Estimating Social Preferences and Gift Exchange with a Piece-Rate Design*

Stefano DellaVigna
UC Berkeley and NBER

John A. List
U Chicago and NBER

Ulrike Malmendier
UC Berkeley and NBER

Gautam Rao
Harvard University and NBER

This version: June 2019

Abstract

We design two field experiments to estimate the nature and magnitude of workers' social preferences towards their employers. Unlike previous gift-exchange field experiments, we vary piece rates in addition to gift treatments. This *piece-rate design* allows us to estimate the elasticity of effort to motivation and in turn identify aspects of the workers' social preferences. The first experiment measures productivity—units of output produced in a fixed amount of time. The second experiment measures a form of labor supply—the willingness to work for extra time. Using the piece-rate treatments, we document that productivity is rather unresponsive to motivation, while labor supply is very responsive. In terms of social preferences, we document, first, that workers provide effort for their employer, but are insensitive to the return to the employer. This result is consistent with models of 'warm glow' or social norms, rather than pure altruism towards the employer. Second, while we do not detect any effect of the gifts in the productivity experiment, we find sizable positive impacts in the labor-supply experiment. We show that, at least in part, this different response to gifts is explained by different elasticities of productivity and labor supply, highlighting the importance of the piece-rate design.

*This paper replaces an earlier draft titled "Estimating Social Preferences and Gift Exchange At Work". 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, Avner Shlain, Jessica Shui, Ferenc Szucs, Jonas Tungodden, and Brian Wheaton for excellent research assistance.

1 Introduction

What motivates workers? Incentives, including career concerns, play a large role. Yet, many jobs do not involve piece rates or other outcome-contingent pay, for example because of difficulties in measuring output or in providing incentives to multi-tasking workers (Holmstrom and Milgrom, 1991). In such cases, workers may still work hard because they care about their contribution to the firm. 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 this pro-social motivation?

The literature points to examples of the role of workers' social preferences towards their employers. Kahneman, Knetsch, and Thaler (1986) provide survey-based evidence on the importance of fairness in wage setting. Krueger and Mas (2004) chronicle a striking case of negative reciprocity at work: a break-down in trust between employer and employees leads to employee retaliation and the production of defective goods leading to hundreds of lives lost.

Overall, however, there is little systematic field evidence about the nature of workers' social preferences towards employers. Do workers' social preferences take into account the employer payoffs, as in pure altruism models (Becker, 1974)? In this case, workers work harder when their effort is of higher value to the employer. Or is the right model one akin to warm glow (Andreoni, 1989, 1990) or models of 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 the Krueger and Mas (2004) findings?

Our paper showcases two field experiments designed to estimate the nature and shape of social preferences at work. We build on gift-exchange field experiments in the spirit of Gneezy and List (2006). In these experiments, employees are hired for a one-time task, to shut down repeated-game incentives. Employees in different treatments are then exposed to different employer actions, such as surprise pay raises (Gneezy and List, 2006), pay cuts (Kube, Marechal, and Puppe, 2013), and in-kind gifts (Kube, Marechal, and Puppe, 2012). The differences in productivity across the treatments provide evidence about gift exchange and reciprocity.

We build on these experiments, but closely link the design to a simple model of social preferences, so as to identify the underlying social preferences. We show that previous field experiments, while providing very valuable qualitative evidence on social preferences at work, 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 effort. Assume for example that an unexpected 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. Alternatively, it could reflect 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 value to the employer

of the worker's effort is unclear. In a model of pure altruism towards the employer, holding all else constant, employee effort increases in the value of the effort. In an alternative model, which we label 'warm glow' (broadly construed to include positive feelings from doing meaningful work, adhering to a social norm of working hard, or signalling prosociality), effort may instead not depend on the employer's actual utility. This distinction could not be tested in prior experiments, since the value of effort was typically unobserved and held constant.

We design two field experiments to address both issues. Like the previous field experiments on gift exchange, our experiments take the form of a one-time hiring opportunity for a clerical task, and involve gift treatments. Unlike the previous experiments, though, our experiments feature what we call a *piece-rate design*. In addition to the gift treatments, the design includes two piece-rate treatments with medium and high piece-rate, as well as a control treatment with no piece rate. Under some simple parameterizations for the cost of effort function, observing effort under three different piece rates allows for identification of the cost of effort function, thus addressing the first issue above. In addition, we address the second issue by informing the workers about the value of their work to the employer or by ensuring that there is a natural measure for the value.

In the first experiment, we measure productivity in a clerical task, as in the previous papers. Over six hours, workers prepare envelopes for multiple charities, and a grocery store. Unlike in previous papers, we adopt a hybrid within-between design. The within dimension is that every 20 minutes workers stuff a batch of envelopes for different employers, with different piece rates. A first charity pays \$7 for 20 minutes, a second charity pays a 20c piece rate per envelope, and a third charity pays a combination: \$3.50 plus a 10c piece rate. To limit any gift response, the schemes are designed to result in equivalent payoffs for a worker of average productivity (35 envelopes in 20 minutes). This piece-rate variation allows us to identify the cost of effort.

We inform the workers about the average per-envelope return to the charity (the employer) based on previous fund-raising returns. We also tell workers (truthfully) that some batches of envelopes raise money for a cause with a one-to-one fundraising match, and thus on average have twice the return to the employer. Further, in some batches—paid training sessions—the workers fold practice envelopes which are not used by the charity, and thus have no return to the employer. To control for order effects, we randomize workers into two different orders for the various batches.

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.[1]

The combination of the traditional between-subject gift treatments with the within-subject

[1]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.

variation in piece rate and return to the employer increases statistical power. Also with power in mind, we hired a large sample of workers, 446, making this, to our knowledge, the largest gift exchange field experiment so far.

We establish three key results from the within-subject part of the experiment. 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 to motivation of 0.1, a point we revisit later. Second, worker productivity increases by 10 percent when the envelopes are utilized by the charity compared to when 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 one-to-one donor match leads to a minimal (and not statistically significant) productivity increase of 1.6 percent. These findings are more consistent with a warm glow model than with pure altruism.

Turning to the between-subject gift treatments, we estimate no statistically significant impact for any of the gift treatments. These null effects are relatively precisely estimated. For example, we can reject that the in-kind gift increases output by more than 2 percent, compared to a 25 percent increase in Kube, Marechal, and Puppe (2012). We discuss the interpretation below.

This first 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 two limitations. First, the within-subject structure is less natural, with the potential to confuse some of the subjects. Second, the precision of the social preference estimates is limited by the fact that the outcome is so inelastic. Even sizable shifts in motivation, 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 designs estimate elasticities of 0.03 (Araujo et al., 2016 and Goerg, Kube, and Radbruch, forthcoming) and 0.04 (DellaVigna and Pope, 2018). This stresses the importance of the piece-rate design, which puts in perspective the observed effects (or lack thereof) on output.

For our second experiment, we would ideally like a simpler, between-subject design that is however as well-powered. To do so, though, requires a task that is significantly more elastic to motivation, and we are aware of no such tasks. Instead, building on the design of Abeler et al. (2011), we measure not how hard workers work in a fixed unit of time—productivity— but how *long* workers are willing to work extra—a form of labor supply.[2] We hire workers for a one-time 2-hour data coding job for $60. After the two hours are over and we paid for the work, 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 let them know that "*unfortunately, we cannot compensate you for this extra time.*" For the two piece-rate treatments, we inform them that "*we will pay you [¢25/ ¢50] for every minute*

[2]DellaVigna and Pope (2019) in an online sample similarly show that this extra-work margin is highly responsive to motivation. In an alternative approach, Goerg, Kube, and Radbruch (forthcoming) estimates an elasticity of 0.35 for a real-effort slider task if subjects are allowed to browse the internet during an experiment.

of work that you do, up to one hour." The workers are highly responsive along this labor supply margin: the average number of extra minutes worked increases from 2.2 in the control group to 12.3 in the medium-piece-rate group to 29.9 in the high-piece-rate group.

The other treatment arms are geared towards identifying the social preferences. In a *monetary gift* group, after the completion of the 2 hours of work and before asking about the extra work, we give the workers an additional $15 as a token of appreciation. In an *in-kind gift* group, we give them a thermos. Further, to study whether any gift exchange would be sufficiently long-lasting, in an *early in-kind gift* group, we gift workers a thermos before they start the 2 hours of work. All told, we have six treatment arms, and we randomize the 300 workers into each arm with equal probability. As a final design piece, within each arm we inform half of the workers that their work has especially high value to the employer before we ask them to do extra work.

The gift treatments raise the average labor supply by 5.6 minutes (monetary gift, s.e.=2.1), 4.2 minutes (in-kind gift, s.e.=2.3), and 6.5 minutes (in-kind early gift, s.e.=2.4). Thus not only is there a gift effect on labor supply, but the effect does not decay quickly over time. In contrast to these sizable gift effects, the qualitative statement on the high value of work to the employer has only a small and not statistically significant impact of 2.2 minutes (s.e.=2.1).

We compare the results of the two experiments. In both designs, experimental variation, whether quantitative or qualitative, in the perceived return to the employer does not seem to affect work-effort sizably (provided the work benefits the employer at all). This evidence fits most obviously with the model of motivation which we labeled 'warm glow'. We do find contrasting results instead on the impact of the gifts. The gifts increase productivity by less than 1 percent in the first experiment, but they increase labor supply by over 200 percent in the second experiment.

We show that this discrepancy has much to do with the vast difference in elasticities between the two tasks. When we translate the gift results into estimates of the reciprocity parameters, we cannot reject ($p = 0.372$) that the impact of monetary gifts on social preferences is the same in experiment 1 ($a_{Gift} = 0.15$, s.e. 0.13) and experiment 2 ($a_{Gift} = 0.30$, s.e. 0.14). The impact of the in-kind gift on social preferences 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 one would infer from 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 piece-rate design, and the structural estimation that it enables.

We see a methodological contribution of our paper as making the case for a *piece-rate design*. As simple as the design is, with the addition of piece-rate treatments, it is not common. None of the gift exchange in the field papers we are aware of employs such design, and only 2 out of 10 behavioral real-effort experiments (not just on gift exchange) published in top-5 journals from 1999 to 2018 employ it. It is also not common among field experiments on productivity, such as Breza, Kaur, and Shamdasani (2018). A piece-rate design enables the researcher to (i) estimate the underlying behavioral parameters and thus allow comparison across experiments with different tasks; (ii) to compute whether a null result is consistent with a sizable behavioral impact (because

the task is insensitive to piece rates), and conversely (iii) to show whether the size of a reduced-form effect (say, of a gift) is plausible. Regarding (iii), we show that some of the findings in previous gift-exchange field experiments imply very large reciprocity effects, assuming an effort elasticity like the one we estimate and the ones in the literature. We recognize that the estimates of the cost-of-effort elasticity obtained under a piece-rate design depend on a set of assumptions, such as the separability of the cost of effort from the rest of the utility function; for example, potential crowd-out of intrinsic motivation with financial incentives in signaling models (Benabou and Tirole, 2006) would bias elasticity estimates downwards.

The piece-rate design allows for estimation of the underlying preferences, as in the growing literature on *structural behavioral economics* (e.g., Laibson, Maxted, Repetto, and Tobacman, 2015; Conlin, O'Donoghue, and Vogelsang, 2007; DellaVigna, List and Malmendier, 2012; DellaVigna, List, Malmendier, and Rao, 2017), reviewed in DellaVigna (2018).

This paper relates to the literature on social preferences at work, providing evidence on workers' *vertical* social preferences towards their employers (e.g., 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). Within the literature on gift exchange in the field, reviewed in Online Appendix Table 1, the main contributions of the paper are the piece-rate design, which allows for estimation of preferences, and the labor supply design in Experiment 2.

This field-experimental literature itself 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 methodologically build a bridge towards this lab design by estimating the cost of effort and specifying the impact on the employer payoff.

2 A Simple Model

We present a simple model of worker effort encompassing both experiments. In experiment 1 we interpret effort as productivity, while in Experiment 2 we interpret it as labor supply. We abstract from specific features, such as the different batches of work in Experiment 1, and return to these features in the structural estimation.

Worker i chooses optimal effort e_i as a function of pay incentives and cost of effort. In addition, the worker has social preferences towards the employer and thus cares about the employer's payoff. 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 piece 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 $C'() > 0$, $C''() > 0$, and $\lim_{e\to\infty} C'(e) = \infty$, guaranteeing the existence of a unique solution.

The third component captures the social preferences: the worker cares about the payoff 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 the ones in Bellemare and Shearer (2011) and Englmeier and Leider (2012c), among others. Maximization problem (1) yields the first-order condition

$$\begin{aligned} p_W + A(Gift, p_E, p_W) - C'(e_i^*) &= 0 \text{ or} && (2) \\ e_i^*(Gift, p_E, p_W) &= C'^{-1}(p_W + A(Gift, p_E, p_W)). && (3) \end{aligned}$$

where $C'^{-1}()$ is the inverse function of $C'()$, which exists and is monotonically increasing by the assumptions above. The second order conditions are satisfied since $-C''(e^*) < 0$. In the following we will assume an interior solution.[3] The optimal effort e^* is increasing in the social preference parameter A and in the piece rate p_W (provided A does not decrease enough in p_W).

Pure Altruism. Under this first interpretation, the worker takes into account the employer's actual *net* return from effort, $p_E - p_W$, and thus $A = \alpha(p_E - p_W)e_i$. 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 the employer. Thus, the social preference term is $A = (\alpha + 1_{Gift}\alpha_{Gift})$.

Warm Glow. Under a second interpretation, the worker simply derives utility from doing his part by exerting effort for his employer, regardless of how the effort translates into payoffs for the employer; thus, $A = a\overline{p_E}$, where $\overline{p_E}$ is a measure of the average return to the employer, and serves purely a normalization purpose. This case is inspired by the idea of warm glow in Andreoni (1989, 1990), where donors derive utility from giving, but not necessarily from the public good itself.[4] This specification also captures, in reduced form, a social norm to put in effort for an employer, signalling (Benabou and Tirole, 2006), or a utility from exerting effort doing meaningful work (Ariely, Kamenica, and Prelec, 2008).[5] As before, we allow for warm glow to change as a result of receiving an unanticipated gift, so $A = (a + a_{Gift}1_{Gift})\overline{p_E}$.

Estimation from Standard Gift Exchange Experiment. Consider a gift exchange experiment à la Gneezy and List (2006). For the combined altruism and warm glow case, taking into account that in these experiments there is no piece rate, the optimal efforts are:

$$e_{Contr}^* = C'^{-1}(\alpha p_E + a\overline{p_E}) \quad \text{and} \qquad (4)$$

[3] A sufficient condition to ensure an interior solution is $C'(0) = 0$ 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.

[4] 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.

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

$$e^*_{Gift} \quad = \quad C'^{-1}\left(\left(\alpha + \alpha_{Gift}\right) p_E + \left(a + a_{Gift}\right)\overline{p_E}\right).$$

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

It is helpful to consider the case of a power cost function $c(e) = ke^{1+\gamma}/(1+\gamma)$, which has a constant elasticity $1/\gamma$ with respect to the return to effort.[6] The solutions under pure altruism are:

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

By dividing through and inverting, we obtain

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

While we cannot back out the altruism parameters without knowledge of the return p_E, we can infer the increase in altruism $(\alpha + \alpha_{Gift})/\alpha$, *provided* one knows the curvature γ. In the quadratic cost-of-effort case ($\gamma = 1$), an x percent increase in 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. A function with this feature is the exponential cost function, $C(e) = k\exp(\gamma e)/\gamma$.[7] In this case, the solutions are

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

We can transform the solution and divide through to obtain

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

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 it is not the case: the x unit increase for the positive gift would require a larger proportional change in altruism (positive reciprocity) compared to the corresponding change in altruism (negative reciprocity) for

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

[7] 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 $\eta_{e,v} = (1/\gamma v) * v/((1/\gamma)\log(v/k)) = 1/\log(v/k)$.

the negative gift. Intuitively, it is harder to increase effort at the margin than to reduce it.

Estimation from 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 the subjects of the return p_E, provided the task allows for it. Identifying the cost of effort, though, requires additional treatments. Unlike the standard gift exchange experiments which do not involve a piece rate, it is useful to experimentally vary the piece rate p_W to identify the cost of effort function. 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 piece rate p_W helps pin down the cost of effort function. Notice, however, that 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 the worker 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 piece rate incentives.[8]

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 piece rate p_W identifies the altruism α :

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

3 Productivity Experiment

3.1 Design and Data

Design Idea. Our 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.

In other ways, though, our design differs from previous experiments. First, we implement a *piece-rate design* by varying the piece rate p_W offered to the workers. Second, we make explicit,

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

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-subjects* unexpected 'gifts' from the employer.

This design idea informed 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, 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, since the workers work for multiple charities, we set different piece rates p_W for each charity for a 20-minute batch: ($7 fixed pay, no piece rate), ($3.5, 10c piece rate), and (no fixed pay, 20c piece rate). The lump-sum pay is set to minimize the gift effects from such piece-rate variation, since the earnings are constant for a person of average productivity (about 35 envelopes per 20 minutes). Finally, to check whether there was anything special about having a charity as employer, one of the batches involved stuffing advertisement mailers for a firm (a grocery store).[9]

Design Details. We hire temporary workers with a Craigslist ad for for a one-time 6-hour job (in 1 day) on Saturdays and Sundays on the University of Chicago campus.[10] The participants are taken to a classroom where a research assistant explains the work following a script. The Becker Center at the 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 their way 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 (*Respond Now*, *Breakthrough Urban Ministries*, and the *Rehabilitation Institute of Chicago*) for 8 batches and an advertising campaign for a local grocery store for 2 batches.

Figure 1 shows Order A, the first of two orders which we randomize. In the first four batches, the participants fold envelopes at the 10-cent piece rate (and $3.50 flat pay), but 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.[11]

[9]The grocery store is not our main employer because we could not find a compelling way to vary the return to the employer.

[10]A typical ad read: '*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.

[11]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

In batches 2 and 3, the 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: '*Thanks to an anonymous donor, [Charity Name] has received a matching grant that will match every dollar raised by these letters 1 to 1 up to $2,000 total. A number of such matching grant campaigns have been run by charities similar to [Charity Name], and 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.[12] In batch 4, the 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.

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.[13] 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, which had to precede the other treatments: batch 8 in order A becomes batch 2 in order B, and so on. While the training sessions remain in rounds 1 and 5, we do switch the pay schemes in the training period between order A and B. The randomization of the order aims to disentangle the treatment effects from confounding effects due to learning and tiredness over the day.[14]

The arrows illustrate four planned experimental comparisons.[15] First, comparing batches 6 and 7 in order A (3 and 4 in order B) illustrates the impact of the piece rate change from 0 cent to 20 cents. There is an 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. Second, comparing batches 5 and 6 provides an estimate of the effect of whether the effort counts at all for the employer, or not (the training). Third, comparing batches 7 and 8 provides evidence on the impact of the quantitative return to the employer. Fourth, comparing batches 3 and 4 in order A (6 and 7 in order B) identify the difference between working for a grocery store versus for a charity.

assume that the Becker Center, which pays for the training batch, does not enter the social preferences of the workers.

[12]The script says: '*A number of such campaigns have been run by charities similar to [Charity Name], and historically, these charities have yielded roughly $0.30 per mailer with such campaigns. Taking account of [Charity Name]'s per-envelope payment for your help today, it expects to get roughly $0.20 for each additional envelope that you prepare*'.

[13]This second training session is justified to the workers by the (slight) difference in materials for the different employer, as well as by restarting after a break.

[14]We did not do a full randomization of the order of the treatments for two reasons. First, the implementation would have been difficult. Second, the choice of just two orders allowed us to maximize power by placing next to each other treatments we intended to compare, like match and no-match (for the same piece rate), thus minimizing the confound of productivity changes over time.

[15]In the pre-registration we emphasize equally comparisons taking place in batches 1-4 and in batches 5-10. For example, the effect of training can also be estimated comparing batches 1 and 2. However, the steep learning by doing oberved in batches 1-4 confounds these comparisons. Thus, we focus on batches 5-10 for the reduced form results. The structural estimates use all the variation in the data, including the early periods.

After batch 8, we randomize the workers in a session into the gift treatments, and send the workers into separate rooms.[16] In the Control treatment, we say that "*in this session and the next, [Charity name] will pay $7 just as it paid in a previous session.*" In the Positive Monetary Gift, we say that "*in this session and the next, [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 "*in this session and the next, [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 "*in this session and the next, [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 with the name of the charity to make clear the gift is coming from the employer. This treatment is modelled on Kube, Marechal, and Puppe (2012). 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. Participants assigned to different treatments are walked to diffrent exits, minimizing the chance of a meeting.

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 then we 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. Between October 2013 and January 2014 we ran 24 sessions with 131 workers.[17] We then stopped 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.[18] After excluding 4 subjects who left the experiment early, the final sample includes 446 workers, the largest sample size that we are aware of among field experiments on gift exchange.

The sample (Column 1 of Online Appendix Table 2) is 52 percent female, covers a wide age range, and overrepresents unemployed individuals.[19] Column 2 shows that productivity is higher 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

[16] We tell the workers '*we will now have to split into a few rooms because our room reservation has expired.*'

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

[18] The sessions are 73 because in one of the sessions one of the letter materials was shown incorrectly, and the RA opted to repeat the session. In the spirit of intent-to-treat and transparency, we also retain this session.

[19] Online Appendix Figure 5 show that our main results are very similar for the subsample of employed individuals.

productivity), with no evidence of 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

Before we turn to the planned comparisons, in Online Appendix Figure 1 we plot average output by batch for the two orders. The confidence intervals, as elsewhere for Experiment 1, are clustered at the session level to account for correlation within a worker over time and across workers in a session.[20] There is substantial learning by doing over the first 4 batches, but from batch 5 onwards, after the lunch break, there are no more obvious gains (nor losses) in productivity. Previewing the key findings, we 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 (20c to 0c) and 8 to 9 in order B (10c to 0c). The figure also shows a response to whether envelopes are used or discarded (holding constant the piece rate, batches 5 and 6), with instead only a very small impact of changes in employer return (batches 7 and 8).[21]

Piece-Rate Response. As Figure 2a shows, increasing the piece rate from 0c to 20c leads to an increase of 4 envelopes (12 percent), a significant difference, if corresponding to a small elasticity. Importantly, the piece-rate effect is not confounded by an income effect, since the flat pay is proportionally lower for the higher piece rate. Figure 2a also provides a comparison to the 10c piece rate, though the batches being compared are not contiguous in this case.

Baseline Social Preferences. Figure 2b presents the impact of the employer return: do workers work more when the return is higher due to a match on the resulting donations? The answer is largely no: the higher-return batches have a (statistically insignificant) 0.6 additional envelopes (1.7 percent), an effect size much smaller than the impact of piece rate. This finding argues 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 worker piece rate is held constant. Productivity is 3.5 envelopes (10 percent) higher when the letters are used, a sizable and statistically significant difference.[22]

Turning to additional evidence, Online Appendix Figure 2a shows that the effort provided does not differ sizably between the charities; thus we pool across the charities. Online Appendix Figure 2b shows that the effort is actually somewhat *higher* when the employer is a firm (a grocery store)

[20] Clustering at the session level produces very similar results to clustering at the individual level, indicating that the within-session correlation of errors is small.

[21] The pattern across rounds and between orders in general lines up well with what one would expect. One exception is the comparison across order A and B in rounds 1 and 4: the productivity is higher in order A despite lower incentives than in order B (10c versus 20c). An imbalance in worker ability does not appear to account for much of this difference, as we discussed above. Furthermore, in the only two rounds in which the two orders have the same treatments and are thus comparable, rounds 9 and 10, the productivity is very similar across the orders. Instead, it appears that workers work harder for a firm (the grocery) than for a charity.

[22] The difference comparing batches 1 and 2, but improvements in productivity over time bias that comparison upward. The structural model utilizes all the rounds and controls for learning across rounds.

versus when it is a charity, holding constant the return to the employer and the piece rate. This suggests that the substantial baseline social preferences we identify towards charities as employers may not over-estimate the social preferences towards employers.

Gift Exchange. Figure 2d shows the 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.

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, without essentially changing the point estimates. 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 increases in Gneezy and List (2006) and in Kube, Marechal, and Puppe (2012). Columns 4-6 show that the results are similar using log output as a measure, the specification implied by a power cost function.

In Online Appendix Figure 3 and Online Appendix Table 3, we also consider the impact of decay of any gift effect, comparing batches 9 and 10. We detect no discernible pattern for the negative gift and the in-kind gift treatments, with suggestive evidence of a gift effect in batch 9 for the positive monetary gift treatment. We also consider how the gift interacts with the return to the employer, comparing a batch with matching donation to a batch without, and detect no interaction. Finally, we consider the heterogeneity in the gift impact. While the differences between treatments are small, there are hints of an output increase with the positive gift at lower quantiles.

As additional robustness, we show that all the key results are parallel using log output (Online Appendix Figure 4), the dependent variable implied by a power cost of effort, as well as restricting to participants who report being employed (Online Appendix Figure 5).

4 Labor Supply Experiment

4.1 Design and Data

Design Idea. The productivity 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 two limitations. First, the within-subject structure is less natural, with the potential to confuse some subjects. Second, the precision of the social preference estimates is limited by the fact that the task is so inelastic. Even sizable shifts in motivation, say due to a gift, would result in fairly small impacts on effort, 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 tasks: previous papers with piece-rate designs estimate elasticities of 0.03 (Araujo et al., 2016 and Goerg, Kube, and Radbruch, forthcoming) and 0.04 (DellaVigna and Pope, 2018).

Our second experiment has a complementary design which takes into account the above limitations. We would ideally like a simpler, between-subject design that is at least as well-powered as our first design. To do so, though, requires a task that is significantly more elastic to effort, and we are aware of no such tasks. Instead, building on the design of Abeler et al. (2011) and Gneezy et al. (2017), we measure not how hard workers work in a fixed unit of time—productivity—, but how *long* workers are willing to work extra—a form of labor supply. In a purely between-subject design, we alter the incentives to provide extra work, and vary the prosocial motivation to do so, and examine how much extra work people supply as the key outcome.

Design Details. Figure 3 shows the design. 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 bring them 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 are over, the research assistant checks back into the offices, pays the 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 to one of 5 conditions. For workers in a control arm, we add: "*Unfortunately, we cannot compensate you for this extra time.*" For the two piece-rate treatments, we inform them 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.*"[23]

The other three treatment arms are 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.

One may be concerned that asking subjects to work longer hours after they were just given a gift may set up a quid pro quo that the subjects may find hard to avoid. We thus 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 check on this, and to study the persistence of any gift effect, a final group, the *early in-kind gift* group received a thermos before the workers started the 2 hours of coding,

[23] We considered having the high piece rate set to overtime (1.5) pay, but individuals staying the full 60 minutes would then have hit the $100 pay that requires IRS reporting, requiring onerous paperwork and creating a likely discontinuity, and possibly non-monotonicity, in the pay.

with the same exact wording as for the *in-kind gift* group.

All told, we have six treatment arms, and we randomize the workers into each arm with equal probability. As a final design piece, we cross-randomize (also between subjects) a statement on the value of work to the employer. Right before asking for extra work, we tell the *high-return* group (but not the *normal-return* group) "*Getting the extra data entered today is really valuable to us.*"

Data. In September 2018 we pre-registered this design, including a power calculation done 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.

The 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 more extra minutes (Column 2). The treatments are balanced on the observables (Columns 3-8).

4.2 Experimental Findings

We present the findings for the number of minutes of extra work, which we pre-specified to be the key variable of interest, but also present additional evidence on the share of workers who agreed to stay extra minutes, as well as on the output produced by the workers.

Piece-Rate Response. Figure 4a documents a very sharp response to the piece rates. The average number of extra minutes worked increases from 2.5 (control) to 15.6 (medium-piece-rate) and 29.9 (high-piece-rate). Not only are these differences highly statistically significant, but they correspond to a highly elastic response: a doubling of the piece rate from 25c to 50c is associated with more than double the number of minutes, indicating an elasticity over 1. We are not aware of any real-effort experiment with productivity as outcome variable that even approaches such an elastic response. In our view, this is a major advantage of this design.

Baseline Social Preferences. In Figure 4b we consider the response to the qualitative manipulation of the return to the employer, the statement "*Getting the extra data entered today is really valuable to us.*" Since this manipulation is cross-randomized, we pool across all six gift and piece-rate arms and compare the workers receiving the high-return statement to the workers who did not. We estimate a small and not statistically significant increase of 2.5 minutes.

Gift Exchange. Figure 4c compares the number of minutes worked in the control group to the minutes worked in the three gift groups. 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-piece-rate treatment. The effect is similar for the monetary gift and the early in-kind gift, and slight 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 time.

Robustness. Figure 5a presents the c.d.f. of the number of extra minutes worked in the 6 main arms. The medium-piece-rate group, compared to the control group, leads to an increase in the share staying extra time (the extensive margin), but even more in the number of extra minutes among those staying longer (the intensive margin). The high-piece-rate treatment is associated

with a large increase along both margins, doubling the share staying for any extra time, with over 30% of workers staying the full 60 minutes. 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, with half of them staying no more than 10 minutes. Thus, the gift triggers extra effort, but in modest numbers.

In Table 2, we present regression results. The benchmark specification is an OLS regression of the number of extra minutes stayed in Column 1, including the high-return condition in Column 2. Column 3 adds controls for gender, fixed effects for age groups, and a fixed effect for the Craigslist sample. The point estimates with controls are similar to the ones without controls, with slightly larger point estimates for the piece-rate and monetary gift treatments. In Columns 4 and 5 we present the results from tobit regressions modeling the censoring at both 0 and 60 minutes. The results are similar, if somewhat less precisely estimated. Then, in Columns 6 and 7 we present marginal effects from probit regressions on an indicator for staying extra minutes. This dichotomous variable, not surprisingly, provides noisier evidence, though the gift effects are still jointly significant (p=0.051 with no controls and p=0.028 with controls).

Online Appendix Figure 6 presents the results for productivity, displaying the number of lines of data coded in the extra minutes, as fraction of the lines coded in the required first 2 hours. (The fraction is 0 for workers who do not stay extra). The results are parallel to the results for extra minutes stayed. Online Appendix Figures 7 and 8 split the key findings for the subsample recruited through Craigslist versus on campus. In both samples, very few subjects stay extra in the control group. The Craigslist sample, though, is more sensitive to both the piece rate treatments and to the gift treatments than the student sample.

5 Structural Estimates

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:

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

The first term in (10) is the cost of effort function $c(e)$, which we consider in two families, power, as used in some previous literature, and exponential. The power cost 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, the cost function is $c(e) = \exp(\gamma e_{i,t})/\gamma$ with $\gamma > 0$. Both cost functions satisfy the desired properties $C'(e) > 0$, $C''(e) > 0$, and $\lim_{e\to\infty} C'(e) = \infty$.[24]

[24]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 in Experiment 1 the lowest effort observed in any round is 7 envelopes and in Experiment 2 we include in the effort the required initial 120 minutes.

The second term, η_i, introduces the heterogeneity as a multiplicative factor on the cost of effort function. The term η_i captures both the impact of observables, X_i, as well as unobservables (a noise term). Importantly, η_i ought to be positive, as a negative draw implies a negative cost of effort. A log-normal distribution for η_i, with a mean that is potentially a function of the observables, satisfies these properties. Specifically, we assume $ln(\eta_i) \sim N(k(X_i), \gamma^2\sigma^2)$.

Given these assumptions, for a power cost of effort function 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 out 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 piece rate p_W) and the pro-social component A. This term also makes clear that $1/\gamma$ is the elasticity of effort to motivation. The second term captures 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, assuming an exponential cost of effort function 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 cost function thus leads to the same specification, except with effort e_i, rather than log effort $\log(e_i)$, as outcome. This simple structural model thus micro-founds reduced-form specifications with outcome variable effort (under exponential cost) or log effort (under power cost).

Productivity Experiment. While this set-up is common to both experiments, there are some differences, especially in the controls $k(X_i)$. In the first experiment, we measure effort for worker i 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.[25] 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 $f(t)$, 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

[25] 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.

gift effects are not biased by some change in the cost of effort in the last rounds.[26] 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 variables. We estimate equation (13) with a non-linear least squares regression, clustering the standard errors by session.

Labor Supply Experiment. In the second experiment we measure effort as the total number of minutes worked, inclusive of the required initial 120 minutes, since those initial minutes presumably contribute to the tiredness. So an individual who stays 30 minutes is coded as exerting effort $e_i = 150$. We assume that the value to the employer p_E of each additional minute is the per-minute cost of an additional worker, that is 50 cents (since we pay \$60 for 2 hours of work).

We model the censoring at $e_i = 120$ (workers who opt not to stay) and at $e_i = 180$ (workers who stay for the full 60 extra minutes) and estimate the model by maximum likelihood.

5.2 Estimates in the Productivity Experiment

Baseline Social Preferences. We first estimate the model using data from batches 1-8, excluding the gift treatments. Column 1 in Table 3 shows the results for the pure altruism case, in which case the social preference weight A equals $\alpha(p_E - p_W)$. The estimated altruism $\hat{\alpha}$ towards the charities as employers is 0.23 (s.e. 0.04), and higher for the grocery store.[27] The curvature of the cost function ($\gamma = 11.1$) implies a low elasticity of effort to the value of work of $1/11.1 = .09$.

To highlight the identification of α, recall 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 in response to a 20-cent increase in the piece rate (that is, $\partial e^*/\partial p_W \simeq 20$). This would imply $\alpha \simeq 0.11$. The estimated α is larger because the degree of altruism is also identified by the effect of training. Moving from the 20-cent training (batch 5 in order A) to a standard 20-cent period (batch 6 in order A), log output should increase by

$$\log(e_{t=6}) - \log(e_{t=5}) = \frac{1}{\gamma}\left[\log\left(.2 + \alpha\left(.3 - .2\right)\right) - \log\left(.2\right)\right].$$

[26] 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.

[27] The difference in α between the charity and the firm may appear surprisingly large. One relevant factor is the fact that we do not run a condition with high return (60 cents) for the grocery store, while we do for the charities. Since the workers do not respond to the higher return, the estimated altruism for the charities is compressed downward, while this is not the case for the firm. Indeed, the difference in social preference weights is much smaller in the warm glow case. As a second factor, given how inelastic the production function is, it takes a large increase in social preference to match even a quite small difference in output.

For $\alpha \simeq .1$, the expression equals approximately $\gamma^{-1}\left[\log\left(1.05\right)\right] \simeq .005$ log points, while the observed response is an order of magnitude larger, 0.1 log points (3.5 envelopes). The altruism model struggles to capture both the response to match and to training, as we further highlight below.

Column 2 shows the warm glow case, with a social preference weight $A = a\overline{p_E}$, where $\overline{p_E}$ is a normalization that denotes the *average* return to the employer, which we take to be 0.3.[28] The estimated curvature γ of the cost of effort function is similar to the one with altruism. The workers exhibit significant warm glow ($\hat{a} = 0.44$, s.e. 0.06), putting weight on the employer equal to about half the weight put on private payoffs. Recall that this model does not predict any response to the charity match, consistent with the data. It predicts a response to training of

$$\log\left(e_{t=6}\right) - \log\left(e_{t=5}\right) = \frac{1}{\gamma}\left[\log\left(.2 + a * 0.3\right) - \log\left(.2\right)\right].$$

For the estimated $\hat{a} \simeq .4$, the right hand side equals $\gamma^{-1}\log\left(1.6\right) = .05$, a predicted 5 percent increase in output from batch 5 to 6, quite close to the observed effect.

We nest the two models in Column 3. The results are striking: the data does not reject the null of no altruism, while the warm glow weights are very similar to the ones in Column 2.

In Columns 5-7 we present the results assuming an exponential cost of effort function, and thus with the number of envelopes prepared, as opposed to the log, as dependent variable. Encouragingly, the results are nearly identical to the results with power cost of effort.

In Figures 2a-c we compare the observed patterns with the predictions of the altruism and warm glow models for the exponential cost specification (Columns 4-5 of Table 3). While both models fit well the response to incentives (Figure 2a), the altruism model has trouble fitting the combination of a small response to the match rate (Figure 2b) and a large response to the training (Figure 2c). The warm glow model instead predicts no response to the match rate by assumption and matches the training effect with a higher level of warm glow a.

In Online Appendix Table 5, we present a set of robustness checks to the specification with warm glow and altruism. The estimates are similar if we model the learning by doing, instead of set of indicators, with quadratic or cubic polynomials. Next, we allow for the altruism model to include only the return to the employer, not taking into account that the piece rate detracts from the employer return, that is, we assume $A = \alpha p_E e_i$. The results are again similar. Finally, instead of assuming that during training periods the workers have no social preferences (since the letters are not used), we allow for warm glow, restricted to be half the size as in the periods where output is used, thus $A = (1/2) * .3e$. The estimated warm glow is now higher and there is some evidence of pure altruism, although the altruism coefficients are smaller than the warm glow ones.

Gift Treatments. In Columns 4 and 8 of Table 3, we use the full data set (batches 1 to 10) and estimate how much the social preference parameters shift (if at all) in response 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

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

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 in Columns 3 and 7, we assume that social preferences are of the warm-glow type. The estimates of baseline social preferences are, not surprisingly, parallel to the estimates using only the first 8 batches. Turning to the reciprocity parameters, we do not find statistically significant evidence of reciprocity for any of the gift treatments. The magnitudes, however, differ across the types of gift, in ways consistent with the reduced-form results: a decrease of social preferences of -0.04 (s.e. 0.12) for the negative gift treatment, as well as for the in-kind gift treatment, -0.09 (s.e. 0.10). For the positive monetary treatments, there is a more positive effect of 0.15 (s.e. 0.13). We return to evaluating these magnitudes below.

In Online Appendix Table 6, we consider a series of robustness checks. First, we consider an 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.[29] Next, we allow for the warm glow effect of a gift, a_{Gift}, to decay to δa_{Gift} in round 10 and estimate substantial decay ($\hat{\delta} = .23$) and a larger and statistically significant effect of the positive monetary gift ($\hat{a}_{Gift} = .37$), consistent with the reduced-form evidence suggestive of some decay. 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 Labor Supply Experiment

Table 4 presents the estimates of the model for the labor supply experiment. Given the evidence supporting the warm glow model, and given that this experiment was not designed to separate altruism from warm glow, we focus on the warm glow model.

In the labor supply, the identification of baseline social preferences is less the focus of the design. Unlike in the productivity experiment, we do not compare effort when the employer benefits to effort in a training condition, which is part of the identification for the social preferences. Still, the baseline warm glow parameter a is in principle identified with the piece rate variation. Yet, Table 4 shows a very wide confidence interval for a. Figure 4a shows the reason for such imprecise identification. The increase in labor supply from the medium-piece rate (25c/minute) to the high piece rate (50c/minute) is about the same size as from the control group (0c/minute) to the low piece rate (25c/minute). Yet, the model predicts that effort should be a concave function of the piece rate, as equations (12) and (11) show. Given that effort instead is about linear in the piece rate (for the three piece rates we considered), the point estimate for a is quite high and we cannot reject the upper bound, as a high value of a contributes to a less concave predicted pattern.

The imprecise identification for the baseline warm-glow parameter a, though, does *not* affect

[29] Online Appendix Figure 9 presents the estimated time path under a quadratic or cubic polynomial, comparing to the benchmark specification. In Online Appendix

Figure 10, 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. The model overall does well in fitting the data.

the identification of the other social preference parameters. 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 sizably, 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 similar if we add controls (Column 2), or estimate the model with log of the number of minutes worked, as implied by the power cost of effort function (Columns 3 and 4).

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 piece-rate design, that it "prices out" the parameter connected to a particular treatment comparison. As Figures 4a and 4c show, the impact on labor supply of the gift treatments, compared to the control, is about half the size of the impact of the medium-piece-rate treatment (25c/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-piece rate, again identifying the underlying social preference.

In Figures 4a-c and in Figure 5b we display the fit of the estimated model in Column 1 of Table 4. Figures 4a-c shows that the estimates fit quite well the observed patterns, but overfit somewhat the number of minutes worked in the control group and in the gift groups. A comparison of Figures 5a and 5b, displaying the full distribution of effort, shows why. In the control group the model slightly underestimates the share of people staying over time, but at the same time it predicts that some participants would stay for longer than we observe. Similarly, for the gift groups the model underpredicts the share staying over time, but overpredicts the share staying more than, say, 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 7 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 a 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: the share staying 0 minutes, 1-29 minutes, and the share staying 60 minutes. For all these three estimators, the estimates for the reciprocity parameters are similar to the benchmark ones.

5.4 Comparison of Estimates in the Two Experiments

The results from the two experiments lead to some common themes. Baseline social preferences are substantial (even though in the labor-supply experiment they are estimated imprecisely). Further, the baseline social preferences do not appear to respond to information, either quantitative or qualitative, on the return to the employer (provided this return is positive), consistent with a "warm glow" model, as opposed to pure altruism.

At the same time, the gift treatments lead to divergent findings: there is no detectable impact in the productivity experiment, but a sizable and statistically significant impact in the labor supply

experiment. More precisely, in Figure 6a we compare the reduced-form gift results across the two experiments in terms of percent effects for the two gift treatments that are in both experiments: the monetary gift and the in-kind gift. Relative to the control group, the monetary gift increases productivity by about 1 percent, while it increases the extra labor supply by over 200 percent. Similarly, there is a wide difference in the effect of the in-kind gift treatments. Online Appendix Figure 11 displays the same results measuring the effect sizes in in standard deviations units.

In Figure 6b we make the same comparison, but in terms of the underlying reciprocity parameters. As we discussed, the reciprocity parameters a_{Gift} are shifts in the warm glow baseline social preferences parameter, which becomes $a + a_{Gift}$ in response to the gift. When seen in terms of these structural parameters, the differences in findings are nowhere near as large. In response to the monetary gift, we estimate $a_{Gift} = 0.15$ (s.e. 0.13) in the productivity experiment and $a_{Gift} = 0.30$ (s.e. 0.14) in the labor supply experiment, and we cannot reject that the two structural parameters are the same ($p = 0.372$). For the in-kind gift, we estimate $a_{Gift} = -0.09$ (s.e. 0.10) in the productivity experiment and $a_{Gift} = 0.18$ (s.e. 0.14) and $a_{EarlyGift} = 0.36$ (s.e. 0.15) in the labor supply experiment. In this case, the difference is statistically significant when comparing 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 the piece-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 labor supply 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 Piece-Rate Design and Structural Estimates

The structural estimates allow us to compare results across our two 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. In one example, in the pre-analysis plan for the labor supply experiment 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 likely comparable to the one in our task. Under this assumption, as Online Appendix Table 8 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 esti-

mates are certainly, to some extent, context specific, as a calibration exercise in Online Appendix Figure 12a we consider how employee effort would respond to piece rate variation (on the x axis) under the assumption of no social preferences ($a = 0$) versus for the estimates of baseline social preferences from the productivity experiment ($\hat{a} = .4$). Under no social preferences, introducing a piece rate has dramatic impacts on output; in the presence of warm glow, instead, effort is quite high even with no piece rate and it increases slowly with the piece rate increases. Given this, the optimal piece rate for firm profits (Online Appendix Figure 12b) is positive (3 cents) without social preferences, but is instead zero under our estimates for warm glow, as the productivity increases for a higher piece rate do not pay off enough. 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. A first finding is that, in both experiments, worker effort does *not* respond to the return to the employer. This is true when the precise quantitative return is specified and varied, as in the productivity experiment, and also when the high return is indicated using a qualitative statement, as in the labor-supply experiment. Thus, workers' social preferences towards the employer do not appear consistent with a model of pure, Beckerian altruism towards the employer. There is evidence, however – particularly from the productivity experiment – of *some* form of prosocial motivation. This is evidenced both by positive effort in the absence of piece-rates, and by the drop in effort when the work output is not used (in the training rounds).

We interpret these findings in a light of a model, which we term 'warm glow', which 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 typically used to model workplace social preferences (e.g. Bandiera, Barankay, and Rasul, 2005), 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 the first experiment (and imprecisely estimated in the second experiment). 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 of importance in the workplace is a classical one, but little previous field evidence pins down the features of baseline social preferences towards the employer.

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 the subjects in these experiments work harder in response to these social incentives, the return to the charity does not affect the worker effort.

Gift Exchange. A second finding is that we find little evidence of gift exchange in the productivity experiment, but substantial evidence of gift exchange in the labor-supply experiment. As we discussed above, we can explain part of the difference in the reduced-form results with the

vast difference in elasticity of the outcome to motivation across the experiments. In particular, the much more elastic labor-supply margin makes it a more powered design to detect the impact of gift exchange. We find it intriguing that in the labor-supply experiment we replicate the early finding in the literature of a positive impact of monetary gifts (e.g., Gneezy and List, 2006), which most 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 labor-supply design employed in our paper, which could further our understanding of social preferences at work.

This still leaves us with a partial puzzle for the lack of gift effects in our productivity design, in particular for non-monetary gifts, for which we can reject that the effect is the same in the two experiments. While we cannot answer this question definitively, we consider a number of possibilities. 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.[30] 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 labor-supply experiment: "*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, in a way that is not as clear in the productivity experiment. Ultimately, we leave it to future literature to conclusively address these possibilities.

Piece-rate design. Finally, this paper makes the case for a *piece-rate design*, which is relatively uncommon in the behavioral literature. We are not aware of papers on gift exchange in the field that employ such design (e.g., papers in Online Appendix Table 1). More broadly, among all real-effort experiments published published in top-5 journals from 1999 to 2018, only 2 out of 10 papers employ this design (Online Appendix Table 9). Field experiments studying productivity at work, such as Breza, Kaur, and Shamdasani (2018), also do not typically employ it.

A piece-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 comparison across experiments with different tasks easier. Moreover, estimating the elasticity of different outcomes will help future research to employ more highly-powered experimental designs, allowing for more precise and cost-effective inference.

[30] For the last 65 workers in the productivity experiment, 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 13 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, as expected.

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.

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

Andreoni, James. 1990. Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. *The Economic Journal*, 100 (401), 464-477.

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

DellaVigna, Stefano, and Devin Pope. 2019. "Stability of Experimental Results: Forecasts and Evidence", NBER Working paper w25858.

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.

Gneezy, Uri, Goette, Lorenz, Sprenger, Charles and Zimmermann, Florian. 2017. "The limits of expectations-based reference dependence." *Journal of the European Economic Association*, 15(4), 861-876.

Goerg, Sebastian J., Sebastian Kube, and Jonas Radbruch. Forthcoming. "The Effectiveness of Incentive Schemes in the Presence of Implicit Effort Costs", *Management Science*.

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.

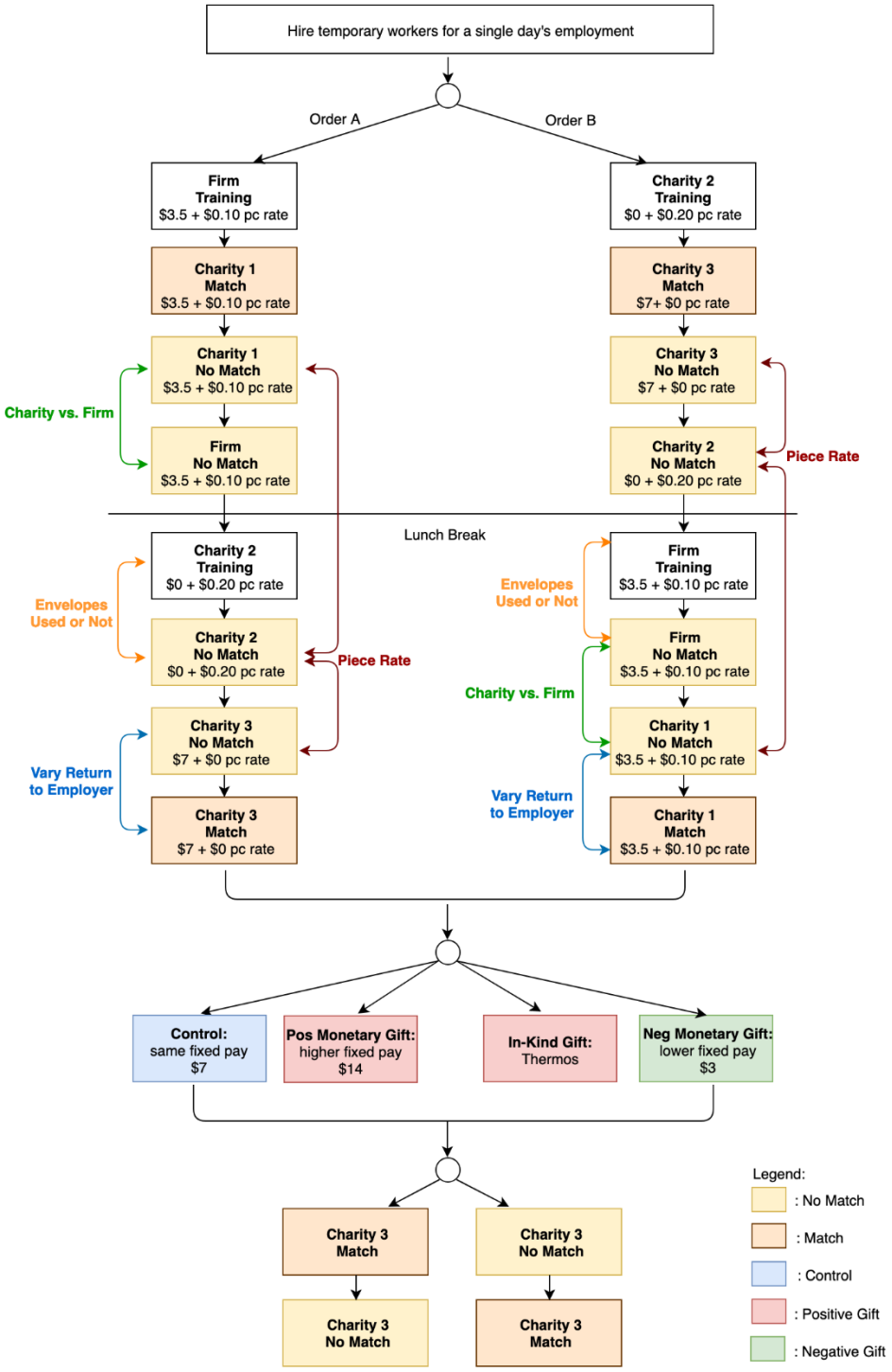
Kube, Sebastian, Michel André Maréchal and Clemens Puppe. 2013. "Do Wage Cuts Damage Work Morale? Evidence From A Natural Field Experiment", *Journal of the European Economic Association* 11 (4), 853–870.

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

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.

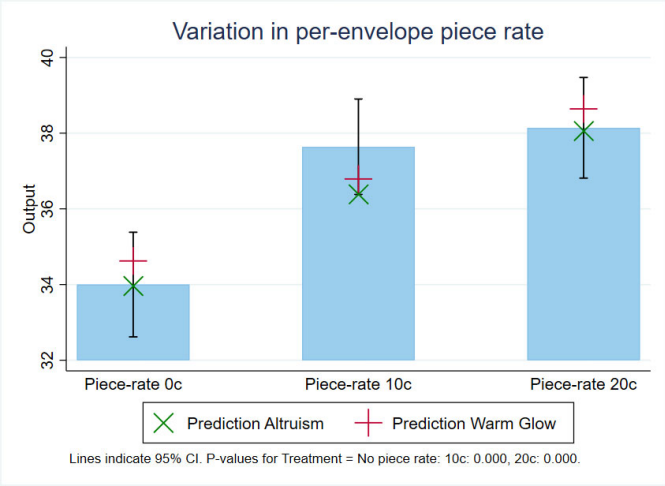
Figure 1. Productivity Experiment: Design



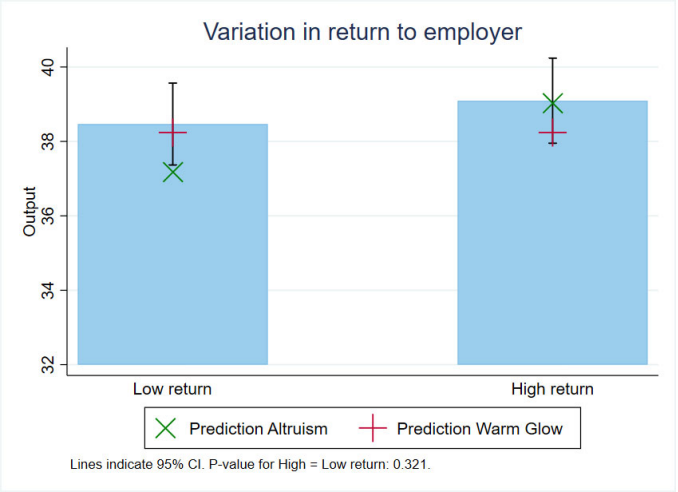
Notes: This figure displays the sequence of the 10 experimental rounds of envelope preparation, each of which lasting 20 minutes, for the productivity experiment. Between each round there is a 10 minute break, except between rounds 4 and 5 when there is a longer break for lunch. Subjects are randomized 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

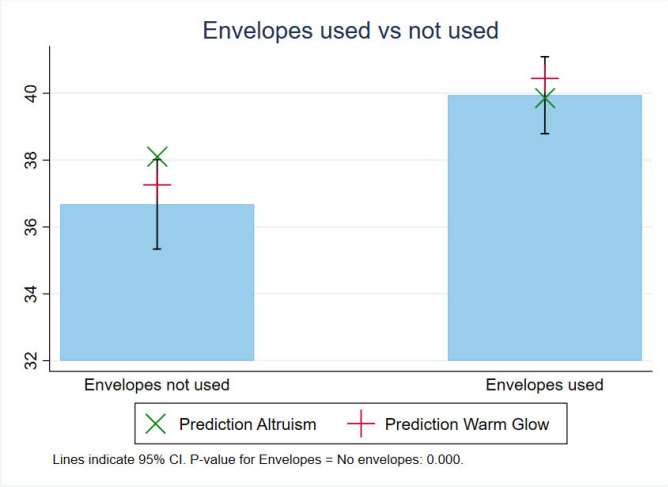
Panel a. Variation in Piece Rate



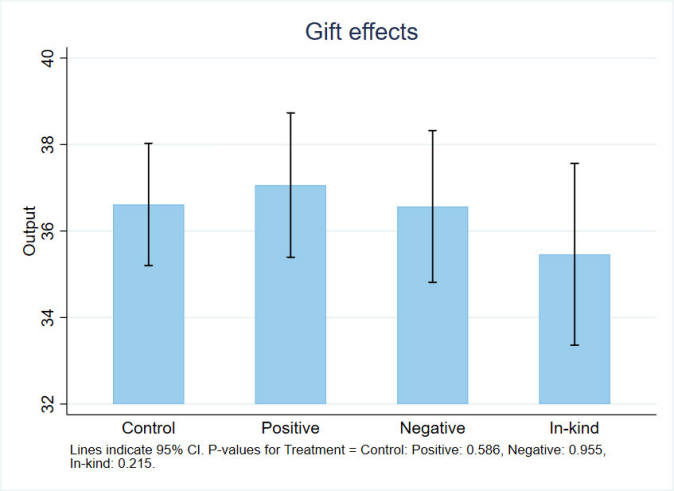
Panel b. Variation in Return to Employer (Match)



Panel c. Consequences to the Employer

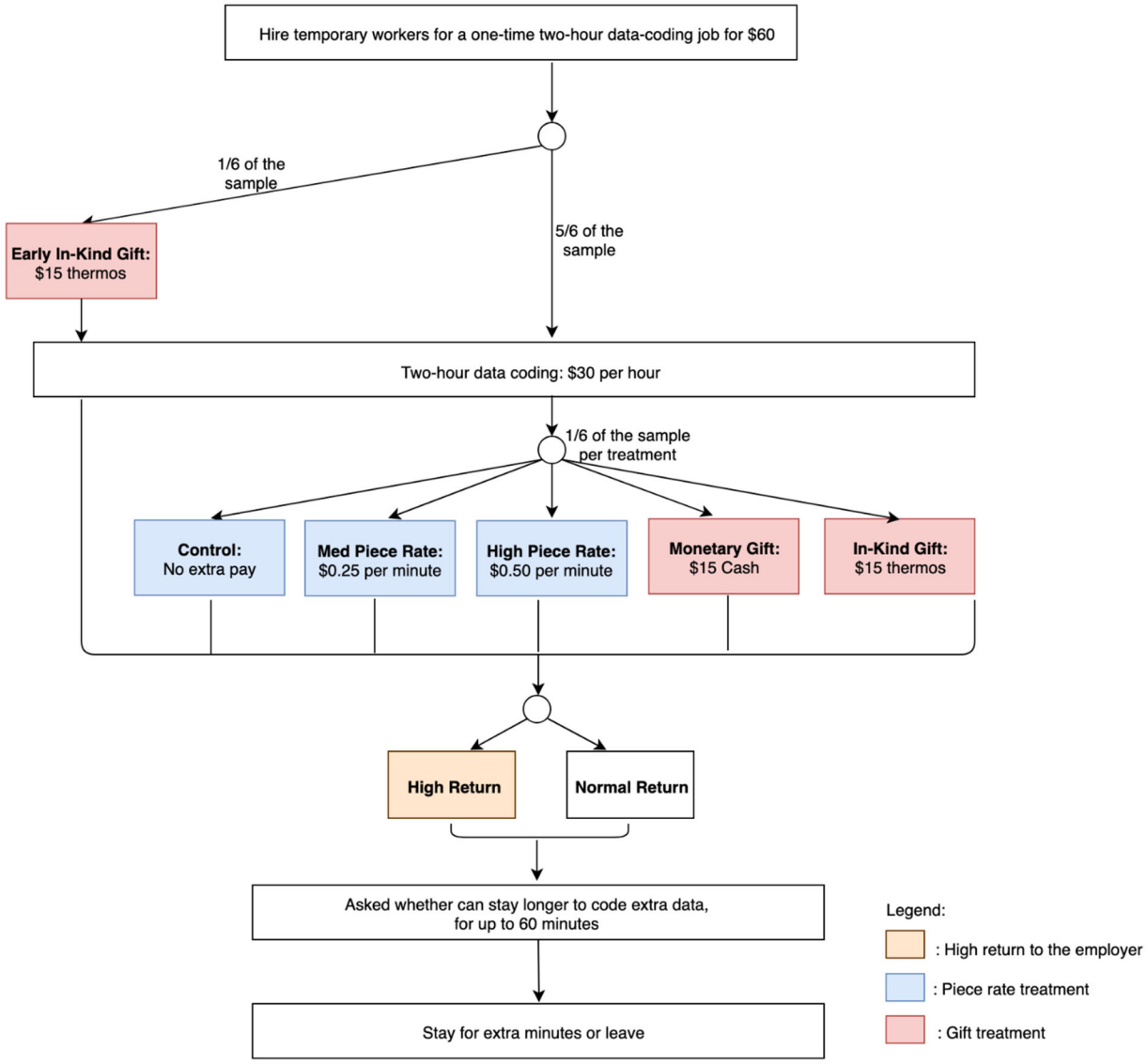


Panel d. Effect of Gift Treatments



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 piece 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 (Table 3, Column 5) or with warm glow (Table 3, Column 6).

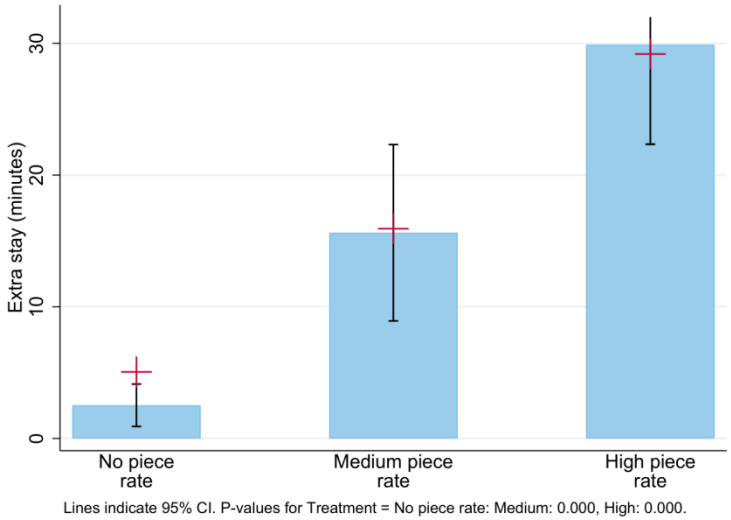
Figure 3. Labor Supply Experiment: Design



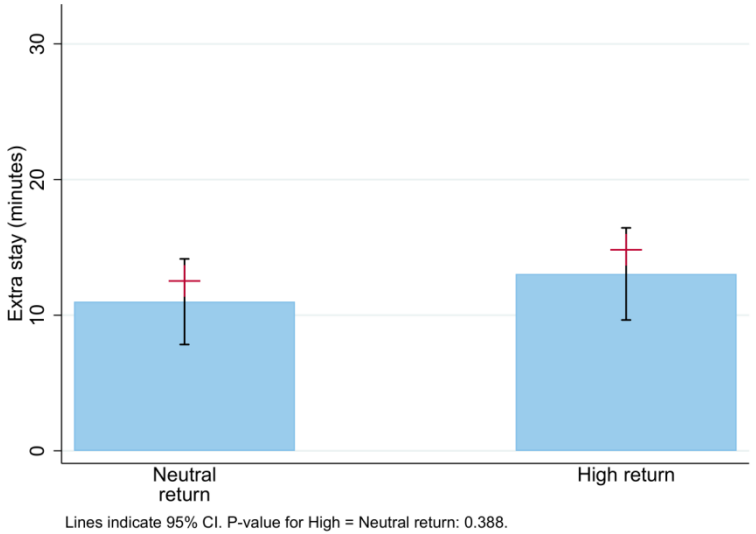
Notes: Figure 3 displays the design for the labor supply experiment.

Figure 4. Labor Supply Experiment: Experimental Findings

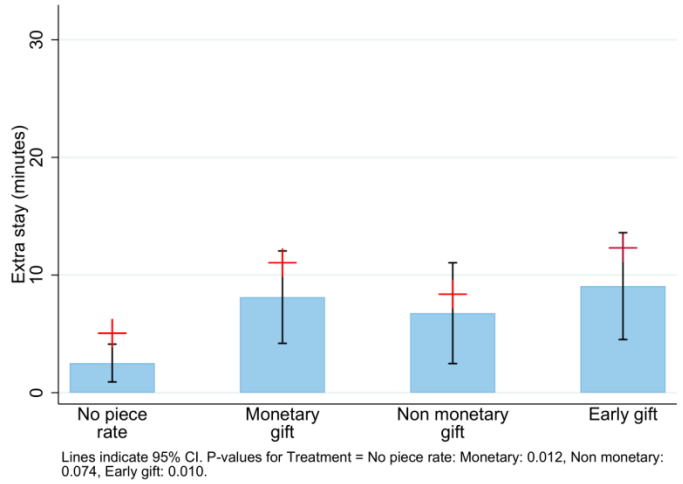
Panel a. Variation in Piece Rate



Panel b. Variation in Return to Employer



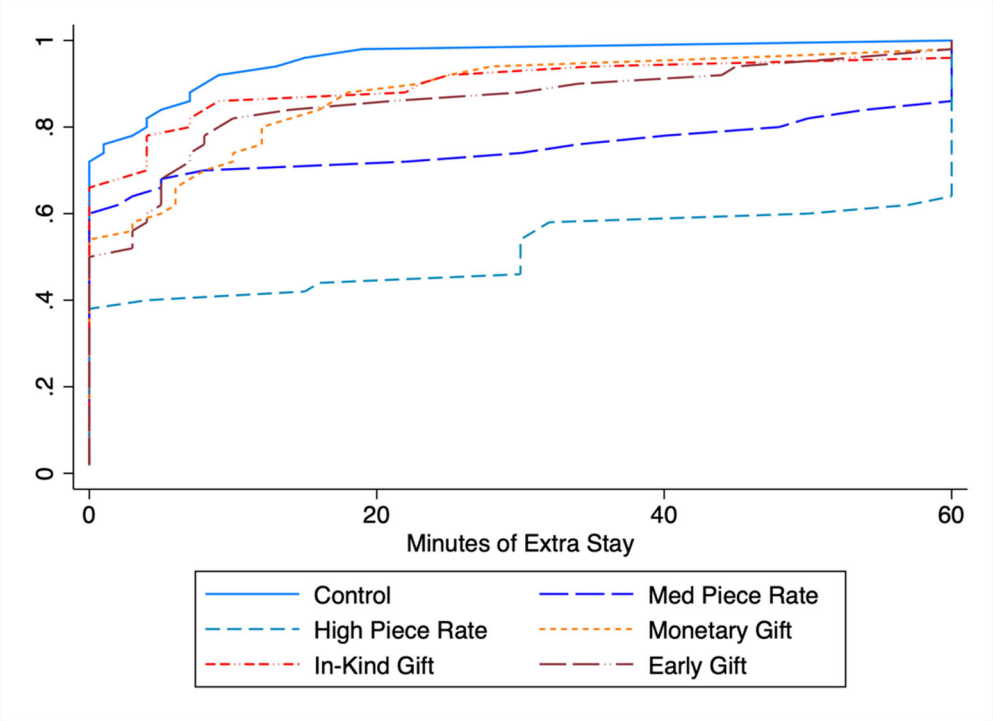
Panel c. Effect of Gift Treatments



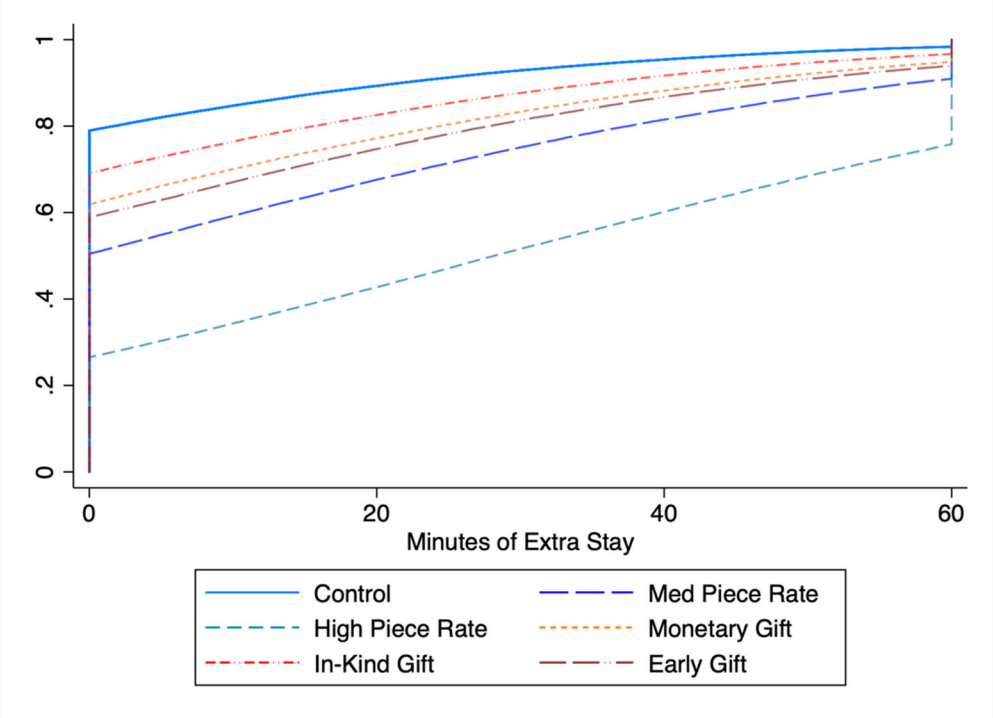
Notes: Figure 4 displays the experimental findings on the number of extra minutes of work in Experiment 2. Figures 4a and 4c compared this form of labor supply across the six main experimental arms. Figure 4b pools across the six experimental arms and compares the (cross-randomized) arms with stated 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. Labor Supply Experiment: Distribution of Effort by Treatment and Model Fit

Panel a. Distribution of Effort (Extra Work) across Treatments



Panel b. Model Prediction of Distribution of Effort across Treatments



Notes: Figure 5a displays the c.d.f. of the number of extra-minutes stayed for the six experimental arms. Figure 5b displays the fit of the model for the specification in Column 1 of Table 4.

Figure 6: Effect of Gift Treatments, Reduced-Form vs. Estimates of Social Preference Parameters

Figure 6a. Comparison of Reduced-Form Effort Change due to Gift

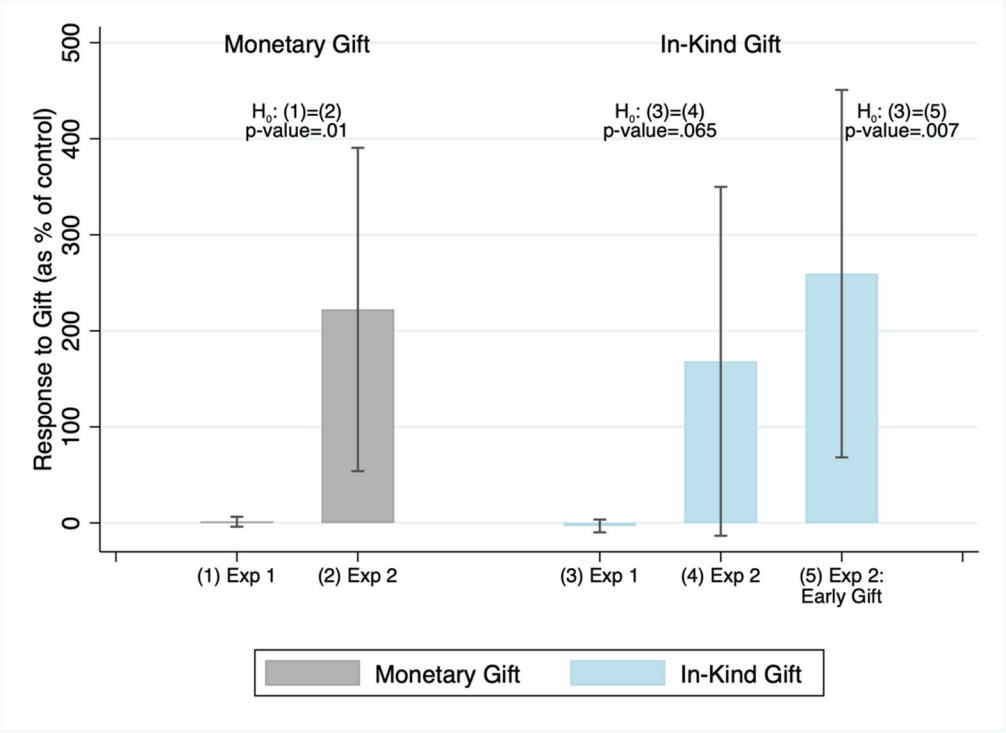
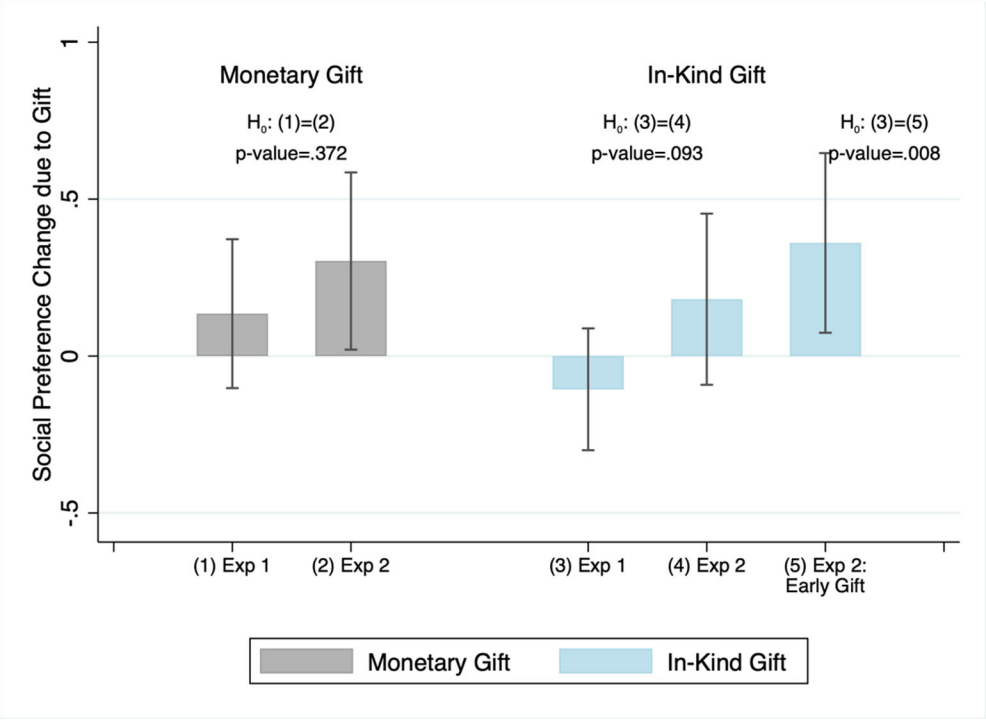


Figure 6b. Comparison of Social Preference Change due to Gift (Reciprocity)



Notes: Figure 6a displays the reduced-form impact of the monetary and in-kind gift in the productivity experiment (labeled "Exp. 1") and in the labor supply experiment (labeled "Exp. 2"). The estimates are shows as percent increases over the average effort in the control group, computed from the estimates respectively, from Table 1, Column 1 and Table 2, Column 1. Figure 6b displays the same comparison for the implied reciprocity estimates from, respectively, Table 3, Column 4 and Table 4, Column 1. The p-values refer to a test of equality of the results comparing across the two 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. Labor Supply Experiment, Findings

Specification:	OLS Regressions			Tobit Regressions		Probit Regressions	
Dependent Variable:	Number of Extra Minutes of Work (0-60)					Indicator for Stay >0 Minutes	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Piece Rate Treatments							
Medium Piece Rate (25c/min)	13.100	13.100	14.011	31.765	34.494	0.131	0.161
Treatment	(3.513)	(3.524)	(3.250)	(11.425)	(10.151)	(0.102)	(0.103)
High Piece Rate (50c/min)	27.380	27.424	28.010	65.287	66.721	0.351	0.392
Treatment	(3.942)	(3.945)	(3.686)	(12.014)	(10.810)	(0.103)	(0.104)
Gift Treatments							
Monetary Gift Treatment	5.600	5.600	7.370	21.678	27.163	0.191	0.243
	(2.164)	(2.132)	(2.522)	(9.284)	(9.267)	(0.102)	(0.106)
In-Kind Gift Treatment	4.240	4.240	4.323	12.914	14.594	0.067	0.074
	(2.335)	(2.317)	(2.481)	(10.154)	(9.597)	(0.103)	(0.109)
In-Kind Gift, Early Delivery	6.540	6.628	6.576	25.355	24.902	0.232	0.253
Treatment	(2.459)	(2.458)	(2.492)	(9.364)	(8.652)	(0.102)	(0.104)
(Crossed) Employer Return Treatment							
Treatment w/ Stated High		2.210	2.320	6.430	5.802	0.051	0.054
Return to the Employer		(2.145)	(1.946)	(6.094)	(5.384)	(0.058)	(0.060)
Constant	2.520	1.415		-38.468			
	(0.820)	(1.285)		(8.827)			
Controls			X		X		X
Hyp.: Gift Treatments = Control	p=0.000	p=0.000	p=0.001	p=0.011	p=0.003	p=0.051	p=0.028
R squared / Pseudo R Squared	0.190	0.193	0.328	0.032	0.068	0.039	0.097
Number of Subjects	300	300	300	300	300	300	300

Notes: Robust standard errors. Columns 3, 5, and 7 include fixed effects for Craigslist sample, gender, and age groups (25-34, 35-44, 45-54, 55+). Columns 6 and 7 report the marginal effects for the probit specification.

Table 3. Productivity Experiment, Social Preference Estimates

Estimation:	**Non-Linear Least Squares**							
Dependent Variable:	**Log (Number of Envelopes in a Batch)**				**Number of Envelopes in a Batch**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Baseline Social Preferences								
Altruism towards Charity	0.230		-0.012		0.253		0.027	
	(0.042)		(0.038)		(0.040)		(0.034)	
Altruism towards Grocery Store	0.759				0.735			
	(0.088)				(0.077)			
Warm Glow towards Charity		0.443	0.461	0.443		0.462	0.422	0.462
		(0.064)	(0.074)	(0.063)		(0.066)	(0.071)	(0.065)
Warm Glow towards Grocery		0.720	0.730	0.720		0.716	0.694	0.716
Store		(0.073)	(0.080)	(0.072)		(0.074)	(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)	11.123	9.440	9.518	9.44	0.293	0.263	0.258	0.263
	(1.449)	(0.747)	(0.841)	(0.737)	(0.030)	(0.018)	(0.019)	(0.018)
Cost of Effort Function:		Power				Exponential		
Std. Deviation of Error Term	0.131	0.130	0.130	0.144	3.994	3.952	3.952	4.318
Std. Dev. of Individual f.e.s * (1/γ)	0.249	0.211	0.014	0.241	8.155	8.158	8.160	8.014
R Squared	0.8346	0.8374	0.8374	0.7915	0.8500	0.8532	0.8532	0.8184
N	3568	3568	3568	4460	3568	3568	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. Labor Supply Experiment, Social Preference Estimates

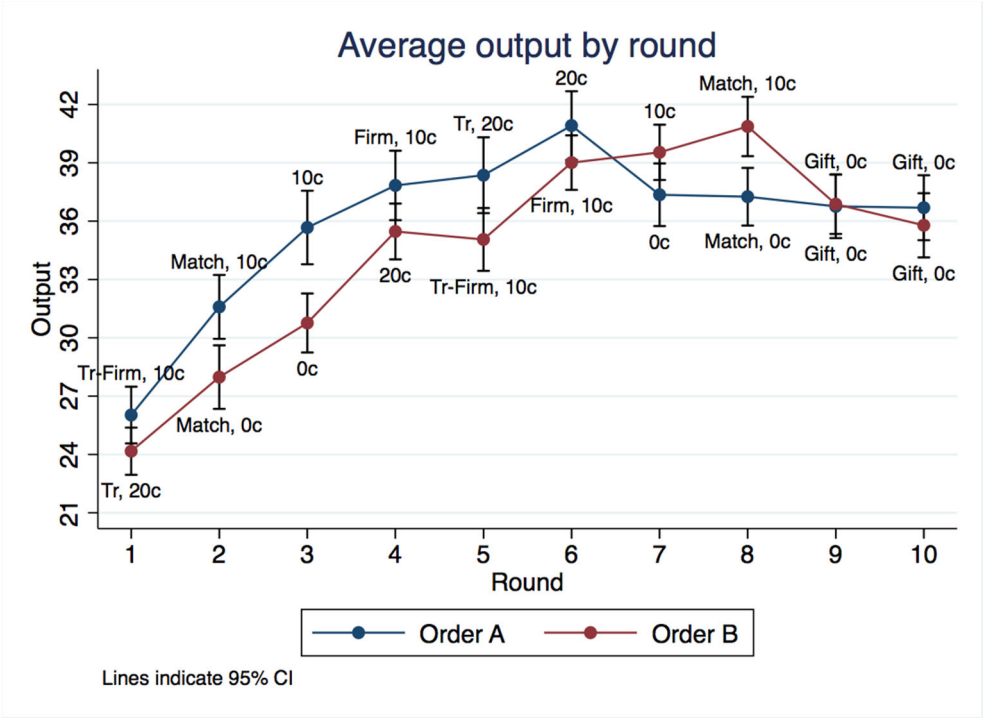
Estimation:	**Maximum Likelihood, Accounting for Censoring at 120 and 180 Minutes Worked**			
Dependent Variable:	**Number of Minutes Worked**		**Log (No. of Minutes Worked)**	
	(1)	(2)	(3)	(4)
Baseline Social Preferences				
Social Preference towards Employer	0.813	0.576	0.806	0.576
	[0.003, 100]	[0.008, 100]	[0.003, 100]	[0.025, 88.43]
Social Preference Change - High Return for Employer	0.109	0.098	0.111	0.100
	(0.093)	(0.084)	(0.096)	(0.085)
Reciprocal Social Preferences				
Social Preference Change -- Monetary Gift	0.303	0.354	0.309	0.360
	(0.144)	(0.142)	(0.140)	(0.126)
Social Preference Change -- In-Kind Gift	0.181	0.186	0.181	0.186
	(0.139)	(0.129)	(0.143)	(0.135)
Social Preference Change -- In-Kind Gift, Early	0.360	0.328	0.365	0.332
	(0.146)	(0.128)	(0.144)	(0.113)
Incidental Parameters				
Cost Function Curvature (γ)	0.007	0.009	1.050	1.324
	[0, 0.079]	[0, 0.053]	[0, 10.08]	[0.009, 5.534]
Std. Deviation of Error Term	45.121	39.401	0.311	0.272
	(4.104)	(3.372)	(0.028)	(0.024)
Cost of Effort Function:	Exponential		Power	
Controls		X		X
Log Likelihood	-637.99	-614.60	-160.58	-137.35
N	300	300	300	300

Notes: Maximum likelihood estimates for the number of minutes stayed (including the initial 120 required minutes). Bootstrap standard deviations are in parentheses and 95% bootstrap confidence intervals are in brackets. The number of bootstrap draws is 1000. Columns 2 and 4 include fixed effects for Craigslist, gender, and, age groups (25-34, 35-44, 45-54, 55+). The upper bound of the baseline social preference parameter is set to be 100 in the estimation.

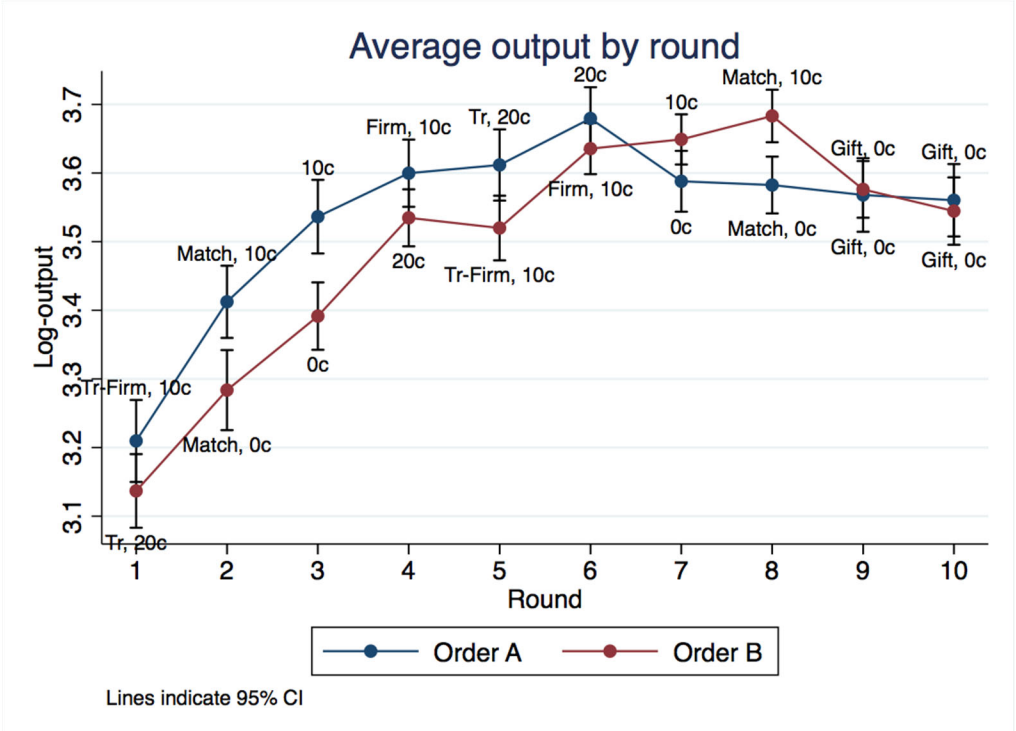
For Online Publication

Online Appendix Figure 1. Productivity Experiment: Average Effort over the 10 Batches, by Order

Online Appendix Figure 1a. Output



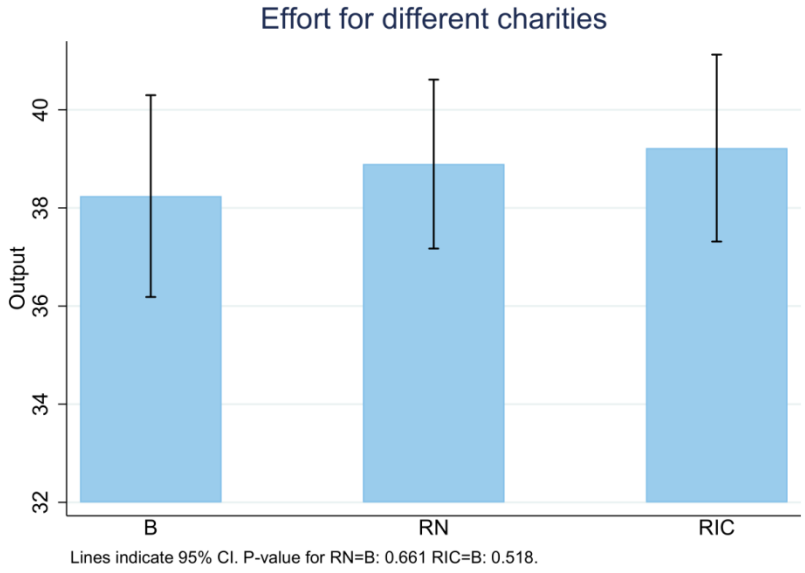
Online Appendix Figure 1b. Log Output



Notes: This figure displays the average output (number of envelopes folded within a 20-minute round) and log output in a batch (round). The figure indicates 95% confidence intervals computed clustering by session, thus allowing for correlation of errors among subjects in a session. Subjects are randomized into Order A or Order B. See Figure 1 for more detailed labeling of the 10 batches in each order. The output for batches 9 and 10 averages across the gift treatments displayed in Figure 1.

Online Appendix Figures 2a-b. Productivity Experiment: Additional Findings

Panel a. Effort Provided For Three Different Charities

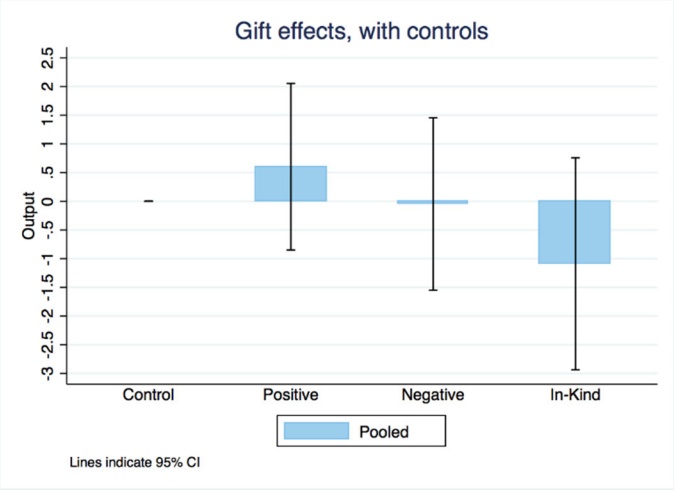


Panel b. Charity Employer versus Grocery Store Employer

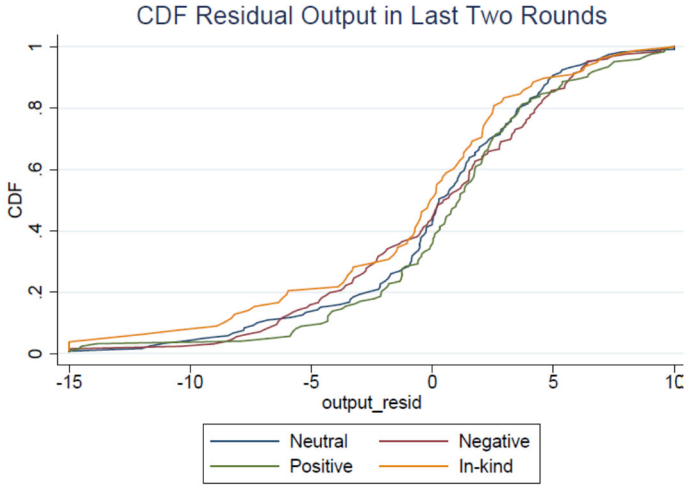
Notes: This figure displays additional experimental results on average output (number of envelopes folded within a 20-minute round). Online Appendix Figure 2a compares productivity across the three different charities used in the experiment. The charities are randomized in a rotating way to take the role of Charity 1, 2, and 3. The comparison uses output in all rounds except for the training rounds. Online Appendix Figure 2b compares output when producing for a charity versus for a firm (a grocery store) holding constant the piece rate at 10 cents and holding constant the perceived return to the employer at 30 cents per envelope. The rounds compared are outlined in Figure 1. The figures indicate 95% confidence intervals computed clustering by session.

Online Appendix Figure 3. Productivity Experiment: Additional Evidence on Gift Treatments

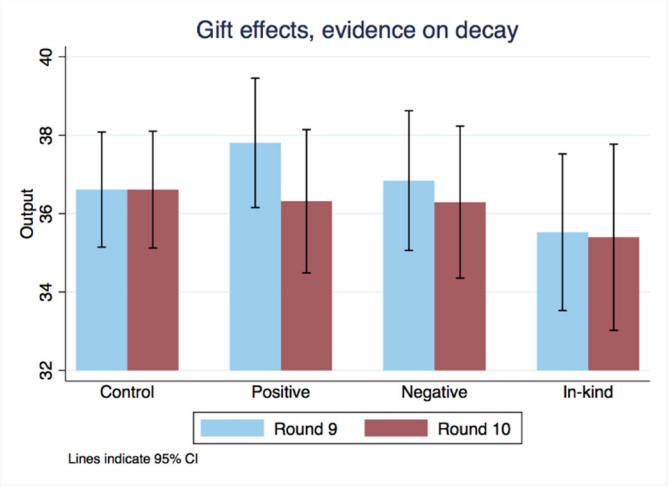
Panel a. Effect of Gift Treatments (With Controls)



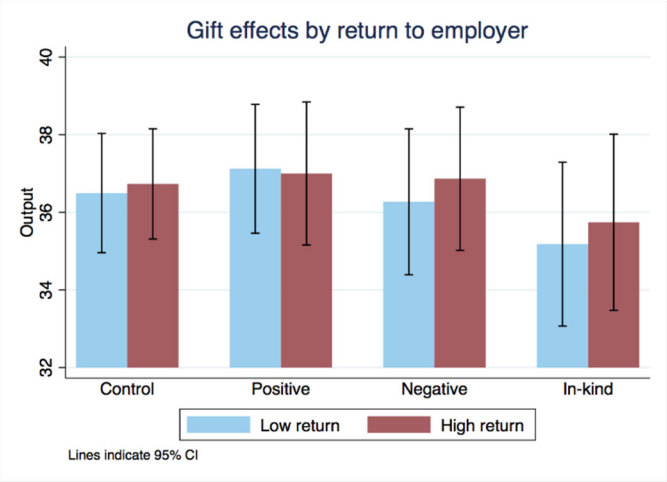
Panel b. Effect of Gift Treatments (With Controls)



Panel c. Evidence on Decay of Gift Effects



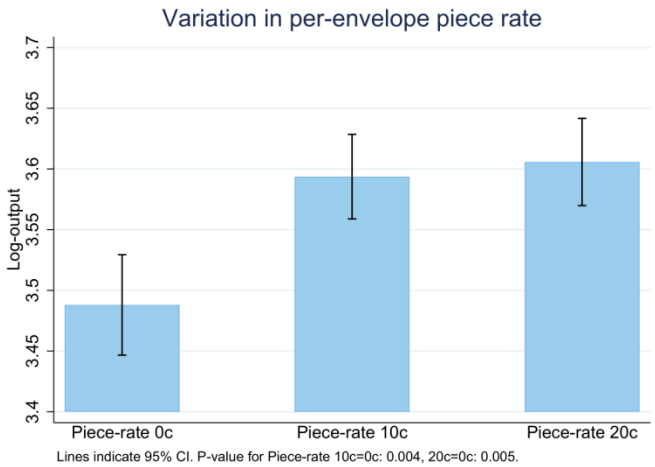
Panel d. Interaction with Return to Employer



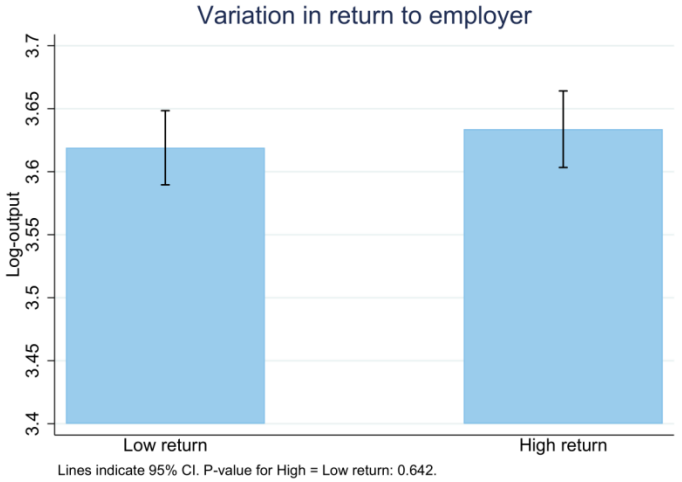
Notes: This figure presents additional results for average output (number of envelopes stuffed in 20 minutes) in the gift treatments in rounds 9 and 10 (see Figure 1). The figures include 95% confidence intervals obtained after clustering for session. Panel a presents the results controlling for average productivity in rounds 5-8 (Table 1, Column 3). Panel b presents the c.d.f. of the worker-level estimated gift effects. (We regress productivity in rounds 9 and 10 on average productivity in rounds 5-8, take the residuals and average the two residuals for each worker.) Panel c examines the possible decay of gift effects. Panel d splits the results by return to the firm: in either round 9 or round 10 (depending on a randomization) the employer earns a higher return due to a charity match.

Online Appendix Figure 4. Findings of Productivity Experiment, Log Output

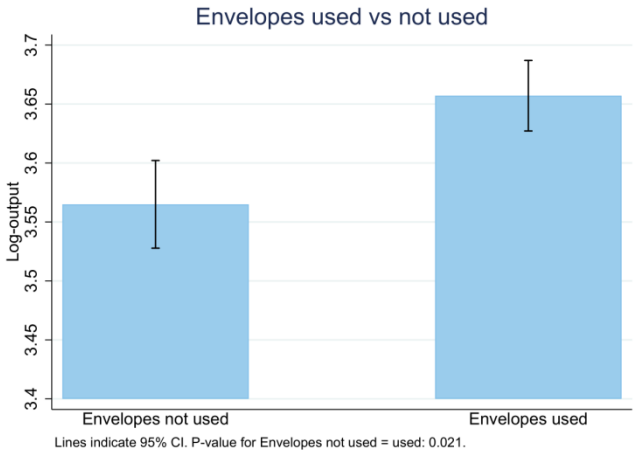
Panel a. Variation in Piece Rate



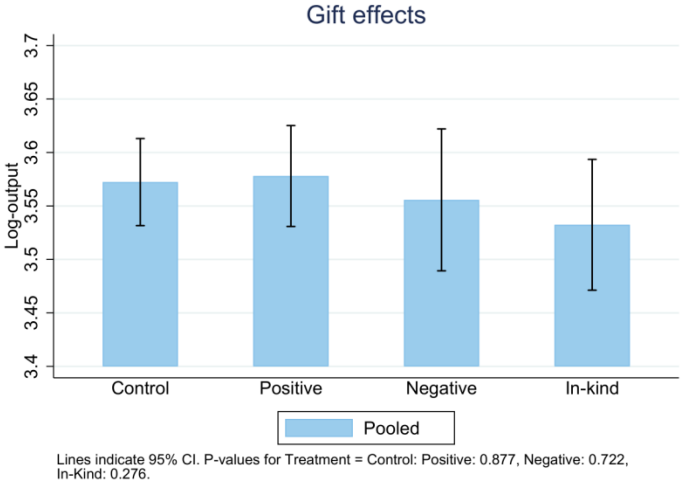
Panel b. Variation in Return to Employer (Match)



Panel c. Consequences to the Employer



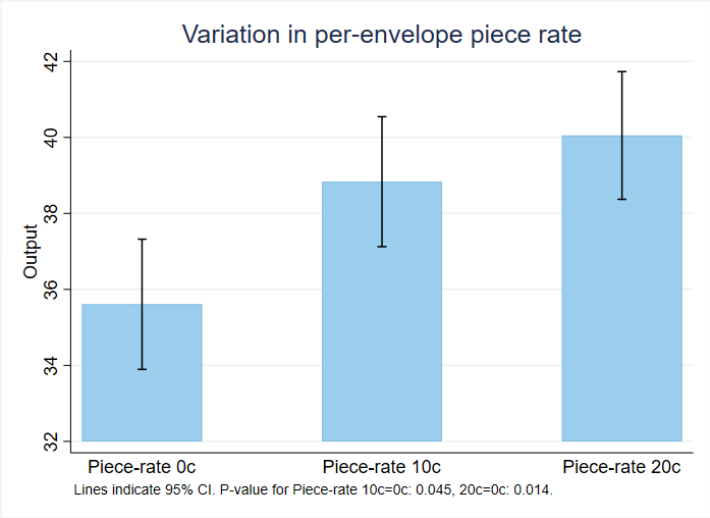
Panel d. Effect of Gift Treatments



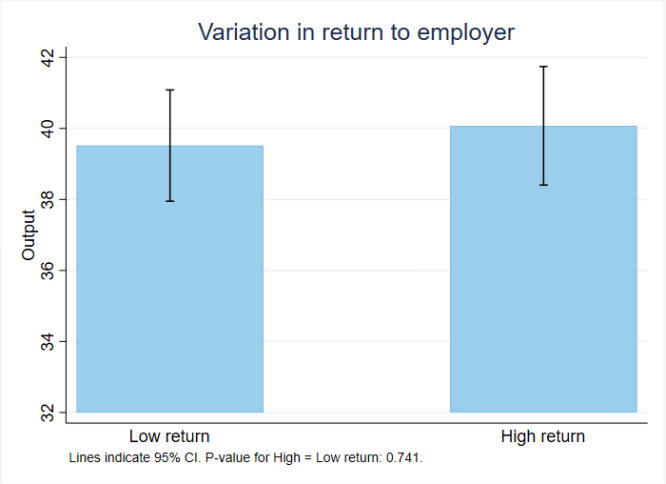
Notes: This figure displays the key findings in Experiment 1 for log output (log of number of envelopes folded within a 20-minute round) rather than output.

Online Appendix Figure 5. Findings of Productivity Experiment, Output, Employed participants only

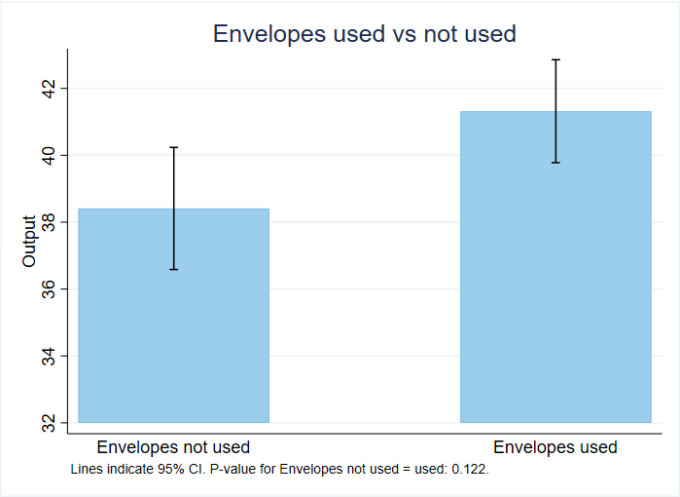
Panel a. Variation in Piece Rate



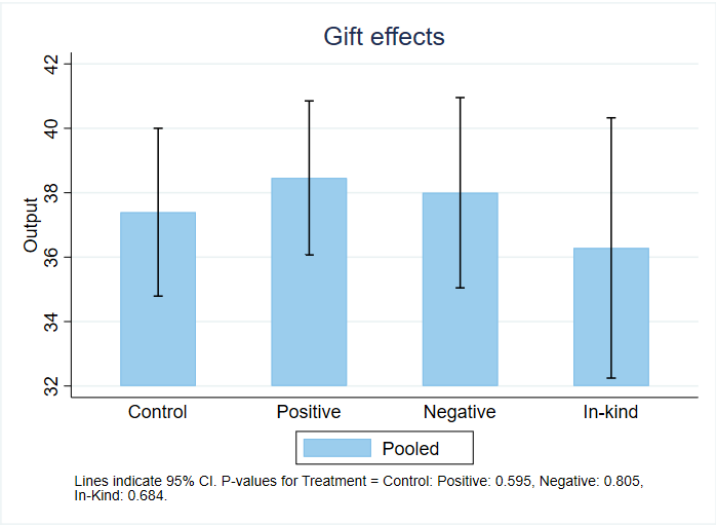
Panel b. Variation in Return to Employer (Match)



Panel c. Consequences to the Employer



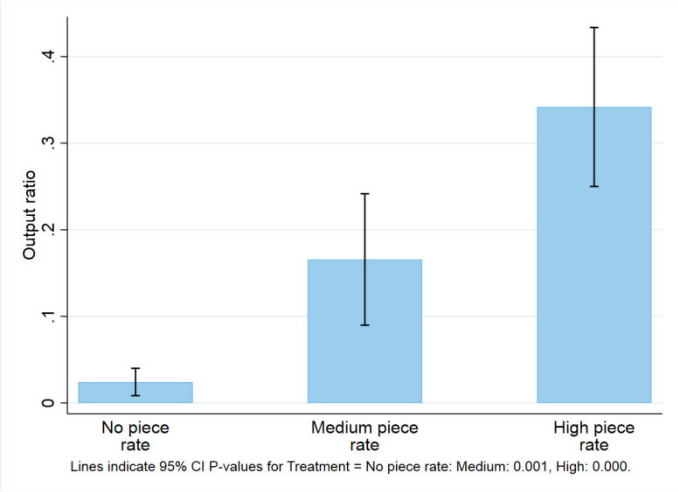
Panel d. Response to Gifts



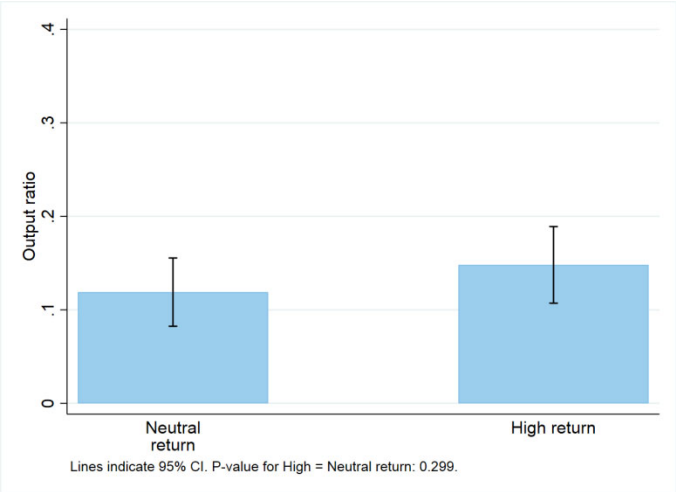
Notes: This figure displays key comparisons of average output (number of envelopes folded within a 20-minute round) including only employed workers.

Online Appendix Figure 6. Findings of Labor Supply Experiment, Output in Extra Minutes (As Fraction of Output in First 120 Minutes)

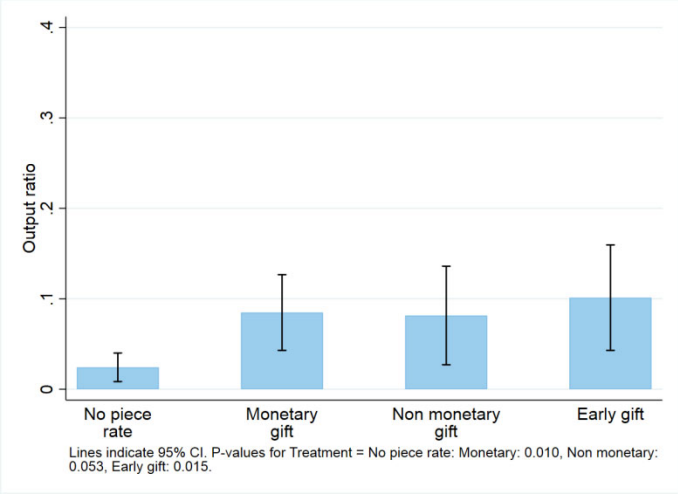
Panel a. Variation in Piece Rate



Panel b. Variation in Return to Employer



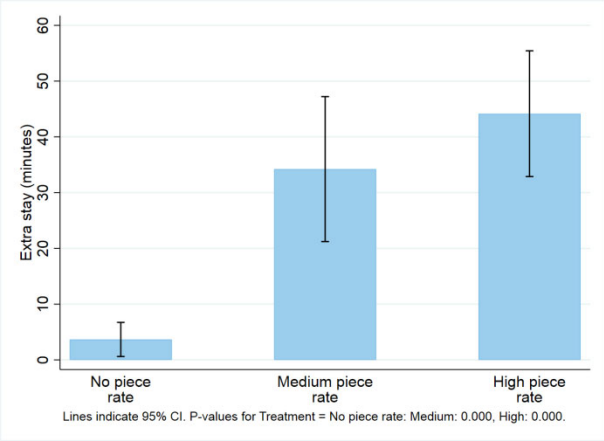
Panel c. Effect of Gift Treatments



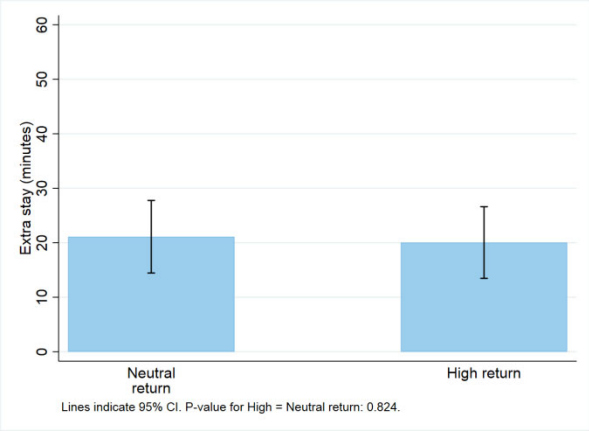
Notes: This figure presents the findings of the labor supply experiment reporting the output (number of lines coded) produced in the extra minutes of work, as fraction of the output produced by that same subject in the initial 120 minutes of work. Output is 0 for subjects who do not stay extra.

Online Appendix Figure 7. Findings of Labor Supply Experiment, Craigslist Participants

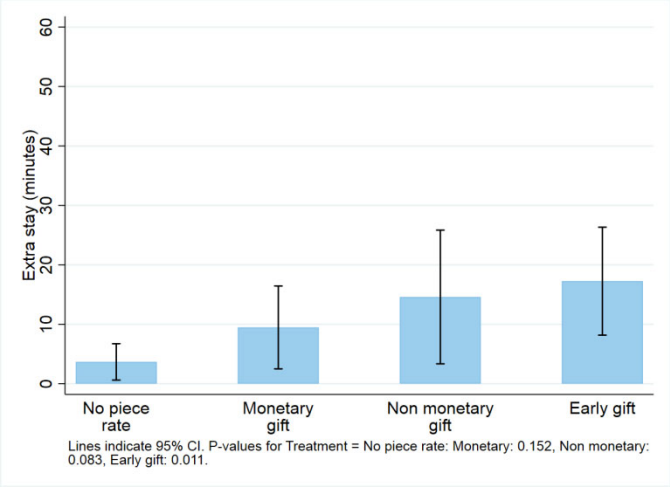
Panel a. Variation in Piece Rate



Panel b. Variation in Return to Employer



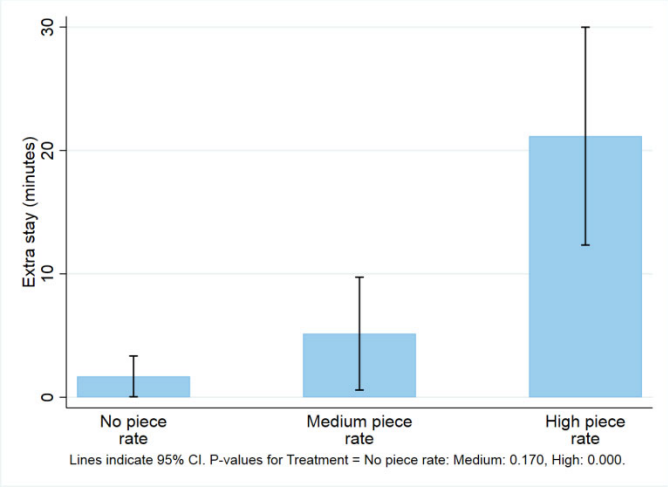
Panel c. Effect of Gift Treatments



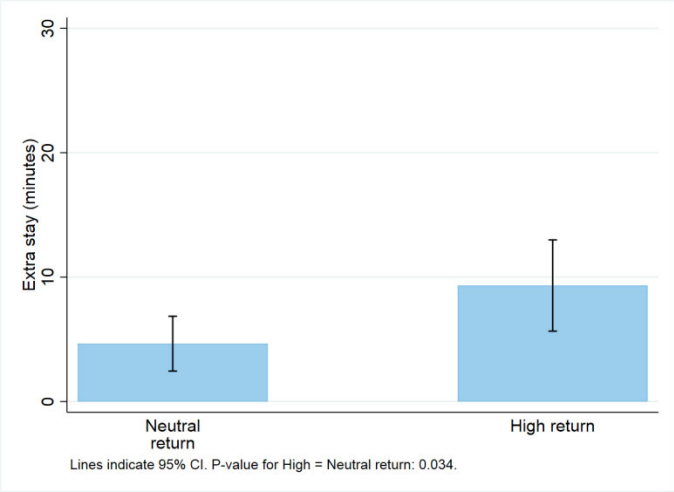
Notes: This figure presents the findings of the labor supply experiment, for the subjects recruited through Craigslist ads.

Online Appendix Figure 8. Findings of Labor Supply Experiment, Student Participants

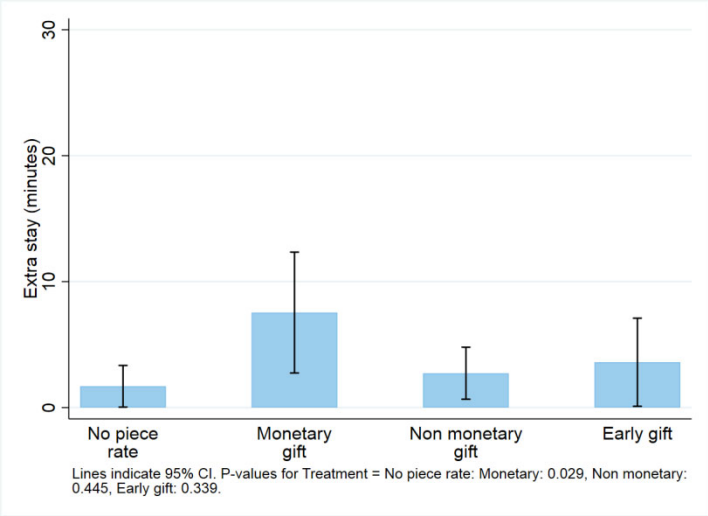
Panel a. Variation in Piece Rate



Panel b. Variation in Return to Employer

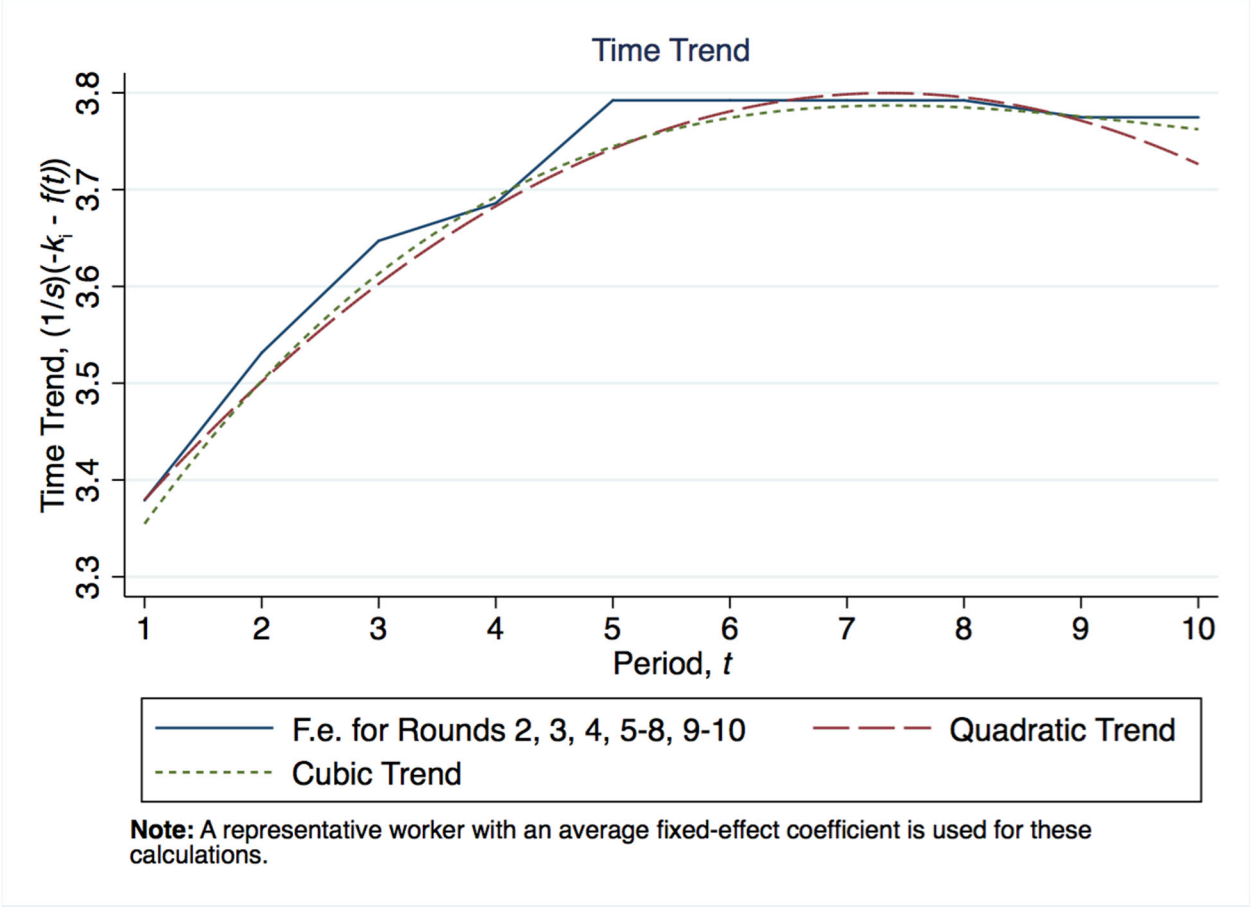


Panel c. Effect of Gift Treatments



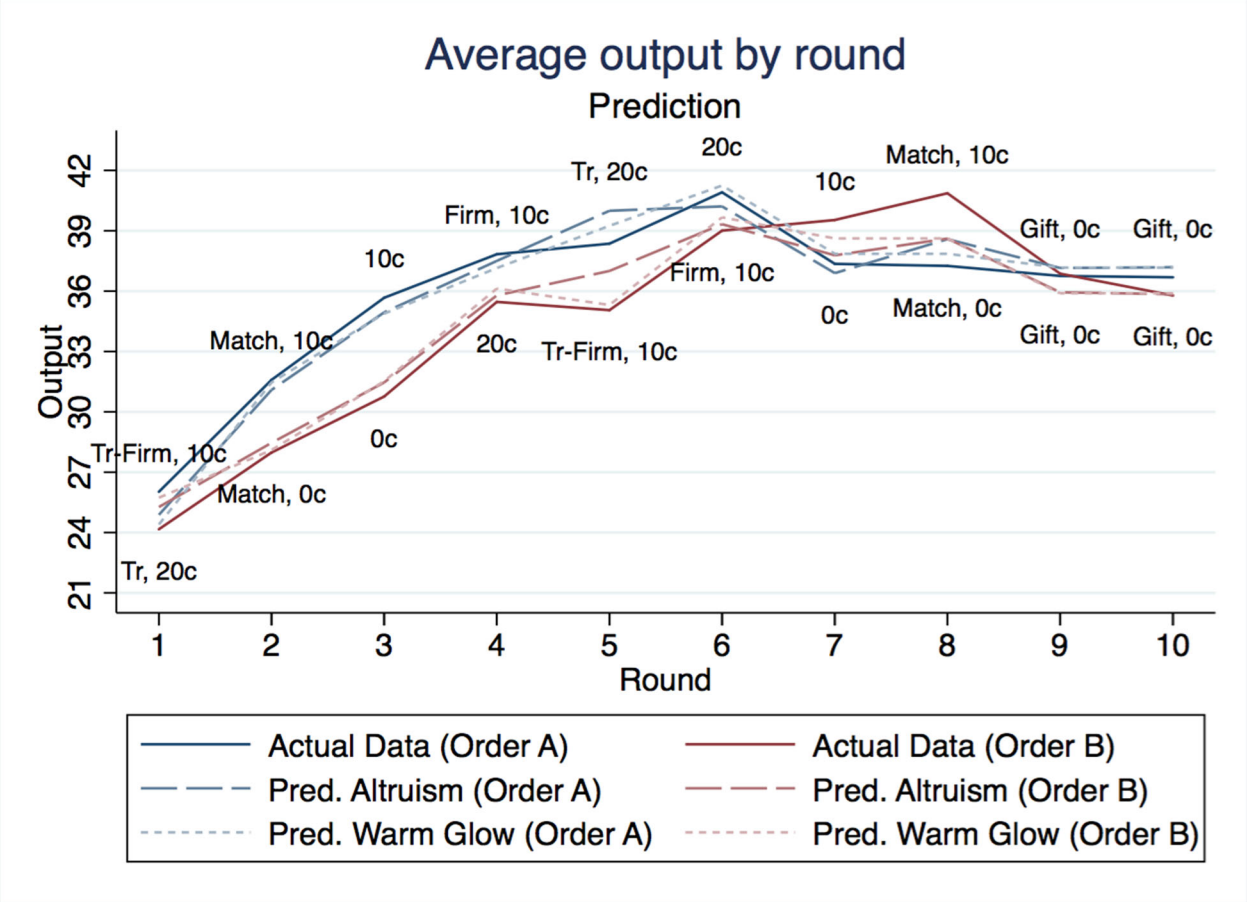
Notes: This figure presents the findings of labor supply experiment, for the subjects who are students.

Online Appendix Figure 9. Productivity Experiment, Estimated Productivity Effects, Different Models



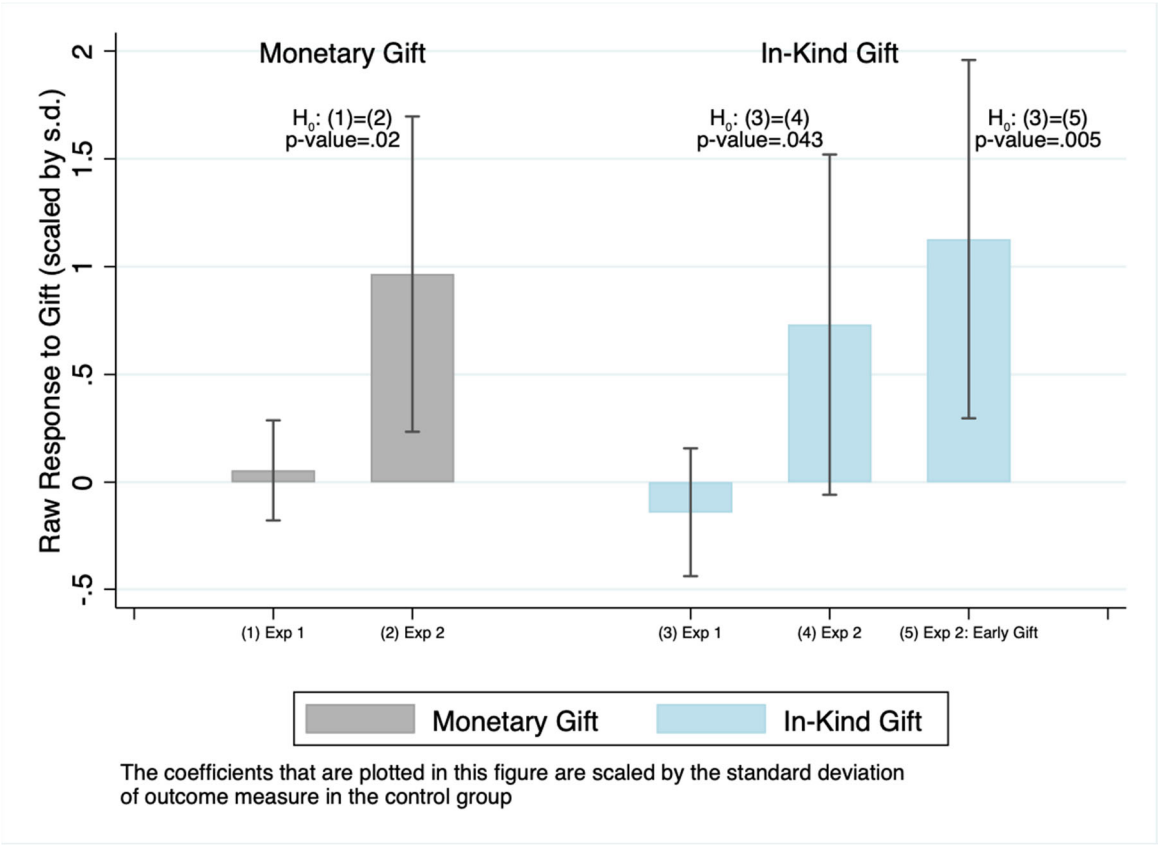
Notes: This figure plots the estimated *(1/gamma)(-k-f(t))* function, that is, how the cost of effort function is estimated to change over time for an individual with representative *k*. The estimated coefficients are from specifications in Table 3, Column 1 (indicators for rounds), and from Online Appendix Table 4, Columns 1 (quadratic polynomial), and 2 (cubic polynomial).

Online Appendix Figure 10. Fit of warm Glow versus Altruism Model, All 10 Rounds, Order A and B



Notes: This figure displays the average output (number of envelopes folded within a 20-minute round) in a round for Order A and Order B, together with the predicted output according to the warm glow model (Column 6 in Table 3) and according to the altruism model (Column 7 in Table 3). See Figure 1 for more detailed labeling of the 10 rounds (batches) in each order. The output for rounds 9 and 10 averages across the gift treatments displayed in Figure 1.

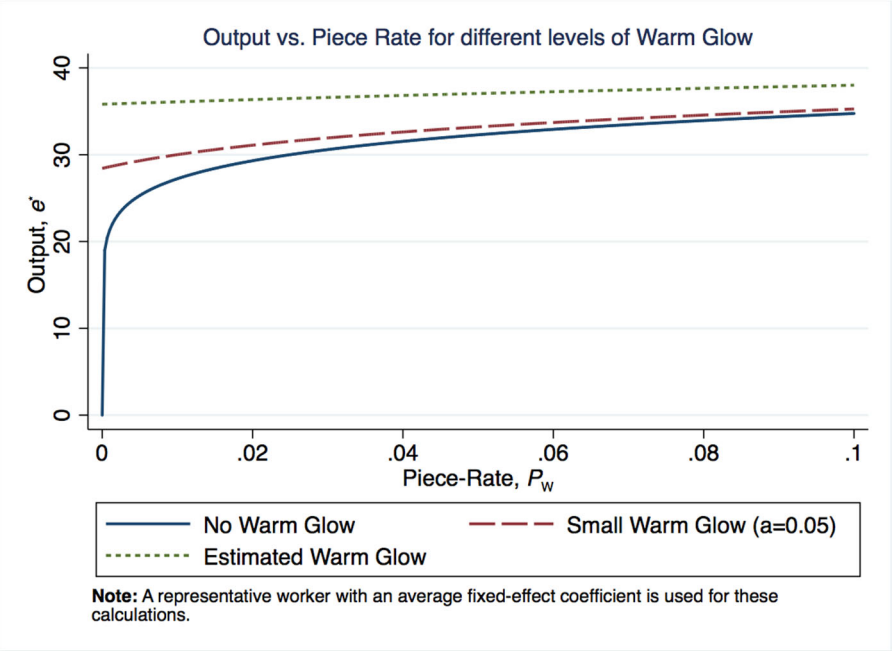
Online Appendix Figure 11. Effect of Gift Treatments, in Standard Deviation Units



Notes: This figure displays the reduced-form impact of the monetary and in-kind gift in the productivity experiment (labeled "Exp. 1") and in the labor supply experiment (labeled "Exp. 2"). The estimates are shows as standard deviation increases over the average effort in the control group, computed from the estimates respectively, from Table 1, Column 1 and Table 2, Column 1. The p-values refer to a test of equality of the results comparing across the two experiments.

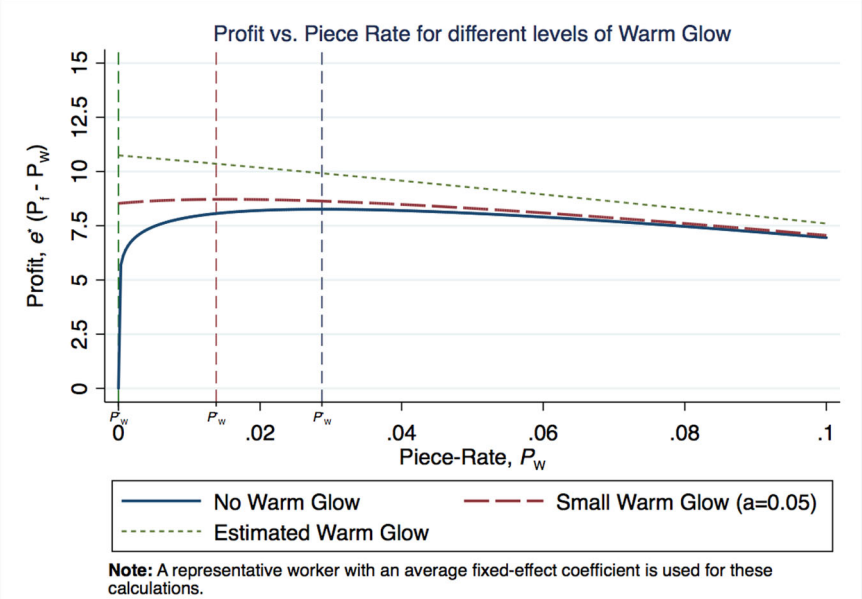
Online App. Figure 12. Productivity Experiment, Optimal Piece Rate for Estimated Social Preferences

Panel a. Effort as Function of Piece Rate



Note: A representative worker with an average fixed-effect coefficient is used for these calculations.

Panel b. Profit Rate as Function of Piece Rate

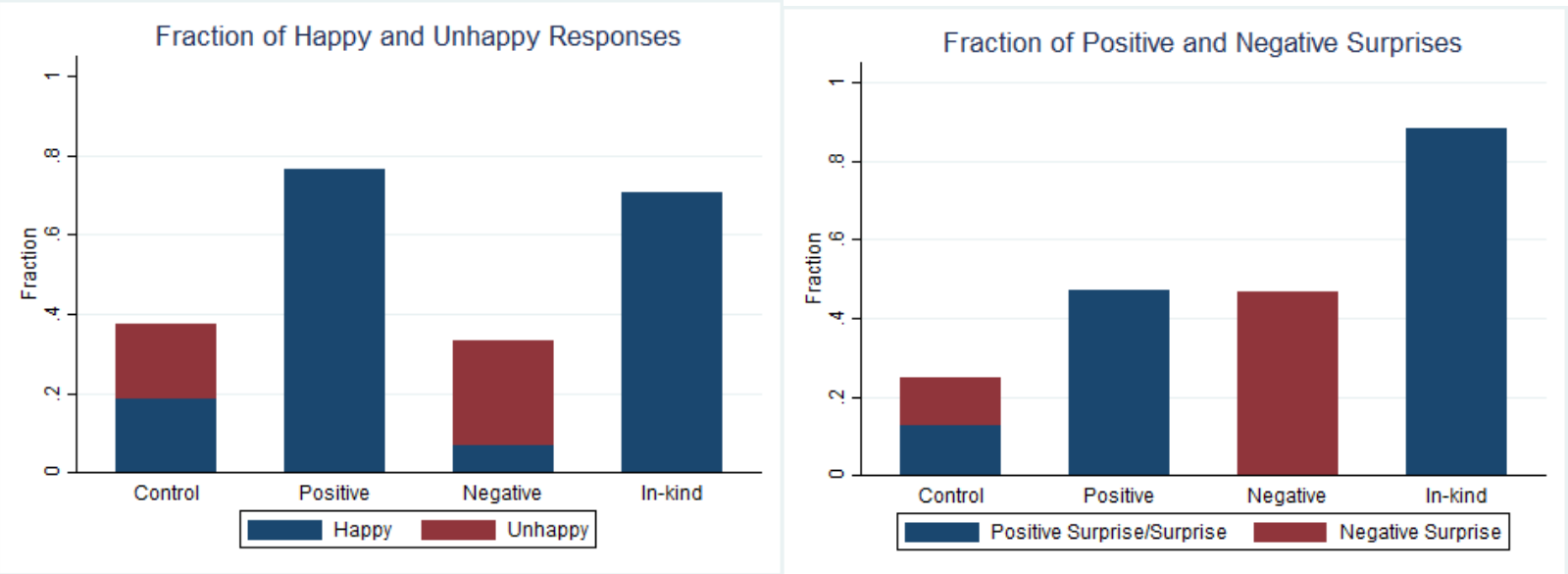


Note: A representative worker with an average fixed-effect coefficient is used for these calculations.

Notes: This figure takes the estimated parameters in the warm-glow specification and predicts the implied effort *e** (Panel a) and profit rate e*(Pf-Pw) (Panel b), for different levels of the piece rate Pw. Specifically, the plots examine the impact on profits of increasing the piece rate holding constant all else (including the lump-sum pay). We take the parameters from Column 2 in Table 3, assuming an individual with an average fixed effect k at the productivity estimated for batches 5-8. The continuous blue line indicates the counterfactual for the case with no social preferences. The dotted green line indicates the curves for the estimated warm glow. The dashed red line indicates the case with warm glow at one tenth of the estimated one, holding all other parameters the same.

Online Appendix Figure 13. Productivity Experiment, Effect of Gift Treatments on Worker Happiness and Surprise

Panel a. Fraction Stating a Happy or Unhappy Reaction **Panel b. Fraction Stating Positive or Negative Surprise**



Notes: This figure presents the average response to a short debriefing questionnaire administered after the end of the productivity experiment. The sample size includes 65 subjects, since the questions were only asked for the last 65 subjects in the experiments. Panel a presents the fraction that indicates being happy and the fraction that indicates being unhappy for each of the various treatments. Panel b indicates the fraction stating a positive surprise versus negative surprise (with the other categories being "as expected" or "none"). For the in-kind treatment, the bar shows the fraction that reported being surprised (we did not ask for the share with negative surprise).

Online Appendix Table 1. Overview of Features of Selected Gift Exchange Papers

Authors (chronologically)	Gift in Treament Condition	Task Assigned	Betweek- or Within-Subject Design? (B/W)	Piece-Rate Design? (Y/N)	Sample Size. Shaded if Larger than 100	Workers Know Return to Employer? (Y/N)	Vary Return to Employer? (Y/N)	Estimate Social Preferences? (Y/N)	Lab or Field Experiment?	Notes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
DellaVigna, List, Malmendier, Rao (2019), *Productivity Experiment*	$7 (100%) increase / $4 (57%) decrease relative to $7 baseline	Folding Charity Envelopes	B for Gift Exchange	Y	446	Y	Y	Y	Field	
Panel A. Real Effort Experiments										
Gneezy and List (2006) *Study 1 - data-entry task*	$8 (67%) increase relative to $12 baseline	Library Book Coding	B	N	19	N	N	N	Field	First design of gift exchange in the field. Value of data entry to employer not clear
Gneezy and List (2006) *Study 2 - door-to-door fundraising*	$10 (100%) increase relative to $10 baseline	Door-to-door Fundraising	B	N	23	Y	N	N	Field	Subjects raise funds for charity and thereby can determine the return to employer
Bellemare and Shearer (2011)	$80 (37%) increase relative to average daily earnings of $215	Tree-Planting	W	N	18	N	N	Y	Field	All tree-planting workers receive $80 bonus on the second of five working days; thus, variation for Gift is Within subject
Hennig-Schmidt, Rockenbach, and Sadrieh (2010), *Study 1 - Data Entry*	DM 2 (10%) / DM 8 (40%) increase relative to DM 20 baseline	Data-Entry Task	B for Gift Exchange	N	103	N	N	N	Field	Examine the effects of peer comparison among workers
Hennig-Schmidt, Rockenbach, and Sadrieh (2010), *Study 2 - Fold Envel.*	EUR 0.25 (10%) increase relative to EUR 2.50 baseline	Folding Envelopes in Lab	B	N	59	Y	N	N	Lab	Return to employer is stated by opportunity costs of outsourcing data-entry task
Englmaier and Leider (2012a)	$5 (38%) increase relative to a $13 baseline	Data-Entry Task	B	N	59	N	Y	N	Field	Experimenters get a "substantial bonus" (worth $10, not known to subjects) if 50% of the work is done by the end of the week
Englmaier and Leider (2012b)	$10 (100%) increase relative to $10 baseline	"Managers" assign 25-minute coding	B	N	192	Y	Y	N	Lab	Subjects in lab exp. assigned to role of *managers* decide pay of $20 or $10 for worker; efficiency of work varies
Kube, Marechal, and Puppe (2012)	EUR 7 (19%) increase or Gift-wrapped thermos relative to EUR 36 baseline	Library Book Coding	B	N	117	N	N	N	Field	Interested in the effect of non-monetary gifts
Kube, Marechal, and Puppe (2013)	EUR 5 (33%) increase / EUR 5 (33%) decrease rel. to EUR 15 base	Library Book Coding	B	N	68	N	N	N	Field	Analyze asymmetric effects of pay raises and cuts
Esteves-Sorenson (2018)	$6 (50%) / $8 (67%) / $12 (100%) increase relative to $12 baseline	Data-Entry Task	B	N	162	N	N	N	Field	Examine several potential confounds of earlier studies
Cohn, Fehr, and Goette (2014)	CHF 5 (23%) increase relative to a CHF 22 baseline	Newspaper Distribution	B for Gift Exchange	N	196	N	N	N	Field	Interested whether fairness considerations drive gift exchange-induced effort increases
Gilchrist, Luca, and Malhotra (2016)	$1 (33%) increase relative to a $3 baseline	Entering CAPTCHAs	B	N	230	N	N	N	Field	Examine the effects of restructuring a portion of the wage as an unexpected gift
Panel B. Stated-Effort Experiments										
Fehr, Kirchsteiger, and Riedl (1993)	Firms post wages, workers can reciprocate according to known effort-cost-schedule	Stated Effort	B		35	Y	N	N	Lab	Test the fair-wage hypothesis in a one-shot setting with a fixed efficiency factor of 126. Return to the employer is given by (126-w)e
Brown, Falk, and Fehr (2004)	Wages determined by an open auction and fixed effort-cost schedule for workers	Stated Effort	B		140	Y	N	N	Lab	Third-party enforceability of contracts and identifiability of workers affects long-term relations, with employer return 10e-w
Kessler (2013)	0/5/10 units as a wage in a bilateral gift-exchange game	Stated Effort	B		44	Y	Y	N	Lab	Varies whether the firm is rich (R=1) or poor (R=0) compared to the worker and whether worker's effort is efficient

Notes: This table contains gift exchange real-effort studies (Panel A) and stated-effort laboratory gift exchange experiments (Panel B) that are categorized according to the following categories: (i) whether they have a piece-rate design; (ii) whether they show the return to the employer or the firm , (iii) whether they vary the return to the employer, and (iv) whether they estimate social preferences structurally. Moreover, the task, amount of the gift, whether the experiment has a between- or within subject design, whether the experiment is a lab or field experiment, and some comments on the feature of the experiment are included. Notice that the sample size refers to the number of subjects in the worker role, i.e., in the laboratory experiments it does not include subjects assigned the role of "firms".

Online Appendix Table 2. Summary Statistics and Covariate Balance, Productivity Experiment

Specification:		OLS Regressions				
	Summary Statistics	Output Predictors	Checks of Randomization			
Dep. Var.:		Average Output	Indicator for Order A	Indicator for Positive Gift	Indicator for Negative Gift	Indicator for In-Kind Gift
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Individual Demographics						
Is employed (self-reported)	0.397	2.022	0.070	0.025	0.012	-0.045
	(0.490)	(0.799)	(0.050)	(0.047)	(0.044)	(0.039)
Female	0.522	2.535	-0.099	0.040	-0.010	0.011
	(0.500)	(0.691)	(0.049)	(0.043)	(0.044)	(0.036)
Age 25-34	0.361	2.959	0.062	-0.110	-0.005	0.063
	(0.481)	(0.875)	(0.058)	(0.060)	(0.052)	(0.048)
Age 35-44	0.191	0.989	0.039	-0.122	-0.063	0.052
	(0.393)	(1.184)	(0.068)	(0.067)	(0.064)	(0.054)
Age 45-54	0.128	-2.122	-0.035	-0.082	-0.160	0.083
	(0.334)	(1.256)	(0.093)	(0.082)	(0.066)	(0.057)
Age 55+	0.058	1.305	0.243	-0.002	-0.012	-0.021
	(0.235)	(1.753)	(0.100)	(0.111)	(0.120)	(0.069)
Has donated to charity	0.691	0.183	-0.131	0.004	-0.000	0.026
(self-reported)	(0.463)	(0.946)	(0.059)	(0.057)	(0.048)	(0.043)
Has volunteered before	0.843	1.159	0.096	-0.043	0.042	0.056
(self-reported)	(0.364)	(1.051)	(0.065)	(0.060)	(0.070)	(0.056)
Mean of Dependent Variable		35.19	0.491	0.276	0.283	0.175
R squared		0.097	0.038	0.017	0.015	0.013
N	N = 446	N = 446	N = 446	N = 446	N = 446	N = 446
Panel B. Index of Demographics						
Predicted Effort Based on			0.004	-0.001	0.013	0.001
Demographics (Col. 2)			(0.010)	(0.008)	(0.008)	(0.007)
R squared			0.000	0.000	0.005	0.000
N			N = 446	N = 446	N = 446	N = 446

Notes: Column 1 in Panel A reports summary statistics on the sample of 446 participants in the experiment. Column 2 in Panel A reports the estimates of an OLS regression of average output (over the 10 rounds) on subject characteristics. Based on the estimate in Column 2 we form an index of predicted productivity based on demographics which we use in Panel B. In Columns 3-6 of Panels A and B we regress the assignment to different conditions (order A/B and assignment to the different gift treatments) on the subject characteristics (Panel A) and on the index of characteristics (Panel B). The standard errors are clustered at the session level.

Onl. App. T. 3. Productivity Experiment, Gift Treatments, Robustness

Specification:	OLS Regressions			
Dependent Variable:	Output in Batches 9 and 10			
Panel A. Measure of Output:	Number of Envelopes Stuffed in 20 Minutes			
Sample:	Batch 9	Batch 10	Match	No Match
	(1)	(2)	(3)	(4)
Gift Treatments				
Positive (monetary) gift	1.350	-0.145	0.428	0.778
Treatment	(0.636)	(0.904)	(0.801)	(0.771)
Negative (monetary) gift	0.226	-0.321	0.133	-0.227
Treatment	(0.738)	(0.949)	(0.840)	(0.859)
Positive In-kind (Thermos) gift	-1.024	-1.155	-0.924	-1.256
Treatment	(0.907)	(1.080)	(1.013)	(0.977)
Control				
Average Output Measure	0.833	0.797	0.834	0.796
In Rounds 5-8	(0.024)	(0.035)	(0.032)	(0.028)
Constant	4.446	5.852	4.537	5.761
	(1.022)	(1.459)	(1.313)	(1.192)
R squared	0.668	0.556	0.622	0.595
N	N = 446	N = 446	N = 446	N = 446
Panel B. Measure of Output:	Log of Number of Envelopes Stuffed in 20 Minutes			
Gift Treatments				
Positive (monetary) gift	0.0390	-0.008	0.008	0.023
Treatment	(0.018)	(0.027)	(0.023)	(0.023)
Negative (monetary) gift	-0.009	-0.027	-0.010	-0.026
Treatment	(0.031)	(0.035)	(0.032)	(0.035)
Positive In-kind (Thermos) gift	-0.027	-0.039	-0.030	-0.036
Treatment	(0.026)	(0.034)	(0.031)	(0.029)
Control				
Average Output Measure	0.8510	0.8120	0.8430	0.8200
In Rounds 5-8	(0.029)	(0.036)	(0.031)	(0.034)
Constant	0.4900	0.6310	0.5230	0.5990
	(0.108)	(0.130)	(0.112)	(0.125)
R squared	0.574	0.473	0.535	0.505
N	N = 446	N = 446	N = 446	N = 446

Notes: Estimates from an OLS regression of output (Panel A) and log output (Panel B) 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.

Online Appendix Table 4. Summary Statistics and Covariate Balance, Labor Supply Experiment

Specification:		OLS Regressions							
	Summary Statistics	Extra Stay Predictors	Checks of Randomization					
Dep. Var.:		Extra Stay	Indicator for Med Piece Rate	Indicator for High Piece Rate	Indicator for Monetary Gift	Indicator for In-Kind Gift	Indicator for Early Gift	Indicator for High Return
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Baseline Productivity	3.598	-0.155	-0.012	-0.009	0.001	0.006	0.036	-0.030
	[1.618]	(0.821)	(0.039)	(0.038)	(0.039)	(0.038)	(0.038)	(0.021)
Craigslist	0.367	7.366	-0.041	-0.104	-0.007	-0.085	0.028	-0.134
	[0.483]	(3.424)	(0.150)	(0.188)	(0.145)	(0.157)	(0.176)	(0.089)
Female	0.497	-0.139	-0.063	0.003	-0.085	-0.005	-0.068	0.010
	[0.501]	(2.253)	(0.105)	(0.110)	(0.105)	(0.103)	(0.106)	(0.059)
Age 25-34	0.237	7.231	-0.042	0.001	-0.051	0.111	-0.022	0.011
	[.426]	(3.106)	(0.140)	(0.173)	(0.136)	(0.139)	(0.157)	(0.081)
Age 35-44	0.097	7.753	-0.160	0.131	-0.445	-0.311	0.020	0.088
	[0.296]	(4.701)	(0.192)	(0.217)	(0.211)	(0.222)	(0.203)	(0.122)
Age 45-54	0.053	15.440	-0.042	0.079	-0.100	0.249	0.233	0.190
	[0.225]	(5.834)	(0.294)	(0.318)	(0.287)	(0.257)	(0.266)	(0.152)
Age 55+	0.023	5.306	-0.042	-0.099	-0.600	-0.085	-0.153	-0.157
	[0.151]	(8.018)	(0.294)	(0.352)	(0.381)	(0.330)	(0.342)	(0.209)
Ho: all the coeffs to jointly be equal to zero		p = 0.000	p = 0.981	p = 0.993	p = 0.326	p = 0.495	p = 0.909	p = 0.547
R squared		0.131	0.016	0.011	0.082	0.065	0.028	0.020
N	N = 300	N = 300	N = 100	N = 100	N = 100	N = 100	N = 100	N = 300

Notes: Column 1 in reports summary statistics on the sample of 300 participants in the experiment. Column 2 reports the estimates of an OLS regression of extra stay on subject characteristics. In Columns 3-8 we regress the assignment to different conditions (assignment to different piece rates and assignment to the different gift treatments) on the subject characteristics. Standard deviations in brackets. Standard errors in parentheses.

Online Appendix Table 5. Productivity Experiment, Baseline Social Preferences, Robustness

Dependent Variable:	Log (Number of Envelopes)				Number of Envelopes in a Round			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Baseline Social Preferences:								
Altruism towards Charity	0.011	-0.096	0.095	0.143	0.003	-0.068	0.120	0.149
	(0.047)	(0.028)	(0.041)	(0.029)	(0.034)	(0.026)	(0.044)	(0.031)
Warm Glow towards Charity	0.392	0.311	0.309	0.842	0.336	0.288	0.291	0.816
	(0.064)	(0.060)	(0.063)	(0.097)	(0.052)	(0.051)	(0.063)	(0.100)
Warm Glow towards Grocery Store	0.587	0.648	0.701	1.236	0.543	0.579	0.690	1.181
	(0.072)	(0.114)	(0.069)	(0.099)	(0.069)	(0.095)	(0.068)	(0.102)
Incidental Parameters:								
Cost Function Curvature (γ)	10.790	15.248	9.260	3.650	0.320	0.404	0.257	0.105
	(0.898)	(1.869)	(0.728)	(0.250)	(0.026)	(0.042)	(0.017)	(0.006)
Cost of Effort Function:	Power Cost of Effort Function				Exponential Cost of Effort Function			
Type of timetrend	Quadratic in Rounds	Cubic in Rounds	Indicators for 2, 3, 4, 5-8	Indicators for 2, 3, 4, 5-8	Quadratic in Rounds	Cubic in Rounds	Indicators for 2, 3, 4, 5-8	Indicators for 2, 3, 4, 5-8
Specification	Benchmark		Altruism term does not include piece rate	Partial Warm Glow During Training	Benchmark		Altruism term does not include piece rate	Partial Warm Glow During Training
Std. Deviation of Error Term	0.130	0.129	0.130	0.129	3.939	3.909	3.947	3.916
Std. Dev. of Individual f.e.s * (1/γ)	0.249	0.249	0.249	0.249	8.153	8.147	8.165	8.161
R Squared	0.8369	0.8405	0.8376	0.8401	0.8541	0.8563	0.8536	0.8558
N	3568	3568	3568	3568	3568	3568	3568	3568

Notes: Specifications are from non-linear least squares regressions as in specification in Section 4, with each observation being a worker-round combination. The sample is restricted to the first 8 rounds. The dependent variable is the log of the number of envelopes produced in that round in Columns 1-4 and is the number of envelopes produced in Columns 5-8. The specifications in Columns 1 and 5 allow for a quadratic function in the round number, while the specifications in Columns 2 and 6 allow for a cubic function in the round. The specifications in Columns 3-4 and 6-7 include indicators for rounds 2, 3, 4, and 5-8. Columns 3 and 7 assume subjects do not take into account that being paid more as piece rate lowers the return to the firm. Columns 4 and 8 assume that there is warm glow (but not altruism) even in the training rounds, assumed to be half the size as in the periods in which the envelopes are used. All specifications allow for both pure altruism towards the firm and 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). All specifications include fixed effects for worker i. 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 γ. The latter ratio indicates the variation in the individual productivity. The standard errors are clustered at the session level.

Online Appendix Table 6. Productivity Experiment, Social Preferences with Gift Treatments, Robustness

Dependent Variable:	**Log (No. Envelopes in a Batch)**					**Number of Envelopes in a Batch**				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Baseline Social Preferences										
Social Preferences towards Charity	0.405	0.343	0.457	0.444	na	0.337	0.307	0.447	0.463	0.187
	(0.043)	(0.043)	(0.057)	(0.063)		(0.032)	(0.035)	(0.055)	(0.065)	(0.043)
Social Preferences towards Grocery Store	0.632	0.539	0.732	0.72	na	0.551	0.506	0.704	0.716	0.797
	(0.064)	(0.062)	(0.068)	(0.072)		(0.058)	(0.062)	(0.065)	(0.073)	(0.108)
Reciprocal Social Preferences										
Social Pref. Change -- Positive Monetary Gift	0.2	0.086	0.065	0.374	na	0.098	0.053	0.041	0.314	0.092
	(0.114)	(0.089)	(0.082)	(0.149)		(0.085)	(0.075)	(0.071)	(0.137)	(0.087)
Social Pref. Change -- Negative Gift	-0.016	-0.076	-0.099	0.032	na	-0.018	-0.047	-0.068	0.067	-0.001
	(0.125)	(0.093)	(0.096)	(0.135)		(0.072)	(0.061)	(0.067)	(0.100)	(0.060)
Social Pref. Change -- In-Kind Gift	-0.074	-0.118	-0.144	-0.044	na	-0.103	-0.118	-0.152	-0.079	-0.062
	(0.097)	(0.072)	(0.080)	(0.099)		(0.072)	(0.060)	(0.074)	(0.089)	(0.056)
Estimated Persistence of Social Preferences From Round 9 to 10				0.233	na				0.246	
				(0.251)					(0.248)	
Incidental Parameters										
Cost Function Curvature (γ)	10.637	11.366	9.039	9.439	na	0.316	0.329	0.257	0.263	0.41
	(0.835)	(0.894)	(0.648)	(0.738)		(0.024)	(0.025)	(0.017)	(0.018)	(0.052)
Cost of Effort Function:	Power					Exponential				
Type of timetrend	Quadratic in Rounds	Cubic in Rounds	Alternative Round Indicators	Standard Round Indicators (rounds 2, 3, 4, 5-8, 9-10)		Quadratic in Rounds	Cubic in Rounds	Alternative Round Indicators	Standard Round Indicators (rounds 2, 3, 4, 5-8, 9-10)	
Specification	Benchmark (Warm Glow)			Estimated Decay of Gift Effect	Altruism (instead of warm glow)	Benchmark (Warm Glow)			Estimated Decay of Gift Effect	Altruism (instead of warm glow)
Std. Deviation of Error Term	0.144	0.144	0.144	0.144		4.308	4.302	4.321	4.315	4.365
Std. Dev. of Individual f.e.s * (1/γ)	0.241	0.241	0.241	0.241		8.015	8.008	7.995	8.012	8.013
R Squared	0.7908	0.7923	0.7912	0.7918		0.8192	0.8197	0.8182	0.8187	0.8144
N	4460	4460	4460	4460		4460	4460	4460	4460	4460

Notes: Specifications are from non-linear least squares regressions, with each observation being a worker-batch combination. The sample includes all 10 batches. The dependent variable is the log of the number of envelopes produced in that round in Columns 1-5 and is the number of envelopes produced in Columns 6-10. All specifications include fixed effects for worker i. Columns 3 and 8 include indicators for batches 2, 3, 4, 5-10. The estimated coefficient on batch 2 is restricted to equal one half of the estimated coefficient in batch 3. Columns 4 and 9 allow for a decay of the warm glow gift parameter in batch 10, to equal deltaa_gift. Thus, delta=1 indicates no decay, delta=0 indicates full decay. The delta does not apply to batch 9. Columns 5 and 10 estimate a model with pure altruism instead of warm glow. The model in Column 5 did not converge. 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 γ. The latter ratio indicates the variation in the individual productivity. The standard errors are clustered at the session level.

Online Appendix Table 7. Labor Supply Experiment, Social Preferences, Robustness

Estimation:	Maximum Likelihood, Accounting for Censoring at 0 and 60 Minutes	Minimum Distance Estimation		Maximum Likelihood, Accounting for Censoring at 0 and 60 Minutes	Minimum Distance Estimation	
		Moments 0', 1'-5', 6'-10', 11'-15',…, 60'	Moments 0', 1'-30', 60'		Moments 0', 1'-5', 6'-10', 11'-15',…, 60'	Moments 0', 1'-30', 60'
Dependent Variable:	Number of Extra Minutes	Number of Minutes Worked		Log (No. Extra Minutes)	Log (No. Minutes Worked)	
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline Social Preferences						
Social Preference towards Employer	0.812	100*	100*	0.400	100*	100*
	[0.002, 100]	[0.000, 100]	[0.000, 100]	[0.001, 100]	[0.000, 100]	[0.000, 100]
Social Preference Change - High Return for Employer	0.109	0.073	0.107	0.129	0.075	0.107
	(0.093)	[-0.180, 0.361]	[-0.184, 0.965]	(0.104)	[-0.191, 0.370]	[-0.175, 0.684]
Reciprocal Social Preferences						
Social Preference Change -- Monetary Gift	0.303	0.435	0.464	0.377	0.435	0.468
	(0.139)	[0.000, 0.735]	[-0.000, 3.483]	(0.168)	[0.001, 0.728]	[0.002, 1.399]
Social Preference Change -- In-Kind Gift	0.181	0.227	0.215	0.204	0.226	0.215
	(0.135)	[-0.158, 0.561]	[-0.207, 0.585]	(0.155)	[-0.163, 0.531]	[-0.205, 0.595]
Social Preference Change -- In-Kind Gift, Early	0.360	0.489	0.508	0.427	0.489	0.511
	(0.144)	[0.027, 0.853]	[0.016, 7.492]	(0.219)	[0.035, 0.931]	[0.057, 2.143]
Incidental Parameters						
Cost Function Curvature (γ)	0.007	0	0	0.146	0	0
	[0, 0.077]	[0, 0.152]	[0, 0.312]	[0, 0.983]	[0, 20.30]	[0, 20.61]
Std. Deviation of Error Term	45.121	54.995	46.711	3.905	0.369	0.330
	(4.265)	(5.220)	(5.279)	(0.312)	(0.034)	(0.035)
Cost of Effort Function:	Exponential			Power		
Log Likelihood / minimum distance	-637.99	0.39	0.35	-399.19	0.37	0.33
N	300	300	300	300	300	300

Notes: Bootstrap standard deviations are in parentheses and 95% bootstrap confidence intervals are in brackets. Columns 1 and 4 report the maximum likelihood estimates using the number of extra minutes worked, not including the required initial 120 minutes. For Column 4, since Log (0 minutes) is undefined, we left-censor the number of extra minutes worked at 1 minute. Minimum distance estimation in Columns 2-3 and 5-6 use the identity matrix as the weighting matrix. The moments used in Columns 2 and 5 are: Share stay 0; Share stay 1-5; … ; Share stay 51-55; Share stay 60. The moments used in Columns 3 and 6 used are: Share stay 0; Share stay 1-30; Share stay 60.

Online Appendix Table 8. Calibration of Reciprocity in Select Gift Exchange Papers

	Gift in Treament Condition	Task Assigned	% Effort Change With Gift	Implied Percent Warm Glow Change (Reciprocity) Due to Gift		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Findings from this paper:						
DellaVigna, List, Malmendier, Rao (2019)	Pay Increase from $7 to $14	Folding Charity Envelopes	2%	34%	.	.
	Pay Decrease from $7 to $3		-2%	-9%	.	.
	Gift of Thermos		-3%	-21%	.	.
Assumption about Cost Function:				Power Cost Function		
Estimated Curvature γ				9.4(0.9)***		
Implied Elasticity				0.11		
Panel B. Selected Previous Findings on Gift Exchange in Field:						
Gneezy and List (2006) *Study 1*	Pay Increase from $12 to $20	Library Book Coding	27% (first 90 min)	846%	230%	61%
Gneezy and List (2006) *Study 2*	Pay Increase from $10 to $20	Door-to-door Fundraising	72% (first 3 hours)	16267%	1405%	196%
Kube, Marechal, and Puppe (2012) *Non-monetary gift condition*	Gift of Thermos	Library Book Coding	25%	715%	205%	56%
Kube, Marechal, and Puppe (2012) *Monetary gift condition*	7 Euro raise (from 36 euro pay)	Library Book Coding	5%	58%	28%	10%
Kube, Marechal, and Puppe (2013)	Pay cut from 15 to 10 euro/hr	Library Book Coding	-20%	-88%	-67%	-36%
Gilchrist, Luca, and Malhotra (2016)	Pay increase from $3 to $4	Entering CAPTCHAs	18%	374%	129%	39%
Cohn, Fehr, and Goette (2014)	Pay increase from 22 to 27 ChF	Newspaper Distribution	3%	32%	16%	6%
Esteves-Sorenson (2018)	Pay Increase from $12 to $20	Enter data	2%	20%	10%	4%
Assumptions about Cost Function:				Power Cost Function		
Assumed Curvature γ				9.4	5.0	2.0
Implied Elasticity				0.11	0.20	0.50

Notes: This table revisits some of the findings in the previous gift exchange experiments in the field, with summary of the key gift treatments and findings in Columns 1-3. Panel A summarizes the effects from this paper: Column 2 reports the findings from Table 4, Column 3, Panel B (on log output). Column 3 reports the results from Table 5, Column 1, taking the ratio of the estimated warm glow change to baseline warm glow. For example, for the positive monetary gift .151/.443=34%. In Panel B we revisit some classic experiments on gift exchange in the field. In Columns 4-6 we compute the implied percent increase in altruism or warm glow implied by the effort increase (or decrease), for a calibrated value of the elasticity of effort. The calibration holds for a power cost of effort function, which is characterized by constant elasticity. Column 4 uses the elasticity estimated for our task (Table 5, Column 1). Columns 5 and 6 report the results assuming higher elasticities.

Online Appendix Table 9. Published Real-Effort Experiments and Piece-Rate Design

Paper	Topic	Experiment (Lab / Field / Online)	Type of Real Effort Task	Piece-Rate Design? (Y/N)	Number of Piece Rates	Notes
	(1)	(2)	(3)	(4)	(5)	(6)
Real Effort Experiments Published in Top-5 Journals from 1999 to 2018						
Gneezy, Rustichini and Niederle (2003)	Competitive Preferences	Lab	Solving mazes	N		
Gneezy and List (2006)	Gift Exchange	Field	Data Entry; Fundraising	N		
Ariely, Bracha and Meier (2009)	Image Motivation	Lab, Field	Typing; Biking	N		Participants face either no incentives or non-linear incentives, but not piece rates
Carpenter, Matthews and Schirm (2010)	Tournaments and Office Politics	Field	Stuffing Envelopes	N		
Abeler, Falk, Goette and Huffman (2011)	Expectations and Effort Provision	Lab	Count number of zeros	N		
Dohmen and Falk (2011)	Incentives and Sorting	Lab	Multiplying numbers	N		
Gill and Prowse (2012)	Disappointment Aversion	Lab	Slider task	N		Participants are stochastically rewarded, with probability of reward increasing in the difference between own effort and a partner's effort. The reward size is varied, but the incentives are not known piece rates
Kube, Marechal and Puppe (2012)	Gift Exchange	Field	Cataloguing Library Books	N		
Augenblick, Niederle and Sprenger (2015)	Time Preferences and Effort	Lab, Online	Data transcription; Tetris	Y	5	Variation in the exchange rate of work between different time periods helps identify the cost of effort
DellaVigna and Pope (2018)	Effort Motivation	Online	Typing	Y	4	

Notes: This table contains real-effort studiespublished in the *American Economic Review, Econometrica,* the *Journal of Political Economy*, the *Quarterly Journal of Economics,* and the *Review of Economic Studies* between 1999 and 2018. We search papers using a search of Google Scholar for papers in these journals and year with the word "real effort" in the text of the paper. We then exclude papers that do not have this feature. It categorizes whether the papers include randomized variation in piece-rates. Two out of ten such published papers we identified include a "piece-rate design".

A Online Appendix A - Related Literature

Online Appendix Table 1 summarizes some of the most related papers in the literature. We identify key features of related papers: (i) the piece-rate design (Column 4); (ii) the sample size (Column 5); (iii) the structural estimation of the social preference parameters (Column 8); and (iv) whether the return to the firm is made explicit and varied experimentally (Columns 6 and 7). We also indicate whether the gift exchange variation is between subjects or within subjects (Column 3) and whether the experiment takes place in a field setting or in the laboratory (Column 9). Panel A documents the most relevant real-effort experiments on gift exchange, including some executed as laboratory experiments, so long as the "work" is real effort and not stated effort.

Regarding the sample size (Column 5), our paper is the real-effort field experiment with the largest sample size this far, though there are other studies that are well-powered (which we somewhat arbitrarily indicate with a sample size above 100). Column 8 documents the fact that there is only one other paper which attempted structural estimation of social preferences in a gift exchange set-up in the field, Bellemare and Shearer (2011). Bellemare and Shearer (2011) has a very nice estimation set-up, which we partly borrow from, such as a power cost of effort, and individual fixed effects. The table clarifies important differences of our work relative to Bellemare and Shearer (2011): (i) (*sample size*) Bellemare and Shearer (2011) estimates the gift exchange effect on a sample size of just 18 workers; (ii) (*within-subject identification*) The identification of gift exchange is based on time-series variation: all workers on a particular day were given a "gift," with no control group on that date. Thus, the identification is based on comparing worker effort on those days versus in the days before (that is, is within subject); (iii) (*returns*) the workers do not know the explicit return to the firm of their effort.

The table also highlights another distinguishing feature of our design: whether the return to the firm was made explicit (Column 6) and varied in the experiment (Column 7). As the table makes clear, few real-effort experiment papers did so (and the list omits a few other gift exchange in the field papers which also do not do so). One of the two Gneezy-List experiments arguably made returns explicit, as the workers were raising money for charity and thus could know the return to their effort (though the return itself was not varied). Also, Englmeier and Leider (2012a) vary the return to the firm by telling people in one case that the experimenters would get "a substantial bonus" if 50% of the work was done by a deadline. Hennig-Schmidt, Sadrieh, and Rockenbach (2010) provide more information on the return to the employer in one of their treatments, and find evidence suggestive of gift exchange only when the return is made clear. Both experiments provide suggestive evidence on the effect of returns, given the relatively small sample size.

A study that both informs workers of the return to the firm, and varies returns across treatments, is Englmeier and Leider (2012b). The paper employs a real-effort task and it has a sizable sample (N = 192). Interestingly, as in our paper, there is no statistically significant response to a gift from the "manager," nor does the response appear to interact with the return to the "manager." We should point to two key differences of this very nice study relative to our work: As the authors themselves emphasize, it is a laboratory experiment, and the "managers" are laboratory subjects assigned to the "manager" role. And this paper does not attempt a structural estimation.

Panel B of the table also shows several of the design features in stated-effort laboratory experiments that our study aims to introduce in the field experiments. Most importantly, the return to the "firm" is made salient, and occasionally also varied. Indeed, a key point in our paper is to show that one can put together the pieces that allow for estimation of preferences in a field setting, as pioneered in the laboratory for stated-effort gift exchange experiments. In this way, our design aims to bridge the gap between the laboratory and field studies, as we say in the paper.