not consistent with Beckerian altruism towards the employer. There is evidence, however, of *some* form of prosocial motivation, given positive effort in the absence of pay-rates, and the drop in effort when the work output is not used in Experiment 1 (in the training rounds).

These findings echo the results in real-effort experiments in which the experimenter donates money to charities as a function of the effort exerted (e.g., Imas, 2014, Tonin and Vlassopoulos, 2015, and DellaVigna and Pope, 2018). While subjects in these experiments work harder in response to these social incentives, the return to the charity does not affect the worker effort.

We interpret our findings in light of a "warm glow" model. Here, warm glow is broadly construed and may also stand in for norms in the workplace ('one needs to put in effort'). We intend it as an alternative to the pure-altruism model that is sometimes used to capture social preferences in the workplace (e.g. Bandiera, Barankay, and Rasul, 2005; Hjort, 2014), and a starting point to better understand social preferences in the workplace. Within this model, we estimate large and economically meaningful baseline social preferences in Experiment 1, sizable ones in Experiment 3, and imprecisely estimated ones in Experiment 2. Taken at face value, this implies that social preferences can play an important role as motivators at work, and have the potential to partially substitute for the role of incentives. The theme that social preferences towards employers may be important in the workplace is a classical one, but little previous field evidence pins down the nature of these preferences.

Gift Exchange. Another insight is that we find little evidence of gift exchange in the productivity experiment, consistent with the papers which find modest or null effects (Cohn, Fehr, and Goette, 2014;Esteves-Sorenson, 2018). Yet we find statistically significant—if quantitatively moderate–evidence of gift exchange in the extra-work experiment. As we discussed above, part of the disparity is due to the vast difference in elasticity of the outcome across the experiments: the

much more elastic extra-work margin makes it possible to detect even modest-sized gift-exchange effects. We find it intriguing that in the extra-work experiments we replicate the early finding in the literature of a positive impact of monetary gifts (e.g., Gneezy and List, 2006), which some recent experiments with a productivity design do not (e.g., Esteves-Sorenson, 2018). We also find an impact of in-kind gifts as in Kube, Marechal, and Puppe (2012), though not larger than the impact of monetary gifts. We hope that more papers will adopt the extra-work design employed in our paper, which could further our understanding of social preferences at work.

What remains is a partial puzzle for the lack of gift effects in Experiment 1, in particular for non-monetary gifts, for which we can reject that the effect is the same as in Experiment 2. We consider a number of possibilities for this discrepancy. A first possibility is that, given the complicated within-person design, the gift treatments did not trigger the required surprise and mood response to induce reciprocation. However, evidence from a short debriefing survey indicates that the gifts in the final batches did induce the intended emotions of happiness and even self-reported motivation.[33] A second possibility is that worker effort towards the end of the longer productivity experiment becomes habitual and unresponsive to incentives. However, we do find significant response to piece-rate variation even in the latter batches. A third possibility is that, for the gifts to have an impact, it helps to explicitly state a request, as we did in the Experiments 2 and 3, e.g., "*Would you be willing to help us enter some more of the data for up to one hour?*" Thus, the employer provides a clear "channel" for the worker to reciprocate. Ultimately, we leave it to future literature to conclusively address these possibilities.

Pay-rate design. Finally, this paper makes the case for a *pay-rate design*, which is relatively uncommon in the behavioral literature (with exceptions such as Augenblick, Niederle, and Sprenger, 2015 and DellaVigna and Pope, 2018). A pay-rate design enables the researcher to map from a reduced-form effect to underlying parameters, allows readers to judge the plausibility of an effect, and makes comparisons across experiments with different tasks easier. Moreover, estimating the elasticity of different outcomes enables researchers to identify more highly-powered experimental designs allowing for more precise and cost-effective inference. In our settings at least, the extra-work designs employed in Experiments 2 and 3 have proven to have substantially higher elasticity; we hope that future papers will further explore this type of design.

[33] For the last 65 workers in Experiment 1, we asked '*How did the pay in the last two periods make you feel? (Check all that apply) [ ] No particular reaction [ ] It made me happy [ ] Felt more motivated and energetic [ ] It was what I expected [ ] Surprised, it was more than I expected [ ] Surprised, it was less than I expected [ ] Felt unhappy [ ] Felt insulted [ ] It was unfair.*' We coded the share that reported being happy or unhappy, as well as the share reporting a positive surprise or a negative surprise. As Online Appendix Figure 14 shows, in the positive gift treatments 70 to 80 percent of subjects report positive mood, compared to 20 percent in the control group and 5 percent in the negative gift group. The results are similar for positive surprise, and are reversed for unhappiness and negative surprise.

References

Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman. 2011. "Reference Points and Effort Provision" *American Economic Review,* 101 (2), 470-492.

Akerlof, George A. 1982. "Labor Contracts as Partial Gift Exchange", *Quarterly Journal of Economics,* 97 (4): 543-569.

Akerlof, George A. and Rachel E. Kranton. 2005. "Identity and the Economics of Organizations" *Journal of Economic Perspectives,* 19 (1), 9-32.

Al-Ubaydli, Omar, Steffen Andersen, Uri Gneezy and John A. List. 2015. "Carrots that look like sticks: Toward an understanding of multitasking incentive schemes" *Southern Economic Journal,* 81(3), 538-561.

Andreoni, James. 1989. "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence." *Journal of Political Economy*, 97 (6), 1447-1458.

Andreoni, James and Charles Sprenger. 2012. "Estimating Time Preferences from Convex Budgets," *American Economic Review*, 102 (7), 3333-56.

Araujo, Felipe A., Erin Carbone, Lynn Conell-Price, Marli W. Dunietz, Ania Jaroszewicz, Rachel Landsman, Diego Lamé , Lise Vesterlund, Stephanie W. Wang, Alistair J. Wilson. 2016. "The slider task: an example of restricted inference on incentive effects". *Journal of the Economic Science Association*, 2(1), 1–12.

Ariely, Dan, Anat Bracha, and Stephan Meier. 2009. "Doing good or doing well? Image motivation and monetary incentives in behaving prosocially." *American Economic Review,* 99(1), 544-55.

Ariely, Daniel, Emir Kamenica, Drazen Prelec. 2008. "Man's search for meaning: The case of Legos" *Journal of Economic Behavior & Organization*, 67 (3–4), 671–677.

Augenblick, Ned, Muriel Niederle and Charles Sprenger. 2015. "Working Over Time: Dynamic Inconsistency in Real Effort Tasks " *Quarterly Journal of Economics*, 130 (3): 1067-1115.

Bandiera, Oriana, Iwan Barankay and Imran Rasul. 2005. "Social Preferences and the Response to Incentives: Evidence from Personnel Data", *Quarterly Journal of Economics,* 120 (3), 917-962.

Becker, Gary S. 1974. "A Theory of Social Interactions" *Journal of Political Economy,* 82(6), 1063-1093.

Bellemare, C. and B.S. Shearer. 2011. "On the Relevance and Composition of Gifts within the Firm: Evidence from Field Experiments," *International Economic Review*, 52 (3), 855-882.

Benabou, Roland, and Jean Tirole. 2003. "Intrinsic and extrinsic motivation." *The Review of Economic Studies* 70, no. 3: 489-520.

Benabou, Roland and Jean Tirole. 2006. "Incentives and Prosocial Behavior" *American Economic Review,* 96 (5), 1652-1678.

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

Breza, Emily, Supreet Kaur, and Yogita Shamdasani. 2018. "The Morale Effects of Pay Inequality." *Quarterly Journal of Economics*, 133 (2), 611-663.

Brown, M., Armin Falk, and Ernst Fehr. 2004. "Relational Contracts and the Nature of Market Interactions," *Econometrica*, 72(3), 747–780.

Carpenter, Jeffrey, Peter Hans Matthews, and John Schirm. 2010. "Tournaments and office politics: Evidence from a real effort experiment." *American Economic Review,* 100(1), 504-17.

Charness, Gary and Peter Kuhn. 2007. "Does Pay Inequality Affect Worker Effort? Experimental Evidence", *Journal of Labor Economics*, 25 (4), 693-723.

Charness, Gary, and Matthew Rabin. 2002. "Understanding social preferences with simple tests." *Quarterly journal of Economics*: 817-869.

Cohn, Alain, Ernst Fehr, Benedikt Herrmann and Frédéric Schneider. 2014. "Social Comparison And Effort Provision: Evidence From A Field Experiment", *Journal of the European Economic Association* , 877–898.

Cohn, Alain, Ernst Fehr, and Lorenz Goette. 2014. "Fair Wages and Effort Provision: Combining Evidence from a Choice Experiment and a Field Experiment", *Management Science*, 61(8), 1777-1794.

Conlin, Michael, Ted O'Donoghue, and Timothy J. Vogelsang. 2007. "Projection Bias in Catalog Orders." *American Economic Review,* 97(4), 1217-1249.

Cullen, Zoe and Ricardo Perez-Truglia. 2018. "How Much Does Your Boss Make? The Effect of Salary Comparisons," working paper.

DeJarnette, Patrick. 2015. "Effort Momentum." working paper.

DellaVigna, Stefano. 2018. "Structural Behavioral Economics" in *Handbook of Behavioral Economics*, Volume 1 (eds. Doug Bernheim, Stefano DellaVigna, and David Laibson), Elsevier.

DellaVigna, Stefano, List, John. A., and Malmendier, Ulrike. (2012). "Testing for Altruism and Social Pressure in Charitable Giving." *Quarterly Journal of Economics*, 127 (1), 1-56.

DellaVigna, Stefano, List, John. A., Malmendier, Ulrike, and Gautam Rao. 2017. "Voting to Tell Others." *Review of Economic Studies*, 84(1), 143-181.

DellaVigna, Stefano, List, John. A., Malmendier, Ulrike, and Gautam Rao. 2021a. "Gift Exchange at Work." AEA RCT Registry. February 26. https://doi.org/10.1257/rct.502-5.0.

DellaVigna, Stefano, List, John. A., Malmendier, Ulrike, and Gautam Rao. 2021b. "Data and code for: "Estimating Social Preferences and Gift Exchange at Work"" American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/***.

DellaVigna, Stefano, and Devin Pope. 2018. "What Motivates Effort? Evidence and Expert Forecasts", *Review of Economic Studies*, 85, 1029–1069.

DellaVigna, Stefano, and Devin Pope. Forthcoming. "Stability of Experimental Results: Forecasts and Evidence", *American Economic Journals: Microeconomics.*

Dohmen, Thomas, and Armin Falk. 2011. "Performance pay and multidimensional sorting: Productivity, preferences, and gender." *American Economic Review,* 101(2), 556-90.

Dufwenberg, Martin, and Georg Kirchsteiger. 2004. "A theory of sequential reciprocity." *Games and economic behavior* 47(2): 268-298.

Englmaier, Florian and Stephen Leider. 2012a."Managerial Payoff and Gift Exchange in the Field" Working paper.

Englmaier, Florian and Stephen Leider. 2012b."Gift Exchange in the Lab - It is not (Only) How Much You Give..." Working paper.

Englmaier, Florian and Stephen Leider. 2012c."Contractual and Organizational Structure with Reciprocal Agents" *American Economic Journal: Microeconomics*, 2012, 4(2), 146-83.

Esteves-Sorenson, Constanca. 2018. "Gift Exchange in the Workplace: Addressing the Conflicting Evidence with a Careful Test," *Management Science*, 64(9), 4365-4388.

Falk, Armin, and Urs Fischbacher. 2006. "A theory of reciprocity." *Games and economic behavior* 54(2): 293-315.

Fehr, Ernst and Simon Gächter, 2000. "Fairness and Retaliation - The Economics of Reciprocity", *Journal of Economic Perspectives*, 14, 159-181.

Fehr, Ernst, Georg Kirchsteiger, and Arno Riedl. 1998. "Gift exchange and reciprocity in competitive experimental markets." *European Economic Review* 42(1): 1-34.

Gilchrist, Duncan S., Michael Luca, and Deepak Malhotra. 2016. "When 3+1>4: Gift Structure and Reciprocity in the Field." *Management Science* 62(9), 2639-2650.

Gill, David, and Victoria Prowse. 2012. "A structural analysis of disappointment aversion in a real effort competition." *American Economic Review,* 102(1), 469-503.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini. 2003. "Performance in Competitive Environments: Gender Differences." *The Quarterly Journal of Economics,* 118(3), 1049-1074.

Gneezy, Uri and John A List. 2006. "Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments," *Econometrica*, 74(5), 1365-1384.

Goerg, Sebastian J., Sebastian Kube, and Jonas Radbruch. 2019. "The Effectiveness of Incentive Schemes in the Presence of Implicit Effort Costs", *Management Science*, 65(9), 3949-4450.

Hennig-Schmidt, Heike, Abdolkarim Sadrieh, and Bettina Rockenbach. 2010. "In Search of Workers' Real Effort Reciprocity—a Field and a Laboratory Experiment" *Journal of the European Economic Association,* 8(4), 817–837.

Hjort, Jonas. 2014. "Ethnic Divisions and Production in Firms", *Quarterly Journal of Economics*, 129(4), 1899-1946.

Holmstrom, Bengt and Paul Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, and Organization*, 7(0), 24-52.

Imas, Alex. 2014. "Working for the "warm glow": On the Benefits and Limits of Prosocial Incentives" *Journal of Public Economics*, 114, 14–18.

Kahneman, Daniel, Jack L. Knetsch and Richard Thaler. 1986. "Fairness as a Constraint on Profit Seeking: Entitlements in the Market" *American Economic Review*, 76(4), 728-741.

Kessler, Judd. 2013. "When Will There Be Gift Exchange? Addressing the Lab-Field Debate With Laboratory Gift Exchange Experiments." Working paper.

Krueger, Alan B. and Alexandre Mas. 2004. "Strikes, Scabs, And Tread Separations: Labor Strife And The Production Of Defective Bridgestone/Firestone Tires," *Journal of Political Economy*, v112(2), 253-289.

Kube, Sebastian, Michel André Maréchal and Clemens Puppe. 2012 "The Currency of Reciprocity: Gift Exchange in the Workplace" *American Economic Review,* 102 (4), 1644-1662.

Kube, Sebastian, Michel André Maréchal and Clemens Puppe. 2013. "Do Wage Cuts Damage Work Morale? Evidence From A Natural Field Experiment", *Journal of the European Economic Association* 11 (4), 853–870.

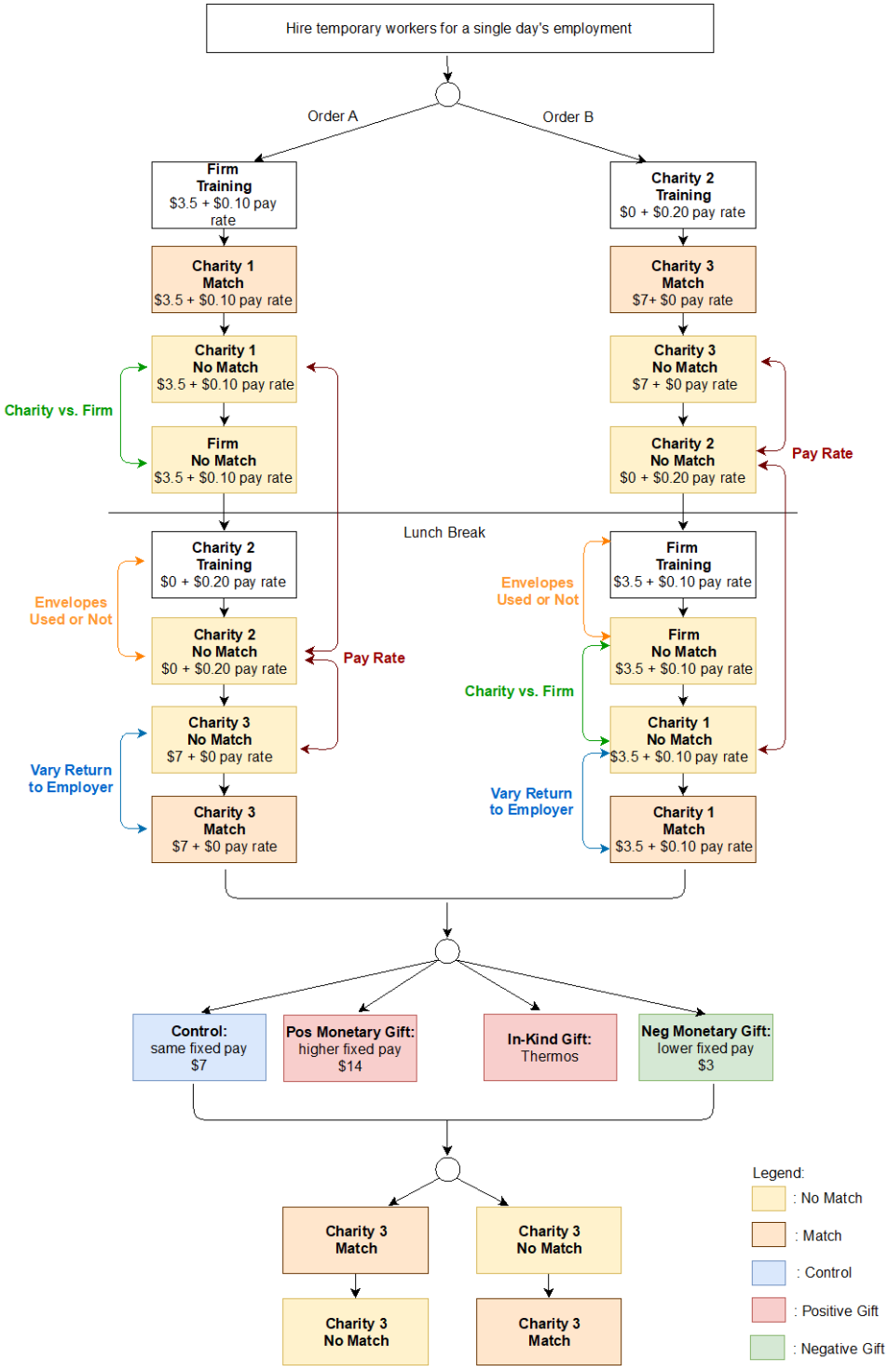
Laibson, David, Peter Maxted, Andrea Repetto, and Jeremy Tobacman. 2015. "Estimating Discount Functions with Consumption Choices over the Lifecycle," Working paper.

Mas, Alexandre. 2006. "Pay, reference points, and police performance." *The Quarterly Journal of Economics* 121.3: 783-821.

Shearer, Bruce. 2004. "Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment" *Review of Economic Studies*, 71 (2): 513-534.

Tonin, Mirco and Michael Vlassopoulos. 2015. "Corporate Philanthropy and Productivity: Evidence from an Online Real Effort Experiment." *Management Science* 61 (8): 1795-1811.

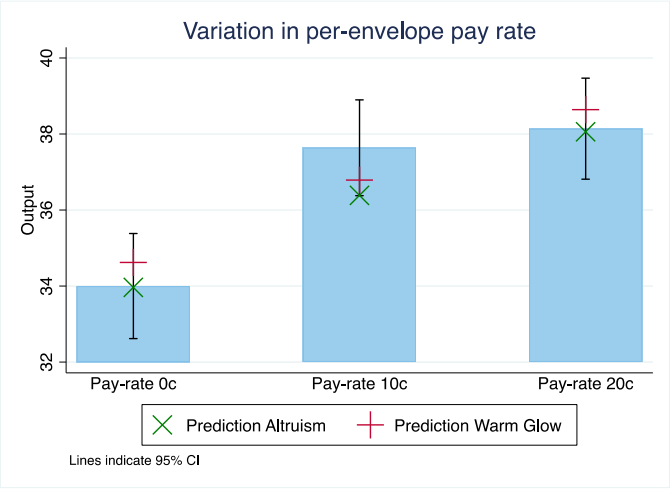
Figure 1. Productivity Experiment: Design



Notes: This figure displays the sequence of the 10 experimental rounds of envelope preparation, each of which lasting 20 minutes, for the productivity experiment. Between each round there is a 10 minute break, except between rounds 4 and 5 when there is a longer break for lunch. Subjects are randomized across sessions into Order A or Order B, as well as into three assignments of charities to be Charity 1, 2, and 3. In rounds 9 and 10, subjects are split within session into four gift exchange treatments (in the first 24 experimental sessions we did not run the in-kind gift treatment). Depending on randomized session assignment, either session 9 or session 10 involves a charity match (high return for the employer). The arrows indicate the main experimental comparisons evaluated in Figure 2a-d and in Online Appendix Figure 1a-b.

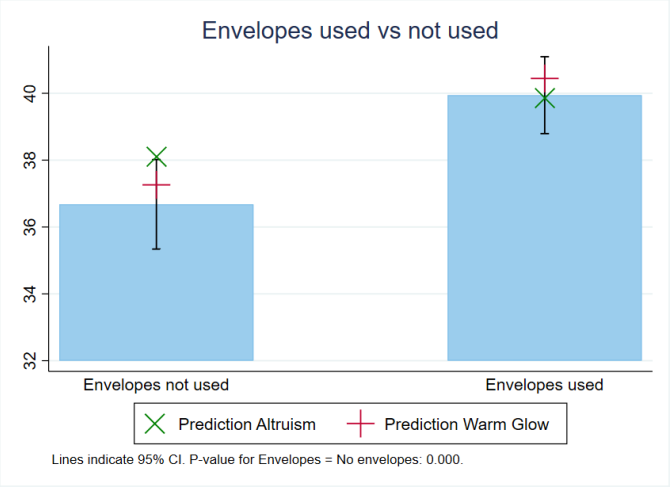
Figure 2. Productivity Experiment: Experimental Findings and Model Fit

Panel a. Variation in Pay Rate

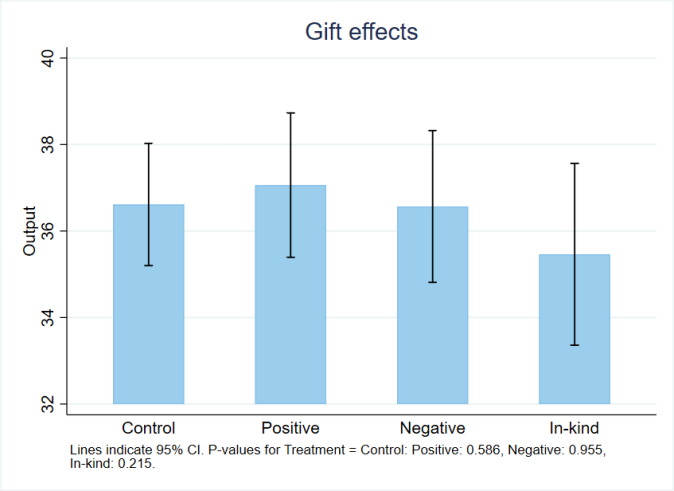


Panel b. Variation in Return to Employer (Match)

Panel c. Consequences to the Employer



Panel d. Effect of Gift Treatments



Notes: Figures 2a-c display key comparisons of average output (number of envelopes folded within a 20-minute round) across batches, as outlined by the arrows in Figure 1. The comparisons average across order A and B. Figure 2a compares the pay rates for 0c, 10c, and 20c (respectively, batches 7, 3, and 6 in Order A and batches 3, 7, and 4 in order B). Figure 2b compares the impact of high return to the employer (charity match) (batches 7 and 8). Figure 2c compares the impact of envelopes being used (batches 5 and 6). Figure 2d presents the key results for the gift exchange treatments in batches 9 and 10. The figures indicate 95% confidence intervals computed clustering by session. Figures 2a-c also indicate the average prediction for the model estimated with altruism (Online Appendix Table 6, Column 3) or with warm glow (Online Appendix Table 6, Column 4).

Figure 3. Extra Work Experiments: Design

Figure 3a. Experiment 2, Design

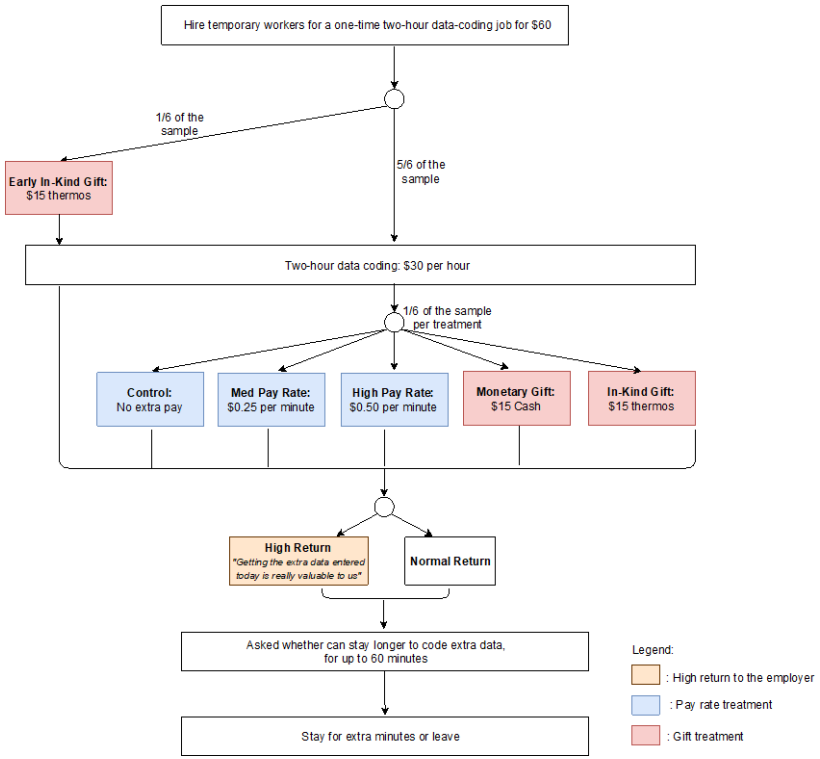
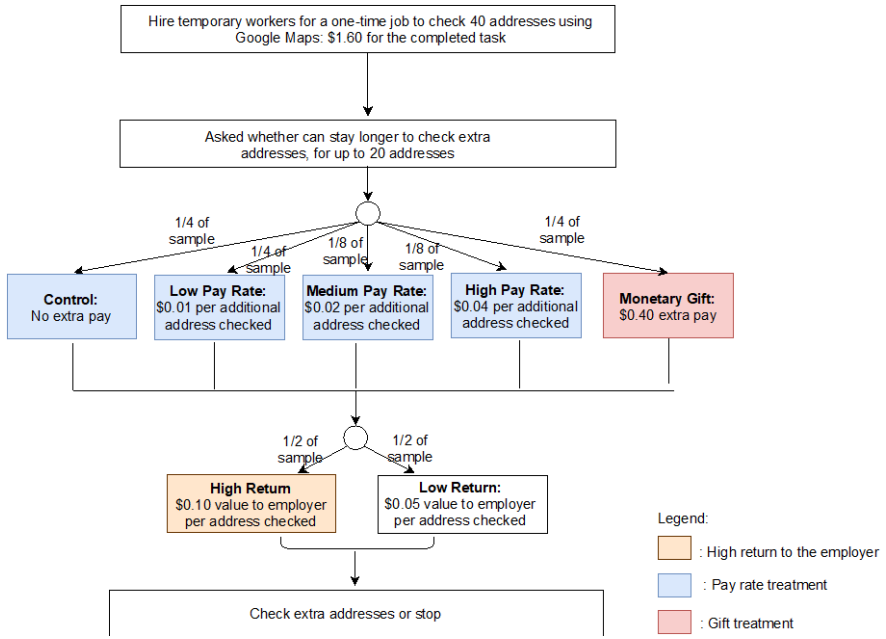


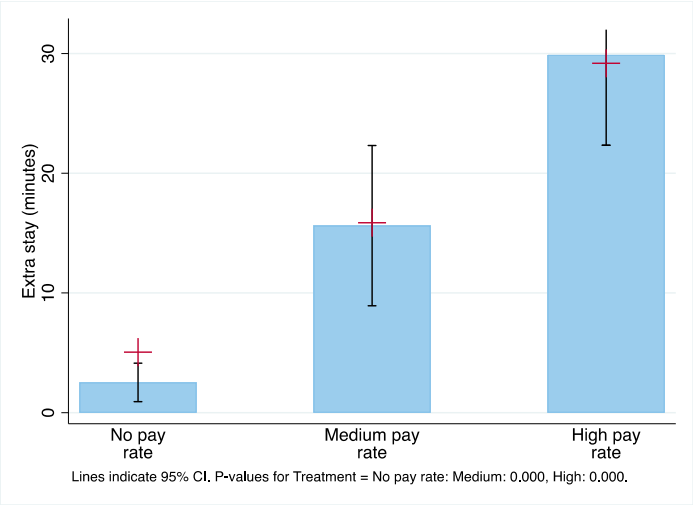
Figure 3b. Experiment 3, Design



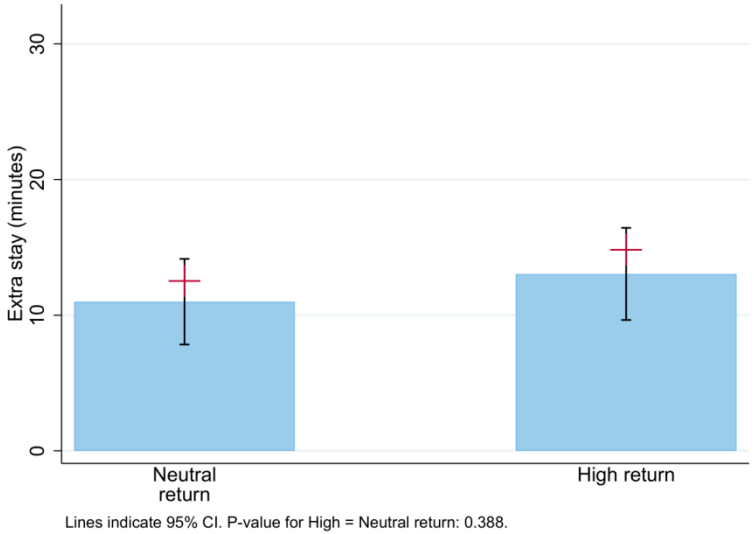
Notes: Figures 3a-b display the design for the extra-work experiments, Experiment 2 and Experiment 3.

Figure 4. Extra Work Experiment 2: Findings

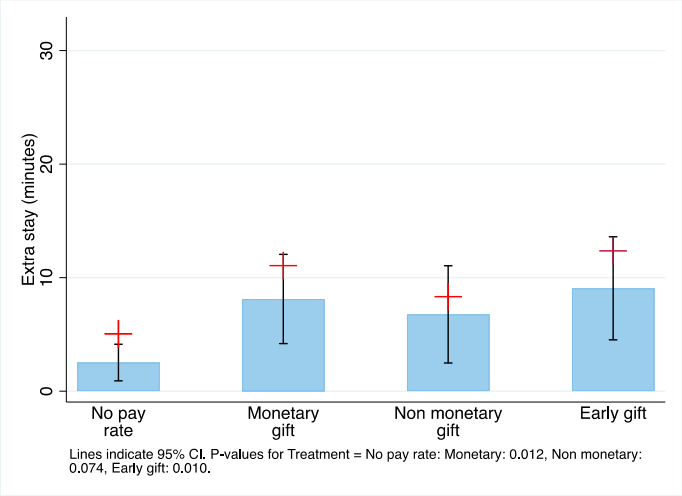
Panel a. Variation in Pay Rate



Panel b. Variation in Return to Employer



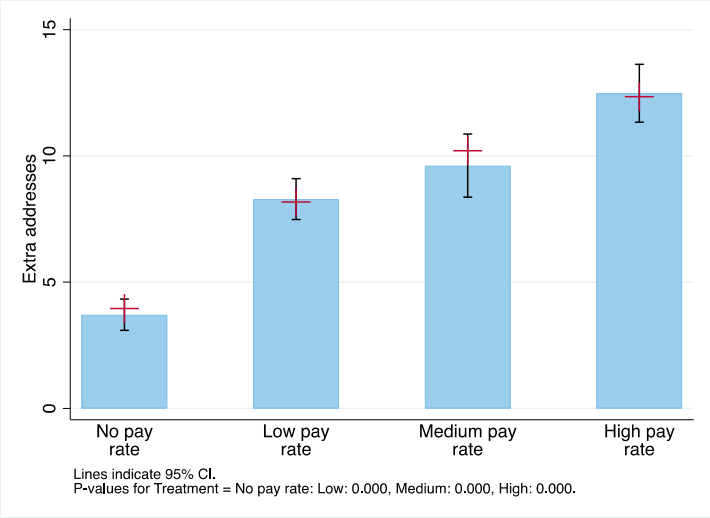
Panel c. Effect of Gift Treatments



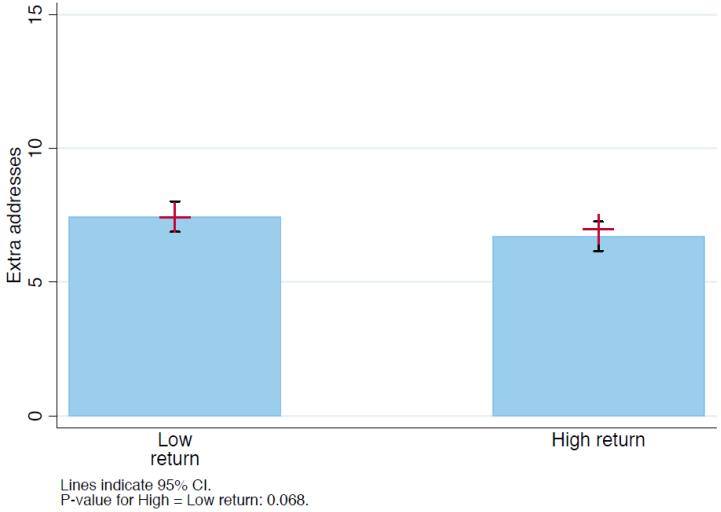
Notes: Figure 4 displays the experimental findings on the number of extra minutes of work in Experiment 2. Panels a and c compare this measure across the six main experimental arms. Panel b pools across the six experimental arms and compares the (cross-randomized) arms with stated high-return for the employer to the control arm. The figures also display as a red cross the model prediction for the parameter estimates in Column 1 of Table 4.

Figure 5. Extra Work Experiment 3: Findings

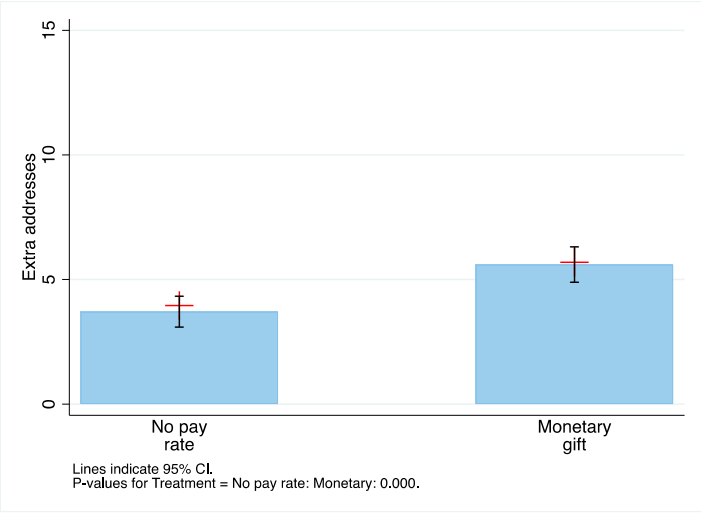
Panel a. Variation in Pay Rate



Panel b. Variation in Return to Employer



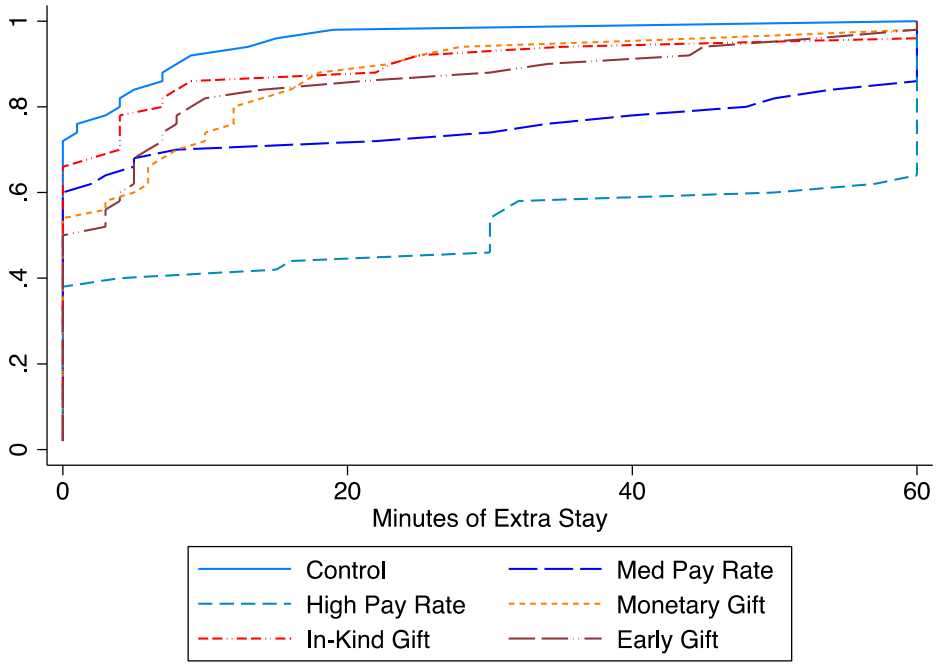
Panel c. Effect of Gift Treatments



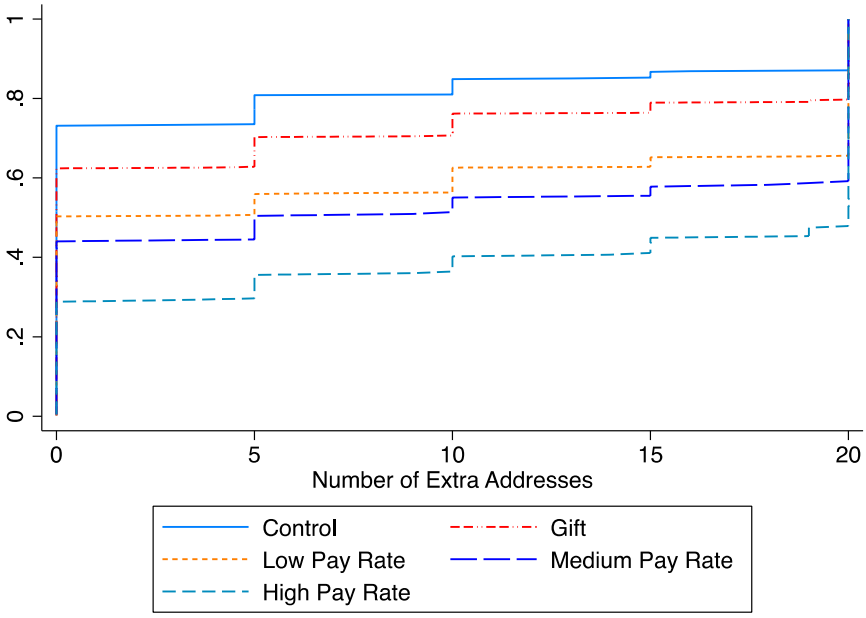
Notes: Figure 5 displays the experimental findings on the number of extra addresses checked in Experiment 3. Panels a and c compared this measure across the main experimental arms. Panel b pools across the experimental arms and compares the (cross-randomized) arms with stated high-return for the employer to the control arm. The figures also display as a red cross the model prediction for the parameter estimates in Column 4 of Table 4.

Figure 6. Extra Work Experiments: Distribution of Effort by Treatment

Panel a. Experiment 2



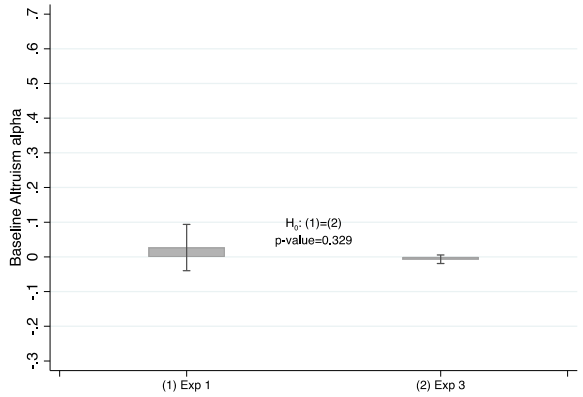
Panel b. Experiment 3



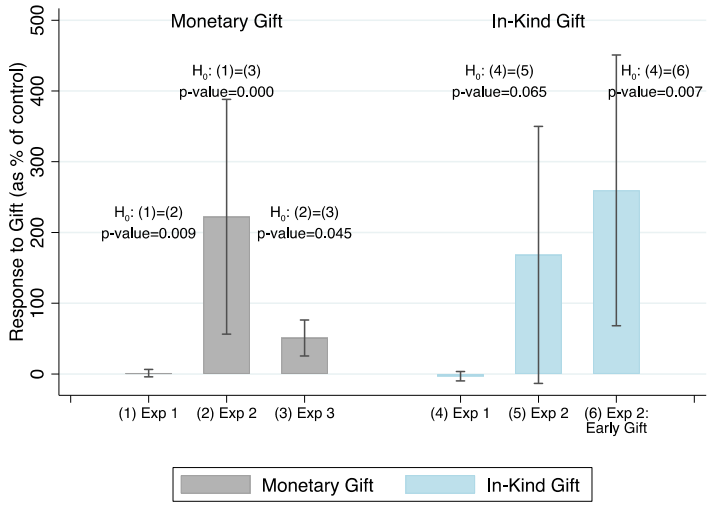
Notes: Panel a displays the c.d.f. of the number of extra-minutes stayed for the six experimental arms in Experiment 2. Panel b displays the c.d.f. of the number of extra addresses coded for the five experimental arms in Experiment 3.

Figure 7: Effect of Gift Treatments, Reduced-Form vs. Estimates of Social Preference Parameters

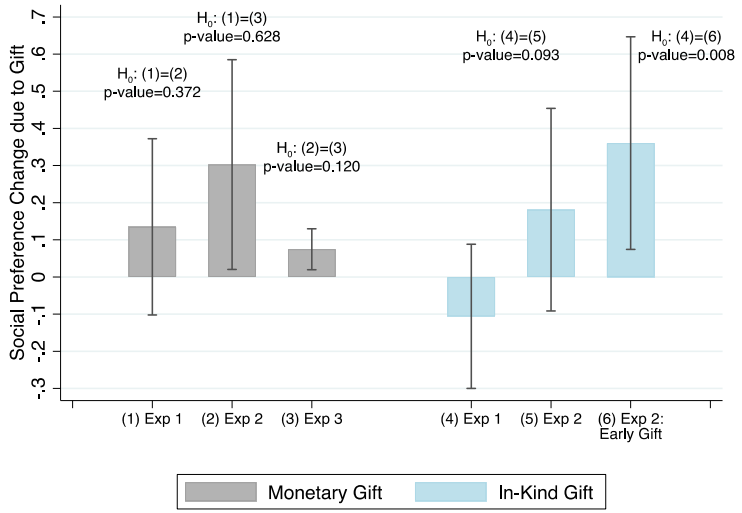
Panel a. Pure Altruism Parameter



Panel b. Reduced-Form Effort Change due to Gift



Panel c. Social Preference Change due to Gift (Reciprocity)



Notes: Panel a presents the estimate from the pure altruism parameter from Table 3, Column 3 and Table 4, Column 3. Panel b displays the reduced-form impact of the monetary and in-kind gift in the productivity experiment (labeled "Exp. 1") and in the extra-work experiments (labeled "Exp. 2" and "Exp. 3"). The estimates are shows as percent increases over the average effort in the control group. Panel c displays the same comparison for the implied reciprocity estimates from, respectively, Table 3, Column 4, Table 4, Column 1, and Table 4, Column 3. The p-values refer to a test of equality of the results comparing across the experiments.

Table 1. Productivity Experiment, Findings for Gift Treatments

Specification:	**OLS Regressions**					
Dependent Variable:	**Output in Batches 9 and 10**					
Measure of Output:	**Number of Envelopes Stuffed in 20 Minutes**			**Log of Number of Envelopes Stuffed in 20 Minutes**		
	(1)	(2)	(3)	(4)	(5)	(6)
Gift Treatments						
Positive (monetary) gift	0.448	0.903	0.603	0.006	0.026	0.015
Treatment	(0.966)	(0.737)	(0.729)	(0.028)	(0.022)	(0.021)
Negative (monetary) gift	-0.046	-0.014	-0.047	-0.017	-0.012	-0.018
Treatment	(0.953)	(0.745)	(0.754)	(0.035)	(0.031)	(0.031)
Positive In-kind (Thermos) gift	-1.152	-1.011	-1.090	-0.040	-0.029	-0.033
Treatment	(1.242)	(0.973)	(0.927)	(0.037)	(0.029)	(0.028)
Controls						
Average Output Measure		0.867			0.782	
In Batches 1-8		(0.028)			(0.028)	
Average Output Measure			0.815			0.831
In Batches 5-8			(0.027)			(0.030)
Constant	36.613	6.223	5.149	3.572	0.826	0.561
	(0.709)	(1.045)	(1.118)	(0.020)	(0.099)	(0.110)
R squared	0.003	0.585	0.608	0.003	0.483	0.519
N	N = 892	N = 892	N = 892	N = 892	N = 892	N = 892

Notes: Estimates from an OLS regression of output (Columns 1 -3) and log output (Columns 4-6) in the final two batches (Batches 9 and 10) on the gift treatments. The omitted category is a Control treatment with no "gift" (pay is the same as previously experienced with the same charity). The standard errors are clustered at the session level.

Table 2. Extra Work Experiments, Findings

Specification:	OLS Regressions		Tobit Regressions		Probit Regressions	
Dependent Variable:	Extent of Extra Work (0-60 Minutes in Exp. 2, 0-20 Addresses in Exp. 3)				Indicator for Extra Work >0	
Experiment:	Exp. 2	Exp. 3	Exp. 2	Exp. 3	Exp. 2	Exp. 3
	(1)	(2)	(3)	(4)	(5)	(6)
Pay Rate Treatments						
Low Pay Rate		4.584		28.557		0.243
Treatment		(0.519)		(3.704)		(0.033)
Medium Pay Rate	13.1	5.900	31.765	35.889	0.131	0.307
Treatment	(3.524)	(0.710)	(11.425)	(4.668)	(0.102)	(0.041)
High Pay Rate	27.424	8.761	65.287	51.003	0.351	0.464
Treatment	(3.945)	(0.664)	(12.014)	(4.800)	(0.103)	(0.042)
Gift Treatments						
Monetary Gift Treatment	5.6	1.890	21.678	12.972	0.191	0.123
	(2.132)	(0.481)	(9.284)	(3.449)	(0.102)	(0.033)
In-Kind Gift Treatment	4.24		12.914		0.067	
	(2.317)		(10.154)		(0.103)	
In-Kind Gift, Early Delivery	6.628		25.355		0.232	
Treatment	(2.458)		(9.364)		(0.102)	
(Crossed) Employer Return Treatment						
Treatment w/ High	2.21	-0.712	6.43	-4.196	0.051	-0.029
Return to the Employer	(2.145)	(0.381)	(6.094)	(2.395)	(0.058)	(0.023)
Constant	1.415	4.068	-38.468	-25.498		
	(1.285)	(0.374)	(8.827)	(3.239)		
Hyp.: Gift Treatments = Control	p=0.001	p=0.000	p=0.011	p=0.000	p=0.051	p=0.000
R squared / Pseudo R Squared	0.193	0.104	0.032	0.035	0.039	0.061
Number of Subjects	300	1,954	300	1,954	300	1,954

Notes: Robust standard errors. Columns 6 and 7 report the marginal effects for the probit specification.

Table 3. Productivity Experiment, Social Preference Estimates

Estimation:				
Dependent Variable:	**Log (No. of Envelopes in a Batch)**		**Number of Envelopes in a Batch**	
	(1)	(2)	(3)	(4)
Baseline Social Preferences				
Altruism towards Employer (Charity)	-0.012		0.027	
	(0.038)		(0.034)	
Altruism towards Employer (Grocery Store)				
Warm Glow towards Employer (Charity)	0.461	0.443	0.422	0.462
	(0.074)	(0.063)	(0.071)	(0.065)
Warm Glow towards Employer (Grocery Store)	0.730	0.720	0.694	0.716
	(0.080)	(0.072)	(0.075)	(0.073)
Reciprocal Social Preferences				
Warm Glow Change -- Positive Monetary Gift		0.151		0.135
		(0.128)		(0.121)
Warm Glow Change -- Negative Gift		-0.042		-0.001
		(0.123)		(0.095)
Warm Glow Change -- In-Kind Gift		-0.095		-0.106
		(0.104)		(0.099)
Incidental Parameters				
Cost Function Curvature (γ)	9.518	9.44	0.258	0.263
	(0.841)	(0.737)	(0.019)	(0.018)
Cost of Effort Function:				
Std. Deviation of Error Term	0.130	0.144	3.952	4.318
Std. Dev. of Individual f.e.s * (1/γ)	0.014	0.241	8.160	8.014
R Squared	0.8374	0.7915	0.8532	0.8184
N	3568	4460	3568	4460

Notes: Specifications are from non-linear least squares regressions, with each observation being a worker-batch combination. The sample is restricted to the first 8 batches in Columns 1-3 and 5-7. The dependent variable is the log of the number of envelopes produced in that round in Columns 1-4 and is the the number of envelopes produced in Columns 5-8. The specifications in Columns 1 and 5 allow for pure altruism towards the employer, in which the worker puts weight alpha on the return to the employer. The specifications in Columns 2 and 6 allow for a form of warm glow, that is, the worker puts a weight on the employer, but on the *average* return (30 cents per envelope), not the actual return (which varies by round). The specifications in Columns 3 and 7 include both altruism and warm glow coefficients, except for the grocery store for which there is no variation in return and thus one cannot separate altruism from warm glow. The specifications in Columns 4 and 8 include also batches 9 and 10. All specifications include fixed effects for worker i as well as indicators for rounds 2, 3, 4, and 5-8; Columns 4 and 8 also include indicators for batches 9-10. The standard deviations listed are the standard deviation of the error term and the standard deviation of the individual fixed effects divided by the curvature *gamma*. The latter ratio indicates the variation in the individual productivity. The standard errors are clustered at the session level.

Table 4. Extra Work Experiments, Social Preference Estimates

Estimation:	**Maximum Likelihood, Accounting for Censoring**					
Experiment:	**Experiment 2**		**Experiment 3**			
Dependent Variable:	**Minutes Worked**	**Log (Minutes Worked)**	**Addresses Checked**		**Log (Addresses Checked)**	
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline Social Preferences						
Warm Glow towards Employer	0.814 [0.002,100]	0.807 [0.002,100]	0.106 (0.072)	0.097 (0.065)	0.106 (0.072)	0.097 (0.065)
Altruism Towards Employer	.	.	-0.007 (0.006)		-0.007 (0.006)	
Warm Glow Change - High Return for Employer	0.109 (0.086)	0.111 (0.089)		-0.025 (0.023)		-0.025 (0.023)
Reciprocal Social Preferences						
Warm Glow Change -- Monetary Gift	0.303 (0.139)	0.309 (0.142)	0.075 (0.028)	0.075 (0.028)	0.076 (0.029)	0.075 (0.028)
Warm Glow Change -- In-Kind Gift	0.181 (0.129)	0.181 (0.135)				
Warm Glow Change -- In-Kind Gift, Early	0.361 (0.142)	0.365 (0.145)				
Incidental Parameters						
Cost Function Curvature (γ)	0.007 [0,0.077]	1.049 [0,11.760]	0.051 (0.014)	0.051 (0.014)	2.504 (0.697)	2.504 (0.697)
Std. Deviation of Error Term	45.121 (3.928)	0.311 (0.028)	43.388 (2.331)	43.388 (2.331)	0.885 (0.047)	0.885 (0.047)
Cost of Effort Function:	Exponential	Power	Exponential		Power	
Log Likelihood	-637.99	-160.58	-2716.61	-2716.61	-1525.77	-1525.77
N	300	300	1954	1954	1954	1954

Notes: Maximum likelihood estimates for the number of minutes stayed (including the initial 120 required minutes) for Experiment 2 and the number of extra addresses check (including the required 40) for Experiment 3. Bootstrap standard deviations are in parentheses and 95% bootstrap confidence intervals are in brackets. The number of bootstrap draws is 1000. The upper bound of the baseline social preference parameter is set to be 100 in the estimation.