Revisiting Gift Exchange: Theoretical Considerations and a Field Test

Constança Esteves–Sorenson*

Yale University

Rosario Macera†

Universidad Católica de Chile

August 2013

Abstract

There is conflicting laboratory and field evidence on the effectiveness of gift exchange—the exchange of above-market wages for above-minimal effort—as an incentive mechanism. We investigate this conflict by theoretically identifying the factors dampening gift exchange in the field—habituation to the gift, fatigue, and small gift size—and by subsequently implementing a field experiment tailored to address them as well as three others: selection of better workers, existence of an effort ceiling, and ambiguous kindness signals. Still, we find no evidence of gift exchange in the field. A subsequent laboratory experiment investigating whether this is driven by a want of reciprocal workers, shows that a substantial portion behaved prosocially in laboratory, but failed to engage in gift exchange in the field. All workers, however, responded to piece rates. Our results favor a classical model of preferences.

JEL Codes: D03, J31, J41, M52.

*constanca.esteves-sorenson@yale.edu.

†rosario.macera@uc.cl. We thank Ben Arsenault, James Floman, Julia Levinson, Tiffany Lin, Andrew Pearlmutter and Beau Wittmer for excellent research assistance. We also thank Stefano DellaVigna, Uri Gneezy, Mitch Hoffman, Lisa Kahn, Botond Kőszegi, Kory Kroft, Michel Maréchal, Matthew Rabin, Frédéric Schneider, Olav Sorenson and Roberto Weber for helpful discussions and suggestions at different stages of this project, and seminar participants at the University of California at Berkeley, University of Warwick, Yale University, and Universidad Católica de Chile for helpful comments. This research was partially funded by Whitebox grants in Behavioral Economics at Yale University and the Institution for Social and Policy Studies at Yale University.

1 Introduction

First proposed by Akerlof (1982), the gift exchange hypothesis, postulating that workers reciprocate above-market wages with above-minimal effort, has garnered mixed empirical support. Laboratory experiments suggest that gift exchange is a powerful incentive mechanism, generating not only involuntary unemployment but also large effort increases in one-shot interactions (e.g., 300%).[1] Persuasive field evidence, however, has found more modest effects: small effort increases (25%-70%) that waned within hours (Gneezy and List (2006)), and for smaller wage increases, effort responses that were not statistically significant (Kube, Maréchal, and Puppe (2012)).[2] Despite gift exchange's importance, the reasons why it is curbed in the field are still unknown.

To shed light on the conflicting laboratory and field evidence on gift exchange, we start by identifying the main factors that could dampen it in the field: workers habituating to the gift, fatigue, and small gift sizes. We jointly formalize these factors in a novel model of expectation-based reference-dependent reciprocity and implement an adequately powered field experiment to test whether gift exchange emerges in the field once we account for them. Despite addressing these factors as well as four others—selection of better workers, existence of an effort ceiling, ambiguous kindness signals, and a want of reciprocal subjects—we find no evidence of gift exchange. Further, we find that prosocial behavior in the laboratory does not translate into that in the field. Workers, however, are sensitive to incentives as they do respond to piece rates.

Our model's focus on these three main factors stems from their ability to explain the strong evidence of gift exchange in the laboratory, but not in the field. First, workers in the field are surprised with a wage increase immediately before they exert effort in a lengthy several-hours task; their reciprocal response can therefore wane as they habituate to the gift in this time span.[3] To the

[1]Gift exchange games document large reciprocal effort magnitudes. For instance Fehr, Kirchsteiger, and Riedl (1993) found that when employers offered wages 140% above the market-clearing level, employees responded with effort 300% above the minimal level. For further laboratory evidence see Fehr, Kirchler, Weichbold, and Gächter (1998), Gächter and Falk (2002), Hannan, Kagel, and Moser (2002), Brandts and Charness (2004), Charness (2004), Maximiano, Sloof, and Sonnemans (2007). Broad features of these games are: anonymous pairing of employers and employees; one-shot interactions lasting a few minutes before re-pairement; employers offering wages first and the employees choosing effort second with choices jointly determining monetary payoffs; common knowledge of cost of effort; payoff functions and wage distributions; wages and effort exchanged in experimental currency.

[2]Though there have been tests for gift exchange in the field (e.g., Bellemare and Shearer (2009)), Gneezy and List (2006) is the most persuasive despite its small sample because they address the two major confounds to gift exchange: threat of dismissal (Shapiro and Stiglitz (1984))—by advertising tasks as a one-time job—and selection of higher quality workers (Weiss (1980))—by contracting at the going market wage. In a first experiment, with a data-entry task, they find that a 67% wage raise increased the number of records entered by 25%, but it disappeared after three hours. In the second experiment, with a fundraising task, a 100% wage raise increased funds raised by 70% only in the first three hours. They concluded that "... there are signs of significant gift exchange in the data ..." (page 1370).

[3]This hypothesis has been speculated in the literature: e.g., Gneezy and List (2006, page 1377) state "an inter-

contrary, subjects in the laboratory never habituate as experimental rounds last only a few minutes. Second, because work in field settings is usually lumped in one day, initial reciprocal effort can fatigue workers resulting in a subsequent waning of productivity.[4] To the contrary, subjects in the laboratory never fatigue over time, as they restart each short round with the same cost-of-effort table written by the experimenter. Third, absence of gift exchange in the field may result from the gift being too small to compensate the worker for the extra effort. To the contrary, cost of effort in the laboratory, which is paid in currency, might not be as onerous as in the field, rendering it easier to respond to gifts.[5] Further, our model predicts that habituation and gift size are irrelevant for classical agents, who exert minimal effort independently of the magnitude of the gifts and whether they are surprising or expected.

Our experimental design starts by addressing these three main untested factors identified by our model. We hire undergraduates for a one-time data-entry task at the going market wage as in Gneezy and List (2006).[6] First, to allow workers to habituate to the gift, we split the work over three weekly shifts.[7] Second, in contrast to previous studies usually packing all work into a single day, this split also ensures that fatigue is not causing a waning of gift exchange as workers can rest between shifts. Third, we randomly assign workers to a Control group and three other main treatments. To address the role of gift size and habituation to the gift, we offer workers in the "OneGift" treatment a surprising and permanent 67% increase at their hourly wage in the beginning of shift one. This gift mimics that in Gneezy and List (2006) and thus, given our similar sample and task, we address a priori whether it is large enough to elicit effort. To further investigate the relevance of the gift size, workers in the "TwoGifts" treatment receive a 50% permanent wage increase in the first shift, followed by a surprising additional increase to 100% in their third shift. As an additional test of reciprocal behavior in the field, we examine the extent of negative reciprocity by testing the effect of unmet gift expectations. Therefore, workers in the "Gift-NoGift" treatment receive the same wage increase as those in the TwoGifts treatment in shifts one and two, but are informed in shift

pretation of our findings [temporary gift exchange] is that our agents' effort levels may simply be adapting to new referentials." We added the text in brackets. Likewise, Bewley (1998, page 478) has also stated that "pay increases normally give only a small and temporary lift to morale."

[4]Gneezy and List's (2006) results suggest that non-separability in the cost of effort—workers become increasingly fatigued—could cause waning of the gift exchange. In their fundraising task where work was split over two consecutive days, the treatment group exerted less effort than the Control on the subsequent day, though it was statistically indistinguishable due to the small sample size (4 and 9 subjects, respectively).

[5]In fact, gift exchange in the laboratory is sensitive to the players' payoff structure (Fehr, Klein, and Schmidt (2007)).

[6]Contracting for a one-time job addresses the major confound of workers increasing effort in response to gifts to avoid dismissal (Shapiro and Stiglitz (1984)) rather than to reciprocate the gift.

[7]We assume, conservatively, that agents habituate to the gift within one week as discussed in Section 2.

three that the gift is "no longer feasible", and thus merely receive the contract wage.[8,9]

Our experimental design addresses four additional factors that can dampen gift exchange in the field, but which have not been jointly addressed in past field research: selection of better workers (who can operate closer to the upper bound of effort and thus not respond to gifts), existence of an effort ceiling in the task (it is too costly to increase effort), ambiguous kindness signals (which can dampen reciprocity if agents do not interpret the wage increase as an intentionally kind action (Charness (2004)) and a want of subjects with social preferences in the worker pool. First, we address selection by hiring at the market wage, as hiring at a higher wage could attract abler workers (Weiss (1980)). Second, to test for an effort ceiling, those in our fifth treatment "PieceRate" receive a surprising per-record piece rate in addition to the contract's market wage in all three shifts. This also allows us to compare the efficiency of gift exchange relative to classical incentives.[10] Third, to ensure that the wage increases are perceived as unambiguously kind and unconditional, they are dispensed in an envelope embossed with the phrase "A Gift for You" immediately before workers start working on each shift. Fourth, we measure the strength of our workers' social preferences by having them play Sequential Prisoner's Dilemma games as in Burks, Carpenter, and Goette (2009) and Schneider and Weber (2012) after the conclusion of the field experiment.

We report three main findings, which favor the predictions of the classical model over social preferences. First, despite addressing the seven factors potentially curbing reciprocal effort in the field and despite our large sample, we find no compelling evidence of gift exchange. Wage raises of 50%, 67% and 100%, whether expected or unexpected, do not engender higher output (number of records) or effort (number of keystrokes) than that of the Control group, which receives no gift. Instead, they elicit slightly fewer records and keystrokes, though these estimates are not statistically significant, while the quality of the output (proportion of correct words inputted) remains approximately unchanged. Further, we cannot document negative retaliation due to expected-but-unmet gifts. Workers in the Gift-NoGift treatment do not withdraw effort versus the Control in the third shift after the removal of the gift.

[8]This treatment, which also stems from our theoretical model, was partially motivated by the finding in Kube, Maréchal, and Puppe (forthcoming) that cuts in the forecasted contract wages reduce effort.

[9]We conducted two additional robustness treatments with the gift of $8 per hour which we describe in Section 2. In all treatments workers received their contract pay of $72 at the end of shift three.

[10]Kube, Maréchal, and Puppe (forthcoming) is the only field paper on gift exchange attempting to ascertain whether the lack of gift exchange was due to an effort ceiling, but their evidence is inconclusive. They hired via a piece rate contract (instead of via the flat market wage contract, as in our case), which induces sorting of higher ability workers (Lazear (2000)). Therefore, it is unknown whether their gift exchange sample, hired at a flat wage, faced an effort ceiling.

Second, an effort ceiling does not account for the absence of gift exchange as, in contrast to gifts, piece rates do motivate workers. Piece rates trigger an increasing level of effort—which becomes marginally statistically significant by the third shift—as well as output, leaving quality generally unchanged. They also more efficiently bring forth effort relative to monetary gifts: though the average expenditure, in excess of the contract wage, in our most expensive gift treatments was almost triple that with piece rates, this did not elicit higher effort whereas the piece rate did.

These two findings contribute to the incentives literature by addressing the effectiveness of gift exchange as an incentive device. The absence of gift exchange in our large sample indicates that persuasive (even if short-lived) small sample field evidence on gift exchange from tests similar to ours (e.g., Gneezy and List (2006)) is fragile, implying that subtle changes in field conditions affect reciprocal effort.[11] This suggests that other explanations for incentives based on above-market wages, such as the threat of dismissal (Shapiro and Stiglitz (1984)) or the selection of better workers (Weiss (1980)), described in Gibbons and Waldman (1999), are more likely.

Further, our findings contribute to the debate on the conflict between laboratory and field evidence on gift exchange by addressing, among others, important lingering issues with the field evidence. One such issue was whether lack of gift exchange was due to gifts being too small to elicit effort. Indeed, in two exploratory papers, Kube, Maréchal, and Puppe, (2012, forthcoming) found that 20% and 33% wage increases for a similar task did not elicit extra effort, whereas 67% gifts in Gneezy and List (2006) did. Further, in some studies (e.g., Kube, Maréchal, and Puppe (forthcoming)) the expected contracting wage is much larger than market wage, possibly inducing the selection of higher productivity workers. Last, whether workers faced an effort ceiling was also untested.[12] Our approach to reconciling the conflicting gift exchange evidence, therefore, departs from other emerging literature on this topic. In contrast to our paper, this literature does not focus on exploring all of the factors that could hinder gift exchange in existing and emerging tests in the field, but rather explores how the addition of miscellaneous contextual elements to gift exchange—such as agents' perceptions of the principal's surplus or of the fairness of the wage—elicit effort. The closest paper to ours in this vein is by Hennig-Schmidt, Sadrieh, and Rockenbach (2010) who test whether salary increases combined with peer salary or principal surplus information generates reciprocal effort, but

[11] There is some evidence of this fragility in the laboratory as well. For example, Charness, Frechette, and Kagel (2004) find that an innocuous change in the laboratory setting—whether subjects in the gift exchange game receive a comprehensive payoff table with the instructions—dampens gift exchange.

[12] See the discussion in footnote 10.

the evidence is inconclusive.[13,14] Though these tests adding miscellaneous factors to gift exchange are valuable avenues for research, they stray from testing the gift exchange hypothesis per se—that workers unconditionally respond to gifts with excess effort (Akerlof (1982))—which is the object of the conflict.[15]

Our third result is that we find no correlation between prosocial behavior in the laboratory and in the field. Though a substantial portion of our workers behaved prosocially in the sequential Prisoner's Dilemma games—on which the laboratory evidence on gift exchange is based—they did not reciprocate the gift with more effort, but they responded to the piece rate. This finding adds to the heated debate on the external validity of laboratory findings, in particular on social preferences. Indeed, while some have found support for a correlation between prosocial behavior in the laboratory and that in the field (e.g., Benz and Meier (2008), Burks, Carpenter, and Goette (2009), Baran, Sapienza, and Zingales (2010)) others have found mixed support (e.g., Karlan (2005)), or none (e.g., List (2006), Stoop, Noussair, and van Soest (2012)).

Finally, we also add to the theoretical literature on social preferences by offering a novel model outlining how expectations may interact with social preferences, formalizing the long held conjecture that reciprocity might be short-lived (Adams (1963), Bewley (1998), Kahneman, Knetsch, and Thaler (1986), Falk (2007), Kube, Maréchal, and Puppe (2012)).[16]

2 The Field Experiment Design

This section describes the field experiment and how it addresses all the factors that could hinder gift exchange in the field. Our design builds on the main tests of gift exchange in the field (Gneezy and List (2006) and Kube, Maréchal, and Puppe (2012; forthcoming)) allowing us to also address the lingering issues with this field evidence.

[13]Their design did not allow them to test how peer compensation interacted with gift exchange. Nor did it enable them to explore the role of gift size in their task or the selection of higher ability workers, due to the above-market-wages contract, in hindering reciprocal effort. As for the increase in productivity when workers receive wage increases together with surplus information, it is similar to that in the Control where workers receive no wage increase or surplus information.

[14]Some of the contextual factors explored in the literature are, for example, the perception of fairness of the wage in tests of the fair wage-effort hypothesis of Akerlof and Yellen (1990) (Cohn, Fehr, and Goette (2012)) and the perception of how much the principal cares about the agent's output (Englmaier and Leider (2012)).

[15]Akerlof, (1982, page 544) states: "On the worker's side, the "gift" given is work in excess of the minimum work standard; and on the firm's side the "gift" given is wages in excess of what these women could receive if they left their current jobs." This claim is unconditional in regard to the perceptions of fairness of the wage (as is the case of the fair wage-effort hypothesis in Akerlof and Yellen (1990)), or in regard to the principal's surplus.

[16]For example, Falk, (2007, page 1510) notes that "One important question, for example, is whether a gift exchange relation can be repeatedly initiated. [...] Once donors get used to getting gifts, they may no longer feel obliged to repay them."; Kube, Maréchal, and Puppe, (2012, page 1657), also point out that "In a dynamic context, workers might become used to receiving gifts on a regular basis and respond less ...".

2.1 The field experiment

We hired a large sample of 194 undergraduates across two campuses in Connecticut to conduct a one-time, six-hour data-entry task split into three, two-hour shifts exactly one week apart, subsequently varying their wage increases and expectations about them. Contracting for a one-time job addressed the major confound of workers increasing effort when receiving a gift to prevent dismissal (Shapiro and Stiglitz (1984)) instead of doing so to reciprocate the gift. Further, separating the six-hour task into three, two-hour weekly shifts ruled out the role of non-separability in the cost of effort and thus fatigue in the ebbing of gift exchange. This, in turn, allowed us to quantify the role of habituation, which we assumed occurred within one week.[17,18]

Similar to previous research, workers built a library database, allowing us to calibrate whether the gift size was large enough to compensate workers for the cost of effort. The gift size is an important concern as the cost-of-effort function for the task is unknown. Uncovering it would require exposing our workers to different piece rates, thus abandoning our between-subjects design, which we view as one of the strengths of our field experiment.[19] To ensure that our gifts would be large enough to elicit effort without abandoning our between-subjects design, we mirrored the structure of the most compelling evidence on gift exchange (Gneezy and List (2006)), by hiring a similar worker sample (US undergraduates) to work on a similar task—entering the author's name, article title, journal, year, volume, issue, and pages numbers into a Windows-based academic article software—while offering the same or higher wage increases (67% or 100% increases from a contract hourly salary of $12 per hour).[20]

Finally, we measured effort by the number of keystrokes pressed, such as characters, backspaces, spaces; output by number of records; and output quality by the number of error-free words entered.[21]

[17] Indeed, an open question in Gneezy and List (2006) was whether the waning of gift exchange in their six-hour data-entry task after three hours could stem from a non-separability in the cost of effort or from workers habituation to the gift as discussed in the Introduction. By separating the task into weekly two-hour shifts, we conservatively ensure separability in the cost of effort: the worker recovers between each weekly shift.

[18] There is little systematic research on how rapidly expectations adjust. A review by Frederick and Loewenstein (1999) on experiments on hedonic adaptation suggests that habituation varies across domains. Some studies suggest that it occurs immediately (e.g., Gill and Prowse (2012)), or within hours (e.g., Card and Dahl (2011)); whereas others point to subjects needing more than a few hours to adapt (e.g., Kube, Maréchal, and Puppe (forthcoming)). For our study we assume that one week is enough time for subjects to adjust their expectations. That is, the information that they will be paid in the following week is no longer a surprise when they do indeed receive the wage increase at the beginning of their shift in the following week.

[19] As we discuss later, the between-subjects design may be crucial for studying gift exchange because, in this design, each subject knows that the principal does not know how much he can produce in the absence of a gift, so he has more freedom to behave selfishly.

[20] In the book data-entry task in Gneezy and List (2006), US undergraduate students entered the title, author, publisher, ISBN number and year of publication. They input similar data in our task.

[21] These data originated from software time-stamping each record and the number of keystrokes to generate it, of

Workers who responded to the campus flyers advertising the task at $12 per hour—the going market wage for this type of job—were randomly assigned to a Control group, four main treatments and two robustness treatments. Hiring at the market wage addresses another major confound of attracting more able workers (Weiss (1980)), which could be closer to an effort ceiling and hence more unresponsive to incentives.[22,23]

To establish the baseline level of effort in the absence of an excess wage, the 47 workers in the ("Control") worked their three two-hour shifts and received the agreed $72 at the end of their third shift.

To explore the extent of gift exchange in the field and that of habituation to the gift in the supply of effort, the 23 workers in the first treatment ("OneGift") received the pleasant news of a wage increase of $8 per hour, to $20 (a 67% raise), for the duration of the contract, immediately before beginning the first shift.

To further investigate the extent of gift exchange and the role of the gift size, the 23 workers in the second treatment ("TwoGifts") also received a pleasant surprise of a substantial, though smaller, hourly wage increase of $6, to $18 (a 50% raise), for the duration of the contract immediately before starting the first shift. However, in shift three, they received the surprise of an additional $6 hourly wage increase (to $24 per hour).

To examine the role of negative departures from expectations in the supply of effort, the 22 workers in the third treatment ("Gift-NoGift") received the same pleasant surprise payments and information in shifts one and two as those in the previous TwoGifts treatment. But, instead of receiving an additional $6 hourly wage increase in shift three, they received a wage cut in this amount, back to the contracted $12 per hour. These workers enjoy the same compensation and information structure in shifts one and two as those in the previous TwoGifts treatment. The only change—the wage cut—is in the third shift.

To ascertain whether any absence of gift exchange could be due to an effort ceiling, the 32 workers in our fourth treatment (PieceRate) received a surprise per-record piece-rate offer for the duration of the contract at the beginning of their first shift, instead of a fixed wage increase, as in the previous treatments. The piece rate was as follows: $\$0 \times x$ if $x \leq 70$; $\$0.05 \times x$ if $70 < x \leq 110$; $\$0.10 \times x$ if

which subjects were unaware.

[22]For example, in other contexts where effort has been proxied by the number of hours worked, higher ability workers, with lower cost of effort, who are closer to the 24-hour threshold, cannot increase their number of hours worked beyond 24 per day.

[23]Source of the market wages: Campus A—student pay scale; Campus B—PayScale.com, Salary.com, accessed 06/2012.

Main Treatments	Beginning of Shift One	Shift Two	Shift Three	End of Shift Three
Control	–	–	–	$72
OneGift	2x $8=$16	2x $8=$16	2x $8=$16	$72
TwoGifts	2x $6=$12	2x $6=$12	4x $6=$24	$72
Gift-NoGift	2x $6=$12	2x $6=$12	-	$72
	End of Shift One	Shift Two	Shift Three	
PieceRate	Piece Rate	Piece Rate	Piece Rate	$72

Table 1: Summary Payments for Main Gift Treatments and PieceRate Treatment

$110 < x \leq 140$; and $\$0.20 \times x$ if $x > 140$ where x is the number of records entered on the shift. Workers collected any piece rate earnings at the end of each shift.

To ensure all subjects interpreted the excess wage as non-contingent on performance and as a voluntary and kind action by the principal (as the principal's volition can affect reciprocal effort (Charness (2004)), they received the excess wage in a gift envelope embossed with the phrase "A Gift for You" immediately before starting the corresponding shift. Finally, subjects in all treatments received their contract pay of $72 at the end of shift three. Table 1 summarizes the weekly shift payments across the different treatments.

In addition to these five main treatments, we implemented two additional robustness treatments to address potential confounds with reciprocity. First, to isolate the effect of surprises on effort and from confounds arising from workers believing—erroneously—that the relationship would continue, the 25 workers in our sixth treatment ("AnticipatedGift"), received the pleasant surprise of an $8 per hour wage increase, to $20 (a 67% raise) for the duration of the contract—just as in the OneGift treatment—with the exception that they received the news of a wage increase *one week in advance* of the first shift.

Second, to explore whether reciprocal effort observed in past research (e.g., Gneezy and List (2006)) could blend subjects' social preferences with their perception that the gift is contingent on performance, the 22 workers in our seventh treatment ("PromisedGift") received the promise of a wage increase equal to that in the OneGift treatment—a permanent raise of $8 per hour to $20—immediately prior to starting their first shift, but only receive this payment at the *end of the last shift*.

Similarly to the main treatments, the wage increases were received in the embossed gift envelope and the contract wage of $72 was paid at the end of the third shift. See Table 2 for a summary of

Robustness Treatments	Beginning of Shift One	Shift Two	Shift Three	End of Shift Three
AnticipatedGift	2x \$8=\$16	2x \$8=\$16	2x \$8=\$16	\$72
PromisedGift	-	-	-	\$72+\$48

Table 2: Summary Payments for Robustness Gift Treatments

the weekly payments across the two robustness treatments.

To avoid any extraneous confounds to our experiment, we prevented researcher and peer effects. All subjects interacted with the same research assistant who was blind to the study's hypothesis so that differences in treatments are not due to unobserved differences in research assistants or demand effects. Subjects also worked alone so that their productivity would not be influenced by that of other workers (e.g., Falk and Ichino (2006), Mas and Moretti (2009)).

After the field experiment concluded, we surveyed our workers to ascertain our sample's heterogeneity in social preferences. Specifically, our field experiment occurred during a pilot and seven subsequent legs between the summer of 2011 and the fall of 2012 and was followed by the online survey with laboratory games in early 2013. See Figure B.1 for an overview of the timing of the treatments across campuses and legs.[24]

Last, we calibrated the sample size ex-ante so that we would have sufficient statistical power, thus addressing the concern that the failure of prior field experiments to document gift exchange could be due to lack of power (e.g., Cohn, Fehr, and Goette (2012)). Namely, we selected the sample size so that we would have at least 80% power to reject the null hypothesis of "no change in effort" in response to gifts in favor of the one-sided alternative of "an increase in effort" at the 5% level. The power calculations focused on the treatments where the conditions facing our subjects in session one should be similar to those in previous research and we achieved 94% power when simply comparing the Control with our pooled OneGift and PromisedGift treatments without pooling any of our other treatments.[25]

[24]Each leg used a new flyer and recruiting assistant. Though we were sampling from very large campuses—our sample of 194 workers originated from undergraduate populations of over 5,000 on campus A and over 8,000 on campus B—we wanted to ensure that past workers would not perceive the new flyers as a continuation of the job encouraging them to reapply or that new applicants would not perceive the job as an ongoing employment opportunity—triggering the reemployment confounds as in Shapiro and Stiglitz (1984)—instead of a one-time opportunity as emphasized during recruitment and the work period.

[25]Specifically, we used the effect sizes and variances of the most similar study to our own (Gneezy and List (2006)) to calibrate the sample size for the OneGift and PromisedGift treatments. In both of these treatments, workers receive the surprise of an \$8 increase in the hourly wage. However, in our OneGift treatment, they collect the gift at the beginning of each shift, whereas in our PromisedGift, they retrieved it at the end of the six hours of work, as in Gneezy and List (2006). Eight records is the approximate effect size in their research for the first two hours of work, which despite the small sample size (10 subjects in the Control or NoGift treatment and nine in the \$8

2.2 Post-Field Experiment Survey

Because social preferences are the key assumption behind gift exchange, after the conclusion of our field experiment we launched an online survey to obtain a measure of our workers' social preferences. This measure allows us to investigate whether any lack of gift exchange observed, on average, in the field could be due to a dearth of workers with social preferences in the sample. Further, it allows us to explore how prosocial behavior in the laboratory translates into that in the field.

The survey included three one-shot Prisoners' Dilemma games, as in Burks, Carpenter, and Goette (2009), and Schneider and Weber (2012). Their goal was to obtain a proxy of the workers' prosocial attitudes, so that we could investigate the role of worker heterogeneity in social preferences on effort responses in our field experiment. For example, we could explore whether the most prosocial subjects in each treatment reciprocated gifts with effort in the field experiment but their responses were diluted by the lack of effort of a majority of non-prosocial subjects in the sample.[26]

These sequential Prisoners' Dilemma games and other similar ones—such as Trust games and Public Good games—have been used to show that abstract laboratory measures of social preferences correlate with prosocial behavior in the field (e.g., Karlan (2005), Benz and Meier (2008), Burks, Carpenter, and Goette (2009), Baran, Sapienza, and Zingales (2010), and Carpenter and Seki (2010)).[27] The underlying assumption in this literature has been that social preferences are an innate trait that manifests itself across laboratory and field domains.

Our procedure was as follows: Subjects received a $10 participation fee together with any gains in the games which could amount to $15. In the first game, subjects chose their action without knowing the opponent's play. In the second and third games, subjects chose their action after the first mover cooperated and defected, respectively. The stakes were equal across all three games and they followed those in the sequential Prisoners' Dilemma and Trust games in Clark and Sefton (2001), and Charness and Rabin (2002), respectively.[28] Each player was randomly and anonymously paired

gift or GIFT treatment) achieved statistical significance at the 5% level in a two-tailed test. Given the approximate variances found in their first 90 minutes of work at $(9.2)^2$ and 15.5^2, for the CONTROL and their $8 GIFT treatment, respectively, our power calculations yielded a sample size for the CONTROL of 40 workers and a sample size for the combined ONEGIFT and PROMISEDGIFT of 30 workers. We surpassed this number: the CONTROL size is 47 workers and the pooled ONEGIFT and PROMISEDGIFT treatments size is 55 workers, achieving ex-ante 94% power. To also reduce the variance in workers' productivity arising from randomizing different data entry packets to students, which reduces power, all workers in a given leg entered the same stack of records.

[26]The survey also included a multiple-choice questionnaire and 11 lotteries. We do not dwell on their description since these results would, for the most part, only have been relevant had gift exchange been observed.

[27]For a review of the types of games used to measure social preferences in the laboratory, see Levitt and List (2007).

[28]The stakes in the games were as follows: If both subjects cooperated they would both receive $4; if both defected they would both receive $1, following Clark and Sefton (2001). The deviation payoff for defecting when the other cooperated was $7.5, following the Trust game in Charness and Rabin (2002) (pages 861 and 862). Finally, the payoff

with another from his university and this pairing determined the payoffs across the games.[29] To ensure subjects' understanding of the games, they played and repeated practice rounds before the actual games.[30]

Using these games to infer an agent's reciprocity offers one important benefit. Unlike other games, such as the Trust game where the action space is continuous (any dollar amount in a given range), each worker's triplet of binary cooperate/defect choices yields an unambiguous taxonomy of the agent's reciprocal behavior. In particular, we can clearly classify subjects into four types: *self-concerned* types, who defect no matter what; *altruists* who cooperate no matter what; *strong reciprocators*, who cooperate as first players but only cooperate as second players if the first player cooperates; and *weak reciprocators*, who defect as first players and only cooperate as second players if the first player cooperates.[31]

Further, our use of the strategy method (where second movers make conditional decisions) instead of the direct-response method (where the second player makes a unique choice after a given observed move of the first player) allows us to collect a broader set of players' choices, without sacrificing their validity. This is because treatment effects found with the strategy method are invariably observed with the direct-response method (Brandts and Charness (2011)).

3 The Model

In this section, we develop a model to formalize the mechanisms leading to the divergent laboratory and field evidence on gift exchange: habituation to the gift, gift size, and separability in the cost of effort. The crucial assumptions of the model lie in the agent's preferences. In particular, we assume preferences have two components: pure reciprocal preferences, which generate gift exchange, and reference-dependent reciprocity, which induces dynamics in reciprocal behavior and thus habituation. To model pure reciprocity we follow Rabin (1993), Charness and Rabin (2002), Gächter and Falk (2002), Dufwenberg and Kirchsteiger (2004), and Falk and Fischbacher (2006); meanwhile, reference-

of cooperating when the other defected was \$0, following Clark and Sefton (2001).

[29] To achieve this pairing, we first had a random sample of students from each university play each of the three games. We then paired our workers with a randomly selected subject from this previously surveyed pool as is customary in the literature.

[30] Subjects faced exactly the same choices in these practice games as in the actual games. Furthermore, subjects could contact a research assistant and ask any questions about the games or the survey in general. Finally, they were unconstrained in the number of practice rounds.

[31] In other games where the action space is continuous, this classification would be ambiguous as it would require the researcher to declare an arbitrary cut-off for classifying subjects as the most reciprocal, such as those in the top quartile or top decile of repayments.

dependent preferences follow the spirit of Kőszegi and Rabin (2006, 2007). Our model also assumes agents experience consumption utility from wages and effort and therefore it nests classical preferences as a particular case.

3.1 Setup

There are three periods. In period $t = 0$, a principal makes a take-it-or-leave-it (TIOLI) offer to the agent. The offer consists of a per-period fixed wage w, in return for exerting effort in periods $t = 1, 2$. Notice that our experimental design in Section 2 considers three working periods rather than two. Considering only two working periods in our model is done for simplicity as it is enough to highlight the mechanisms under study. Given the offer, the agent forms expectations about the wage he will receive and the effort he will exert in periods one and two. Denote the period-zero wage expectations for future period t as $\tilde{w}_{0t} \in \mathbb{R}_+$ and the effort plans as $\tilde{e}_{0t} \in \mathbb{R}_+$ for $t = 1, 2$. At the end of the period the agent accepts or rejects the contract.

At the beginning of period $t = 1$, the principal announces a permanent per-period wage increase of $g_1 \geqslant 0$. The promised wage for period one, w_1, is now $w + g_1$. Following our experimental design, immediately after receiving the news, the agent receives the gift g_1 and executes period-one effort, e_1. At the end of the period, the agent updates his wage expectations for period two, $\tilde{w}_{12}$, updates his period-two effort plans, $\tilde{e}_{1,2} \in \mathbb{R}_+$, and receives his fixed period-one wage, w.

At the beginning of period $t = 2$, the principal announces an additional second-period wage increase $g_2 \geqslant 0$, which is granted in excess of g_1. If $g_2 = 0$, no further announcement is made. The promised second-period wage, w_2, is now $w + g_1 + g_2$. Following our experimental design, immediately after receiving the news, the agent receives both gifts $g_1 + g_2$ and executes period-two effort, e_2. At the end of the period, the agent receives his fixed period-one wage, w, and the principal-agent relationship credibly ends. Figure 1 summarizes the timing of the principal-agent relationship.

Preferences. We now describe the agent's preferences, which corresponds to the model of reference-dependent social preferences in Esteves-Sorenson and Macera (2012). Since our main purpose is to derive testable predictions to implement in the field, we assume simplifying functional forms that highlight the underlying mechanisms. We start by introducing the components of the utility function sequentially.

(1) *Consumption Utility from Wages and Effort.* Because our experiment involves small stakes, we assume agents are risk neutral in wages. To allow for the possibility that agents exert more than zero effort when paid a fixed wage, we assume that the cost of effort corresponds to a quadratic

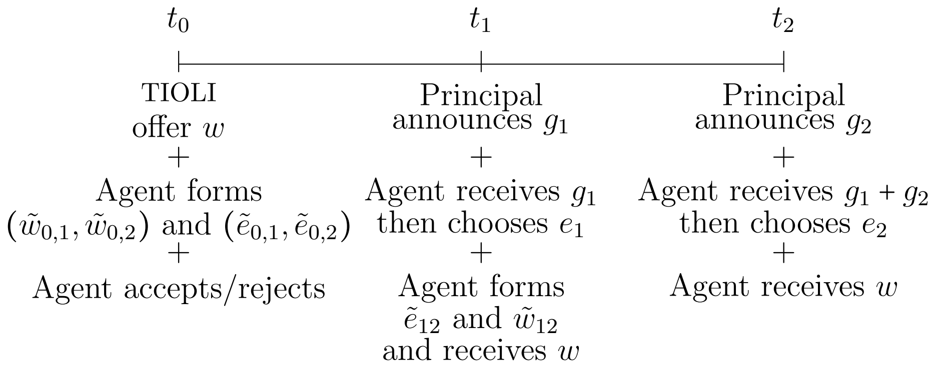


Figure 1: Timing of the Principal-Agent Interaction

function with an interior stationary point $\underline{e}$, with associated cost $C \in \mathbb{R}_+$. The period-t utility flow of a classical agent corresponds to

$$w_t - \frac{1}{2}(e_t - \underline{e})^2 - C \tag{1}$$

for $t = 1, 2$.[32] Notice that equation (1) assumes that consumption utility is separable across time and consumption domains. The temporal separability of effort comes from our experimental design where periods are distant and thus previous-period effort should not affect the marginal cost of present effort. This is an important feature of our model, implying that any dynamics in effort will not arise from fatigue.

(2) *Pure Reciprocity.* Agents who have a pure reciprocity component in their preferences experience utility from responding to kind actions with kindness and responding to unkind actions with unkindness. As it is standard in the reciprocity literature, to formalize the agent's utility function, we start defining these kindness functions. The utility the agent gets from the principal's kindness towards him corresponds to $K_a(w_t) = w_t - \bar{W}$, the distance between the period-t wage and a fair wage, $\bar{W}$. Equivalently, the agent's kindness towards the principal corresponds to $K_b(w_t) = B_t - \bar{B}$, the distance between the principal's payoff B_t, and a fair payoff, $\bar{B}$. We assume the principal's payoff has a simple linear structure $B_t = be_t - w_t$, where b is the constant marginal benefit of effort. To define the fair payoffs we take advantage of the structure of the principal-agent interaction and assume that the fair actions are those implied by the mutually agreed upon initial contract. The fair wage $\bar{W}$ corresponds thus to w, meanwhile, the fair effort corresponds to the effort the agent would exert if the contract is honored in periods one and two, that is $\underline{e}$. This implies $\bar{B} = b\underline{e} - \bar{W} = b\underline{e} - w$.

[32] Despite the fact that the gifts are granted at the beginning of the period and the fixed wage is granted at the end, notice that the utility flow lumps both in w_t. This implies we are assuming that the utility flow is realized when the gift announcement is made, and that there is no discounting within a period. Since gifts are credible and they are not contingent on effort, this assumption is without loss of generality.

The period-t utility from pure reciprocity corresponds to

$$\alpha K_a(w_t) K_b(w_t) \tag{2}$$

for $t = 1, 2$. If the agent perceives that the principal has been kind by paying him a higher period wage than the fair wage ($K_a > 0$), then his best response is to exert effort so to increase the principal's payoff above the principal's fair payoff ($K_b > 0$). Conversely, if the principal has been mean by paying the agent a wage lower than the fair one ($K_a < 0$), then the agent reciprocates by exerting less effort than that needed to reach the principal's fair payoff ($K_b < 0$). The parameter α denotes the importance of reciprocal utility relative to consumption utility from wages and effort.

(3) *Reference-Dependent Reciprocity and Reference-Dependent Effort.* When agents have reference-dependent preferences, they experience utility from departures from a reference point in all consumption domains. In our model, this implies that agents will experience reference-dependent reciprocity as well as reference-dependent cost of effort.[33]

Our model posits that agents with reference-dependent reciprocal preferences experience utility from responding to unexpected kindness with further kindness and responding to unexpected unkindness with further unkindness. Their reference point, therefore, corresponds to the expectation they formed in the previous period about the principal's (and their own) kindness. This idea of the reference point being the agent's expectations about outcomes follows Kőszegi and Rabin (2006, 2007, 2009).[34] The same idea applies to the cost of effort: agents experience reference-dependent utility whenever the cost of effort is greater or smaller than the expected cost of effort.

The period-t utility from reference-dependent preferences corresponds to

$$\eta_k \alpha \mu\Big(K_a(w_t) - K_a(\tilde{w}_{t-1,t})\Big)\mu\Big(K_b(w_t) - K_b(\tilde{w}_{t-1,t})\Big) + \eta_e \mu\Big(-\frac{1}{2}(e_t - \underline{e})^2 + \frac{1}{2}(\tilde{e}_{t-1,t} - \underline{e})^2\Big) \tag{3}$$

for $t = 1, 2$, where $\tilde{w}_{t-1,t}$ and $\tilde{e}_{t-1,t}$ correspond to the expectation formed in period $t-1$ about period-t wages and effort respectively, and μ corresponds to the Kahneman and Tversky (1979) value function capturing the asymmetry between gains and losses, asymmetry they called loss aversion. To isolate the role of loss aversion, we assume that μ is piece-wise linear, $\mu(x) = x$ if $x > 0$, $\mu(0) = 0$, and

[33] This implies they will also experience reference-dependent utility from wages. However, this term does not depend on effort, and thus, for tractability, we omit it from the preference structure.

[34] Direct evidence that the agent's reference point corresponds to his recent expectations about outcomes can be found in, for example, Ericson and Fuster (2011), Crawford and Meng (2011), Abeler, Falk, Goette, and Huffman (2011), Card and Dahl (2011), Pope and Schweitzer (2011), and Gill and Prowse (2012).

$\mu(x) = \lambda x$ if $x < 0$ where $\lambda > 1$ is the loss aversion parameter capturing the strength to which losses hurt more than same-sized gains please.

The first term in (3) represents reference-dependent reciprocity. It captures the idea that the agent reciprocates the principal's kindness only if this kindness is greater than the expectation the agent formed in period $t-1$ about how kind the principal would be towards him in period t, that is, the expected kindness $K_a(\tilde{w}_{t-1,t})$. If $K_a(w_t) > K_a(\tilde{w}_{t-1,t})$ the agent's best response is to exert effort to increase his own kindness towards the principal above expectation ($K_b(w_t) > K_b(\tilde{w}_{t-1,t})$). Equivalently, if the principal's kindness is lower than expected, the agent's best response is to exert effort to reciprocate with less kindness than expected. The second term in (3) represents reference-dependent cost of effort: the agent compares the actual cost of effort with the cost of effort he planned in the previous period to experience in the current period.[35]

As usual the parameters $\boldsymbol{\eta} = (\eta_k, \eta_e)$ represent the importance of the reference-dependent components of the utility function in the reciprocity and effort domains, respectively. Importantly, notice that the reference-dependent reciprocity term is also multiplied by α, the weight of pure reciprocal preferences in equation (2). This highlights an important aspect of the Kőszegi and Rabin (2006) preferences: reference-dependent utility is proportional to consumption utility. This implies that if agents do not experience pure reciprocal preferences, they will not experience reference-dependent reciprocity either.

This model can distinguish three types of preferences, from most general to particular. First, *Reference-Dependent Reciprocal preferences* whenever $\alpha > 0, \boldsymbol{\eta} > \mathbf{0}$. Second, *Pure Reciprocal preferences* whenever $\alpha > 0, \boldsymbol{\eta} = \mathbf{0}$ and *Classical preferences* whenever $\alpha = 0, \boldsymbol{\eta} = \mathbf{0}$. In all cases, total utility U_t corresponds to the summation of utility flows across periods, where, for simplicity, we assume there is no time discounting.[36]

Equilibrium. To study the behavior of these agents in our experimental set up we need to distinguish two cases: the case of full surprise and the case without surprises. Following Kőszegi and Rabin (2006), in environments with no surprises reference-dependent agents behave as consumption-utility maximizers. Intuitively, because agents form plans rationally, absent uncertainty or informa-

[35] One can argue that this model mixes two layers of reference-dependent utility as kindness itself can be thought of as a reference-dependent phenomena where the reference point corresponds to the fair payoffs. We argue, however, that these two reference-dependent phenomena are by no means substitutes, as each will reflect different hedonic experiences associated to a given wage: the comparison with a fair payoff corresponds to the (dis)satisfaction of a (un)fair outcome, meanwhile the comparison with the expected outcome reflects the cognitive (dis)satisfaction of getting a (worse)better outcome than expected. Furthermore, as shown on Section 3.2, they have different implications on the agent's behavior.

[36] In our model, because periods are unrelated, exponential discounting would not affect our predictions.

tion arrival, their expectations always match their actions and reference-dependent utility is zero.[37] Whenever fully unexpected information arrives, this new information will cause departures from expectations. Because agents do not foresee this information and because they choose effort immediately after receiving the gift, we assume expectations do not adapt and thus agents maximize conditional on the reference. Furthermore, because they do not foresee information arrival and payments are fixed, they form plans as a consumption utility maximizer. See Definition 1 in Appendix C for a formal definition of the equilibrium.

3.2 Predictions

We start considering the case where the principal honors the contract and pays w in all periods. Proposition 1 describes the evolution of effort for subjects in the Control.

Proposition 1 *(Behavior of the* Control*)*
Assume $g_1 = g_2 = 0$ *and* $\tilde{w}_{0,1} = \tilde{w}_{0,2} = w_1 = \tilde{w}_{1,2} = w_2 = w$. *Then, agents with classical, pure reciprocal and reference-dependent reciprocal preferences exert* $e_1 = e_2 = \underline{e}$.

Proposition 1 says that, whenever there are no gifts—either expected or unexpected—agents with any preference structure exert the least-cost effort in all periods. Intuitively, agents with classical preferences have no incentives to exert more than the least-cost effort because the wage is fixed. Agents with reciprocal preferences also exert the least-cost effort, because absent gifts there is no scope for reciprocity. Finally, agents with reference-dependent preferences also behave as consumption utility maximizers as there are no departures from expectations.

We now consider our first treatment: the case where the principal surprises the agent with a permanent wage increase at the beginning of period one, before he begins working on the task.

Proposition 2 *(Behavior of the* OneGift *Treatment)*
Assume $g_1 > 0$, $g_2 = 0$, $\tilde{w}_{0,1} = \tilde{w}_{0,2} = w$ *and* $w_1 = \tilde{w}_{1,2} = w_2 = w + g_1$. *Then, agents with classical preferences exert* $e_1 = e_2 = \underline{e}$, *meanwhile, agents with pure reciprocal preferences exert* $e_1 = e_2 > \underline{e}$. *Agents with reference-dependent reciprocal preferences exert* $e_1 > \underline{e}$ *and* $e_2 > \underline{e}$, *where* $e_1 > e_2$ *if* $\eta_e < \frac{\eta_k}{\lambda}$.

Proposition 2 says that when the agent is surprised by a permanent wage increase ($w_1 = w_2 = w + g_1$) and the agent anticipates no further increases of the wage ($\tilde{w}_{1,2} = w + g_1$), agents with classical preferences do not reciprocate the gift. The intuition is straightforward: since the gift is not tied to

[37] In Appendix C we show this formally in Lemma 1.

performance, it fails to provide incentives and the agent exerts the least-cost action in every period. Further, agents with pure reciprocal preferences display a flat path of effort. To see this, notice that pure reciprocal agents in every period solve,

$$\max_{e_t} w + g_1 + \alpha g_1 (b(e_t - \underline{e}) - g_1) - \frac{1}{2}(e_t - \underline{e})^2 - C \quad \Rightarrow \quad e_t^* = \underline{e} + \alpha g_1 b > \underline{e}$$

and thus the agent exerts the same amount of effort in each period, which is higher than that exerted by classical agents. Finally, Proposition 2 shows that agents with reference-dependent reciprocal preferences do reciprocate the gift, but this reciprocity wanes in time. The mechanism is as follows: In period zero—because the contract is credible—he forms effort plans as a consumption utility maximizer and thus $\tilde{e}_{0,1} = \tilde{e}_{0,2} = \underline{e}$. Once in period one he is surprised by the gift and thus chooses e_1 to maximize his period-one utility flow

$$\max_{e_1} w + g_1 + \alpha g_1 (b(e_t - \underline{e}) - g_1) + \eta_k \alpha g_1 \mu\big(b(e_1 - \underline{e}) - g_1\big) - \frac{1}{2}(e_1 - \underline{e})^2 - C + \eta_e \mu\Big(-\frac{1}{2}(e_1 - \underline{e})^2\Big)$$

since $K_a(w_1) - K_a(\tilde{w}_{0,1}) = (w + g_1 - w) - (w - w)$ and $K_b(w_1) - K_b(\tilde{w}_{0,1}) = b(e_1 - \underline{e}) - g_1$. The necessary and sufficient first-order condition for an interior maximum corresponds to

$$e_1^* = \underline{e} + \frac{\alpha g_1 b + \eta_k \alpha g_1 \mu'\big(b(e_1^* - \underline{e}) - g_1\big) b}{(1 + \lambda \eta_e)} > \underline{e} \tag{4}$$

where $e_1 > \underline{e}$ because $\mu' \in \{1, \lambda\} > 0$ and $b > 0$.

Intuitively, there are two reference-dependent forces at play after the surprising gift. First, a pleasant surprise from the gift, that triggers reciprocity and which constitutes a marginal benefit. Second, a negative surprise from having to exert more effort than expected. This reference-dependent marginal benefit plus that from pure reciprocity, outweigh the marginal costs and the agent exerts more effort than classical agents. At the end of the first period the agent updates his expectations: he knows the principal will grant him a gift and thus she will be kind towards him. Because the agent does not expect any further arrival of information, he forms his plans as a consumption utility maximizer and thus sets $\widetilde{e}_{1,2} = \underline{e} + \alpha g_1 b$. Once in period two, no further information arrives and the agent behaves as an agent with pure reciprocal preferences. As a consequence, $e_2^* = \underline{e} + \alpha g_1 b$. Proposition 2 shows that whenever the weight of the negative surprise from having to exert more effort than expected is small enough relative to the weight of the pleasant surprise from reciprocating the unexpected gift, then the second-period effort will be smaller than that in the first period and effort wanes over time.[38]

Three comments are key to Proposition 2. First, notice that the increasing and then decreasing pattern of effort occurs under the assumption that the cost of effort is separable across periods. This implies that the decrease in effort after the first-period reciprocity is not due to effort becoming increasingly costly. Second, equation (4) shows that the size of the divergence between e_1^* and $\bar{e}$ is increasing in g_1. If the gift is arbitrarily small, then e_1^* is arbitrarily close to $\underline{e}$. Intuitively, when the gift is small the associated marginal benefit—originating from reference-dependent reciprocity—is also small. The marginal cost of effort, however, does not depend on the gift size. This implies that for a given marginal cost of effort a larger gift triggers a stronger behavioral response. This highlights that the excess wage must be sizable to trigger a significant effort response. Finally, for reference-dependent social preferences, the most important assumption behind Proposition 2 is that the agent's expectations do not adapt immediately after receiving the gift: it is the divergence between the actual wage and the expected one that triggers incentives from the reference-dependent component of the utility function.

We now explore the case of two fully unexpected wage increases. Proposition 3 compares the effort path of agents with classical and pure reciprocal preferences to those of agents with reference-dependent reciprocal preferences.

Proposition 3 *(Behavior of the* TwoGifts *Treatment)*
Suppose $g_1 > 0$ and $g_2 > 0$. Further, assume $\tilde{w}_{0,1} = \tilde{w}_{0,2} = w$, $w_1 = \tilde{w}_{1,2} = w + g_1$ but $w_2 = w + g_1 + g_2$. Then agents with classical preferences exert $\underline{e} = e_1 = e_2$, agents with pure reciprocal preferences exert $\underline{e} < e_1 < e_2$ and agents with reference-dependent reciprocal preferences exert $e_1 > \underline{e}$ and $e_2 > \underline{e}$.

Proposition 3 states that when agents have classical preferences and are surprised a second time by a further period-two wage increase ($g_2 > 0$), they do not respond to either gift: their effort path is flat. As before, because gifts are not tied to performance, they do not affect agents' behavior. To the contrary, agents with pure reciprocal preferences will display an increasing path of effort because their behavior depends on the absolute magnitude of the gift. Further, Proposition 3 shows that agents with pure reference-dependent preferences exert more effort than classical agents in both periods. The intuition mimics that in the first period in Proposition 2: because gifts are unexpected, they both cause a positive departure from expectations that push towards exerting more effort than

[38]That is, simple algebra shows that $\eta_e < \frac{\eta_k}{\lambda}$ implies $\frac{\alpha g_1 b + \eta_k \alpha g_1 \mu'\left(b(e_1^* - \underline{e}) - g_1\right)b}{(1+\lambda\eta_e)} > \alpha g_1 b$.

absent the gift.[39]

Proposition 2 and Proposition 3 jointly shed light as to what should we expect in the three-period case in our TwoGifts treatment. If agents have reciprocal preferences and a gift is granted in periods one and three, we should observe an initial kick in effort (as predicted by Proposition 2), then a decrease in effort in the second period (as expectations adapt and there are no further gifts) and a renewed kick in effort in the third period once the second surprising gift arrives (as predicted by Proposition 3).

Proposition 3 raises one important question: what happens if the agent anticipates the wage increase? In Proposition 5 in Appendix D we show that the result of a decreasing path of effort holds even if the agent expects to receive g_2 with a positive probability. Intuitively, the realization of the gift still implies a departure from the agent's—now stochastic—expectations. Such a deviation thus still triggers incentives towards high effort, even though of a smaller magnitude than in the full surprise scenario, as the departure is only likely (rather than certain) from the agent's perspective.[40]

From this discussion a related question arises: what happens if the agent expects the gift and he does not receive it? To answer this question our last treatment explores the case when the agent receives a permanent gift in period one—just as in OneGift treatment—but the principal is not able to grant the gift in period two (Gift-NoGift treatment).

Proposition 4 *(Behavior of the* Gift-NoGift *Treatment)*
Assume $g_1 > 0$, $g_2 = 0$, $\tilde{w}_{0,1} = \tilde{w}_{0,2} = w$, $w_1 = \tilde{w}_{1,2} = w + g_1$ but $w_2 = w$. Then in the second period agents with classical and pure reciprocal preferences exert effort as in Proposition 1; meanwhile, agents with reference-dependent reciprocal preferences exert $e_1 > \underline{e} > e_2$.

For agents with classical preferences, the result in Proposition 4 is obvious: these agents exert the least-cost effort independently of the granting or withdrawal of gifts. When agents have pure reciprocal preferences they do not engage in negative reciprocity either, as the gift withdrawal does not violate the fairness of the contract. The interesting insights come with reference-dependent social preferences. Agents with reference-dependent reciprocal preferences retaliate expected-but-unfulfilled

[39] With reference-dependent reciprocal preferences, whether the effort path is increasing or decreasing in time depends on the relative size of the gifts. In particular, is easy to show that the path will be increasing if both gifts are big enough and $\frac{g_2}{(g_1 - g_2)} > \eta\lambda$.

[40] The implications over the agent's preferences are not trivial. When the agent expects the gift his reference point is no longer fixed but stochastic. Thus he forms contingent effort plans, which he latter must carry through because of the rational expectations assumption. Moreover, the anticipation of the gift has a negative impact over the agent's acceptance-rejection decision. We show that, since equilibrium expected reference-dependent utility is negative and decreasing in loss aversion, period zero total expected utility may be lower in the case the agent anticipates the gift relative to the non-gift when the agent is sufficiently loss averse.

payments: Because they are expecting the principal to be kind and this expected kindness is not met, they reciprocate by exerting less effort than they would if they had classical or pure reciprocal preferences.

4 Results

This section presents the results of our field experiment and laboratory games. Section 4.1 presents some important features of our data, which will support the analysis. Section 4.2 presents our main results. Overall, we find no evidence of gift exchange: subjects do not reciprocate monetary gifts of any magnitude, expected or unexpected, with higher effort. They do, however, respond to piece rates. This suggests a classical model of preferences is a better fit for our workers' preferences relative to a model of social preferences.

Finally, section 4.3 presents the disaggregate results splitting the sample according to prosocial types gleaned from the Prisoner's Dilemma games in our post-field experiment survey. We find an absence of correlation between prosocial behavior in the laboratory and the field. Though many of our workers display prosocial behavior in the laboratory, they fail to reciprocate gifts in the field. They behave similarly to workers behaving non-prosocially in the laboratory.

4.1 Preliminaries

We start with a few noteworthy features of our data. First, attrition is unlikely to bias our results as it was small and uncorrelated with any specific treatment.[41] All workers completed the three required shifts except for 13 (6.7% of our sample), who came to the first shift but missed one or two of the subsequent sessions. They were distributed as follows: five in the Control, five in the Gift-NoGift, one in the AnticipatedGift, and two in the PromisedGift treatments.

Second, there is significant variation in the amount of output and effort across campuses, suggesting the need to control for unobserved campus differences. For instance, the average number of records entered by campus A students per two-hour shift in the Control was 105 across all shifts, whereas that of campus B students was a little more than half at 56. The number of keystrokes follows the same pattern at 21,775 and 13,188, respectively (see column (5) in Tables 3 and 4, respectively).

Third, because there is little variation in the average quality of records per campus and treatment, the analysis focuses on the quantity of output (records) and effort (keystrokes). Specifically, the

[41] Analysis available upon request.

average proportion of error-free words entered by campus A students per two-hour shift in the Control was 0.98 across all shifts, whereas it was 0.97 for campus B students (see column (5) in Tables 3 and 4, respectively). Further, the proportion of error-free words inputted ranges generally between 97%-98% across all treatments.[42]

Fourth, there is learning across both campuses, necessitating comparisons between the different treatments and the Control by session to control for learning about the task. Specifically, the average number of records entered by the Control on campus A increases from 95 to 112 between shifts one and two, stabilizing thereafter. The pattern for campus B is similar: the average number of records entered by the Control on campus B increases from 51 to 57 between shifts one and two, stabilizing thereafter (see columns (2)-(4) in Tables 3 and 4, respectively).

Fifth, the data is quite noisy, requiring pooling the treatments and campuses to garner a more precise picture of the output and effort responses across treatments. For example, though workers received the same information and payments in shifts one and two of the TwoGifts and Gift-NoGift treatments, the average number of records entered in shift one versus the Control in each of these treatments has an inconsistent pattern. For campus A, for example, the average number of records in session one for the TwoGifts treatment was 7% lower than that in the Control at 89 records, whereas it was 10% higher in the Gift-NoGift treatment at 105 records, despite the identical information and payments for session one. See Table 3, column (2).

4.2 Field Experiment Results

Our first result is that we cannot document gift exchange: workers in the several gift treatments do not generate more output or effort than those in the Control. Namely, the unadjusted average of records entered for the gift treatments—all treatments except the PieceRate— though higher than that of the Control for all shifts (see Table 5, Panel I, columns (2)-(4)) is generally not significant. Further, the excess average records in the gift treatments declines substantially when we estimate the differences between the gift treatments and the Control within campuses (see Table 5, Panel I, columns (5)-(7)).[43] Last, workers in the gift treatments produce less than the Control when we estimate the differences between each of the gift treatments and the Control within each leg, on each campus, to address any unobserved specific campus and leg factors influencing these productivity

[42]The one exception is the PieceRate treatment at campus B, where the quality declined slightly to 95% of words correctly inputted.

[43]The specific functional form for this adjustment is: $records_{i,s,t} = \beta_{1,1} + \beta_{1,2}T_1S_2 + \beta_{1,3}T_1S_3 + \sum_{t=2}^{7}\sum_{s=1}^{3}\beta_{t,s}T_t.S_s + campus_c + \epsilon_{i,s,t}$; where i=subject 1 through 194, s=session 1,2,3; t=treatment (t=1, ..., 7).

differences. This negative difference, however, is not statistically significant (see Table 5, Panel I, columns (8)-(10)).[44] The pattern is starker for effort: average keystrokes in the gift treatments are in general lower than those in the CONTROL, instead of higher, though not statistically significant (see Table 5, Panel II, columns (2)-(10)).

Our second result is that—in contrast to gifts—piece rates always induce positive output and effort. Workers in the PIECERATE treatment input five to eight more records than those in the CONTROL. Although these point estimates are not statistically significant, the estimates for effort are. Namely, in contrast to workers in gift treatments, those in the PIECERATE increasingly press more keys, in an effort to input more records, and this difference in effort becomes marginally statistically significant in the third shift. See Table 5, Panels I and II, columns (2)-(10). See also Figure 2.

It is important to note that these significance levels and those in subsequent analyses are for one-sided right-tailed tests, consistent with the power calculations. That is, we calibrated the sample size to be able to reject the null of gifts not increasing output in favor of a one-sided alternative that they do, as explained in Section 2. Further, the standard errors are clustered by individual to address the serial correlation in outcome measures across shifts for a given individual (Bertrand, Duflo, and Mullainathan (2004)).

These contrasting results between gifts and piece rates do not change once we pool similar treatments into three groups to increase the precision in our estimates: (i) the \$8 surprise wage increases to \$20 per hour (ONEGIFT and PROMISEDGIFT treatments), (ii) the \$6 surprise wage increase to \$18 per hour in shifts one and two (TWOGIFTS and GIFT-NOGIFT treatments), and (iii) all the \$8 wage increase treatments (ONEGIFT, ANTICIPATEDGIFT and PROMISEDGIFT treatments). The point estimates of the average number of records and keystrokes entered versus the CONTROL for these groups, within campuses and legs, is zero or slightly negative instead of positive, though not statistically significant despite the reduction of the standard errors arising from the larger pooled sample. In contrast, the PIECERATE treatment yields positive point estimates, in both output and effort. Further, the increase in effort versus the CONTROL becomes statistically significant despite the much smaller sample size in this treatment relative to any of the three pooled gift treatments described above (Table 6, Panels I and II, columns (2)-(10)) and see also Figure 3). For a breakdown of the comparison of the records and keystrokes in the pooled gift treatments with PIECERATE by

[44] The specific functional form for this adjustment is: $records_{i,s,t} = \beta_{1,1} + \beta_{1,2}T_1S_2 + \beta_{1,3}T_1S_3 + \sum_{t=2}^{7}\sum_{s=1}^{3}\beta_{t,s}T_t.S_s + campus_c \times leg_l + \epsilon_{i,s,t}$; where i=subject 1 through 194, s=session 1,2,3; t=treatment (t=1, ..., 7).

campus, please see, respectively, Appendix Tables B.1 and B.2.

Further, it is worth noting that the average piece rate expenditure per subject was much lower but elicited much more effort than that with our largest gifts. The average piece rate expenditure of $14 per subject, beyond the contracted $12 per hour wage, consistently raised effort. In contrast, the average additional expenditure of $48 in the $8 gift treatments, in addition to the same contracted wage, elicited negligible or slightly negative output and effort responses, though statistically insignificant.

Our third result is that we cannot document loss of output or effort in response to expected-but-unfulfilled gifts. Agents do not reduce output or effort versus Control in the third shift of the Gift-NoGift treatment when the expected gift of $6 is withdrawn and workers merely receive the contracted $12 hourly wage. The average number of records and keystrokes when the $6 gift is withdrawn is the same or slightly above the Control, instead of below, and these estimates are not statistically significant. See Table 5, Panels I and II, column (10).

4.3 Post-Field Experiment Results on Worker Heterogeneity

We now describe two other important results in the relationship between the manifestation of social preferences in the laboratory and in the field arising from our post-field experiment survey.

Of our 194 workers, 139 participated in an online survey with the three sequential Prisoner's Dilemma games after the conclusion of the field experiment, representing 72% of our sample. As described previously, each workers triplet of choices allowed us to classify him into eight distinct prosocial types arising from the combination of the two decisions in each of the three games.

Of these types we picked two extremes, which concentrate the majority (73%) of respondents: *Most Prosocial* and *Least Prosocial*. We categorized as *Most Prosocial* those subjects who behaved the most prosocially in these games: *altruists*—who cooperated as first and second movers, independently of whether the first player cooperated or defected—or *strong reciprocators*—who always cooperated unless the first mover defected. In our sample, 57 workers (41% of respondents) behaved as either *altruists* (10 workers) or *strong reciprocators* (47 workers). See Table 7, columns (4)-(7). In contrast, we classified subjects as *Least Prosocial* if they behaved the least prosocially in these games: they were *self-concerned*, defecting both as first movers and second movers independently of whether the first mover cooperated or defected. In our sample, 45 workers (32%) behaved as *Least Prosocial*. See Table 7, columns (8)-(9).

We now introduce our fourth result: *Most Prosocial* workers in the Prisoner's Dilemma games did not behave more prosocially in the field by increasing productivity and effort versus the Control. Not only is their productivity and effort not statistically significant, but also it is similar to that of *Least Prosocial* workers. For example, of the 70 workers in all the larger gift treatments (a wage increase of \$8 to \$20 per hour), 48 responded to the survey (see Table 7, column (2), rows (3) and (6)-(7)). Of these, 22 behaved as *Most Prosocial* and 14 as *Least Prosocial* (see columns (6) and (8), rows (3) and (6)-(7)). When comparing output and effort of these workers versus those in the Control, we find that this difference is not statistically significant. Further, this pattern of productivity is similar to that of workers who were also in these large gift treatments but behaved as *Least Prosocial*. See Table 8, Panels I and II, columns (5)-(12), row (4). See also Figure 4.

We use all the workers in the Control as a benchmark as this group does not receive any gifts (or piece rates). Therefore, agents' behavior in this group should not differ according to their prosocial type as social preferences are not triggered.

Our fifth result is that, in contrast to the gift treatments, both *Most Prosocial* and *Least Prosocial* workers increased their productivity when receiving a piece rate and this increase is statistically significant. Specifically, among the workers who answered the survey and who belong to the PieceRate treatment, those who behaved as *Most Prosocial* in the games increase their productivity and effort versus the Control and this increase in statistically significant. And this behavior is similar to those in the PieceRate treatment who behaved as *Least Prosocial*. See Table 8, Panel I and II, columns (5)-(12), row (3).

These results are robust to selection into the survey. Despite the generally high response rate to the survey (72% overall and 65% or above in all treatments, except the AnticipatedGift with 50%), one might worry that selection into the survey might bias our results. That is, there exist unobserved factors that are correlated both with selection into the survey and responses to the games. For example, those workers with stronger social preferences may both select at higher rates into the survey and behave more prosocially in the games. This, however, would not invalidate our results: our sample of respondents would just have a higher proportion of *Most Prosocial* types. The conclusion that these *Most Prosocial* types do not behave prosocially in the field, as their productivity and effort is not higher than the Control in the gift treatments, but higher in the PieceRate, would still hold. The same is true for workers with weaker social preferences selecting into the survey: we would have fewer *Least Prosocial* types answering the survey, but conditional on their responding it is valid to correlate their behavior in the laboratory games with their behavior in the field experiment.

Another selection-related concern is that self-selection into the survey is correlated with the treatment assignment. For example, workers who received larger gifts could be more prone to answering the survey. This would not bias our results as these workers still reveal their type in the games—whether they are most or least prosocial—and the correlation between their type and their behavior in the field would still enable us to test whether most prosocial workers in the laboratory behave prosocially in the field.

In sum, selection into the survey may influence the distribution of prosocial workers in our final sample—as inferred by their answers in the Prisoner's Dilemma games—and their distribution within treatment. However, conditional on these workers selecting into the survey and revealing their prosocial type through the games, it does not invalidate our analysis of whether those revealed as *Most Prosocial* in laboratory games behave more prosocially in the field experiment and any differently than those revealed as *Least Prosocial*.

5 Conclusion and Discussion

We present and test a unifying model reconciling the divergent laboratory and field evidence on gift exchange. Our model highlights three potential sources for the conflicting evidence: habituation to the gift, fatigue, and inadequate gift size. We implemented an ex-ante adequately powered field experiment to test our model, addressing the factors above, as well as four others, which could dampen gift exchange in the field: selection of more able workers (Weiss (1980)), the existence of an effort ceiling for the task, ambiguity as to the kind intentions of the principal (e.g., Charness (2004)), and a small proportion of workers with social preferences in our sample. We find that workers do not respond to gifts, even after addressing all these confounding factors, but do respond to piece rates. A test of the heterogeneity of social preferences in our sample shows that though many of our workers behave prosocially in laboratory games, they do not engage in gift exchange in the field.

There are two possible explanations for why we do not find gift exchange in the field in contrast to the laboratory. First, the laboratory evidence on gift exchange can conflate social preferences with the desire to appear reciprocal or with experimenter demand effects, as noted by Levitt and List (2007) and Zizzo (2010). Namely, when participants' selfish actions in the laboratory are shrouded by uncertainty as to whether they were intentional or caused by factors beyond their control, or when participants enjoy a higher degree of anonymity versus other players or the experimenter, their prosocial behavior declines (e.g., see Levitt and List (2007) for a review of these studies). This may explain the prosocial behavior in gift exchange games, as well as in our sequential Prisoner's Dilemma games (of which the gift exchange game is a version) and not in our field experiment.

Specifically, in our games, subjects know that their actions are observable to the experimental team that effectuates the payments. In contrast, in our between-subjects field experiment, where aggregate subjects' behavior is compared to that of a control group, each subject knows that the principal is unaware of his productivity in the absence of a gift, enjoying room to behave more selfishly. This between-subjects design is an advantage over a within-subjects one since comparing performance pre- and post-gift for the same subjects could have led to an artificial finding of gift exchange as workers would know that the principal knows their productivity in the absence of the gift.

Second, social preferences triggered by monetary payments may emerge in a very small fraction of the population. Therefore, firms' workers described in Akerlof (1982) to motivate the gift exchange hypothesis may engage in gift exchange because their firms screened on these preferences. One such screening tool is years-long probation, which Esteves-Sorenson, Macera, and Pohl (2013) explore with data from two Portuguese firms in order to quantify the contribution of different efficiency wages mechanisms—threat of dismissal, gift exchange, and selection of higher ability workers—to productivity in a repeated interaction environment.[45]

Our ancillary finding of no retaliation for expected-but-unfulfilled wage increases may be due to agents only retaliating wage cuts when these violate the contract. This hypothesis dovetails the idea of contracts as reference points, as in Hart and Moore (2008), and the evidence with Kube, Maréchal, and Puppe (forthcoming) who find that workers retaliate cuts versus the forecasted contract wage.[46] This suggests that the contract plays a role in shaping agents' expectations and therefore their reference point.[47]

Finally, it is important to note that though our results do not support gift exchange in the workplace, there is evidence that workers do reciprocate non-monetary gifts, such a thermoses (e.g., Kube, Maréchal, and Puppe (2012)). This suggests that social preferences do exist, but they are triggered by the nature of the gift. It is an open question, however, whether workers habituate to these gifts causing a decline of reciprocity over time. This is a prediction of our model that has yet to be tested.

[45]In this case, workers enjoy ample time to habituate to the higher wage, and therefore gift exchange arises from the consumption utility from reciprocity instead of solely from departures from expectations.

[46]Kube, Maréchal, and Puppe (forthcoming) found that workers hired at a forecasted wage of 15 euros per hour for a one-time data entry task but informed that they would be paid only 10 euros per hour upon arriving at work, retaliated with less effort than the Control, which received the forecasted contract wage of 15 euros per hour.

[47]The Hart and Moore (2008) model postulates agents in flexible contracts—such as those in Kube, Maréchal, and Puppe (forthcoming), where wage in the contract is a forecast—feel aggrieved when they receive a lower wage than that to which they felt entitled, resulting in a decline in performance. Our contract, in contrast, is rigid in that it promises $12 per hour, a promise which is never violated, thus not triggering retaliation.

References

Abeler, J., A. Falk, L. Goette, and D. Huffman (2011): "Reference Points and Effort Provision," *American Economic Review*, 101, 470–492.

Adams, J. (1963): "Toward an Understanding of Inequity," *Journal of Abnormal and Social Psychology*, LXVII, 422–36.

Akerlof, G. (1982): "Labor Contracts as Partial Gift Exchange," *The Quarterly Journal of Economics*, pp. 543–569.

Akerlof, G., and J. Yellen (1990): "The Fair-Wage Hypothesis and Unemployment," *Quarterly Journal of Economics*, 105(2), 255–83.

Baran, N., P. Sapienza, and L. Zingales (2010): "Can We Infer Social Preferences from the Lab? Evidence from the Trust Game," *NBER Working Paper 15654*.

Bell, D. (1985): "Disappointment in Decision Making Under Uncertainty," *Operations Research*, 33(1), 1–27.

Bellemare, C., and B. Shearer (2009): "Gift Giving and Worker Productivity: Evidence from a Firm-Level Experiment," *Games and Economic Behavior*, 67(1), 233–244.

Benz, M., and S. Meier (2008): "Do People Behave in Experiments as in the Field? Evidence from Donations," *Experimental Economics*, 11(3), 268–281.

Bertrand, M., E. Duflo, and S. Mullainathan (2004): "How Much Should We Trust Differences-In-Differences Estimates?," *Quarterly Journal of Economics*, 119(1), 249–275.

Bewley, T. (1998): "Why not cut pay?," *European Economic Review*, 42(3-5), 459–490.

Brandts, J., and G. Charness (2004): "Do Labour Market Conditions Affect Gift Exchange? Some Experimental Evidence," *The Economic Journal*, 114(497), 684–708.

——— (2011): "The Strategy Versus the Direct-Response Method: a First Survey of Experimental Comparisons," *Experimental Economics*, 14(3), 375–398.

Burks, S., J. Carpenter, and L. Goette (2009): "Performance Pay and Worker Cooperation: Evidence from an Artefactual Field Experiment," *Journal of Economic Behavior and Organization*, 70, 458–469.

Card, D., and G. B. Dahl (2011): "Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior," *Quarterly Journal of Economics*, 126, 103–143.

Carpenter, J., and E. Seki (2010): "Do Social Preferences Increase Productivity? Field Experimental Evidence From Fishermen in Toyama Bay," *Economic Inquiry*, 49(2), 612–630.

Charness, G. (2004): "Attribution and Reciprocity in an Experimental Labor Market," *Journal of Labor Economics*, 22(3), 665–688.

Charness, G., G. Frechette, and J. Kagel (2004): "How Robust is Laboratory Gift Exchange?," *Experimental Economics*, 7(2), 189–205.

Charness, G., and M. Rabin (2002): "Understanding Social Preferences with Simple Tests," *Quarterly Journal of Economics*, 117(3), 817–869.

Clark, K., and M. Sefton (2001): "The Sequential Prisoner's Dilemma: Evidence on Reciprocation," *The Economic Journal*, 111(468), 51–68.

Cohn, A., E. Fehr, and L. Goette (2012): "Fair Wages and Effort: Evidence from a Field Experiment," Lucca, Italy, 2011. Centre for Economic Policy Research, Meeting: The Role of Experiments for the Advancement of Effective Labour Legislation.

Crawford, V., and J. Meng (2011): "New York City Cabdrivers Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income," *American Economic Review*, 101(5), 1912–1932.

Dufwenberg, M., and G. Kirchsteiger (2004): "A Theory of Sequential Reciprocity," *Games and Economic Behavior*, 47(2), 268–298.

Englmaier, F., and S. Leider (2012): "Managerial Payoff and Gift Exchange in the Field," *CESifo working paper No. 3707.*

Ericson, K., and A. Fuster (2011): "Expectations as Endowments: Evidence on Reference-Dependent Preferences from Exchange and Valuation Experiments," *Quarterly Journal of Economics*, 126, 1879–1907.

Esteves-Sorenson, C., and R. Macera (2012): "Reference-Dependent Social Preferences," *Working Paper.*

Esteves-Sorenson, C., R. Macera, and V. Pohl (2013): "Efficiency Wages and Its Mechanisms: Empirical Evidence," *Working Paper.*

Falk, A. (2007): "Gift exchange in the field," *Econometrica*, 75(5), 1501–1511.

Falk, A., and U. Fischbacher (2006): "A Theory of Reciprocity," *Games and Economic Behavior*, 54(2), 293–315.

Falk, A., and A. Ichino (2006): "Clean Evidence on Peer Effects," *Journal of Labor Economics*, 24(1), 38–57.

Fehr, E., E. Kirchler, A. Weichbold, and S. Gächter (1998): "When Social Norms Overpower Competition: Gift Exchange in Experimental Labor Markets," *Journal of Labor Economics*, 16(2), 324–351.

Fehr, E., G. Kirchsteiger, and A. Riedl (1993): "Does Fairness Prevent Market Clearing? An Experimental Investigation," *The Quarterly Journal of Economics*, 108(2), 437–459.

Fehr, E., A. Klein, and K. Schmidt (2007): "Fairness and Contract Design," *Econometrica*, 1, 121–154.

Frederick, S., and G. Loewenstein (1999): "Hedonic Adaptation," *in "Scientific Perspectives on Enjoyment, Suffering and Well-Being." eds Kahneman, Daniel. and Diener, E. and Schwarz, N. (Russel Sage Foundation).*

Gächter, S., and A. Falk (2002): "Reputation and Reciprocity: Consequences for the Labour Relation," *Scandinavian Journal of Economics*, 104, 1–26.

Gibbons, R., and M. Waldman (1999): "Careers in Organizations: Theory and Evidence," *in "Handbook of Labor Economics." eds: Ashenfelter, O. and Card, David. (Elsevier)*, 3.

Gill, D., and V. Prowse (2012): "A Structural Analysis of Disappointment Aversion in a Real Effort Competition," *American Economic Review*, 102(1), 469–503.

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

Hannan, R., J. Kagel, and D. Moser (2002): "Partial Gift Exchange in an Experimental Labor Market: Impact of Subject Population Differences, Productivity Differences, and Effort Requests on Behavior," *Journal of Labor Economics*, 20(4), 923–951.

Hart, O., and J. Moore (2008): "Contracts as Reference Points," *Quarterly Journal of Economics*, 123(1), 1–48.

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

Kahneman, D., J. Knetsch, and R. Thaler (1986): "Fairness as a Constraint on Profit Seeking: Entitlements in the Market," *The American Economic Review*, 76(4), 728–741.

Kahneman, D., and A. Tversky (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47(2), 263–291.

Karlan, D. (2005): "Using Experimental Economics to Measure Social Capital and Predict Financial Decisions," *American Economic Review*, 95(5), 1688–1699.

Kőszegi, B., and M. Rabin (2006): "A Model of Reference-Dependent Preferences," *Quarterly Journal of Economics*, 121(4), 1133–1165.

——— (2007): "Reference-Dependent Risk Attitudes," *American Economic Review*, 97(4), 1047–1073.

——— (2009): "Reference-Dependent Consumption Plans," *American Economic Review*, 99(3), 909–936.

Kube, S., M. Maréchal, and C. Puppe (2012): "The Currency of Reciprocity: Gift Exchange in the Workplace," *American Economic Review*, 102(4), 1644–1662.

——— (forthcoming): "Do Wage Cuts Damage Work Morale? Evidence From a Natural Field Experiment," *Journal of the European Economic Association*.

Lazear, E. (2000): "Performance Pay and Productivity," *American Economic Review*, pp. 1346–1361.

Levitt, S., and J. List (2007): "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?," *Journal of Economic Perspectives*, 21(2), 153–174.

List, J. (2006): "The Behavioralit Meets the Market: Measuring Social Preferences and Reputation Effects in Actual Transactions," *Journal of Political Economy*, 114(1).

Loomes, G., and R. Sugden (1986): "Disappointment and Dynamic Consistency in Choice under Uncertainty," *Review of Economic Studies*, 53(2), 271–282.

Mas, A., and E. Moretti (2009): "Peers at Work," *American Economic Review*, 99(1), 112–145.

Maximiano, S., R. Sloof, and J. Sonnemans (2007): "Gift Exchange in a Multi-Worker Firm," *The Economic Journal*, 117(522), 1025–1050.

Pope, D. G., and M. E. Schweitzer (2011): "Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes," *American Economic Review*, 101, 129–157.

Rabin, M. (1993): "Incorporating Fairness into Game Theory and Economics," *American Economic Review*, pp. 1281–1302.

Schneider, F., and R. Weber (2012): "Long Term Commitment and Cooperation," *Unpublished Manuscript, Univeristy of Zurich.*

Shapiro, C., and J. Stiglitz (1984): "Equilibrium Unemployment as a Worker Discipline Device," *The American Economic Review*, 74(3), 433–444.

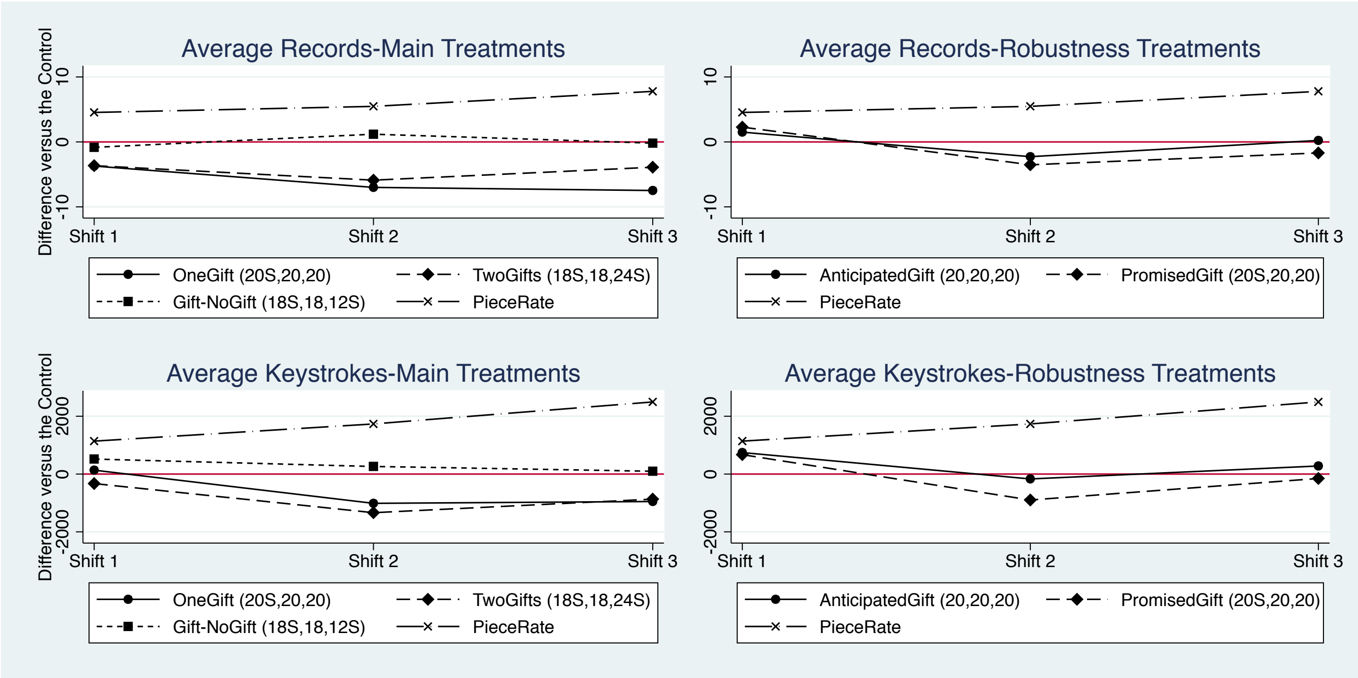
Stoop, J., C. N. Noussair, and D. van Soest (2012): "From the Lab to the Field: Cooperation Among Fishermen," *Journal of Political Economy*, 120(6), 1027–1056.

Weiss, A. (1980): "Job Queues and Layoffs in Labor Markets with Flexible Wages," *Journal of Political Economy*, 88(3), 526–538.

Zizzo, D. (2010): "Experimenter Demand Effects in Economic Experiments," *Experimental Economics*, 13, 75–98.

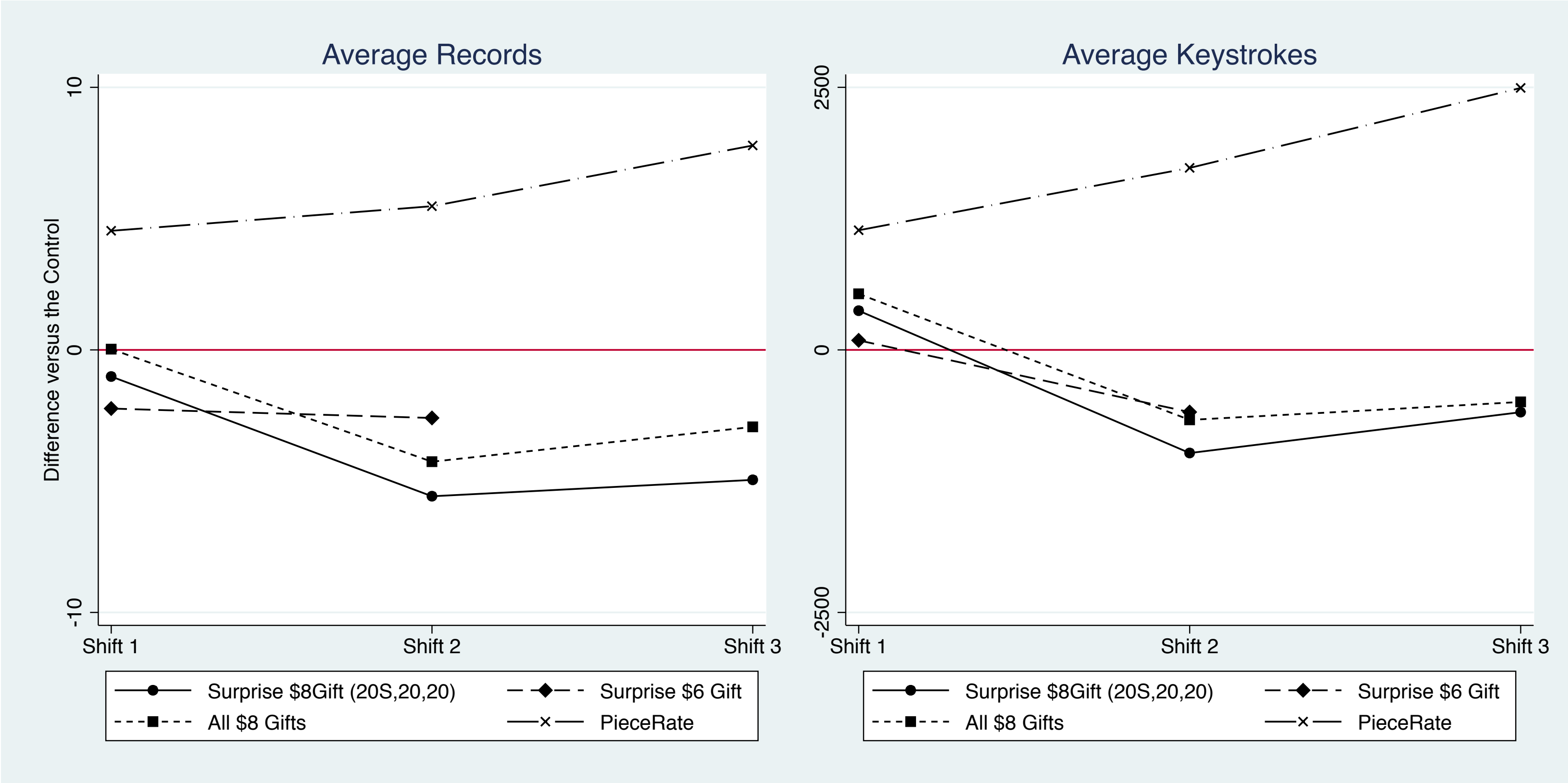
A Figures and Tables

Figure 2: Average Records and Keystrokes within Campus and Leg—All Treatments



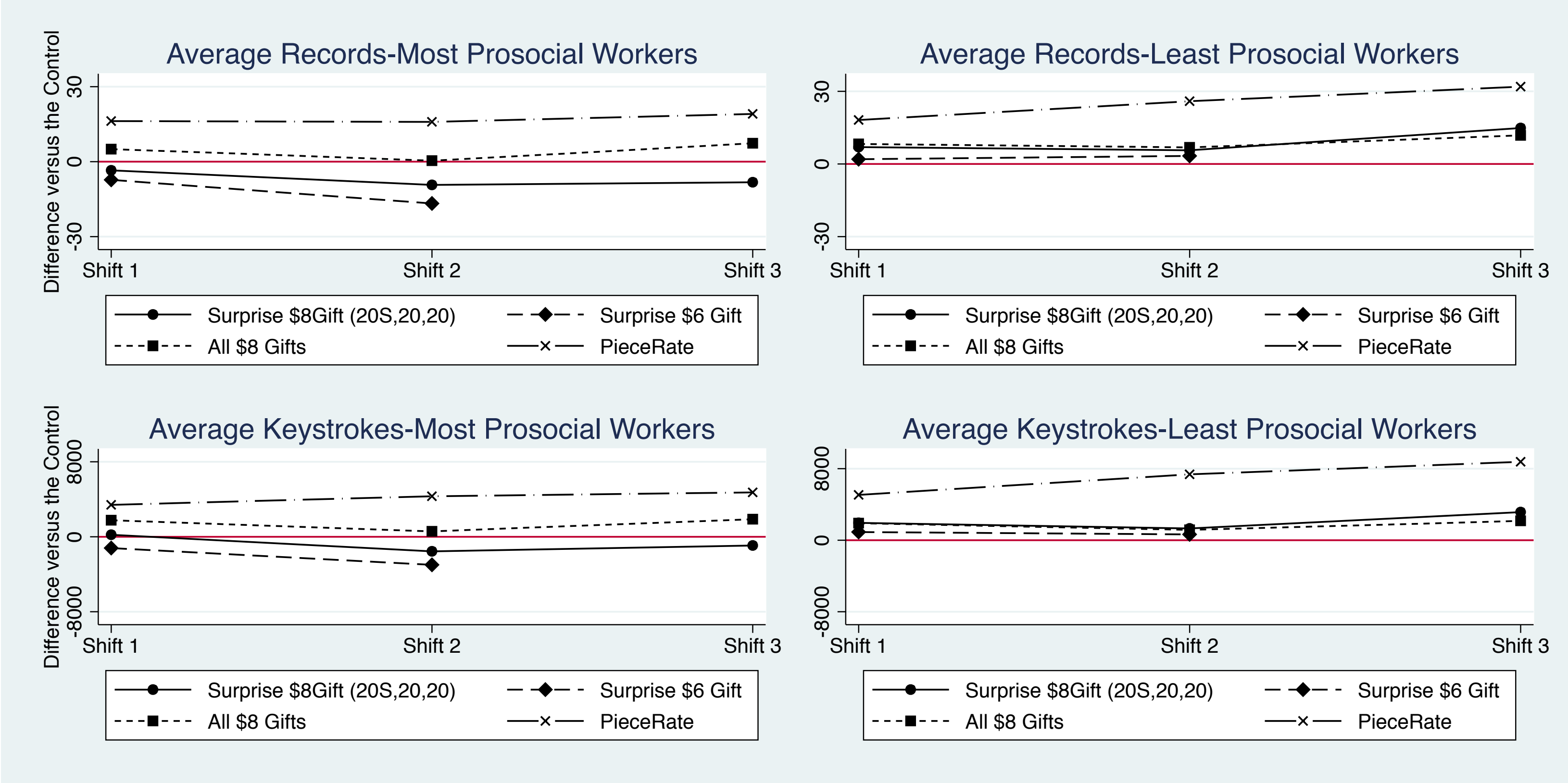
Notes: (1) The left panel top represents the average number of records entered by workers in the main gift treatments and how they compare against the average number of records entered by workers offered the piece rate: (i) OneGift, where workers receive a surprise wage increase of \$8 to \$20 per hour for the duration of the contract, (i) TwoGifts, where workers receive two surprise wage increases of \$6: one in shift one, increasing their wage to \$18 per hour and one in shift three increasing their wage to \$24 per hour, (iii) Gift-NoGift, where workers receive a surprise wage increase of \$6 in shift one, increasing their wage to \$18 per hour, and a surprise wage decrease in shift three decreasing their wage to the contracted \$12 per hour, (iv) and the PieceRate treatment where workers are offered a piece rate upon arriving at the work site. (2) The right top panel represents the average number of records entered by workers in the robustness gift treatments and how they compare against the average number of records entered by workers offered the piece rate. The robustness gift treatments are: (i) AnticipatedGift, where workers receive a surprise wage increase of \$8 to \$20 per hour for the duration of the contract, but where they received the news about the wage increase one week in advance of the start of work, (ii) PromisedGift where workers receive a surprise wage increase of \$8 to \$20 per hour for the duration of the contract, but where this wage increase is only paid at the end of the contract, instead of at the beginning of each shift, as in all other gift treatments. (3) The bottom left panel coveys the same information as the top left panel (see (1)) but for keystrokes instead of records. (4) The bottom right panel conveys the same information as the top right panel (see (2)) but for keystrokes instead of records. Standard error bars were omitted for visual clarity.

Figure 3: Average Records and Keystrokes within Campus and Leg—Pooled Gift Treatments



Notes: (1) The left panel represents the average number of records entered by workers when pooling all the gift treatments and how they compare versus the PieceRate treatment. (1) The *Surprise $8Gift* pools the two treatments when workers receive a surprise in shift one that they are going to receive a wage increase of $8 to $20: (i) OneGift, where workers receive a surprise wage increase of $8 to $20 per hour for the duration of the contract, and (ii) PromisedGift where workers receive a surprise wage increase of $8 to $20 per hour for the duration of the contract, but where this wage increase is only paid at the end of the contract, instead of at the beginning of each shift, as in all other gift treatments; (2) *All $8 Gifts* pools the three treatments with $8 gifts: the previous treatments (i) and (ii) as well as the AnticipatedGift treatment, where workers receive a wage increase of $8 to $20 per hour for the duration of the contract, but where they receive the news about the wage increase one week in advance of the start of work; (3) *Surprise $6 Gift*, which pools the treatments where workers receive a $6 wage surprise wage increase to $18 in shift one which remains in place in shift two: (i) TwoGifts and (ii) Gift-NoGift. (2) The right panel conveys the same information as the left panel but for the average number of keystrokes pressed instead of records entered. Standard error bars were omitted for visual clarity.

Figure 4: Average Records and Keystrokes within Campus and Leg By Social Preferences Type—Pooled Gift Treatments



Notes: The top left panel depicts the average number of records entered by *Most Prosocial* workers when pooling all the gift treatments, and how they compare versus the PieceRate treatment. The *Most Prosocial* workers are those who behaved the most prosocially in the three sequential Prisoner's Dilemma games: (i) either by cooperating as a first mover, cooperating as a second mover if the first mover cooperated, and cooperating as second mover even if the first player defected (CCC), or (ii) by only defecting when the first mover defected (CCD). The treatment (1) *Surprise $8Gift* pools the two treatments when these workers receive a surprise in shift one that they are going to receive a wage increase of $8 to $20: (i) OneGift, where these workers receive a surprise wage increase of $8 to $20 per hour for the duration of the contract, and (ii) PromisedGift where these workers receive a surprise wage increase of $8 to $20 per hour for the duration of the contract, but where this wage increase is only paid at the end of the contract, instead of at the beginning of each shift, as in all other gift treatments; (2) *All $8 Gifts* pools the three treatments with $8 gifts: the previous treatments (i) and (ii) as well as AnticipatedGift, where these workers receive a wage increase of $8 to $20 per hour for the duration of the contract, but where they receive the news about the wage increase one week in advance of the start of work; (3) *Surprise $6 Gift*, which pools the treatments where workers receive a $6 wage surprise wage increase to $18 in shift one which remains in place in shift two: (i) TwoGifts and (ii) Gift-NoGift. The top right panel conveys the same information as the top left panel, but for the *Least Prosocial* workers: those who behaved the least prosocially in the sequential Prisoner's Dilemma games, by defecting as a first mover, defecting as a second mover even if the first mover cooperated, and defecting as a second mover even if the first player defected (DDD). The bottom left and right panels convey the same information at the top left and right panels, respectively, but for the number of keystrokes pressed instead of the number of records entered. Standard error bars were omitted for visual clarity.

Table 3: Summary statistics for campus A—Records, Keystrokes and Proportion of Error-Free Words per Two-Hour Shift

	Number of Subjects	Means per shift across all legs			Statistics across all shifts and legs			
		Shift One	Shift Two	Shift Three	Mean	SD	Min	Max
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Number of Records Entered per Two-Hour Shift	111							
Control (12, 12, 12)	26	95	112	109	105	30	40	161
OneGift (20S, 20, 20)	12	97	105	106	103	33	46	191
TwoGifts (18S, 18, S24)	15	89	100	100	96	28	47	162
Gift-NoGift (18S, 18, 12S)	15	105	125	120	116	30	66	192
Anticipated Gift (20S, 20, 20)	18	94	101	103	99	41	40	214
PromisedGift (20S, 20, 20)	10	118	131	133	127	34	69	189
PieceRate	15	91	109	111	103	29	51	170
Number of Keystrokes Entered per Two-Hour Shift	111							
Control (12, 12, 12)	26	20,593	22,734	22,085	21,775	6,638	9,335	37,722
OneGift (20S, 20, 20)	12	19,452	19,637	19,907	19,665	5,931	8,061	30,645
TwoGifts (18S, 18, S24)	15	17,584	18,817	18,397	18,266	4,673	11,475	27,100
Gift-NoGift (18S, 18, 12S)	15	19,994	22,320	21,459	21,196	5,135	12,460	31,703
Anticipated Gift (20S, 20, 20)	18	18,115	19,044	19,008	18,716	6,911	7,800	37,828
PromisedGift (20S, 20, 20)	10	22,218	23,509	24,590	23,439	4,549	14,426	30,681
PieceRate	15	23,865	27,379	28,493	26,535	7,092	14,632	42,200
Proportion of Error-Free Words per Two-Hour Shift	82							
Control (12, 12, 12)	13	0.98	0.98	0.98	0.98	0.02	0.92	1.00
OneGift (20S, 20, 20)	10	0.97	0.98	0.98	0.98	0.01	0.93	0.99
TwoGifts (18S, 18, S24)	11	0.98	0.98	0.99	0.98	0.01	0.95	1.00
Gift-NoGift (18S, 18, 12S)	9	0.98	0.99	0.99	0.99	0.01	0.95	1.00
Anticipated Gift (20S, 20, 20)	12	0.98	0.99	0.98	0.98	0.02	0.94	1.00
PromisedGift (20S, 20, 20)	12	0.98	0.98	0.98	0.98	0.01	0.95	1.00
PieceRate	15	0.98	0.98	0.98	0.98	0.01	0.93	0.99

Notes: (1) Each treatment name is followed by the hourly wage per shift, in parentheses, where the "S" indicates whether the wage increase to the shown amount was a surprise. (2) The proportion of error-free words per two-hour shift is the proportion of words entered that were correctly spelled. (3) Columns (5)-(8) show summary statistics within treatment, but across all shifts and campuses. (4) The number of subjects on campus A for which we have data on the proportion of error-free words is lower than that for which we have data on the number of records and keystrokes. This was due to a software recording error in leg 2, which did not record the exact text the subjects entered, just their number of records and keystrokes. Given our large sample and little variation in the quality of the output this not a major issue in our analysis.

Table 4: Summary statistics for campus B—Records, Keystrokes and Proportion of Error-Free Words per Two-Hour Shift

	Number of Subjects	Means per shift across all legs			Statistics across all shifts and legs			
		Shift One	Shift Two	Shift Three	Mean	SD	Min	Max
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Number of Records Entered per Two-Hour Shift								
	83							
Control (12, 12, 12)	21	51	57	59	56	19	25	107
OneGift (20S, 20, 20)	11	58	67	63	63	18	35	101
TwoGifts (18S, 18, S24)	8	65	70	74	70	21	36	102
Gift-NoGift (18S, 18, 12S)	7	50	53	56	53	11	32	80
Anticipated Gift (20S, 20, 20)	7	84	88	94	89	38	33	169
PromisedGift (20S, 20, 20)	12	57	56	56	56	17	29	106
Piece Rate	17	50	58	59	56	24	17	117
Number of Keystrokes Entered per Two-Hour Shift								
Control (12, 12, 12)	21	12,644	13,436	13,497	13,188	4,683	6,298	26,183
OneGift (20S, 20, 20)	11	12,761	13,549	12,679	12,996	2,739	7,881	17,725
TwoGifts (18S, 18, S24)	8	13,468	12,923	14,084	13,492	3,477	3,792	18,237
Gift-NoGift (18S, 18, 12S)	7	12,184	12,004	11,800	12,006	3,610	7,868	19,239
Anticipated Gift (20S, 20, 20)	7	17,900	17,538	18,525	17,988	5,393	9,083	28,955
PromisedGift (20S, 20, 20)	12	12,065	10,851	10,643	11,212	3,504	4,983	25,106
PieceRate	17	13,734	14,854	14,771	14,461	5,623	6,455	28,526
Proportion of Error-Free Words per Two-Hour Shift								
Control (12, 12, 12)	21	0.97	0.97	0.97	0.97	0.02	0.91	0.99
OneGift (20S, 20, 20)	11	0.96	0.97	0.97	0.96	0.02	0.87	0.99
TwoGifts (18S, 18, S24)	8	0.97	0.97	0.97	0.97	0.02	0.91	0.99
Gift-NoGift (18S, 18, 12S)	7	0.97	0.97	0.97	0.97	0.01	0.94	0.99
Anticipated Gift (20S, 20, 20)	7	0.97	0.97	0.97	0.97	0.02	0.93	1.00
PromisedGift (20S, 20, 20)	12	0.96	0.97	0.97	0.97	0.02	0.93	1.00
PieceRate	17	0.94	0.96	0.96	0.95	0.03	0.86	0.99

Notes: (1) Each treatment name is followed by the hourly wage per shift, in parentheses, where the "S" indicates whether the wage increase to the shown amount was a surprise. (2) The proportion of error-free words per two-hour shift is the proportion of words entered that were correctly spelled. (3) Columns (5)-(8) show summary statistics within treatment but across all shifts and campuses.

Table 5: Average Records and Keystrokes per Weekly Two-Hour Shift

PANEL I: RECORDS		Unadjusted			Adjusted per campus			Adjusted per campus X leg		
		Shifts			Shifts			Shifts		
	N	One	Two	Three	One	Two	Three	One	Two	Three
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	194									
Control (12,12, 12)	47	75	86	85	-	-	-	-	-	-
		(5)***	(6)***	(5)***						
Difference vs. Control										
OneGift	23	3	1	0	4	1	1	-4	-7	-7
(20S, 20,20)		(7)	(9)	(9)	(5)	(7)	(7)	(7)	(8)	(8)
TwoGifts	23	5	4	6	1	-2	0	-4	-6	-4
(18S, 18, 24S)		(7)	(9)	(8)	(6)	(7)	(7)	(6)	(7)	(7)
Gift-NoGift	22	12	14	14	6	8	7	-1	1	0
(18S, 18, 12S)		(9)	(11)	(10)	(7)	(8)	(6)	(7)	(8)	(6)
PieceRate	32	-6	-4	-2	-2	-2	1	5	5	8
		(7)	(9)	(9)	(5)	(6)	(6)	(7)	(8)	(8)
Robustness Gift Treatments										
AnticipatedGift	25	16	12	15	8	3	7	2	-2	0
(20, 20, 20)		(9)**	(10)	(10)	(9)	(10)	(10)	(8)	(9)	(9)
PromisedGift	22	9	6	7	13	8	10	2	-4	-2
(20S, 20, 20)		(10)	(12)	(12)	(7)**	(8)	(8)	(8)	(8)	(8)
PANEL II: KEYSTROKES										
Control (12,12, 12)	47	16,964	18,409	18,077	-	-	-	-	-	-
		(1,024)**	(1,155)***	(1,106)***						
Difference vs. Control										
OneGift	23	-712	-1,684	-1,627	-536	-1,577	-1,533	134	-1,014	-950
(20S, 20,20)		(1,492)	(1,665)	(1,734)	(1,180)	(1,376)	(1,416)	(1,302)	(1,431)	(1,494)
TwoGifts	23	-812	-1,642	-1,180	-1,695	-2,595	-2,145	-330	-1,337	-867
(18S, 18, 24S)		(1,336)	(1,672)	(1,474)	(1,204)	(1,416)*	(1,301)	(1,212)	(1,406)	(1,277)
Gift-NoGift	22	545	300	331	-579	-635	-894	519	263	96
(18S, 18, 12S)		(1,682)	(1,945)	(1,765)	(1,371)	(1,446)	(1,284)	(1,403)	(1,489)	(1,307)
PieceRate	32	1,672	2,316	3,097	2,156	2,853	3,639	1,139	1,732	2,494
		(1,719)	(1,948)	(2,105)*	(1,333)*	(1,473)**	(1,630)**	(1,766)	(1,916)	(1,971)*
Robustness Gift Treatments										
AnticipatedGift	25	1,090	195	795	-343	-1,213	-721	744	-170	279
(20, 20, 20)		(1,568)	(1,799)	(1,781)	(1,644)	(1,778)	(1,821)	(1,570)	(1,672)	(1,720)
PromisedGift	22	-284	-1,531	-793	438	-1,054	-329	670	-899	-154
(20S, 20, 20)		(1,770)	(1,986)	(2,079)	(1,268)	(1,306)	(1,382)	(1,449)	(1,404)	(1,494)

Notes: (1) Each treatment name is followed by the hourly wage per shift, in parentheses, where the "S" indicates whether the wage increase to the shown amount was a surprise. (2) Standard errors clustered by worker. (3) *Significant at the 10% level, **Significant at the 5% level, ***Significant at the 1% level in one-tailed tests $H_0 : \beta_i = 0$ versus $H_a : \beta_i > 0$ (consistent with power calculations). Details on the specification are described in the body of the paper.

Table 6: Average Records and Keystrokes per Two-Hour Shift—Pooled by Gift Treatment

PANEL I: RECORDS										
		Unadjusted			Adjusted per campus			Adjusted per campus X leg		
		Shifts			Shifts			Shifts		
	N	One	Two	Three	One	Two	Three	One	Two	Three
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	194									
Control (12,12, 12)	47	75	86	85	-	-	-	-	-	-
		(5)***	(6)***	(5)***						
Difference vs. Control										
Surprise $8Gift (20S, 20, 20)	45	6	3	4	9	4	5	-1	-6	-5
(OneGift and PromisedGift)		(7)	(8)	(8)	(5)**	(6)	(6)	(6)	(7)	(6)
Surprise $6Gift	45	9	9	-	3	3	-	-2	-3	-
(18S,18, 24S or 12S)		(7)	(8)		(5)	(6)		(6)	(6)	
(TwoGifts and Gift-NoGift)										
Piece Rate	32	-6	-4	-2	-2	-2	1	5	6	8
		(7)	(9)	(9)	(5)	(6)	(6)	(7)	(8)	(8)
All $8Gifts (20,20, 20)										
(OneGift, PromisedGift and	70	9	6	8	9	4	6	0	-4	-3
AnticipatedGift)		(6)	(8)	(8)	(5)**	(6)	(6)	(6)	(6)	(6)
PANEL II: KEYSTROKES										
Control (12,12, 12)	47	16,965	18,409	18,077	-	-	-	-	-	-
		(1,016)***	(1,147)***	1,098)***						
Difference vs. Control										
Surprise $8Gift (20S, 20, 20)	45	-503	-1,611	-1,229	-60	-1,328	-959	373	-983	-593
(OneGift and PromisedGift)		(1,356)	(1,517)	(1,549)	(1,044)	(1,144)	(1,175)	(1,193)	(1,216)	(1,259)
Surprise $6Gift	45	-149	-739	-	-1,148	-1,682	-	85	-593	-
(18S,18, 24S or 12S)		(1,285)	(1,507)	0	(1,084)	(1,202)	0	(1,113)	(1,201)	0
(TwoGifts and Gift-NoGift)										
Piece Rate	32	1,672	2,316	3,097	2,155	2,852	3,638	1,139	1,733	2,494
		(1,707)	(1,934)	(2,090)*	(1,323)*	(1,463)**	(1,618)**	(1,753)	(1,902)	(1,956)*
All $8Gifts (20,20, 20)										
(OneGift, PromisedGift and	70	66	-973	-496	-161	-1,287	-872	534	-667	-249
AnticipatedGift)		(1,244)	(1,403)	(1,397)	(1,047)	(1,138)	(1,159)	(1,110)	(1,127)	(1,158)

Notes: (1) Each treatment name is followed by the hourly wage per shift, in parentheses, where the "S" indicates whether the wage increase to the shown amount was a surprise. The only exception is for the "All $8Gifts (20, 20, 20)" treatments, which only shows the hourly wage in parentheses. (2) Standard errors clustered by worker. (3) *Significant at the 10% level, **Significant at the 5% level, ***Significant at the 1% level in one-tailed tests $H_0 : \beta_i = 0$ versus $H_a : \beta_i > 0$ (consistent with power calculations). Details on the specification are described in the body of the paper.

Table 7: Distribution of Prosocial Types Among Field Experiment Workers—Overall and by Treatment

		Workers in Field Experiment	Workers responding to Survey								
					Prosocial type						
					Most prosocial				Least Prosocial		Others
		Total	Total	Response Rate	(CCC)	(CCD)	Total	Prop. Resp.	(DDD)	Prop. Resp.	Total Others
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1)	N	194	139	0.72	10	47	57	0.41	45	0.32	37
(2)	Control	47	34	0.72	2	10	12	0.35	14	0.41	8
(3)	OneGift (20S, 20, 20)	23	15	0.65	0	8	8	0.53	5	0.33	2
(4)	TwoGifts (18S, 18, S24)	23	17	0.74	1	3	4	0.24	9	0.53	4
(5)	Gift-NoGift (18S, 18, 12S)	22	18	0.82	1	7	8	0.44	5	0.28	5
(6)	Anticipated Gift (20S, 20, 20)	25	21	0.84	1	8	9	0.43	6	0.29	6
(7)	PromisedGift (20S, 20, 20)	22	11	0.50	1	4	5	0.45	3	0.27	3
(8)	PieceRate	32	23	0.72	4	7	11	0.48	3	0.13	9

Notes: (1) Each treatment is followed by the hourly wage per shift where "S" indicates whether the wage increase from $12 to the displayed amount was a surprise. (2) *Most Prosocial* workers are those who behaved the most prosocially in the three sequential Prisoner's Dilemma games: (i) either by cooperating as a first mover, cooperating as a second mover if the first mover cooperated, and cooperating as second mover even if the first player defected (CCC), (ii) or by only defecting when the first mover defected (CCD). The *Least Prosocial* workers are those who behaved the least prosocially in these games: by defecting as a first mover, defecting as a second mover even if the first mover cooperated, and defecting as a second mover even if the first player defected (DDD).

Table 8: Records and Keystrokes by High and Low Prosocial Type—within Campus and Leg

PANEL I : RECORDS	Total Sample				Most Prosocial vs. Control				Least Prosocial vs. Control			
		Shifts				Shifts				Shifts		
	N	One	Two	Three	N	One	Two	Three	N	One	Two	Three
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Total	194				83				72			
Control (12,12, 12)	47	-	-	-	47	-	-	-	47	-	-	-
Difference vs. Control												
(1) Surprise $8Gift	45	-1	-6	-5	13	-4	-9	-8	8	7	6	15
(20S, 20, 20)		(6)	(7)	(6)		(9)	(12)	(10)		(10)	(10)	(9)*
(OneGift and PromisedGift)												
(2) $6Gift	45	-2	-3	-	12	-7	-17	-	14	2	3	-
(18S, 18, 24S and 12S)												
(TwoGifts and Gift-NoGift)		(6)	(6)			(9)	(8)**			(8)	(8)	
(3) PieceRate	32	5	6	8	11	16	16	19	3	18	26	32
		(7)	(8)	(8)		(8)**	(10)*	(11)**		(11)*	(12)**	(14)**
(4) All $8Gifts (20, 20, 20)	70	0	-4	-3	22	5	0	7	14	8	7	12
(OneGift, PromisedGift and AnticipatedGift)		(6)	(6)	(6)		(8)	(9)	(10)		(9)	(9)	(9)

PANEL II: KEYSTROKES												
Total	194				83				72			
Control (12,12, 12)	47	-	-	-	47	-	-	-	47	-	-	-
Difference vs. Control												
(1) Surprise $8Gift	45	373	-983	-593	13	212	-1,554	-935	8	1,942	1,316	3,138
(20S, 20, 20)		(1,193)	(1,216)	(1,259)		(1,772)	(2,010)	(1,990)		(2,066)	(1,766)	(1,642)*
(OneGift and PromisedGift)												
(2) $6Gift	45	85	-593	-	12	-1,003	-2,997	-	14	926	654	-
(18S, 18, 24S and 12S)		(1,113)	(1,201)			(1,653)	(1,619)*			(1,490)	(1,530)	
(TwoGifts and Gift-NoGift)												
(3) PieceRate	32	1,139	1,733	2,494	11	3,404	4,320	4,734	3	5,064	7,364	8,769
		(1,753)	(1,902)	(1,956)		(2,007)**	(2,447)**	(3,084)*		(2,848)**	(3,351)**	(3,421)***
(4) All $8Gifts (20, 20, 20)	70	534	-667	-249	22	1,760	574	1,871	14	1,901	1,158	2,162
(OneGift, PromisedGift and AnticipatedGift)		(1,110)	(1,127)	(1,158)		(1,538)	(1,610)	(1,709)		(1,818)	(1,760)	(1,707)

Notes: (1) Each treatment is followed by the hourly wage per shift where "S" indicates whether the wage increase from $12 to the displayed amount was a surprise. (2) *Most Prosocial* workers are those who behaved the most prosocially in the three sequential Prisoner's Dilemma games: (i) either by cooperating as a first mover, cooperating as a second mover if the first mover cooperated, and cooperating as second mover even if the first player defected (CCC), (ii) or by only defecting when the first mover defected (CCD). The *Least Prosocial* workers are those who behaved the least prosocially in these games: by defecting as a first mover, defecting as a second mover even if the first mover cooperated, and defecting as a second mover even if the first player defected (DDD). (3) Standard errors clustered by worker (4) *Significant at the 10% level, **Significant at the 5% level, ***Significant at the 1% level in one-tailed tests $H_0 : \beta_i = 0$ versus $H_a : \beta_i > 0$.

B Online Appendix - Additional Figures and Tables

Figure B.1: Timing of field experiment and post-field experiment survey

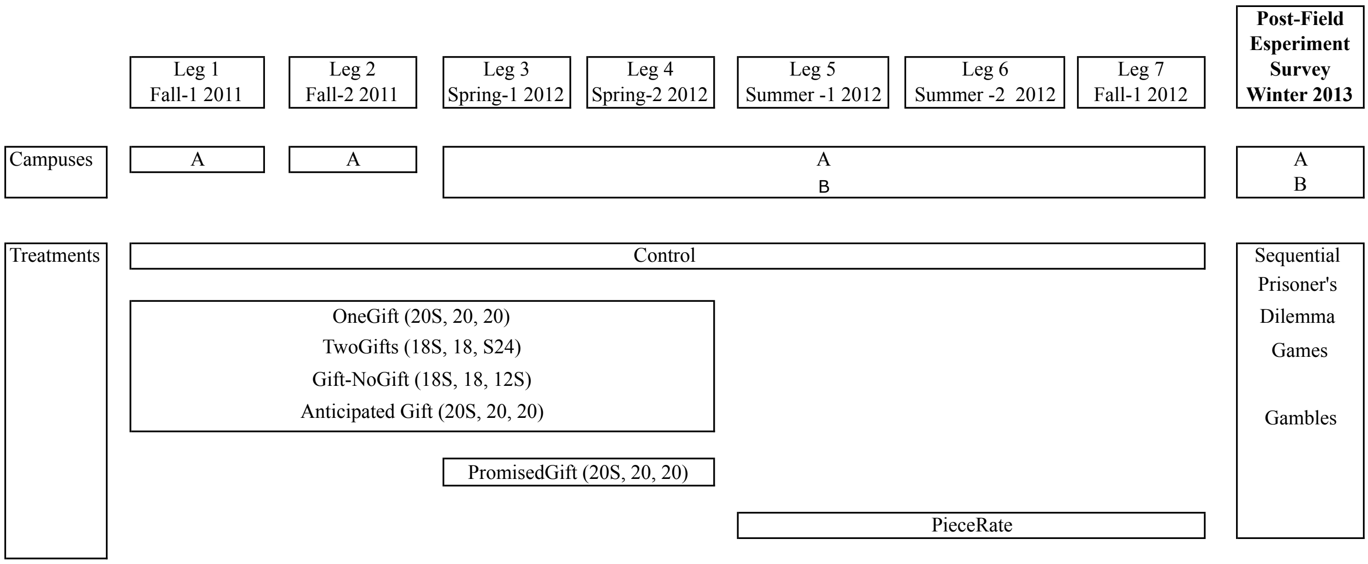


Table B.1: Average Records Inputted per Two-Hour Shift Pooled by Gift treatment—by Campus

	N	Unadjusted			Adjusted per campus X leg		
		Shift One	Shift Two	Shift Three	Shift One	Shift Two	Shift Three
		(1)	(2)	(3)	(7)	(8)	(9)
PANEL I: CAMPUS A	111						
Control (12,12, 12)	26	95	112	109	-	-	-
		(6)***	(6)***	(6)***			
Difference versus the control							
Surprise $8Gift (20S, 20, 20)	22	11	4	10	-1	-6	-2
(OneGift and PromisedGift)		(8)	(11)	(10)	(9)	(10)	(9)
Surprise $6Gift (18S, 18, 24S or 12S)	30	2	-1	-	-6	-7	-
(TwoGifts, Gift-NoGift)		(8)	(9)		(7)	(8)	
Piece Rate	15	-5	-3	2	8	11	15
		(9)	(9)	(10)	(13)	(14)	(15)
All $8Gift treatments (20,20, 20)							
(OneGift, PromisedGift and	40	6	-2	3	-6	-12	-8
AnticipatedGift)		(8)	(9)	(9)	(7)	(8)	(8)
PANEL II: CAMPUS B	83						
Control (12,12, 12)	21	51	57	59	-	-	-
		(5)***	(6)***	(5)***			
Difference versus the control							
Surprise $8Gift (20S, 20, 20)	23	7	4	1	2	0	-4
(OneGift and PromisedGift)		(5)	(11)	(6)	(8)	(8)	(8)
Surprise $6Gift (18S, 18, 24S or 12S)	15	7	5	-	5	4	-
(TwoGifts, Gift-NoGift)		(6)	(7)		(8)	(9)	
Piece Rate	17	-1	1	0	1	2	2
		(6)	(8)	(8)	(7)	(9)	(8)
All $8Gift treatments (20,20, 20)							
(OneGift, PromisedGift and	30	13	11	9	9	7	5
AnticipatedGift)		(6)**	(7)	(7)	(8)	(9)	(9)

Notes: (1) Each treatment name is followed by the hourly wage per shift, in parentheses, where the "S" indicates whether the wage increase to the shown amount was a surprise. The only exception is for the "All $8Gift (20, 20, 20)" treatments, which only shows the hourly wage in parentheses. (2) Standard errors clustered by worker. (3) *Significant at the 10% level, **Significant at the 5% level, ***Significant at the 1% level in one-tailed tests $H_0 : \beta_i = 0$ versus $H_a : \beta_i > 0$ (consistent with power calculations).

Table B.2: Average Keystrokes Inputted per Two-Hour Shift Pooled by Gift treatment—by Campus

		Unadjusted			Adjusted per campus X leg		
	N	Shift One	Shift Two	Shift Three	Shift One	Shift Two	Shift Three
		(1)	(2)	(3)	(7)	(8)	(9)
PANEL I: CAMPUS A	111						
Control (12,12, 12)	26	20,593	22,734	22,086	-	-	-
		(1,016)***	(1,147)***	(1,098)***			
Difference versus the Control							
Surprise $8Gift (20S, 20, 20)	40	116	-1,338	-50	509	-1,117	218
(OneGift and PromisedGift)		(1,671)	(1,885)	(1,946)	(1,735)	(1,782)	(1,842)
Surprise $6Gift (18S, 18, 24S or 12S)	30	-1,804	-2,291	-	-361	-1,041	-
(TwoGifts, Gift-NoGift)		(1,589)	(1,735)		(1,468)	(1,505)	
Piece Rate	15	3,272	4,645	6,407	1,513	2,713	4,465
		(2,086)	(2,262)**	(2,433)***	(3,225)	(3,549)	(3,577)
All $8Gift treatments (20,20, 20)	40						
(OneGift, PromisedGift and		-1,052	-2,363	-1412	-455	-1,899	-941
AnticipatedGift)		(1,589)	(1,746)	(1,784)	(1,483)	(1,476)	(1,529)
PANEL II: CAMPUS B	83						
Control (12,12, 12)	21	12,644	13,436	13,497	-	-	-
		(1,016)***	(1,147)***	(1,098)***			
Difference versus the Control							
Surprise $8Gift (20S, 20, 20)	30	-246	-1,338	-1836	509	-1,117	218
(OneGift and PromisedGift)		(1,256)	(1,885)	(1,277)	(1,735)	(1,782)	(1,842)
Surprise $6Gift (18S, 18, 24S or 12S)	15	224	-942	-	1,195	45	-
(TwoGifts, Gift-NoGift)		(1,307)	(1,631)	0	(1,671)	(1,934)	
Piece Rate	17	1,089	1,419	1,275	812	1,074	870
		(1,612)	(1,738)	(1,932)	(1,794)	(1,970)	(2,097)
All $8Gift treatments (20,20, 20)	30						
(OneGift, PromisedGift and		1,038	53	-179	2,169	1,229	980
AnticipatedGift)		(1,291)	(1,383)	(1,420)	(1,644)	(1,683)	(1,772)

Notes: (1) Each treatment name is followed by the hourly wage per shift, in parentheses, where the "S" indicates whether the wage increase to the shown amount was a surprise. The only exception is for the "All $8Gifts (20, 20, 20)" treatments, which only shows the hourly wage in parentheses. (2) Standard errors clustered by worker. (3) *Significant at the 10% level, **Significant at the 5% level, ***Significant at the 1% level in one-tailed tests $H_0 : \beta_i = 0$ versus $H_a : \beta_i > 0$ (consistent with power calculations).

C Online Appendix - Equilibrium

Definition 1 *(The Agent's Equilibrium Behavior)*

An executed effort e_t *and a plan for future effort* $\tilde{\boldsymbol{e}}_{\boldsymbol{t},\boldsymbol{\tau}}$ *correspond to an equilibrium for the agent if*

(i) $e_t \in \operatorname*{argmax}_{e'_t \in E} EU_t(e'_t, w_t | \tilde{e}_{t-1,t}, \tilde{w}_{t-1,t})$ *and*

(ii) $\tilde{\boldsymbol{e}}_{\boldsymbol{t},\boldsymbol{\tau}} \in \operatorname*{argmax}_{\boldsymbol{e}'_t \in E^{PE}} EU_\tau(\boldsymbol{e}'_t, \tilde{\boldsymbol{w}}_{\boldsymbol{t},\boldsymbol{\tau}} | \boldsymbol{e}'_t, \tilde{\boldsymbol{w}}_{\boldsymbol{t},\boldsymbol{\tau}})$ $\tau > t$

where (1) E *corresponds to the set of all possible efforts and* E^{PE} *corresponds to the set of all efforts that constitute an equilibrium for periods* $\tau > t$ *given* $\tilde{\boldsymbol{w}}_{\boldsymbol{t},\boldsymbol{\tau}}$, (2) $\tilde{e}_{t-1,t}$ *constituted an equilibrium in period t-1 and* (3) $\tilde{\boldsymbol{w}}_{\boldsymbol{t},\boldsymbol{\tau}}$ *and* $\tilde{w}_{t-1,t}$ *are rationally formed given the economic environment.*

Part (i) in Definition 1 says that in each period the agent exerts effort so as to maximize his utility given the period-t expectations about effort and wages he formed in period $t-1$. Part (ii) ensures that when forming his effort plans for the future periods—effort plans that will determine his expectations about the principal's payoff and kindness—he will only choose a plan that he knows he will be willing to follow given the outcome expectations these plans generate.[48]

One nice implication of this solution concept in an environment with fixed wages is stated in the following lemma.

Lemma 1 *(Optimal Effort Plans)*

Suppose the agent expects his future wage to be a known fixed wage. Then, he forms his effort plans as a consumption utility maximizer.

Proof Lemma 1

We prove it for period zero. The proof for the updating at the end of period one is analogous. Furthermore, because the agent's problem is separable across periods, we focus on period-one plans, independently of period-two plans. When the agent anticipates certainty, his period-zero wage expectations for period one are $\tilde{w}_{0,1} = z$ where z is any constant. According to Definition 1 he solves

$$\max_e \Big\{ z + \alpha(z-w)(be-z) + \eta\mu\Big((z-w)-(z-w)\Big)\mu\Big(be-z-(be-z)\Big) - \frac{1}{2}(e-\underline{e})^2 - C + \eta_e\mu\Big(-\frac{1}{2}(e-\underline{e})^2 + \frac{1}{2}(e-\underline{e})^2\Big)\Big\}$$

[48] This equilibrium concept differs from that in Kőszegi and Rabin (2009) in that it does not impose that future implemented efforts match the agent's current plans. As it will be clear from Section 3.2 this more lax solution concept is capable of dealing with total unexpected arrival of information (rather than just uncertainty resolution) by defining the equilibrium "point wise."

Because $\mu(0) = 0$, this reduces to $\max_e \left\{z + \alpha(z-w)(be-z) - \frac{1}{2}(e-\underline{e})^2 - C\right\}$, which is the same problem a consumption utility maximizer with pure reciprocal preferences would solve. ■

D Online Appendix - Proofs and Extra Proposition

Proof of Proposition 1

The result for all three preferences comes from noticing that in this environment all three solve the same maximization problem in every period: $\max_{e_t} w - \frac{1}{2}(e_t - \underline{e})^2 - C$ since $K_a = 0$ and $\mu(0) = 0$ because wages are fixed. The solution to this maximization problem corresponds to $e_t^* = \underline{e}$ for all t.

■

Proof of Proposition 2

Classical preferences are obvious because the gift does not depend on effort and thus they solve $\max_{e_t} w + g_1 - \frac{1}{2}(e_1 - \underline{e})^2 - C$ in all periods. Pure reciprocal preferences and reference-dependent reciprocity are derived in the text. ■

Proof of **Proposition 3**

We focus on period two as period one mimics Proposition 2. Classical preferences are obvious because the gift does not depend on effort for any gift size. Agents with pure reciprocal preferences exert $e_2^* = \underline{e} + \alpha(g_2 + g_1)b > \alpha g_1 b = e_1^*$. Agents with reference-dependent reciprocal effort solve

$$\max_{e_2} w + g_1 + g_2 + \alpha(g_1 + g_2)(b(e_2 - \underline{e}) - (g_1 + g_2)) + \eta_k \alpha \mu(g_2)\mu(b(e_2 - \underline{e}) - g_2) - \frac{1}{2}(e_2 - \underline{e})^2 + \eta_e \mu(-\frac{1}{2}(e_2 - \underline{e})) \quad (5)$$

with first-order condition

$$e_2^* = \underline{e} + \frac{\alpha(g_1 + g_2)b + \eta_k \alpha g_2 \mu'(b(e_2^* - \underline{e} - \alpha g_1 b) - g_2)b}{(1 + \eta_e \lambda)} > \underline{e} \quad (6)$$

■

Proof of **Proposition 4**

In period two the agent solves,

$$\max_{e_2} w + \eta_k \alpha \mu(-g_1)\mu\Big(b(e_2 - \underline{e} - \alpha b g_1) + g_1\Big) - \frac{1}{2}(e_2 - \underline{e})^2 - C + \eta_e \mu\Big(-\frac{1}{2}(e_2 - \underline{e})^2\Big)$$

since $K_a(w) = w - w = 0$, $K_a(w) - K_a(\tilde{w}_{1,2}) = (w - w) - (w + g_1 - w) = -g_1$ and $K_b(w) - K_b(\tilde{w}_{1,2}) = (be_2 - w) - (b\tilde{e}_{1,2} - (w + g_1)) = b(e_2 - \tilde{e}_{1,2}) + g_1$ and $\tilde{e}_{1,2} = \underline{e} + \alpha b g_1$. The necessary and sufficient first-order condition for an interior maximum corresponds to

$$e_2^* = \underline{e} - \frac{\lambda\eta_k\alpha g_1\mu'\big(b(e_2^* - \underline{e} - \alpha b g_1) + g_1\big)b}{(1 + \lambda\eta_e)} < \underline{e}$$

where the last equality follows from $\mu \in \{1, \lambda\}$ and $b > 0$.

Preamble to Extra Proposition 5

We now consider the case where agents in the OneGift treatment make a probabilistic assessment about the period-two payment. We show that in this case the agent still increases effort above the Control.

For simplicity we assume there are two levels of effort: high (where the agent works) and low (where the agent does not work), where $e_H = 1$ with $c(1) = c$ and $e_L = 0$ with $c(0) = 0$ and w.l.o.g we omit wages in levels and set $w = g_1 = 0$ to shorten utility specifications. Notice that under these assumptions the Control extorts effort zero because wages are fixed. The rest of the set up follows that in the main text.

Following our experimental design, the timing is as follows: At the end of period one the agent makes plans about the effort he will exert in period two. To form these plans he assigns probability $0 < \pi < 1$ to receiving the gift $G_2 \in \{0, g_2\}$ at the beginning of the second shift. There are, therefore, four possible contingent effort plans: work for any realization of G_2, work only if $G_2 = g_2$ (and therefore not work if $G_2 = 0$), do not work only if $G_2 = g_2$ (and therefore work if $G_2 = 0$), and never work for any realization of G_2.

Following Kőszegi and Rabin (2006), the agent's reference point corresponds to his rational expectations about outcomes. In our set up, this reference point corresponds to one of the agent's contingent effort plans described above. The fact that these plans constitute both, the reference and the actual action, poses a circularity problem: preferences are used to chose the reference plan but at the same time the reference plan determines the shape of the preference structure because agents have reference-dependent preferences.

To solve the circularity problem, Kőszegi and Rabin (2006) propose to use a *Personal Equilibrium* or PE. This solution concept takes advantage of the rational expectations assumption: agents can only form expectations they are willing to follow once these expectations constitute the reference point. An action therefore constitutes a PE if, once planned and implemented, it provides the highest utility relative to any possible deviation from the plan. For instance, in our framework, if the agent chooses to exert effort for any realization of w_2, then implementing this plan must give him higher utility than deviating to low effort—having planned to work—in case $w_2 = 0$ is observed at the beginning of period two. Whenever a given effort plan obeys this rational expectation condition, the plan it is said to be *credible* and it constitutes a personal equilibrium.

There could be, however, several plans that are credible. For instance, both, work for any wage and work only if $w_2 = w$ can be credible, in the sense that the agent will actually exert effort for any w_2 and $w_2 = w$, respectively, once these constitute the reference point. In this situation the agent chooses the credible effort plan that gives him the highest reference-dependent utility. Such plan constitutes what Kőszegi and Rabin (2006) call a *Preferred Personal Equilibrium* or PPE.[49]

Personal Equilibrium is not the only possible solution concept. Kőszegi and Rabin (2007) also propose a *Choice-Acclimating Personal Equilibrium* or CPE. As PPE, CPE proposes that the reference point corresponds to the agent's expectations about outcomes (one of the effort plans described above). Contrary to the PPE, however, CPE does not make sure that the equilibrium effort plan is credible—by comparing the equilibrium utility with the utility of all possible deviations from the equilibrium plan—but rather it assumes it: given that agents have rational expectations CPE imposes that an agent will implement the effort plan that constitutes his reference and then he will chose among these plans the one that provides the reference-dependent highest utility. For instance, to check whether "work" for any piece rate realization is a CPE, the agent computes the utility of working for any piece rate (conditional on this being the plan) and compares it against the utility that planning and implementing any other contingent effort plans provides.[50,51]

PPE and CPE are adequate in different temporal contexts. In particular, PPE is adequate whenever plan formation and its implementation are simultaneous or take place within a short period of time. Meanwhile, CPE is adequate whenever there is a time lag between the plan formation stage and their implementation. The crucial assumption in a CPE, therefore, is that this time lag will allow the agent to commit to his effort plans and thus implement them once planned. PPE, to the contrary, does not assume commitment (but rather imposes credibility) and thus it allows for short

[49]Notice in our main model forming expectations and choosing actual actions given expectations are simple problems because of two reasons. First, because plans were made under certainty and therefore they were computed through a straight consumption utility maximization, and second because departures from expectations happened right before effort was exerted, giving the agent no time to update his expectations.

[50]This implies that it can be the case that a plan that is not credible constitutes a CPE.

[51]The CPE is actually a very similar concept to the solution concepts proposed by Bell (1985) and Loomes and Sugden (1986). The difference is that in these papers the agent's reference point is assumed to be the lottery's certainty equivalent, meanwhile in Kőszegi and Rabin (2006) agents compare each possible scenario with each possible realization of the reference point.

time distances between planning and implementation.[52]

Because in our experimental design there is a week between the payment news and no information arrives between the end of period one—when period-two effort plans are made—and the second-period effort decision, we use the CPE as a solution concept.

Proposition 5 *(Stochastic Gift)*
Suppose $\tilde{w}_{1,2} = w + G_2$ *with* $G_2 \in \{0, g_2\}$, $g_2 > 0$ *where* $\mathbb{P}(G_2 = g_2) = \pi$. *Then,* $e_2 = 1$ *whenever* $G_2 = g_2$.

Proof We show that work if $G_2 = g_2$ and don't work if $G_2 = 0$ provides greater utility relative to never work and work if $G_2 = 0$ and don't work if $G_2 = g_2$. To see this, start noticing that the fair outcome in this framework is to exert effort zero and be paid fixed w.

The utility if the agent works whenever $G_2 = g_2$ and don't work whenever $G_2 = 0$, $U(e = 1 \text{ if } G_2 = g_2 | e = 1 \text{ if } G_2 = g_2)$, corresponds to

$$\pi\alpha g_2(b - g_2) - \pi\frac{c^2}{2} + \pi(1-\pi)(1+\lambda^2)\eta g_2(b - g_2) + \pi(1-\pi)(1-\lambda)\eta\frac{c^2}{2} \tag{7}$$

The utility if the agent never works, $U(e = 0 \text{ for all } G_2 | e1 \text{ for all } G_2)$, corresponds to

$$-\pi\alpha g_2^2 - \pi(1-\pi)\lambda\eta\frac{g_2^2}{2} < \tag{8}$$

which is obviously smaller than $U(e = 1 \text{ if } G_2 = g_2 | e = 1 \text{ if } G_2 = g_2)$.

Finally, the utility if the agent work whenever $G_2 = 0$ and don't work whenever $G_2 = g_2$, $U(e = 1 \text{ if } G_2 = 0 | e = 1 \text{ if } G_2 = 0)$, corresponds to

$$-\pi\alpha g_2(b + g_2) - (1-\pi)\frac{c^2}{2} - 2\pi(1-\pi)\lambda\eta g_2(b + g_2) + \pi(1-\pi)(1-\lambda)\eta\frac{c^2}{2} \tag{9}$$

which is also smaller than $U(e = 1 \text{ if } G_2 = g_2 | e = 1 \text{ if } G_2 = g_2)$.

■

[52] This is why Kőszegi and Rabin (2006) say that in a PPE the reference point corresponds to the agent's *recent* rational expectations.