Gift Exchange in the Workplace:
Addressing the Conflicting Evidence with a Careful Test

Constança Esteves–Sorenson[*]   Rosario Macera[†]

Yale University   Universidad Católica de Chile

January 2015

Abstract

Gift exchange, postulating that workers derive utility from reciprocating above-market wages with above-minimal effort, holds the promise that workers will exert extra effort even in the absence of performance-contingent rewards. Though extensively tested, this hypothesis has received mixed support. We identify the several confounds that could have led to the conflicting evidence—prosocial signaling, insufficient wage raises, an effort ceiling, fatigue, selection of high-productivity workers, peer effects, reemployment concerns, and small samples—and implement a comprehensive test dealing with them. Our test comprises a field experiment hiring workers for a data-entry job, followed by laboratory games assessing their prosocial behavior. After addressing these confounds, we find that workers did not engage in gift exchange—they did not increase effort in response to fixed wage raises—but raised effort in response to a piece rate. Further, workers who behaved prosocially in laboratory games did not behave prosocially in the field.

JEL Codes: D03, J31, J41, M52.

Keywords: Incentives, Gift Exchange, Field Experiments, Laboratory Experiments.

As a consequence of sentiment for the firm, the workers acquire utility for an exchange of "gifts." ... 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 [workers] could receive if they left their current jobs.

<div align="right">Akerlof (1982), pages 543-544</div>

1   Introduction

Proposed by Akerlof (1982), gift exchange has important implications for the provision of incentives. By postulating that workers derive intrinsic utility from reciprocating above-market wages with above-minimal effort, it holds the promise that workers will exert excess effort even in the absence of performance-contingent rewards, such as performance pay, reemployment, promotion or a favorable market reputation. It could thus obviate not only the need for costly monitoring technologies and performance evaluations to motivate workers, but also the use of repeated interactions, as agents would exert high effort even in one-shot settings.

Given its importance, gift exchange has been extensively tested, receiving mixed support. Laboratory evidence suggests it is a powerful incentive mechanism, yielding large pay-effort elasticities. For example, Fehr, Kirchsteiger, and Riedl (1993), the most cited laboratory test of gift exchange, found that a 140% fixed wage raise led to a 300% effort increase (a 2.14 elasticity). However, other tests, in particular field experiments, found lower or negative elasticities. For example, a 37% raise in Bellemare and Shearer (2009) increased effort by 11%-14% (a 0.30-0.38 elasticity), while a similar 33% raise in Kube, Maréchal, and Puppe (2012) lowered effort by -0.3% (a -0.01 elasticity), though this magnitude was statistically insignificant.

The conflicting gift-exchange elasticities have led to an extensive debate. Levitt and List (2007) argued that the considerable laboratory elasticities could be inflated because they reflect not only subjects' true prosocial preferences but also their desire to be viewed as prosocial by others. In contrast, Cohn, Fehr, and Goette (forthcoming) argued that small samples and thus low power could explain the failure to find statistically significant positive gift-exchange elasticities. And recently, Card, DellaVigna, and Malmendier (2011) pointed out that gift-exchange estimates could be dampened by the onerousness of some tasks, which might curb substantial effort increases.

We contribute to this discussion first by identifying the main factors that could have inflated or deflated gift-exchange elasticities in prior tests, leading to the conflicting evidence, and second by implementing a comprehensive test dealing with them. Our research design has two parts: a between-subjects field experiment in which workers were unaware they were part of a study, and later, laboratory games in which their prosocial behavior was assessed. Our aim was to find more accurate elasticities by tackling the three issues noted above—prosocial signaling, small samples and

<div align="center">1</div>

an effort ceiling which bounds effort responses—as well as five others: wage raises that are insufficient to compensate workers for higher effort; fatigue; selection of high-productivity workers with baseline output close to the highest feasible for a task; reemployment concerns promoting effort increases unrelated to reciprocity; and peer effects.

After addressing all the factors that could have biased gift-exchange's pay-effort elasticities, we cannot document gift exchange: we find statistically insignificant negative to small positive elasticities of -0.08 to 0.06. In contrast, a cheaper piece-rate scheme yielded much larger elasticities, reaching a conservative 0.40-0.60. Further, workers who behaved prosocially in the laboratory did not do so in the field.

In our field test, we recruited 194 students for a one-time data-entry job—the most used worker-job combination in field tests of gift exchange—for the going market wage, and randomly assigned them to four conditions varying additional compensation. Students from two universities were hired for $12 per hour to digitize an academic library for six hours split into three two-hour weekly shifts. After hiring, we randomly assigned 47 workers to the CONTROL, where they received no wage raises; 70 to a 67%RAISE condition, where they received a raise to $20 per hour prior to starting the task; 45 to a 50%-100%RAISE condition, where they received a raise to $18 per hour prior to starting the task and a further surprise raise to $24 per hour before the third shift; and 32 to a PIECERATE condition, where instead of the fixed wage raise, workers received an additional per-record piece rate before starting the job, representing an average 21% additional compensation.

This field test addressed the previously identified confounds in gift-exchange studies. First, our between-subjects design prevented prosocial image concerns from inflating gift-exchange estimates by shrouding individual workers' actions. Between-subjects tests, such as ours, compare effort between the group receiving a wage raise before the job (treatment) and the group not receiving it (control). Thus each worker in the treatment labors only under the wage raise. As a result, his baseline effort (effort in the absence of the raise) is unknown to others (e.g., the principal) and, therefore, it is also unknown whether he is behaving selfishly by not increasing effort under the raise. By enabling each worker in the treatment to behave selfishly without detection, between-subjects designs ensure that effort increases are not due to agents' desiring to avoid a selfish image but rather due to gift exchange (i.e., the intrinsic taste for reciprocating the wage). In contrast, in the two other tests of gift exchange—laboratory and within-subject field tests—effort in the absence of the raise is observable and thus selfishness can be easily detected. In laboratory studies, each worker's minimum effort is known to other players and/or the experimenter, while in within-subject field tests, each worker's

2

effort is observed before and after the raise. Thus, in either, others can see if the worker increased effort after the raise, giving him an additional motivation to boost effort: to be favorably viewed as prosocial instead of unfavorably as selfish.

Second, we used substantial fixed wage raises of 50%, 67% and 100%, which match or exceed those in other field tests, to allay concerns that raises might be too small to elicit effort and thus hinder gift exchange. Further, we also compared effort under these raises to that under the piece rate, which averaged only a 21% increase in pay. If effort does not increase under the large raises but rather under the cheaper piece rate, then lack of gift exchange cannot be due to insufficient raises.

Third, the piece rate scheme also allowed us to ascertain whether an effort ceiling could dampen responses to fixed wage raises, by observing whether effort increases occurred under the piece rate.

Fourth, we divided the six-hour task into three two-hour shifts to minimize the role of fatigue in the waning of reciprocal responses to wage raises. Workers who increased effort in a given two-hour shift to reciprocate the raise had time to rest until the next shift, so as to continue exerting high effort. Further, splitting the job over three weeks, also enabled workers to learn over time and to ponder how to become more productive from shift to shift.

Fifth, hiring at the market wage curbed the depression of gift exchange by the selection of high-productivity workers. Above-market wages might attract these workers, whose output may be close to the highest feasible for a task, preventing them from lifting it substantially to reciprocate a raise.

Sixth, the one-time job discouraged workers from increasing effort to be rehired at the higher wage, thus preventing reemployment concerns from inflating gift-exchange estimates. Seventh, workers labored alone and received no peer information to preclude peer effects from biasing performance.

Eighth, our large sample addressed the dual issue posed by small samples: low power for detecting true gift exchange elasticities; or, alternatively, estimating elasticities that are higher than the true one, as only those estimates which happen to be large enough by luck are able to be statistically significant, despite the large standard errors induced by the small sample.

In the second part of our study, the same workers were invited to play Sequential Prisoner's Dilemma (SPD) after the conclusion of the field experiment, as during it workers had to be unaware they were part of a study. These SPD games measure prosocial behavior in the laboratory, as in Burks, Carpenter, and Goette (2009) and Schneider and Weber (2012). We tested whether workers who behaved prosocially in these games, where each individual's prosocial actions could be observed, also had increased effort in response to their wage raise during the field experiment, where this prosocial action could not be observed due to our between-subjects design, as discussed.

Our main finding is that large wage raises of 50%, 67%, and a further raise to 100% yielded statistically insignificant effort changes of -4% to 4% (elasticities of -0.08 to 0.06) in the most favorable specification to gift exchange. However, the cheaper piece-rate scheme, peaking at only 30% extra pay by the third shift, increased effort by a conservative 18% (an elasticity of 0.60). Further, those who behaved the most prosocially in the laboratory did not behave distinctly prosocially on the job, exhibiting similar statistically insignificant negative to small positive effort changes of -3% to 7%.

This paper explains how the conflicting pay-effort elasticities in gift-exchange tests could be due to one or the combination of several confounds, which we then address in a single test. Though others have attempted to tackle some of these factors, such as an effort ceiling (Kube, Maréchal, and Puppe (2013)) and small samples (e.g., Cohn, Fehr, and Goette (forthcoming)), to our knowledge, none has jointly dealt with all the confounds explored here, as we document in Section 2.

Our findings thus further the debate on whether gift exchange is an efficient incentive mechanism. Prior large elasticities observed in gift-exchange games in the laboratory and the more modest elasticities seen in within-subject tests in the field have kindled the promise that principals, in particular those in settings where worker performance cannot be easily monitored or objectively assessed, may not need to engage in costly monitoring and/or subjective performance evaluations embedded in long-term relationships to elicit effort (see Gibbons (2005) for an overview of these standard incentives). Namely, workers could exert high effort even in the absence of these tools. Our results, showing that unconditional wage raises did not increase effort, while a piece rate did, suggest instead that prior positive elasticities may stem from one or more of the confounds addressed in our test.

2 Prior Gift-Exchange Tests in the Workplace and Potential Confounds

This section focuses on the most cited field and laboratory studies with gift-exchange tests in the workplace, all testing whether workers increased effort in response to fixed wage raises that were not conditional on performance and in one-shot interactions. These two features are used to ensure that excess effort is due to workers' intrinsic preference for reciprocating raises rather than to either the pursuit of higher performance-contingent pay or of reemployment at the higher wage, respectively.

We show that similar wage raises have yielded conflicting effort responses in this research, describe how this could stem from unanticipated confounds associated with reasonable features of these tests, and briefly note how we addressed these confounds. Panels I and II in Table 1 summarize field and laboratory tests, respectively, and Panel III surveys real-effort laboratory experiments that have accompanied field tests. Table 2 summarizes these tests' potential confounds, which we discuss below.

(1) *Prosocial signaling when others can observe the worker's minimum or baseline effort and the resulting advantage of between-subjects designs.* Much of the evidence for gift exchange has arisen in the laboratory, starting with the influential work by Fehr, Kirchsteiger, and Riedl (1993), introducing the Gift-Exchange Game and showing that fixed wage raises of 140% over the market clearing level increased workers' effort beyond the minimum by 300%, an elasticity of 2.14 (Table 1, Panel II, row (1)). The authors created four labor markets in the laboratory, each with 13-15 students randomly assigned to the role of employer or employee. Employers offered wages for three minutes, which workers accepted or rejected. Once hired, workers chose effort from the same cost-of-effort table— where the minimal effort of 0.1 units corresponded to a cost of effort of zero, for example—and each worker's effort choice was observable by the principal and the experimenters. There was no worker-level heterogeneity in cost of effort, as all workers had the same table. Each one-shot game, from the wage bids to the workers' effort choices, lasted ten minutes. Subjects played twelve one-shot games to enable learning, but their trading partner was kept anonymous to minimize reputation confounds.

An extensive experimental literature building on Fehr, Kirchsteiger, and Riedl (1993) has since shown large elasticities for gift-exchange games: employers offered excess wages of over 100% and workers chose excess effort of over 200% (Table 1, Panel II, columns (3) and (4)). These games shared generally similar features: subjects, usually students, were randomly assigned to be employers or employees and respectively given a common-knowledge profit and cost-of-effort function. The employer offered the wage first, in a publicly observable bid or a private offer to a randomly matched worker, and the employee chose effort second, with choices jointly determining payoffs in experimental units. These one-shot interactions lasted a few minutes before subjects were re-paired.

These large laboratory elasticities may, however, conflate gift exchange with subjects' desire to be perceived as prosocial. A review of prosocial behavior in laboratory experiments by Levitt and List (2007) documented that selfish behavior increased when it was harder for others (e.g., experimenters or other players) to observe each player's actions. In contrast, prosocial behavior increased when it was easier to ascertain which action each player had taken. This shift from selfish to prosocial behavior when actions could be observed suggested that subjects cared about being viewed as prosocial, instead of as selfish. Workers in gift-exchange games could thus have chosen higher effort not only because of an intrinsic taste for reciprocating a raise but also to signal their prosociality to experimenters and others who could observe their effort response to the raise.

Beyond possibly inflating gift-exchange elasticities in the laboratory, prosocial signaling may also overstate estimates in within-subject field tests. In field tasks, individual cost of effort is heteroge-

neous and unobservable, as some workers find the task harder than other workers. Thus, some studies test for gift exchange by observing each worker's effort before and after a raise, a within-subject test. Though this approach is reasonable, it introduces the confound of prosocial signaling: the worker knows the principal knows his baseline effort in the absence of the raise and thus whether the worker is behaving prosocially by increasing effort following the raise. Thus, the worker may lift his effort not only to reciprocate but also to signal prosociality, inflating the estimates.

Between-subjects tests of gift exchange prevent prosocial image concerns from inflating effort responses and thus may yield lower elasticities than those in laboratory and in within-subject tests. As discussed, between-subjects tests for gift exchange compare a treatment group that receives the raise with a control group that does not. Therefore, the baseline effort of each "treated" worker— effort in the absence of the raise—is unknown and thus whether he exceeded his baseline under the raise is also unknown. Thus each worker can behave selfishly by not lifting effort without worrying that others will detect this selfishness. As a result, effort increases in between-subjects tests are not inflated by workers' desiring to avoid a selfish image.

The confound of prosocial signaling may thus partially explain why between-subjects field tests document lower effort responses to the same wage raise than laboratory tests. For example, in the between-subjects field experiment in Gneezy and List (2006), the most cited gift-exchange field test, a 67% increase in the fixed wage increased output by 27% in the first 1.5 hours (an elasticity of 0.4), which was significant at the 5% level in a one-tailed test (Table 1, Panel I, row (1)). Effort waned thereafter, however, resulting in a statistically insignificant increase of 2% over the six hours of the task (an elasticity of 0.03). In contrast, the same 67% raise in the Bilateral Gift-Exchange Game played in the laboratory in Fehr, Kirchler, Weichbold, and Gächter (1998) yielded a much larger effort increase of 50%-100% and thus a substantially larger elasticity of 0.74-1.5.[1] Gneezy and List (2006) hired students for $12 per hour for a one-time job of digitizing the holdings of a library for six hours and randomly assigned them to two groups. The 10 in the control received the agreed wage whereas the 9 in the treatment received a surprise 67% raise to $20 per hour. Output was the number of records inputted (e.g., the author's name, title of book etc.).

The prosocial signaling confound may also partially explain why between-subjects designs yield lower effort responses than within-subject tests. For example, Kube, Maréchal, and Puppe (2013) found, in their between-subjects test, that a 33% raise did not increase output, but rather decreased it

---

[1]See Appendix D, Panel II, bullet (3.5) documenting the effort response to a 67% wage raise. Table 1, Panel II, row (3) summarizes this game, where many of the raises exceeded 67%, resulting in an average wage raise of 215%.

slightly by a statistically insignificant -0.3% (an elasticity of -0.01), as documented in Table 1, Panel I, row (7). In contrast, the within-subject test in Cohn, Fehr, and Goette (forthcoming), where workers' productivity was observed with and without a smaller wage raise of 23%, documented a statistically significant 3% increase in productivity, with an upper bound of 14% for a subsample (elasticities of 0.13 and 0.61, respectively). Specifically, in Kube, Maréchal, and Puppe (2013), in a setup similar to Gneezy and List's (2006), students digitized library holdings for six hours, a one-time job for a projected 15 euros per hour. The 25 subjects in the control received the agreed pay, whereas the 22 in the treatment received a surprise 33% raise to 20 euros per hour. Meanwhile, in the within-subject test in Cohn, Fehr, and Goette (forthcoming) documented in row (8), 196 workers distributed newspapers in three-hour shifts for 22 CHF per hour where, unbeknown to them, their wages would alternate between 22 CHF and 27 CHF during the study's four weeks. Effort increased by 3%. They subsequently surveyed workers on the amount by which they had felt underpaid and had them play a laboratory game assessing their reciprocity. Of the 61% who answered, 65% were labeled as reciprocal and 35% as nonreciprocal. Only those who answered, felt underpaid, and were reciprocal, responded to the raise: they increased effort by 2.8% for each CHF of underpayment, yielding an upper bound of 14% for those who felt underpaid by 5 CHF. These estimates, however, could be inflated by prosocial signaling as most workers' effort was observed pre- and post-raise.

We addressed the confound of prosocial signaling by using a between-subjects test. We further investigated whether this signaling could inflate gift-exchange elasticities by asking workers to play SPD games after the field experiment. Since gift-exchange games are also SPD games in that the first mover cooperates/defects by offering/not offering excess wages and the second mover cooperates/defects by exerting/not exerting excess effort, we tested whether prosocial actions by each worker in our SPD games, which could be observed (e.g., by the experimental team to process the payments), corresponded to those in the field, where they could not.

(2) *Size of fixed wage raise.* In contrast, past gift exchange elasticities may have been deflated by insufficient wage raises that did not compensate workers for their marginal cost of effort. For example, though Gneezy and List (2006) found that a 67% raise boosted effort by 27% in the first 1.5 hours of their 6-hour data-entry task, Kube, Maréchal, and Puppe (2013) showed that a smaller 33% raise did not increase effort in the first 1.5 hours on a similar job, rather it lowered effort by 10% (though this magnitude was statistically insignificant). To prevent small raises from inhibiting effort responses, we offered 50%, 67% and 100% fixed wage raises, matching or exceeding those in prior field studies: raises on data-entry jobs with student-workers, such as ours, have ranged from

19% to 67% and those in other field tests have ranged from 10% to 100% (Table 1, Panel I).

(3) *Effort ceiling.* Gift exchange in prior tests could also have been depressed by an effort ceiling: workers may have wanted to reciprocate a raise, but the task was so onerous that they quickly reached an infinite marginal cost of effort. For example, those receiving the 67% raise in Gneezy and List (2006) may have been not been able to increase effort beyond the observed 27% for the first 1.5 hours, bounding the elasticity at 0.4. In contrast, players receiving the same raise in the Bilateral Gift-Exchange Game in Fehr, Kirchler, Weichbold, and Gächter (1998) increased effort by 50%-100%, yielding approximately two-to-four times larger elasticities, as the workers cost-of-effort function could have been designed by the experimenter to minimize bounded effort responses.

We assessed whether an effort ceiling could curb gift exchange by running a piece-rate treatment testing whether it was feasible to increase effort in our task: we offered a subsample an additional per-record piece rate instead of a fixed wage raise. Our test thus differs from that in Kube, Maréchal, and Puppe (2013) who also explicitly tested whether an effort ceiling could explain the absence of gift exchange in their study. They found that workers hired via a piece-rate contract produced more than those hired with a fixed-wage contract and subsequently given a 33% fixed raise. This evidence, though suggestive, had been viewed as inconclusive as piece-rate contracts attract higher-productivity workers than fixed-wage contracts (Lazear (2000)). The higher output of the piece-rate hires could thus have been due to their higher productivity vis-à-vis the fixed-wage hires rather than due to the absence of an effort ceiling for the fixed-wage hires. We ensured that workers receiving fixed wage raises and the piece rate were similar as both selected into the same fixed-wage contract.

(4) *Fatigue.* Fatigue could have also deflated gift-exchange estimates, as workers became tired of exerting higher effort in response to a raise. For example, Gneezy and List (2006) showed that the bulk of the effort response to the 67% raise in their 6-hour data-entry task occurred in the first 1.5 hours. Similarly, in a companion field experiment with a fundraising task, they found that most of the response to the 100% fixed wage raise occurred in the first 3 hours of the 6-hour job. Of the 23 students hired at $10 per hour to raise funds for a charity, 13 (treatment) received a 100% raise to $20 per hour and 10 (control) did not. Those in the treatment raised 72% more funds in the first 3 hours than those in the control (Table 1, Panel I, row (2)). However, effort waned thereafter, translating into an average increase of 38% over the 6 hours, which was significant at the 10% level in a one-tailed test. The authors tested if this waning could be due to fatigue by inviting workers to raise more funds the next day, after resting. However, only 4 and 9 subjects in the control and treatment, respectively, returned the next day, yielding low power for detecting a difference.

Similarly, Bellemare and Shearer (2009), showed that most of an effort increase occurred on the day workers received a raise. In a seven-day field experiment, 18 tree planters received a one-time lump-sum of $80 on the second day they planted. This amounted to a 37% raise over the average daily earnings per worker of $215 and increased effort by 11%-14% (Table 1, Panel I, row (3)).

This spike-followed-by-waning in productivity in these studies could be due to fatigue. Namely, after increasing effort initially, workers became too tired (as past effort increased their subsequent marginal cost of effort) deflating effort over the total duration of these tasks.

To prevent fatigue from weakening gift exchange and to enable workers to learn and become comfortable with our study's task over time, we split it into three two-hour shifts exactly one week apart. Thus, workers who became tired on one shift from reciprocating the raise had one week to recover. Further, workers could think about how to improve productivity over the weekly interval.

(5) *Selection of higher-productivity workers.* Selection of higher-productivity workers may have also curbed gift exchange, as their baseline output may be close to the upper bound for a task preventing them from lifting it substantially after a raise. Hiring at above-market wages may attract these workers if reservation wages and ability are positively correlated (Weiss (1980), Bewley (1999)).

Selection of abler workers could thus be another reason (besides prosocial signaling) why Kube, Maréchal, and Puppe (2013) found that a 33% wage raise did not increase effort but Cohn, Fehr, and Goette (forthcoming) found that a 23% raise did. The former hired at 15 euros per hour, almost twice the market wage of 8 euros, potentially attracting a high proportion of higher-output workers (who would not have applied for the job at the lower market wage). Thus, workers receiving the 33% raise in the treatment group might have had little room to increase effort over that of similar high-productivity workers in the control. Workers in the latter study, however, were hired at the market wage, so they may have had more leeway to increase productivity.[2]

To prevent the selection of high-ability workers from deflating gift-exchange estimates, we hired at the market wage. This offers the further advantage of more closely following the gift-exchange hypothesis, whereby a natural proxy for the workers' outside option is the market wage.

(6) *Peer effects.* Some tests may have found negative though statistically insignificant gift-exchange elasticities due to peer effects. For example, Hennig-Schmidt, Sadrieh, and Rockenbach (2010) found that offering peer wage information together with a 40% wage raise decreased effort by 28%, though

---

[2]Cohn, Fehr, and Goette (forthcoming) hired at the market wage because they argued that hiring at above-market wages might curb effort increases to subsequent wage raises through a different mechanism than the selection of high-productivity workers: the perception of the fairness of the wage. The higher the hiring wage, the fairer it will be perceived to be and thus the lower the effort response to subsequent wage raises. This is because workers *do not* reciprocate raises to wages viewed as fair, as proposed by the fair wage-effort hypothesis in Akerlof and Yellen (1990).

this decline was statistically insignificant. They hired 103 students to transcribe abstracts for 20 DM per hour, for two one-hour sessions, each one month apart. Workers received the agreed hourly wage in the first session, whereas in the second session a random subsample of 23 received a surprise wage raise of 40% and information that peers were only receiving a 10% raise. Their increase in productivity from the first to the second hour was 28% *lower* than that for the 24 subjects in the control (Table 1, Panel I, row (5)). Wage-raise subjects could have lowered effort—though the magnitude was statistically insignificant—to punish the principal for behaving unfairly towards similar workers by giving them a substantially lower raise, in line with Fehr and Fischbacher (2004), who show that subjects punish unfair actions towards a third party.[3]

To avoid potential peer-effects confounds in our test, we had subjects work in isolation and gave them no information about other workers, such as their wages.

(7) *Reemployment confounds.* In contrast, some elasticities may have been inflated by reemployment concerns. For example, Bellemare and Shearer (2009), showed that planters who returned the following year responded more to the 37% fixed raise, by planting more trees than those who did not. Also, contrary to Kube, Maréchal, and Puppe (2013) where the surprise 33% raise did not increase performance, the same surprise raise in Gilchrist, Luca, and Malhotra (2014) increased performance by 26% among frequent oDesk workers, but it did not do so among first-timers (Panel I, row (9)). They hired a random sample of 168 workers (asking a \$2-\$3 hourly wage) at \$3, for a four-hour transcription task to be completed within one week. They gave a subsample of 58 a surprise raise from \$3 to \$4 per hour and informed them that the job was expected to be one-time. The raise increased effort only among frequent oDesk workers. One reason why only these workers responded to the raise could have been reputation concerns. In oDesk, workers' prior job history, such as the number of jobs, hourly wages and employer ratings are available for future employers. Thus, frequent workers could have increased effort following the raise not only to reciprocate but also to earn a good rating for the higher paid job, a valuable signal of their higher marginal product to future employers. Reputation would naturally have a higher value for the frequent workers—their heavy usage of oDesk suggests it is a substantial source of current and future earnings—than for first-timers.[4]

To avoid reemployment confounds, our workers were reminded during recruiting and the job execution that it was one-time.

---

[3]Cohn, Fehr, Herrmann, and Schneider (2014), however, show evidence that this punishment may not always occur, in a field experiment with teams of two workers.

[4]This paper contains another treatment, in which workers select into a \$4 per hour contract and perform the task at this wage. This treatment is thus outside of the scope of this review, which focuses on tests of gift exchange: whether workers reciprocate fixed wage raises with higher effort.

(8) *Small samples.* Last, small samples can also yield diverging evidence on gift exchange, by either having low power to identify effort increases or by overstating them. For example, Cohn, Fehr, and Goette (forthcoming) point out that the smaller and thus lower-power sample in Kube, Maréchal, and Puppe (2012) could account for their diverging findings. Though these studies offered similar raises (23% and 19%, respectively) and obtained similar effort increases (3% and 5%, respectively) their statistical significance differed. Cohn, Fehr, and Goette (forthcoming) found that the 3% increase in effort was statistically significant in their sample of 196 workers. In contrast, Kube, Maréchal, and Puppe (2012) found that the 5% increase in effort was statistically insignificant in their smaller sample of 69 workers (Table 1, Panel I, row (6)). They hired students for the one-time job of digitizing library holdings for three hours for 12 euros per hour. The 35 in the control received the agreed wage whereas the 34 in the treatment received a 19% raise.[5]

Small samples could, alternatively, overstate elasticities, as only estimates that are high, by chance, manage to be statistically significant despite the large standard errors induced by the small sample.[6] For example, the small-sample experimental conditions in Gneezy and List (2006), with 13 or fewer workers, could have inflated their statistically significant short-run elasticities. Similarly, in a real-effort laboratory experiment accompanying their field test, Hennig-Schmidt, Sadrieh, and Rockenbach (2010) found that a 10% wage raise supplemented by surplus information (the value of the task to the principal) for a 19-student treatment yielded an increase in effort by 29% vis-à-vis the 10-student control, who received no raise or surplus information (Table 1, Panel III, row (2)). Students stuffed envelopes for two fifteen-minute sessions, receiving a show-up fee of 1.50 euros and a wage of 2.5 euros per session. All workers received 2.5 euros in session one. In session two, a 19-subject subsample received a 10% raise and surplus information. They raised their output vis-à-vis session one by 12.9 envelopes, whereas those in the control raised it by 10. This 29% magnitude, if statistically significant, could overstate the true elasticity, due to the very small control sample.[7]

We prevent these two issues, arising particularly in very small samples, by having one of the largest samples in field tests. Our CONTROL, 67%RAISE and 50%-100%RAISE start at 47, 70 and 45 workers, respectively. A simple power calibration, fully described in Appendix Section B, shows we had ex-ante 98% power to reject the null hypothesis, at the 5% level, that a 67% wage raise would not increase effort in favor of the one-sided alternative that it do so by 20%.

---

[5]Kube, Maréchal, and Puppe (2012) also show that workers responded to gifts in kind. Though the domain of non-monetary incentives is important, it is outside this review's scope, which focuses on the provision of monetary incentives.

[6]See Button, Ioannidis, Mokrysz, and et al. (2013) on the statistical properties of estimates from small samples.

[7]The statistical significance of this result is not reported.

11

Table 1: Overview of Studies of Gift Exchange in the Workplace

PANEL I: Field Studies (In chronological order)	Task (1)	Design (2)	Sample Sizes (3)	% Wage Increase (4)	% Effort Response (5)	Elasticity (6)
(1) Gneezy and List (2006)	Data entry (6 consecutive hours in one day).	Between subjects	9 student workers in wage-raise ("Gift") treatment; 10 in control ("noGift").	67% ($8 per-hour raise relative to $12 per-hour base).	2% vs. control (whole 6 hours; not significant); 27% vs. control (first 1.5 hours; significant at 5%); 11% vs. control (first 3 hours; not significant); All one-tailed tests.	0.03 (whole 6 hours); 0.40 (first 1.5 hours); 0.16 (first 3 hours).
(2)	Door-to-door fundrasing (3 hours pre-lunch and 3 hours post-lunch).	Between subjects	13 student workers in wage-raise ("Gift") treatment; 10 in control ("noGift").	100% ($10 per-hour raise relative to $10 per-hour base).	38% vs. control (whole 6 hours; significant at 10%); 72% vs. control (first 3 hours; significant at 5%); 6% vs. control (second 3 hours; not significant); All one-tailed tests.	0.38 (whole 6 hours); 0.72 (first 3 hours); 0.06 (second 3 hours).
(3) Bellemare and Shearer (2009)	Tree planting (7 days spread over two weeks).	Within subjects	18 workers of a tree-planting firm.	37 % ($80 one-day raise relative to $215 average daily earnings).	11% -14% (significant at 5%)[1].	0.30-0.38
(4) Hennig-Schmidt, Rockenbach and Sadrieh (2010)	Data entry (2 hours; 1-hour shift per month; wage raise in second hour).	Within and between subjects	25 student workers in wage raise ("F10") treatment; 24 in control ("F0").	10% (2 DM per-hour raise relative to 20 DM per-hour base).	-76% decrease in effort across shifts vs. control (not significant).	-7.6
(5)		Within and between subjects	23 student workers in wage raise ("F40 peer") treatment; 24 in control ("F0").	40% (8 DM per-hour raise relative to 20 DM per-hour base) + Peers' wage information.	-28% decrease in effort across shifts vs. control (not significant).	-0.69
(6) Kube, Maréchal and Puppe (2012)	Data entry (3 consecutive hours in one day).	Between subjects	34 student workers in wage raise ("Money") treatment; 35 in control ("Baseline").	19% (€7 raise relative to €36 pay, from the €12 per-hour base).	5% vs. control (not significant).	0.26
(7) Kube, Maréchal and Puppe (2013)	Data entry (6 consecutive hours in one day).	Between subjects	22 student workers in wage raise ("PayRaise") treatment; 25 in control ("Baseline").	33% (€5 per-hour raise relative to €15 per-hour base announced as a "projected" wage).	-0.3% vs. control (whole 6 hours; not significant); -10%, 1%, 0.2% and 7% vs. control (1st, 2nd, 3rd and 4th 1.5 hours, respectively; none significant).	-0.01 (whole 6 hours); -0.33, 0.03, 0.01 and 0.21 (1st, 2nd, 3rd and 4th 1.5 hours, respectively).
(8) Cohn, Fehr and Goette (2014)	Newspaper distribution (average of 6.5 three-hour shifts in 4 weeks).	Within and between subjects	196 workers of a promotion agency commissioned to distribute newspapers.	23% (5 ChF per-hour raise relative to 22 ChF per-hour base).	3% (full sample; significant at 5%); 14% (reciprocal subjects who felt underpaid by 5 ChF; significant at 5%).	0.13 (full sample); 0.61 (reciprocal subjects who felt underpaid by 5 ChF (upper bound)).
(9) Gilchrist, Luca and Malhotra (2014)	CAPTCHA entry in oDesk (4 hours spread as desired across 7 days).	Between subjects	58 oDesk workers in wage raise ("3+1") treatment; 110 in control ("3").	33% ($1 per-hour raise relative to $3 per hour base).	18% vs. control (full sample; significant at 5%); 26% vs. control frequent workers; significant at 5%); 3% (first-time workers; not significant).	0.55 (full sample); 0.79 frequent workers); 0.09 (first-time workers).

Notes: Most cited published field studies, with the exception of that in row (9), the most recent working paper containing a gift-exchange test. All tests are two-tailed unless otherwise stated. "vs." is an abbreviation for "vis-à-vis". "Not significant" means not significant at either the 5% or 10% levels. [1]Estimate of the percentage increase in effort as the fixed-effects model only identifies level (instead of percentage) changes in effort. "Within and between-subjects" describe tests where workers' effort was observed pre- and post-raise for a given wage sequence in one condition and pre- and post-raise for a different wage sequence in another condition. "Significant at 5%" means significant at the 5% level. "Significant at 10%" means significant at the 10% level. Appendix D documents the sources for this table.

Table 1: Continued

PANEL II: Most Cited Laboratory Studies (In chronological order)	Task (1)	Sample Sizes (2)	% Wage Increase (experimental currency) (3)	% Effort Response (experimental currency) (4)	Elasticity (5)
(1) Fehr, Kirchsteiger and Riedl (1993)	Employers make public wage bids. Upon accepting, workers choose effort.	Four 2-hour sessions, each with 8 to 9 student workers and 5 to 6 student employers.	140%: 72 units average wage over 30 units minimum (equilibrium) wage.	300%: 0.4 average effort over 0.1 minimum (equilibrium) effort.	2.14
(2) Fehr, Kirchsteiger and Riedl (1998)	Buyers make price offers. Upon accepting, sellers choose a quality level.	Approx. two 3-hour sessions, each with 9 to 12 student sellers and 6 to 8 student buyers ("Reciprocity" treatment).	147%: 74 units average wage over 30 units minimum (equilibrium) wage.	250%: Approx. 0.35 average effort over 0.1 minimum (equilibrium) effort.	1.70
(3) Fehr, Kirchler, Weichbold and Gächter (1998)	Employer-Worker random matching. Employer makes wage offer. Upon accepting, worker chooses effort.	Four 2-hour sessions with Austrian soldiers as subjects, each with 10 workers and 10 employers ("Bilateral GE" treatment).	215%: Approx. 63 units average wage over 20 units minimum (equilibrium) wage.	260%: Approx. 0.36 average effort over 0.1 minimum (equilibrium) effort.	1.21
(4)	Employers make public wage bids. Upon accepting, workers choose effort.	Four 2-hour sessions with Austrian soldiers as subjects, each with approximately 9 to 12 workers and 6 to 8 employers ("GE Market" treatment).	195%: Approx. 59 units average wage over 20 units minimum (equilibrium) wage.	290%: Approx. 0.40 average effort over 0.1 minimum (equilibrium) effort.	1.54
(5) Gatcher and Falk (2002)	Employers make public wage bids. Upon accepting, workers choose effort.	Three 2-hour sessions, each with 10 student workers and 10 student employers ("One-Shot" treatment).	248%: Approx. 73 units average wage over 21 units minimum (equilibrium) wage.	310%: 0.41 average effort over 0.1 minimum (equilibrium) effort.	1.25
(6) Brown, Falk and Fehr (2004)	Employers make public or private wage bids. Upon accepting, workers choose effort.	Four 1.5-hour sessions, each with 10 student workers and 7 student employers ("Incomplete Contract Random" treatment).	380%: Approx. 24 units average wage over 5 units minimum (equilibrium) wage.	230%: Approx. 3.3 average effort over 1 minimum (equilibrium) effort.	0.61
PANEL III: Companion Real-Effort Laboratory Experiments	Task	Sample Sizes	% Wage Increase	% Effort Response	Elasticity
(1) Hennig-Schmidt, Rockenbach and Sadrieh (2010)	Folding letters and enveloping them in two 15-minute shifts separated by a 5-minute break; wage raise in second shift.	10 students in wage raise ("L10") treatment; 10 in control ("L0").	10%: €0.25 per 15 minutes over €2.5 per 15 minutes base wage.	-1% decrease in effort across shifts vs. control (not significant).	-0.10
(2)		19 students in wage raise ("L10 surplus") treatment; 10 in control ("L0").	10%: €0.25 per 15 minutes over €2.5 per 15 minutes base wage + Surplus information.	29% increase in effort across shifts vs. control (significance not reported).	2.90

Notes: Only the most cited studies of gift exchange in labor markets were included in Panel II. These are either laboratory tests focusing solely on gift exchange or that contain a gift-exchange treatment. The percentage wage increase in column (3) is the percentage wage increase versus the equilibrium wage when workers have money maximizing preferences (selfish preferences). The percentage effort increase in column (4) is the increase in effort versus the minimum effort (equilibrium effort) when workers have pure money maximizing preferences (selfish preferences). These studies also analyze the wage-effort relationship by regressing effort against a wage offer where the positive coefficient on the wage (signifying that increases in the wage are positively correlated with increases in effort) is always statistically significant at least at the 5% level using a two-tailed test. "Approx." is an abbreviation for "Approximately". Appendix D documents the sources for this table.

Table 2: Summary of Potential Confounds in Studies of Gift Exchange in the Workplace

PANEL I: Field Studies (In chronological order)	Risk of prosocial signaling? (Has a within-subjects design)	%Wage raise large enough to yield a significant effort increase?	Test for effort ceiling?	Potential fatigue? (All work in one or adjacent days)	Potential selection? (Workers hired above the average market wage)	Potential peer effects?	Potential reemployment concerns?	Small sample? (Conditions with fewer than 20 subjects)	Laboratory test assessing heterogeneity in prosocial preferences?
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) Gneezy and List (2006): Data-entry task	No	67% raise led only to a short-term effort increase	No	Yes	Unclear (market wage not declared).	No	No	Yes	No
(2) Gneezy and List (2006): Fundrasing task	No	100% raise led to a 38% effort increase, but bulk (72%) in short term.	No	Inconclusive (extra 3-hour shift next day has a small sample: 9 subjects in treatment and 4 in control).	Unclear (market wage not declared).	No	No	Yes	No
(3) Bellemare and Shearer (2009)	Yes	37% raise led to 11%-14% effort increase.	No	Yes	Unclear (market wage not declared).	No	Yes (on-going relationship for some workers).	Yes	No
(4) Hennig-Schmidt, Rockenbach and Sadrieh (2010): "F10" treatment	Yes	10% raise led to no effort increase.	No	No	Yes	No	No	No	No
(5) Hennig-Schmidt, Rockenbach and Sadrieh (2010): "F40 peer" treatment	Yes	40% raise led to no effort increase.	No	No	Yes	Yes	No	No	No
(6) Kube, Maréchal and Puppe (2012)	No	19% raise led to no effort increase.	Indirect test (gift in kind increased effort)[1]	Yes	Unclear (market wage not declared).	No	No	No	No
(7) Kube, Maréchal and Puppe (2013)	No	33% raise led to no effort increase.	Inconclusive test (subjects recruited using the piece rate, inducing selection).	Yes	Yes ("projected" wage used in recruitment is above market wage).	No	No	No	No
(8) Cohn, Fehr and Goette (2014)	Yes	23% raise led to 3% effort increase.	No	No	No	No	No	No	Yes (Moonlighting game).
(9) Gilchrist, Luca and Malhotra (2014)	No	33% raise led to no effort increase for first-time workers, but a 26% increase for frequent ones.	No	Unclear (4 hours spread across 7 days as desired).	Unclear (subjects hired at their oDesk reservation wage of $2-$3 per hour).	No	Yes (signaling to future employers via publicly viewable reviews).	No	No

Notes: Column (9) shows whether the field study also contains a laboratory test assessing heterogeneity in prosocial preferences by workers in the field experiment. Even though this is not a confound per se, it is an important feature with the goal of finding a relationship between laboratory and field behavior, similarly to that in our test.[1] This paper provides an indirect test of an effort ceiling in that it shows that gifts in kind were able to increase effort beyond the observed statistically insignificant 5% estimated with the wage raise.

Table 2: Continued

PANEL II: Most Cited Laboratory Studies (In chronological order)	Risk of prosocial signaling?	%Wage raise large enough to yield a significant effort increase?	Test for effort ceiling?	Potential fatigue?	Potential selection?	Potential peer effects?	Potential reemployment concerns?	Small sample? (Conditions with fewer than 20 subjects)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) Fehr, Kirchsteiger and Riedl (1993)	Yes	140% raise led to a 300% effort increase.	n/a	No[1]	No[2]	No	No[3]	No
(2) Fehr, Kirchsteiger and Riedl (1998)	Yes	147% raise led to a 250% effort increase.	n/a	No[1]	No[2]	No	No[3]	No
(3) Fehr, Kirchler, Weichbold and Gächter (1998): "Bilateral GE" treatment	Yes	215% raise led to a 260% effort increase.	n/a	No[1]	No[2]	No	No[3]	No
(4) Fehr, Kirchler, Weichbold and Gächter (1998): "GE Market" treatment	Yes	195% raise led to a 300% effort increase.	n/a	No[1]	No[2]	No	No[3]	No
(5) Gatcher and Falk (2002)	Yes	248% raise led to a 310% effort increase.	n/a	No[1]	No[2]	No	No[3]	No
(6) Brown, Falk and Fehr (2004)	Yes	380% raise led to a 230% effort increase.	n/a	No[1]	No[2]	No	No[3]	No

PANEL III: Companion Real-Effort Laboratory Experiments

	Risk of prosocial signaling?	%Wage raise large enough to yield a significant effort increase?	Test for effort ceiling?	Potential fatigue?	Potential selection?	Potential peer effects?	Potential reemployment concerns?	Small sample?
(1) Hennig-Schmidt, Rockenbach and Sadrieh (2010): "L10" treatment	Yes	10% raise led to no effort increase vs. control.	No	Yes	No	No	No	Yes
(2) Hennig-Schmidt, Rockenbach and Sadrieh (2010): "L10 surplus" treatment	Yes	10% raise + Surplus information led to a 29% effort increase vs. control (significance not reported).	No	Yes	No	No	No	Yes

Notes: Only the most cited studies laboratory studies focusing on gift exchange or with a gift exchange treatment pertaining to labor markets were included in Panel II. "n/a" and "vs." stand for "not applicable" and "vis-à-vis", respectively. [1] In the laboratory, effort is defined by the experimenter in an cost-of-effort table, which is fixed across rounds and sessions; therefore, effort in one round does not affect the marginal cost of effort in subsequent rounds. [2] The cost-of-effort table, written by the experimenter, does not vary across subjects. [3] Reemployment concerns are minimized by having subjects play a series of one-shot rounds.

3 Research Design

This section describes our study and elaborates on how it addresses the confounds discussed in Section 2. The first part of the study comprised a field experiment implemented between the fall of 2011 and the fall of 2012 in two legs at university A (an elite national university) and then in five further legs both there and also at university B (a local state university). Recruiting workers from two campuses allowed us not only to garner a larger sample but also to see whether we would observe similar patterns across these two different samples.

The second part was a laboratory test in the winter 2013, after the conclusion of the field experiment. We invited workers to play SPD games and asked their consent to use their responses for research, as is standard in laboratory tests. Because we were required to ask for consent to use workers' field experiment data for research before they participated in the SPD games, we only implemented the games after the last wave of the field experiment had concluded. The workers therefore remained unaware had been partaking in a study while on the job ensuring the findings' external validity.[8] We now further describe the study's two parts.

3.1 The Field Experiment

(1) *Recruitment and task.* We hired 194 students from the two universities to create a bibliographic database for a department at one of the universities. Workers entered data on academic articles (e.g., title, authors, journal, year, volume, issue and pages) using bibliographic software. Campus flyers advertised the job as a one-time sequence of three weekly two-hour shifts. The hiring wage was $12 per hour, the standard wage for data entry.[9] Subjects worked alone and did not know that all characters they entered—including spaces between words, backspaces, etc.—were recorded.

(2) *Between-subjects design and the treatments.* After hiring in each leg, workers at a given campus were randomly assigned to four main conditions (e.g., in leg 3, those at campus A were randomly assigned to each condition and those at campus B were similarly, separately, assigned). The 47 workers in the CONTROL performed the task and received $72 at the end of the third shift.

The 70 in the 67%RAISE condition received a 67% raise to $20 after being hired but before starting the first shift. This condition results from the aggregation of three subtreatments—67%SURPRISERAISE, 67%ANTICIPATEDRAISE and 67%PROMISEDRAISE—varying the timing of the information about the raise (immediately before or one week before the first shift) and when the

---

[8]Asking for consent to use the field experiment data before the implementation of the SPD games was a Human Subjects requirement. No worker forbade the use of any data.

[9]Source for market wages (06/2012): student pay scale (Campus A) and PayScale.com, Salary.com (Campus B).

raise was paid (at the beginning or at the end of each shift). Given that the distribution of outcomes for these three subtreatments was not statistically different (see Results section 4.1.1), we aggregated them into a single condition 67%RAISE.[10]

The 45 in the 50%-100%RAISE condition received a 50% raise to $18 per hour after being hired but before they started their first shift, and a subsample of these workers (23) received an additional surprise wage raise of 100%, to $24 per hour in the third shift.[11]

The 32 in the PIECERATE condition were offered a per-record piece rate for the duration for the contract, instead of a fixed wage raise, before they started their first shift. Importantly, the piece rate was offered *after* hiring at the fixed $12 per hour market wage had occurred, as an add-on to this wage, thus not used to recruit workers. We therefore avoided selection of higher-ability workers into piece-rate contracts (Lazear (2000)), ensuring that workers in this treatment were similar to those in the fixed wage-raise treatments. The piece-rate scheme was piecewise convex, where x is the number of records per shift: $\$0 \times x$ if $x < 70$; $\$0.05 \times x$ if $70 \leq x \leq 110$; $\$0.10 \times x$ if $110 < x \leq 140$; and $\$0.20 \times x$ if $x > 140$. These workers collected their piece-rate earnings at the end of each shift. For an overview of the timing of the treatments across campuses and legs see Figure A.1.

The fixed wage raises, offered after recruitment at the market wage but before the start of the job, were dispensed in a gift envelope embossed with the phrase "A Gift for You" at the start of each weekly shift whereas the contract wage of $72 was paid at the end of the third shift, upon the job's completion. Table 3 shows the total compensation per shift for the different conditions. For an overview of the protocol for the treatments, see Appendix E.

(3) *Ensuring external validity and addressing confounds.* The external validity of the findings is ensured by subjects' being unaware that they were part of a study while executing the task and that software tracked the characters they entered.

---

[10]Specifically, the 67%SURPRISERAISE and 67%ANTICIPATEDRAISE subtreatments tested whether the pleasant surprise of a wage raise increased reciprocal effort. To this end, those in the 67%SURPRISERAISE received the news of the wage raise immediately before starting work on the first shift whereas those in the 67%ANTICIPATEDRAISE received this news one week in advance of the first shift (but after recruiting), so that the wage raise would not be a surprise. In both these treatments we paid the raise at the beginning of each shift to convey that it was not conditional on performance, as we worried that if it were given at the end, workers might erroneously perceive the raise as conditional on performance, thus inflating gift exchange estimates. The 67%PROMISEDRAISE variation tested for this confound: whether wage raises promised at the beginning of the first shift, but paid at the end of the contract—instead of being paid upfront as in the previous two treatments—could be misconstrued as contingent on performance, thus artificially enhancing effort. Our concerns were unfounded: we found that paying the raise upfront at the beginning of each shift or promising it at the beginning of the task and paying it at the end yielded the same outcome, a result also found in other studies (e.g., Kube, Maréchal, and Puppe (2013)).

[11]We ran one additional treatment, in which the remaining subset of the workers in this condition received a wage cut in the third shift, instead of a wage increase, returning to the $12 per hour contract wage (thus receiving a $6 per hour wage cut). We do not report these results as they are outside the scope of this paper, which focuses on effort responses to wage increases over the contracted wage, rather than wage declines to the contracted wage.

17

Treatments	Shift One	Shift Two	Shift Three
CONTROL	$2 \times \$12$	$2 \times \$12$	$2 \times \$12$
67%	$2 \times \$20$	$2 \times \$20$	$2 \times \$20$
50%-100%	$2 \times \$18$	$2 \times \$18$	$2 \times \$24$
PIECERATE	$2 \times \$12$ +Piece Rate	$2 \times \$12$ +Piece Rate	$2 \times \$12$ +Piece Rate

Table 3: Compensation per two-hour weekly shift

(3.1) *Dealing with the confound of prosocial signaling.* The between-subjects design enabled subjects to behave selfishly without fear of being viewed as egoistic. Each worker in the fixed wage raise conditions knew that the principal had not observed how much he/she could produce in the absence of the raise, as each worker's effort was only observed after the raise (and we inferred gift-exchange from comparing outcomes in these conditions, as a group, against those in the CONTROL). Thus, each worker in the wage raise conditions could behave selfishly by not increasing effort versus his/her personal baseline without fear of being detected. An exception was shift three in the 50%-100%RAISE, where a subsample received the additional raise of 100% of the contract wage. These workers knew that the principal had observed how much they could produce with a 50% raise, so they might increase effort after the additional raise not only to reciprocate it but also to signal their nonselfishness.

(3.2) *Dealing with the confound of insufficient wage raises.* We aimed to compensate workers for their marginal cost of effort with 50%, 67%, and 100% raises, matching or exceeding those in prior data-entry field tests with students, the most used task-worker combination in field tests of gift exchange. Still, we worried that data entry was so tiring that even large raises would be insufficient to elicit extra effort—i.e., that our workers faced an effort ceiling.

(3.3) *Dealing with the confound of an effort ceiling.* The PIECERATE condition tested whether workers hired at the $12 per hour contract wage and later offered the per-record piece rate (instead of a fixed raise) increased effort, thus ruling out an effort ceiling. Importantly, the piece rate was a cheaper incentive than the fixed wage raises, requiring only a 21% extra outlay over the hiring wage. If workers raise effort with the piece rate but fail to do so following fixed wage raises, then the absence of gift exchange is likely not due to an effort ceiling.

(3.4) *Dealing with fatigue.* To minimize the role of fatigue in dampening gift-exchange estimates, we split the work into three two-hour shifts, each exactly one week apart. Thus, if in a given shift workers were tired from exerting higher effort to reciprocate the raise, they had one week to recover.

(3.5) *Dealing with selection of higher-productivity workers.* We ensured that gift-exchange elasticities

were not curbed by the selection of higher-ability workers by hiring at the market wage. Hiring at more than the market wage of \$12 per hour could have attracted a higher proportion of these workers, if reservation wages are positively correlated with ability (Weiss (1980), Bewley (1999)).

(3.6) *Dealing with peer effects.* Subjects worked in isolation to avoid peer effects on performance (Falk and Ichino (2006) and Mas and Moretti (2009)) and were not given information about the participation or wages of other workers.

(3.7) *Dealing with reemployment confounds.* Upon the students' recruitment and throughout the six hours, we emphasized that the task was a one-time job.

(3.8) *Dealing with small samples.* We curbed the dual issues posed by small samples—low power to detect gift-exchange elasticities or the risk of overstating them—by having one of the largest between-subjects samples in a gift-exchange study. Our CONTROL group alone started at 47 subjects; the 67%RAISE, at 70; and the 50%-100%RAISE, at 45. In particular, this sample had 98% power to reject the null of no gift exchange in favor of the one-sided alternative that a 67% raise would increase effort, as shown in the simple power calibration in Appendix B. Further, if workers increased effort in any statistically significant way in the PIECERATE, which had fewer subjects, but failed to do so in the wage-raise treatments, then lack of gift exchange is very unlikely due to low power.

Beyond assuring the external validity of the results and addressing prior confounds, the design enhanced ex ante the potential for gift exchange by framing the fixed wage raise as a voluntary kind action by the principal that was not conditional on performance. As described, the raise was offered in an envelope embossed with the phrase "A Gift for You" at the beginning of each shift. Though framing the raise as a gift is not a necessary condition for gift exchange—see Akerlof (1982), Fehr, Kirchsteiger, and Riedl (1993), Gneezy and List (2006), for example—Charness (2004) showed that the perception of a principal's volition boosts reciprocal effort.

Lastly, we avoided differences in research assistants or demand effects from biasing results by having all subjects interact with the same research assistant who was blind to the research hypothesis. Thus differences in treatments' outcomes cannot be due to different assistants or demand effects.

3.2 Further Testing for Prosocial Signaling: Post-Field Experiment Laboratory Games

To test whether each workers prosocial behavior in the laboratory, which could be observed by others (e.g., the experimental team), correlated with that in the field experiment, where individual prosocial actions could not be observed by the principal or any third party due to our between-subjects design, we invited workers to play laboratory games in the semester after the conclusion of the field experiment. The invitation included a consent form to participate and asked subjects

to play three one-shot SPD games as in Burks, Carpenter, and Goette (2009) and Schneider and Weber (2012).[12] Subjects received a \$10 participation fee and any gains made in the games, which could amount to up to \$15. In the first SPD game, subjects chose their action without knowing the opponent's play. In the second and third games, they chose their action after the first mover cooperated and defected, respectively. The stakes were equal in the three games and followed those in the SPD and trust games in Clark and Sefton (2001), and Charness and Rabin (2002), respectively. Each player was randomly and anonymously paired with another from his/her university and this pairing determined the payoffs. Subjects played practice rounds before the actual games to ensure they understood them.[13]

The three SPD games offer a less ambiguous taxonomy of agents' prosocial type, in contrast to other games, such as gift-exchange games, where the action space is continuous (any effort choice in a given range). In our SPD games, each worker's triplet of binary cooperate/defect choices corresponds to one of eight types, of which three are our focus. On one extreme are the *altruists*, who cooperate no matter what (play Cooperate, Cooperate, Cooperate) followed by *conditional cooperators*, who cooperate as first movers, but only cooperate as second movers if the first player cooperates (play Cooperate, Cooperate, Defect). At the opposite extreme are the *selfish*, who defect no matter what (play Defect, Defect, Defect).[14] In contrast, in games with a continuous action space, classification of workers into prosocial types would be more ambiguous as it would necessitate more arbitrary cut-offs, such as defining those in the top decile or quartile of reciprocal effort as prosocial.

The importance of this classification is that if behavior in the games reflects an underlying prosocial preference, then the most prosocial workers in the laboratory, the *altruists* and *conditional cooperators*, should increase effort in response to wage raises in the field experiment, if feasible. Namely, *conditional cooperators*, which comprise most of prosocial types in these games, reciprocate the cooperative behavior of the first mover in the SPD games. Thus, they should reciprocate the cooperative behavior of the first mover in the field experiment—the principal, who increased the

---

[12]These games and other similar ones (e.g., trust games and public good games) have also been used to test whether prosocial behavior in the laboratory correlates with that 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)). For a review of laboratory games assessing prosocial behavior see Levitt and List (2007).

[13]Our SPD games use the strategy method (where second movers make conditional decisions) instead of the direct-response method (where the second player makes a unique choice after observing the first-player move). This method lets us collect a broader set of players' choices without sacrificing their validity because treatment effects found with the strategy method are invariably observed with the direct-response method (Brandts and Charness (2011)). For details of the protocol, stakes of the game, pairing between subjects, etc., see Appendix E.2.

[14]This terminology is adapted from Schneider and Weber (2012), who call their conditional cooperators "optimistic conditional cooperators" and their selfish players "pessimistic selfish."

wage—by increasing effort. Further, the wage raise should also lift effort among the remainder of prosocial workers, the *altruists*, as they always cooperate as second movers. However, if reciprocal behavior in the games reflects, to a large extent, the desire to be perceived favorably by others as prosocial instead of unfavorably as selfish, then there should be little or no correlation between the behavior in the laboratory, where prosocial or selfish actions by each individual worker can be observed, and in the between-subjects field test, where they cannot.

Importantly, because of their natural correspondence to our field experiment and to gift-exchange games, SPD games enabled us to test the correlation between laboratory and field behavior without alerting subjects to our purpose, which could bias their responses. Gift-exchange games are SPD games in which the principal cooperates/does not cooperate by offering/not offering above-market wages, and having observed this, workers cooperate/defect by increasing/not increasing effort beyond their minimum. Our field experiment is setup like a gift-exchange game and thus like a SPD game: the principal cooperates by offering a raise above the \$12-market wage and workers cooperate/defect by increasing/not increasing effort over the baseline: the effort observed in the CONTROL. We thus inferred our workers' prosocial behavior without implementing a gift-exchange game, which given its similarity to our field experiment, could signaled our intent.

4 Study Results

This section discusses the results of the field experiment and the post-experiment laboratory SPD games. After dealing with all the confounds that could have biased previous gift-exchange elasticities, we find no evidence of gift exchange: large fixed wage raises of 50% and 67%, and a further raise of 100%, elicited statistically insignificant negative or small positive effort responses, yielding elasticities of -0.08 to 0.06, in the most favorable specification to gift exchange. The additional piece-rate payment, though entailing a maximal expenditure of only 30%, increased effort did by up to 18%-19%, yielding an substantially larger elasticity of up to 0.63. Moreover, though a substantial portion of workers behaved prosocially in the laboratory, they did not do so in the field: they did not distinctly increase effort in response to the wage raises.

4.1 Field Experiment Results

4.1.1 Sample and Effort Measure

Sample. As described, we recruited 194 participants at two universities, A and B. Slightly more than half (57%) came from A where we received clearance to run the study earlier. Workers within each university and leg were randomly assigned, without their knowledge to each condition. We started

with forty-seven students in the CONTROL, 70 in the 67%RAISE, 45 in the 50%-100%RAISE and 32 in the PIECERATE. As described, 67%RAISE pooled three subtreatments giving workers a 67% raise—67%SURPRISERAISE, 67%ANTICIPATEDRAISE and 67%PROMISEDRAISE—given that their performance was not statistically different within each campus and leg (see Appendix Table A.1).[15] Of the 45 workers in the 50%-100%RAISE, a random subsample of 23 received a further surprise raise to 100% of the contract wage in the third shift.

Effort measure. We used the number of characters inputted per subject, instead of the number of records, as the former more closely approximates effort: some records may have longer titles and more coauthors, and so require more characters, for example. Further, number of characters is the measure used in other studies (e.g., Kube, Maréchal, and Puppe (2012)). Nonetheless, Appendix Table A.3 shows that the final set of results using number of characters as an outcome—those in section 4.1.3—are similar to those using number of records (a noisier measure of effort) or number of correct words inputted (a measure of quality).

4.1.2 Descriptive Statistics

Disaggregated overall statistics. Table 4 documents that workers in the 67%RAISE and 50%-100%RAISE exerted lower effort than those in the CONTROL. However, the those in the PIECERATE increased effort. Column (1) shows that workers in the CONTROL entered an average of 17,591 characters over 131 worker-shifts. Column (2) documents, however, that those in the 67%RAISE and 50%-100%RAISE entered -1% and -2% fewer characters than those in the CONTROL over their respective 207 and 111 worker-shifts. The finding that wage raises induce slight effort decreases is consistent with that of previous tests of gift exchange (e.g., the decrease in effort by -10% and -0.03% in the first 1.5 hours and the whole 6 hours, respectively, in Kube, Maréchal, and Puppe (2013) described in Section 2). In contrast, column (2) shows that those in the PIECERATE inputted 15% more characters than those in the CONTROL over their 90 worker-shifts. This effort increase results from only an average incremental 21% compensation (an average per-worker piece-rate expenditure of $5.1 per two-hour shift in addition to the base wage of $24).

Disaggregated statistics per campus. The finding that the piece rate elicited more effort but the

---

[15]The distributions are statistically indistinguishable between the three treatments within each campus, leg and shift, with only a minor exception: when including two campus B outliers in the 67%ANTICIPATEDRAISE, whose effort was twice the average effort across these three treatments in campus B. Excluding these two outliers from the Kruskall-Wallis test renders the distributions of the three treatments at campus B statistically indistinguishable. Including them, in contrast, leads to the rejection that these three distributions are the same for campus B in shifts one (at the 10% level) and three (at the 4% level). Yet we include the two outliers in the analysis because they increase the average effort of the 67%RAISE relative to the CONTROL, thus biasing our estimates in favor of gift exchange. Despite this, we observed that the 67%RAISE yielded no increases, as we document later.

fixed wage raises did not holds not only overall but also across campuses. Table 5, Panel A, column (1), shows that the average number of characters inputted by the CONTROL at campus A, an elite university, was 21,382. Column (2) shows, however, that fixed wage raises did not increase effort, which was lower by -5% and -9% than in the CONTROL in the 67%RAISE and 50%-100%RAISE, respectively. However, the piece rate increased effort by 25%. This differential response to wage raises and the piece rate is similar for campus B, a local university. As shown in Panel B, column (1), students in the CONTROL inputted 13,240 characters, displaying lower baseline productivity than those at campus A. Though it is unsurprising that the baseline performance for students at the state school was lower than that of those at the elite school, all responded in the same differential way to fixed raises and the piece rate. As column (2) shows, the 67% and 50%-100% raises had a mild and mixed impact on effort (+2% and -2% over the CONTROL, respectively) whereas the piece rate raised effort by 10%.[16]

Disaggregated statistics per shift. The piece rate not only elicited more effort than fixed wage raises across campuses, but also within shifts. Figure 1 shows that the piece rate increased the number of characters inputted vis-à-vis the CONTROL in every shift, despite the much smaller average additional expenditure per worker-shift of 10%, 24% and 30% in shifts one, two and three, respectively (that is, an average per worker piece-rate expenditure of $2.4, $5.8 and $7.3 in shifts one, two and three, respectively, in addition to the $24 per-shift base pay). In contrast, fixed wage raises of 50%, 67% and 100% over the $24 per-shift base pay, in general, did not.

Although the pattern of no response to fixed wage raises, in contrast to piece rates, holds across campus and shifts, it stems from raw average differences across conditions. Given that we randomly assigned workers to the several conditions within campus and leg, we need to control for unobserved time-invariant determinants of performance at each campus and during each leg. Further, it is also important to analyze how effort progresses over time: whether effort increases from one shift to the other as workers rest and have an opportunity to reflect on how to improve their productivity for the following shift, and whether the additional fixed raise in the third shift to 100% in the 50%-100%RAISE elicited extra effort. We thus proceed by analyzing effort across shifts, controlling for the above-mentioned factors as well as for time-invariant shift heterogeneity, which could affect the per-shift difference between the treatments and the CONTROL.

---

[16]See Appendix Table A.2 for additional and more detailed summary statistics for the whole sample and per campus.

4.1.3 Empirical Method and Results

This section documents that after controlling for unobserved campus, leg and shift time-invariant heterogeneity, we still observe that fixed wage raises did not increase effort over time whereas the piece rate did by up to 18%-19%.

Empirical method. To compare our elasticities to those in previous research, we used the natural log of characters as our dependent variable. Also, because within each leg, and within each campus, we randomly assigned workers to each condition, the correct specification estimates differences between the conditions within each leg, campus and shift and pools them. Thus, we estimate the natural log of the characters entered by a subject i, in t_1 (Control), t_2 (67%Raise), t_3 (50%-100%Raise), t_4 (PieceRate) in campus c, leg l, and shift s as follows:

$$\ln(\text{characters})_{i,t,s,c,l} = \alpha_{1,1} + \alpha_{1,2}t_1s_2 + \alpha_{1,3}t_1s_3 + \sum_{\tau=2}^{4}\sum_{j=1}^{3}\beta_{\tau,j}t_\tau s_j + \psi_c \times \psi_l \times \psi_s + \epsilon_{i,s,t,c,l} \qquad (1)$$

The interaction of campus, leg and shift fixed effects ($\psi_c \times \psi_l \times \psi_s$) captures unobservable time-invariant, campus, leg and shift determinants of outcomes. Campus fixed effects control, for example, for unobserved different propensities to perform, or respond to incentives, between campuses, which could affect the difference between the treatments and the Control within each campus. For example, students at one campus may be more productive (as is the case with campus A), may have more taste for reciprocity or more ability to increase effort than students at the other. Leg fixed effects control, for example, for unobserved different conditions across legs (e.g., students may be more tired in a leg occurring closer to final exams than in another occurring early in the term), which could influence both their baseline performance and how they respond to incentives within each leg. Shift fixed effects control, for example, for unobserved learning across shifts.

The interaction of campus, leg, and shift conservatively addresses whether these unobservables could occur differentially within each leg, campus, and shift. For example, the tiredness of students working during a leg close to final exams at a given campus could undermine both effort at a shift and learning across shifts. Thus they could respond less to incentives per shift, within that campus and leg. This fixed effects interaction further allows us to estimate differences between the treatments and the Control, within campus, leg, and shift in line with the previously described random assignment within campus and then to pool these differences.

The causal parameters of interest are the $\beta_{t,s}$ on the interaction of the treatment and shift dummy variables. They pool the percentage differences in characters between the treatments and

24

the CONTROL within a campus, leg and shift. For example, $\beta_{2,1}$ identifies the percentage difference in characters between t_2 (67%RAISE) and the CONTROL for shift one, by pooling all these differences within each campus and leg. The parameter $\alpha_{1,1}$ is the natural log of characters for the base category, the CONTROL in shift one, which is not separately identified from the fixed effects, as usual.

To account for serial correlation in worker effort across shifts (serial correlation in $\epsilon_{i,s,t,c,l}$), we cluster the standard errors at the worker level (Bertrand, Duflo, and Mullainathan (2004)).[17]

Further, all tests comparing the treatments to the CONTROL are one-tailed following, for example, Gneezy and List (2006), the most influential gift exchange field study. One-tailed tests are used because gift exchange makes the one-sided prediction of an increase in effort in response to raises. Further, the sample size was partially determined by our power calculations (Appendix B), which yielded a reasonable minimum size needed to reject the null in a one-tailed test. We also use one-tailed tests for the PIECERATE, so as to apply the same standard of statistical significance applied for the wage-raise treatments. Further, reasonable piece rates have been shown to not decrease effort.[18]

We now describe the results from specification 1, comparing effort between the treatments and the CONTROL per shift, controlling for campus, leg, and shift unobservables. We build up to these results by first documenting estimates unadjusted for fixed effects.

Unadjusted estimates per shift. Columns (1)-(3) in Table 6 show that, in keeping with previous results, fixed wage raises did not increase effort. The evidence suggests that piece rates, however, not only increased effort, but accelerated it over time. Column (1), shows that, in the first shift, the unadjusted percentage difference in effort between the wage-raise treatments and the CONTROL was small at 2%, whereas it was 8% for the PIECERATE. Column (2) shows that, in the second shift, the difference between the wage-raise treatments and the CONTROL became negative at -3%, whereas it increased to a conservative 12% in the PIECERATE. Column (3) shows that, in the third shift, the difference between 67%RAISE and the CONTROL was still negative at -1%, whereas there was no difference between the 50%-100%RAISE and the CONTROL. However, the difference between the PIECERATE and the CONTROL enlarged to a conservative 15%, hinting that the piece rate accelerated learning, whereas fixed wage raises did not.[19],[20] These estimates are statistically insignificant, however, due to

---

[17]Clustering at the worker level is more conservative than clustering at the campus, or the leg, or at another broader level of clustering.

[18]Gneezy and Rustichini (2000) found that very small piece rates depressed effort whereas reasonable piece rates increased effort, as expected.

[19]This is consistent with anecdotal evidence volunteered by subjects in the PIECERATE treatment to the research assistant: They spent their time between the weekly shifts mulling over how to increase performance to earn higher earnings on the following shift.

[20]The stated percentage increases for the PIECERATE of 12% and 15% in shifts two and three, respectively, are actually conservative lower bounds for the estimated percentage increases for these shifts, which are $e^{(0.12)} - 1 = 0.13$

the large standard errors induced by the model's low explanatory power (R^2 of 0.02), which did not include unobserved campus differences in productivity, for example (workers in campus campus A tend to have productivity than those in campus B). We add these controls as they not only reduce the standard errors but also reduce the bias of the estimates.

Estimates per shift adjusted for campus fixed effects. Campus fixed effects takes into account unobserved differences in workers' baseline productivity across the two campuses. Including campus fixed effects is critical given the previous evidence that workers in campus A tend to be more productive than those in campus B. Campus fixed effects thus ensure that differences between the treatments and the CONTROL are estimated within campus (treatments vis-à-vis CONTROL in campus A and treatments vis-à-vis CONTROL in campus B) and pooled. Relying solely on the previous raw means to draw conclusions would be incorrect, as these raw means simply average effort across all workers irrespective of their campus membership and their distribution across the different experimental conditions. Thus, we would be, for example, comparing the performance of the lower productivity workers in campus B in the treatments to the performance of the higher productivity workers in campus A in the CONTROL.

Columns (4)-(6) show that adding campus fixed effects both reduces the bias of the estimates and increases the fit of the model considerably (R^2 of 0.35) as, as expected, campus membership explains a great deal of worker performance. The better fit of the model reduces the standard errors substantially for most estimates, which become more precise. Nonetheless, the general pattern they convey remains similar to that described above.

Specifically, column (4) documents that, in the first shift, workers' effort in the 67%RAISE and 50-100%RAISE was 1% and -4% lower than that in the CONTROL, respectively, though these estimates are not statistically significant. In contrast, effort in the PIECERATE was 11% higher than that of the CONTROL in the first shift and statistically significant at the 10% level. Column (5) documents that, in the second shift, those in the 67%RAISE and 50%-100%RAISE undersupplied effort by -6% and -9% compared to those in the CONTROL, respectively, though these magnitudes are, again, not statistically significant. In contrast, those in the PIECERATE further increased effort vis-à-vis the CONTROL by 14%, which is statistically significant at the 5% level. Finally, column (6) shows that, in the third shift, those in the 67%RAISE and 50-100%RAISE stabilized at slightly less effort than

---

and $e^{(0.15)} - 1 = 0.16$, respectively. This is because though the natural log specification approximates well small percentage changes (thus the estimates for the 67%RAISE and 50%-100%RAISE correspond to actual estimated percentage changes, such as $e^{(0.02)} - 1 = 0.02$ and $e^{(-0.03)} - 1 = -0.03$), it underestimates percentage increases when these are large.

those the CONTROL at -3% and -6%, respectively, though these estimates are, again, not statistically significant. In contrast, those in the PIECERATE further increased their effort versus the CONTROL by 19%, which is significant at the 1% level.

Nonetheless it could be that, beyond time-invariant campus unobservables, unobserved differences across legs and shifts could bias results. Thus we add the full set of controls for time-invariant campus, leg and shift unobservable heterogeneity, as laid-out in specification 1.

Estimates per shift adjusted for campus, leg and shift fixed effects. Including interactions of unobserved time-invariant campus, leg and shift heterogeneity increases the model fit slightly (the R^2 increases to 0.38). This reduction in the error variance is not enough to offset the loss of power induced by the large increase in the number of fixed effects (from 2—one for each campus—to close to 40). This decreases the precision of some estimates—as is the case with the PIECERATE—as their standard errors increase by at least 12%, yet it favorably raises gift-exchange estimates, which reach a high of 4% in one shift. We now describe these magnitudes and discuss their implications.

Column (7) documents that, in the first two-hour shift, workers in the 67%RAISE and 50%-100%Raise increased effort by 4% (the best estimate for gift exchange) and 2% compared to those in the CONTROL, respectively, though these estimates are statistically insignificant. They result nonetheless in very small and statistically insignificant elasticities for the first two hours of 0.06 and 0.04, respectively, which are within the range of those in the literature, in particular for data-entry tasks employing students, which include statistically insignificant elasticities of -0.69 for the first one-hour shift in Hennig-Schmidt, Sadrieh, and Rockenbach (2010) and of -0.33 for the first 1.5 hours in Kube, Maréchal, and Puppe (2013) as well as the significant elasticity of 0.40 in the first 1.5 hours in Gneezy and List (2006), all discussed in Section 2. Those in the PIECERATE increased effort by 6% vis-à-vis the CONTROL with an additional expenditure of 10% (an average per-worker piece-rate payment of $2.4 in shift one in addition to the $24 base pay). This point estimate of 6%, yielding an elasticity of 0.60, is larger than that for the wage-raise treatments, though not statistically significant.

Column (8) documents that, in the second shift, however, workers in the 67%RAISE and 50-100%RAISE slowed down, exerting -2% and -4% less effort, respectively, than workers in the CONTROL, though these magnitudes are not statistically significant. These negative, though statistically insignificant, elasticities of -0.03 and -0.08, respectively, fall within those in the prior literature noted above and in Section 2. In contrast, workers in the PIECERATE further increased effort relative to the CONTROL, to 10%, though this estimate is not statistically significant due to the larger standard error of 0.10—which is more than 25% higher than that for the wage-raise treatments—induced by

the smaller sample and the larger number of fixed effects. Nonetheless, this 10% point estimate of effort corresponded to an increase in 24% in compensation (an average per-worker piece-rate payment of $5.8 in shift two in addition to the $24 base pay), much smaller than the fixed wage raises, resulting in a pay-effort elasticity of 0.41.

Column (9) shows that, in the third shift, those in the 67%RAISE and 50%-100%RAISE continued exerting less effort than those in the CONTROL, inputting -4% and -3% fewer characters, though these magnitudes are statistically insignificant. As before, these negative but statistically insignificant elasticities of -0.06 and -0.03, respectively, dovetail with prior research as discussed above and in Section 2. However, workers in the PIECERATE further boosted effort relative to the CONTROL to 18%, which is statistically significant at the 10% level (instead of at the 1% level, as in the 19% estimate with only campus fixed effects) despite the larger standard errors induced by the numerous campus, leg and shift fixed effects. This further increase in effort to 18% corresponded in a 30% increase in compensation (an average per-worker piece-rate payment of $7.3 in shift three in addition to the $24 base pay), still substantially smaller than the fixed wage raises, yielding a pay-effort elasticity of 0.60 (and of 0.63 for the 19% increase in effort in the previous specification). We note also that these 18% and 19% estimates are conservative lower bounds on the percentage increases: the estimates of 0.18 and 0.19 correspond to 21% and 20% effort increases, respectively, as discussed in footnote 20.

4.1.4 Discussion of the Field Experiment Results

Could the absence of gift exchange be due to an effort ceiling? Our results suggest not: after hiring workers at the $12 wage, those subsequently offered the piece rate increased effort versus the CONTROL but those offered fixed raises did not. Our result thus complements that of the only study trying to directly assess whether an effort ceiling could have curbed gift exchange (Kube, Maréchal, and Puppe (2013)).

Could the absence of gift exchange be due to the fixed wage raises being too small to elicit effort? An open question arising from prior tests was that workers' failure to reciprocate a wage raise could have been due to its being too small to elicit effort, in particular in data-entry tasks similar to ours (e.g., raises of 19%, 33% and 40% in Kube, Maréchal, and Puppe (2012), Kube, Maréchal, and Puppe (2013) and Hennig-Schmidt, Sadrieh, and Rockenbach (2010), respectively, elicited negative to small positive statistically insignificant effort). Our evidence suggests that the absence of gift exchange is not due to small wage raises. We find that the wage raise conditions offering 50%, 67% and 100% raises beyond the contract wage did not elicit effort. In contrast, the PIECERATE condition elicited

extra effort, which grew to 18%-19% in the third shift, despite only entailing a maximal expenditure of 30% over the per-shift base wage of $24.

Could the absence of gift exchange be due to low power? Our findings also suggest not. Our simple power calibrations show that our large sample of 47 workers in the CONTROL and 70 in the 67%RAISE condition had 98% power to reject, at the 5% level, the null hypothesis that a 67% wage raise does not increase effort in favor of the one-sided alternative that it does by 20% per two-hour shift. This calibration was conservatively based on the widely cited study by Gneezy and List (2006), which had a similar wage raise, task and worker sample as ours and showed that workers increased effort by 27% following a 67% raise during the first 1.5 of the task, statistically significant at the 5% level in a one-tailed test. Though we achieve smaller standard errors than this prior research due to our larger sample size, our point estimates are much smaller than 20%, ranging from -4% to 4%, (Table 6, columns (7)-(9)). We are thus unable to document gift exchange, even in the short term.

Nonetheless, because we worried that the short-term effect size of 27% in Gneezy and List (2006), on which we based our power calculations, could be potentially inflated due to the risk of overstatement in small samples (see section 2), the PIECERATE condition helped us assess the power of our gift exchange test. The PIECERATE detected statistically significant effort increases despite its having a smaller sample than the 67%RAISE and 50%-100%RAISE conditions, whereas the wage-raise treatments did not detect any statistically significant effort increases across any of the specifications. Thus the absence of gift exchange is unlikely due to insufficient sample sizes.

Could the absence of gift exchange be due to no opportunity to learn? Our findings suggest not. Though workers in the 67%RAISE and 50%-100%RAISE increased their effort slightly by a non-significant 4% and 2% in the first shift relative to the CONTROL, respectively, they slowed down over time, reducing effort versus the CONTROL by -2% to -4% in subsequent shifts, though these magnitudes are also not statistically significant. The finding of the negative but statistically insignificant magnitudes is consistent with similar prior findings in the literature, as discussed previously. The piece rate, in contrast, seems to have motivated workers to learn to improve productivity from shift to shift to gain higher piece rate earnings, resulting in a statistically significant effort increase of 18%-19% by the third shift.

Relatedly, the piece rate seems to have also motivated workers to persist on the task, though this effect is only marginally significant at the 10% level. Of the starting 194 workers, 15 did not complete all three shifts, of which most (7) were in the CONTROL. The remaining 8 were split as follows: 2 in the 67%RAISE, 2 in the 50%-100%RAISE and 4 in the PIECERATE. As shown in Appendix

Section C, a regression analysis of sample attrition, there was no statistically differential attrition between workers in the wage-raise conditions and those in the CONTROL within each campus and leg. In contrast, 11% fewer workers in the PIECERATE missed shifts two and three vis-à-vis the CONTROL, within each campus and leg, though this effect is only marginally significant. This persistence accords with anecdotal evidence that subjects viewed the kinked piece-rate scheme as a challenge, returning each week with the goal of surpassing prior performance and reaching the next kink in earnings.

Did workers engaged in gift exchange by increasing output quality? Our evidence suggests not. Table A.3, columns (7)-(9), shows that the number of correct words inputted in the wage raise conditions was, in general, slightly *lower* than that in the CONTROL, though not statistically significant. In contrast, the number of correct words inputted by subjects in the PIECERATE was higher than that in the CONTROL, but also statistically insignificant. Thus, those in the PIECERATE also did not produce lower-quality output, consistent with prior research (Bandiera, Barankay, and Rasul (2005)).

We now discuss whether workers who behaved prosocially in the laboratory games, where their individual actions could be observed, behaved similarly in the field, where they could not. Though our field experiment suggests that prosocial signaling may have arrested the decline in effort in the third shift for the 50%-100%RAISE—the only case in which an individual worker's effort was observed pre- and post-raise—the evidence is inconclusive as the estimates are statistically insignificant.[21]

4.2 Correlation Between Behavior in the Laboratory and the Field Experiment

This section documents that workers who behaved more prosocially in SPD games, did not particularly behave more prosocially during the field experiment. The best gift exchange estimates for this select subsample of prosocial workers show increases in effort of 7%, -3% and -1% in shifts one, two and three, respectively, which are not substantially different from those in the overall sample.

4.2.1 Main Results

SPD games participants and classification into prosocial types. Table 7, column (2) shows that a substantial portion of workers (72% of the 194 workers, or 139) participated in the SPD games after the conclusion of all the field experiment's waves. This high participation rate, did not, in general, vary substantially across conditions: 71% of the workers from the wage-raise treatments participated,

---

[21]Specifically, Table 6, column (9) shows that those in the 50%-100%RAISE increased effort in the third shift by 1% (to -3% vis-à-vis the CONTROL) after having declined by 6% (from +2% to -4% vis-à-vis the CONTROL), in shifts one and two, as shown in columns (7) and (8), respectively. In contrast, those in the 67%RAISE, where their baseline effort was unknown, and thus each could refrain from raising effort without fear of being viewed as selfish, showed a continuing decline in effort versus the CONTROL: from 4%, to -2% and to -4% in shifts one, two and three, respectively, as shown in columns (7)-(9). These estimates are, however, statistically insignificant and thus insufficient as a basis for conclusions.

and 72% and 74% from the PIECERATE and CONTROL, respectively (column (3)).[22]

Table 7, row 1, columns (7) and (9), show that the two opposite extremes in prosociality concentrate the majority of respondents at 73%: 41% behaved as *Prosocial* and 32% as *Selfish*. The *Prosocial* type adds the two types who behaved the most prosocially in these games: the 10 *altruists*, who always played cooperate (CCC), and the 47 *conditional cooperators*, who always cooperated except when the first mover defected (CCD), as documented in columns (4) and (5), respectively. At the opposite extreme were the 45 *Selfish* players who always defected (DDD) (row (1), column (8)). This distribution of types lies within that of other laboratory experiments. For example, a review by Fehr and Fischbacher (2002, page C6) pointed out that 40% or more of responders in gift-exchange games, which are SPD games, exhibited prosocial behavior. And, for example, Cohn, Fehr, and Goette (forthcoming) classified 35% of workers as selfish in the post-experiment game described in Section 2.

We now analyze whether the subsample of workers who behaved most prosocially in the games and were in the wage-raise conditions during the field experiment engaged in gift exchange by exerting higher effort than the CONTROL.

Sample for analysis. Table 7, row (2), documents that of the 82 workers in the 67%RAISE and 50-100%RAISE participating in the SPD games, 34 (41%) behaved as *Prosocial*. We thus test whether this subset of 34 workers—who received the fixed wage raises, participated in the SPD games, and behaved most prosocially in them—exerted more effort than those in the CONTROL. The idea is that the average effort increase in the wage-raise treatments vis-à-vis the CONTROL for this subset of workers should be higher than that previously documented for all workers in the wage-raise treatments in Table 6, columns (7)-(9), because the full sample in the wage-raise conditions contained both the most prosocial types as well as the remaining types with weaker prosocial preferences (e.g., as selfish), which may have diluted average effort responses to raises.

Empirical method and results. We use specification 1 but restrict the workers in the wage-raise and piece-rate treatments to be only the *Prosocial* ones. Also, we aggregate the *Prosocial* workers in the two wage-raise treatments into one group—the 34 discussed above—so as to have a large enough sample size from which to draw conclusions. Thus, instead of comparing the effort of all workers in the treatments to those in the CONTROL as in section 4.1.3, we compare only the effort of the most prosocial workers in the treatments to that in the CONTROL. As before, we control for time-invariant unobserved campus, leg and shift heterogeneity.

---

[22]For a further breakdown of participation by wage raise treatment, see Appendix Table A.4.

Table 8, columns (4)-(6), show that workers who behaved the most prosocially did not particularly increase effort in response to the wage raise during the field experiment. Column (4) shows effort higher than that in the CONTROL of 7% in shift one, though this estimate is statistically insignificant. Further, columns (5) and (6) show a subsequent slow down to -3% and -1% less effort than the CONTROL in shifts two and three, respectively, though these estimates are also statistically insignificant. These estimates' statistical insignificance, despite the large wage raises, and their similarity to those using the whole sample of workers in the wage-raise treatments, documented in columns (1)-(3), imply that this subset of workers were as non-responsive to wage raises as the whole sample.[23]

Our results suggest that prosocial behavior in the laboratory, which could be observed by others, did not translate into prosocial behavior in the field experiment—an increase in effort after a wage raise—where the between-subjects design enabled selfish behavior without detection.

4.2.2 Robustness Checks on the Correlation Between Laboratory and Field Results

Did Prosocial workers face an effort ceiling? It could be that *Prosocial* workers in the games were those who faced an effort ceiling during the field experiment and this is why they failed to increase effort following the raise. We tested this hypothesis by assessing whether *Prosocial* workers in the PIECERATE were able to increase effort. Table 7, row (3), shows that of the 23 workers in the PIECERATE who participated in the SPD games (a 72% response rate), 11 (48%) behaved as *Prosocial*. Table 8, columns (4)-(6) show that these workers increased effort substantially, by 15%-24% in comparison to the CONTROL, an effect statistically significant at the 10% level, though this sample is too small to allow conclusions. This pattern suggests, nonetheless, that an effort ceiling did not prevent *Prosocial* workers receiving the fixed wage raises from increasing effort.

Could the lack of effort response by Prosocial workers be due to lower-effort workers selecting disproportionately into the SPD games? It could also be that participants in the SPD games tended to be workers who exerted low effort during the field experiment. This could have caused the finding of no effort increases by *Prosocial* workers in the wage-raise treatments vis-à-vis the CONTROL, because average effort in the CONTROL could have been inflated by including both game participants and

---

[23]These results do not change substantially if we exclude the two *altruists* from the 47 workers in the CONTROL. These two workers could be exerting higher effort even under no wage raise given that they cooperate independently of the action of the first mover (receiving a raise or not). They could thus inflate the effort in the CONTROL and bias downwards the estimates of gift exchange in the wage-raise treatments and the PIECERATE. Excluding them actually results in similar though slightly lower estimates to wage raises (e.g., a decline of about 1% in all shifts) suggesting these two *altruists* exerted lower effort than the average worker in the CONTROL. Thus the already low estimates above are nonetheless biased in favor of gift exchange.

non-participants. Restricting the CONTROL to only the 34 workers who played the games, instead of the original 47, when estimating specification 1 allows for effort responses by *Prosocial* workers in the fixed wage-raise treatments to be higher than that of a potentially lower-effort CONTROL. Table 8, columns (7)-(9) document, however, that the estimates are slightly smaller than those obtained previously, with effort in the wage-raise treatments ending at a statistically insignificant -3% relative to the CONTROL in the third shift. Thus the non-increase in effort by *Prosocial* workers in the wage raise conditions relative to those in the CONTROL holds conditional of participation in the games. For a detailed sample decomposition per shift and condition for the estimation in Table 8, see Appendix Table A.5.

5 Conclusion

We identify the main factors that could have led to the conflicting evidence on gift exchange tests—prosocial signaling, insufficient wage raises, the existence of an effort ceiling, fatigue, selection of high-productivity workers, peer effects, reemployment concerns, and small samples—and implement a comprehensive test aiming to uncover more accurate elasticities by dealing with these confounds.

We found that, after addressing all these confounds, workers did not respond to fixed wage raises, though they responded to a cheaper piece rate. Our post-field experiment Sequential Prisoner's Dilemma games showed that even workers who behaved most prosocially in the laboratory games did not behave particularly prosocially during the field experiment.

Our findings suggest that merely raising wages—even if this raise is framed as a voluntary gift by the principal—is not sufficient to trigger reciprocal effort in the workplace. Perhaps gift exchange requires sustained long-term interactions between agents and principals in order for workers to develop the "sentiment" towards the firm described by Akerlof (1982). Or perhaps firms screen for reciprocal workers, anticipating that only these types of workers will raise effort in response to wage raises. These and other conditions for gift exchange are explored in Esteves-Sorenson, Pohl, and Freitas (2013), which analyzes worker performance over several years in two firms paying above-market fixed wages and using years-long probation periods to screen workers.

References

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.

BANDIERA, O., I. BARANKAY, AND I. RASUL (2005): "Social Preferences and the Response to Incentives: Evidence From Personnel Data," *Quarterly Journal of Economics*, 120(3), 917–962.

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.*

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. F. (1999): *Why Wages Don't Fall During a Recession.* Harvard University Press.

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

BROWN, M., A. FALK, AND E. FEHR (2004): "Relational Contracts and the Nature of Market Interactions," *Econometrica*, 72(3), 747–780.

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.

BUTTON, K. S., J. P. A. IOANNIDIS, C. MOKRYSZ, AND ET AL. (2013): "Power failure: why small sample size undermines the reliability of neuroscience," *Nature Reviews Neuroscience*, 14(5), 365–376.

CARD, D., S. DELLAVIGNA, AND U. MALMENDIER (2011): "The Role of Theory in Field Experiments," *Journal of Economic Perspectives*, 25(3), 39–62.

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., 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.

COHEN, J. (1988): *Statistical Power Analysis for The Behavioral Sciences.* Psychology Press, New York, NY, second edition edn.

COHN, A., E. FEHR, AND L. GOETTE (forthcoming): "Fair Wages and Effort: Evidence from a Field Experiment," *Management Science.*

COHN, A., E. FEHR, B. HERRMANN, AND F. SCHNEIDER (2014): "Social Comparison and Effort Provision: Evidence from a Field Experiment," *Journal of the European Economic Association*, 12(4), 877–898.

ESTEVES-SORENSON, C., V. POHL, AND E. FREITAS (2013): "Efficiency Wages and Its Mechanisms: Empirical Evidence," *Working Paper*.

FALK, A., AND A. ICHINO (2006): "Clean Evidence on Peer Effects," *Journal of Labor Economics*, 24(1), 38–57.

FEHR, E., AND U. FISCHBACHER (2002): "Why Social Preferences Matter-The Impact of Non-Selfish Motives on Competition, Cooperation and Incentives," *Economic Journal*, pp. 1–33.

FEHR, E., AND U. FISCHBACHER (2004): "Third-Party Punishment And Social Norms," *Evolution and Human Behavior*, 25(2), 63–87.

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., G. KIRCHSTEIGER, AND A. RIEDL (1998): "Gift Exchange and Reciprocity in Competitive Experimental Markets," *European Economic Review*, 42(1), 1–34.

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

GIBBONS, R. (2005): "Incentives Between Firms (And Within)," *Management Science*, 51(1), 2 – 17.

GILCHRIST, D., M. LUCA, AND D. MALHOTRA (2014): "When 3+1>4: Gift Structure and Reciprocity in the Field," *Harvard Business School Working Paper*.

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.

GNEEZY, U., AND A. RUSTICHINI (2000): "Pay Enough or Don't Pay at All," *Quarterly Journal of Economics*, 115(3), 791–810.

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.

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

KUBE, S., M. MARÉCHAL, AND C. PUPPE (2012): "The Currency of Reciprocity: Gift Exchange in the Workplace," *American Economic Review*, 102(4), 1644–1662.

——— (2013): "Do Wage Cuts Damage Work Morale? Evidence From a Natural Field Experiment," *Journal of the European Economic Association*, 11(4), 853–870.

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.

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

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

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

Figures and Tables

Figure 1: Raw Average of Characters Inputted per Shift Across the Four Conditions
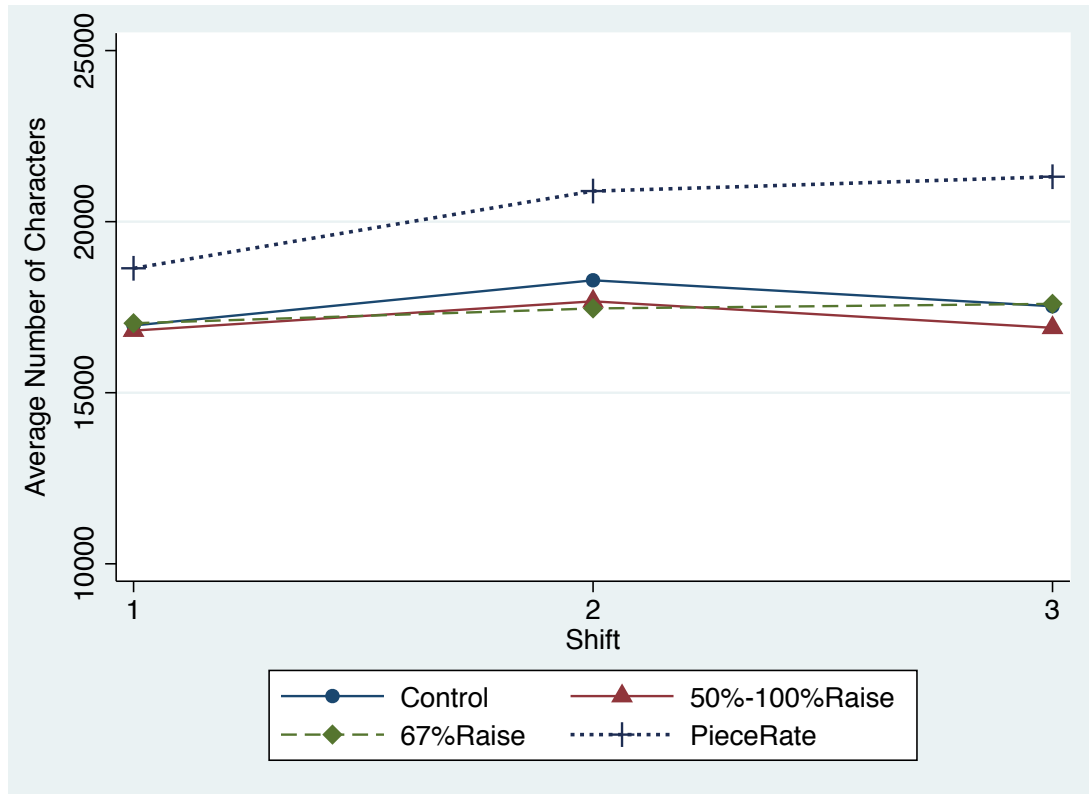
Note: PieceRate workers received an average per-worker piece-rate payment of $2.4, $5.8 and $7.3 in shifts one, two and three, respectively, in addition to the $24 per-shift base pay. This represents an additional average per-worker compensation of 10%, 24% and 30% in shifts one, two and three, respectively, which is smaller than that arising from the wage-raise treatments: 67% in all shifts for the 67%Raise and 50% for shifts one and two, and 100% for shift three, in the 50%-100%Raise.

Table 4: Average Number of Characters Inputted per Condition and Worker Sample

	Number of characters inputted per subject			
	Average across all worker-shifts	Ratio relative to the Control	Total workers in shift one	Total number of worker-shifts
	(1)	(2)	(3)	(4)
Control	17,591	1.00	47	131
67%Raise	17,361	0.99	70	207
50%-100%Raise	17,164	0.98	45	111
PieceRate	20,301	1.15	32	90
			194	539

Notes: Column (1) documents the average number of characters entered across all the worker-shifts in the CONTROL and the treatments. For example, workers in the CONTROL inputted 17,591 characters across all their shifts, whereas those in the 67%RAISE inputted 17,361 characters. Column (2) depicts the ratio of the number of characters inputted versus the CONTROL. For example, those in the 67%RAISE treatment inputted 0.99 = 17,361/17,591 characters relative to those in the CONTROL. Column (3) documents the number of recruited workers and therefore the number of workers in the first shift for each treatment. The sample for 50%-100%RAISE treatment though starting at 45 workers, only had 23 subjects in the third shift has only as subsample of 23 workers received the surprise wage raise to 100% of the market wage at the beginning of the third shift. Column (4) documents the number of worker-shift observations to account for missing shifts. For example, the 131 worker-shifts in the CONTROL result from 46 workers in shift one (the character-recording software did not record the characters for one worker), 45 in shift two (as two workers attrited in shift two) and 40 in shift three (as an additional 5 workers attrited in shift three). The 207 worker-observations in the 67%RAISE result from 70, 69 and 68 workers in shifts one, two and three, as one worker attrited in shift two and an additional worker attrited in shift three. The 111 worker-shift observations for the 50%-100%RAISE treatment results from 45 and 43 workers in shifts one and two, respectively (two attritted in shift two) and the additional subsample of 23 workers receiving the additional raise up 100%. The 90 worker-shift observations in the PIECERATE treatment result from 31 workers in shift one, 31 in shift two (as the character-recording software did not capture one worker's characters in shift one and another worker's characters in shift two) and 28 in shift three (as four workers attrited in shift three).

Table 5: Average Number of Characters Inputted per Condition per Campus, and Worker Sample per Campus

	Number of characters inputted per subject		Total workers in shift one	Total number of worker-shifts
	Average across all worker-shifts	Ratio relative to the Control		
	(1)	(2)	(3)	(4)
Panel A: Campus A				
Control	21,382	1.00	25	70
67%Raise	20,217	0.95	40	117
50%-100%Raise	19,363	0.91	30	75
PieceRate	26,812	1.25	15	42
			110	304
Panel B: Campus B				
Control	13,240	1.00	21	61
67%Raise	13,498	1.02	30	88
50%-100%Raise	12,977	0.98	15	38
PieceRate	14,500	1.10	16	48
			82	235

Notes: Panel A, column (1) shows the average number of characters inputted across the four conditions for campus A. For example, workers in the CONTROL, in campus A entered an average of 21,382 characters in their shifts whereas those in the 67%RAISE entered 20,217. Column (2) shows the ratio of the number of characters inputted versus the CONTROL. For example, those in the 67%RAISE treatment in campus A inputted $0.95 = 20,217/21,382$ characters relative to those in the CONTROL in campus A. Column (3) documents the number of recruited workers and therefore the number of workers in the first shift for each treatment in campus A. For example, the CONTROL started with 25 workers in shift one. Column (4) represents the total number of worker shift-observations. A decomposition of the number of workers per shift for the whole sample and per campus is in Appendix Table A.2. Panel B shows the same information as Panel A, but for campus B.

38

Table 6: Characters Inputted Across the Different Treatments Relative to the CONTROL

	Dependent Variable: ln(Number of Characters per Subject)								
	Unadjusted			Within campus			Within campus X leg X shift		
	Shifts			Shifts			Shifts		
	One	Two	Three	One	Two	Three	One	Two	Three
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Difference vs. Control									
67%Raise	0.02	-0.03	-0.01	0.01	-0.06	-0.03	0.04	-0.02	-0.04
	(0.07)	(0.08)	(0.08)	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)
50%-100%Raise	0.02	-0.03	0.00	-0.04	-0.09	-0.06	0.02	-0.04	-0.03
	(0.08)	(0.09)	(0.08)	(0.06)	(0.07)	(0.07)	(0.07)	(0.08)	(0.08)
PieceRate	0.08	0.12	0.15	0.11	0.14	0.19	0.06	0.10	0.18
	(0.10)	(0.10)	(0.12)	(0.08)*	(0.08)**	(0.09)***	(0.09)	(0.10)	(0.12)*
Constant	9.66	9.73	9.70	9.40	9.48	9.45	9.74	9.70	9.69
	(0.06)***	(0.06)***	(0.06)***	(0.06)***	(0.06)***	(0.06)***	(0.04)***	(0.03)***	(0.04)***
Campus fixed effects	-	-	-	Yes	Yes	Yes	-	-	-
CampusXlegXshift fixed effects	-	-	-	-	-	-	Yes	Yes	Yes
R-squared		0.02			0.35			0.38	
Number of workerXshift observations		539			539			539	

Notes: Columns (1)-(3) document the percentage difference between the raw means of the three treatments versus the CONTROL by using specification 1 without the fixed effects. The estimates for shifts two and three are also obtained using specification 1 without the fixed effects, but using the CONTROL in shifts two and three, respectively, as the baseline category, for ease of exposition. We also change the baseline this way for shifts two and three for the remaining analysis in this table. Columns (4)-(6) control for unobserved time-invariant campus heterogeneity by only having campus fixed effects in specification 1, instead of campus, leg and shift fixed effects. Columns (7)-(9) control for unobserved time-invariant determinants of performance within campus, leg and shift, with campus, leg and shift fixed effects, as outlined in specification 1. Standard errors clustered by worker. *Significant at the 10% level, **Significant at the 5% level, ***Significant at the 1% level. All tests are one-tailed.

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

	Workers in Field Experiment	Workers who Responded to Survey		Breakdown Of Respondents by Prosocial Type					
				Prosocial				Selfish	
	Total	Total	Response Rate	Altruists	Conditional Cooperators	Total	Prop. of Respondents	Total	Prop. of Respondents
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) All workers	194	139	0.72	10	47	57	0.41	45	0.32
(2) Wage raise treatments (67%Raise and 50%-100%Raise)	115	82	0.71	4	30	34	0.41	28	0.34
(3) PieceRate	32	23	0.72	4	7	11	0.48	3	0.13
(4) Control	47	34	0.74	2	10	12	0.35	14	0.41

Notes: Column (1) documents the distribution of workers across the whole field experiment. Column (2) documents the number of workers who responded to the survey in total and by condition. We aggregate the 67%Raise and 50%-100%Raise condition into row (2) to garner a larger sample size of at least 30 subjects. For a detailed breakdown of prosocial types by these two conditions, please see appendix table A.4. Column (3) documents the response rates to the survey for the whole sample and by condition. It is thus the ratio of column (2) over column (1). Column (4) documents the number of workers who behaved as *altruists* in the SPD games for the whole sample and by condition. They cooperated as a first mover, cooperated as a second mover if the first mover cooperated, and cooperated as second mover even if the first player defected (CCC). Column (5) documents the number of workers who behaved as *conditional cooperators* in the SPD games for the whole sample and by condition. They cooperated as a first mover, cooperated as a second mover if the first mover cooperated, but did not cooperate as a second mover if the first player defected (CCD). Column (6) documents the number of workers who behaved as *Prosocial* for the whole sample and by condition. Column (7) documents the proportion of prosocial types among all the respondents, both for the whole sample and by condition. It is thus the ratio of column (6) over (2). Column (8) documents the number of workers who behaved as selfish players in the SPD games for the whole sample and by treatment. They defected as a first mover, defected as a second mover even if the first mover cooperated, and defected as a second mover even if the first player defected (DDD). Column (9) documents the proportion of *Selfish* types among all the respondents, both for the whole sample and by treatment. It is thus the ratio of column (8) over (2). The remainder workers' were at neither end of this spectrum.

Table 8: Characters Inputted Across the Different Treatments Relative to the Control by Prosocial Type—within Campus, Leg and Shift

	Dependent Variable: ln(Characters inputted by subjects)								
				Prosocial in Treatments					
Sample:	Total Sample			Prosocial Workers in Treatments Versus Full Control			Prosocial Workers in Treatments Versus Survey Respondents in Control		
	Shifts			Shifts			Shifts		
	One	Two	Three	One	Two	Three	One	Two	Three
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Difference vs. Control									
Wage Raise treatments (67%Raise and 50%-100%Raise)	0.04 (0.07)	-0.03 (0.07)	-0.04 (0.07)	0.07 (0.08)	-0.03 (0.08)	-0.01 (0.10)	0.05 (0.10)	-0.04 (0.10)	-0.03 (0.11)
PieceRate	0.06 (0.09)	0.10 (0.10)	0.18 (0.12)*	0.15 (0.11)*	0.19 (0.14)*	0.24 (0.17)*	0.10 (0.12)	0.15 (0.16)	0.22 (0.19)
Constant	9.66 (0.06)***	9.73 (0.06)***	9.70 (0.06)***	9.73 (0.05)***	9.56 (0.05)***	9.75 (0.05)***	9.86 (0.10)***	9.78 (0.06)***	9.67 (0.07)***
CampusXlegXshift fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared		0.38			0.44			0.40	
Number of subjectsXsession		539			251			218	

Notes: Columns (1)-(3) document the percentage difference between the combined wage-raise treatments (67%Raise and 50%-100%Raise) and the PieceRate versus the Control for the whole sample for each of the three shifts. Columns (4)-(6) document these percentage differences for the subsample of workers in the treatments who behaved most prosocially in the SPD games (*Prosocial* workers), relative to the whole Control. Columns (7)-(9) document, as a robustness check, the same percentage differences for subsample of the *Prosocial* workers in the treatments relative the subsample of workers in the Control who responded to the survey. Standard errors clustered by worker. *Significant at the 10% level, **Significant at the 5% level, ***Significant at the 1% level. All tests are one-tailed.

A Additional Figures and Tables

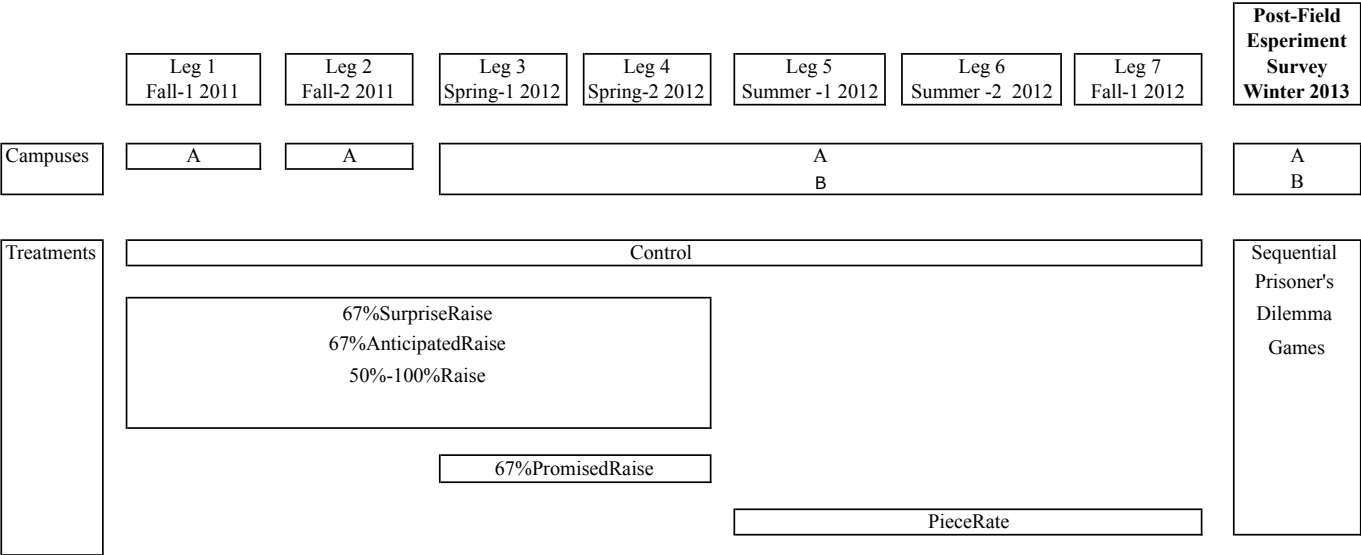Figure A.1: Timing of field experiment and post-field experiment survey

	Shift One (p-value) (1)	Shift Two (p-value) (2)	Shift Three (p-value) (3)
Panel A: Campus A			
(1) 67%SurpriseRaise=67%AnticipatedRaise= =67%PromisedRaise	0.32 (N_1=12, N_2=18, N_3=10)	0.27 (N_1=12, N_2=18, N_3=10)	0.27 (N_1=12, N_2=17, N_3=10)
Panel B: Campus B			
Without the two high productivity outliers in the 67%AnticipatedRaise			
(2) 67%SurpriseRaise=67%AnticipatedRaise= =67%PromisedRaise	0.21 (N_1=12, N_2=5, N_3=12)	0.41 (N_1=12, N_2=5, N_3=11)	0.14 (N_1=12, N_2=5, N_3=11)
With the two high productivity outliers in the 67\%AnticipatedRaise			
(3) 67%SurpriseRaise=67%AnticipatedRaise= =67%PromisedRaise	0.10 (N_1=12, N_2=7, N_3=12)	0.17 (N_1=12, N_2=7, N_3=11)	0.04 (N_1=12, N_2=7, N_3=11)

Notes: This table presents the results of the Kruskall-Wallis test for whether the 67%SURPRISERAISE, 67%ANTICIPATEDRAISE and 67%PROMISEDRAISE samples were drawn from the same population (against the alternative that they were not), within a given campus and shift, per leg. Panel A, row (1) tests for whether the samples for the 67%SURPRISERAISE, 67%ANTICIPATEDRAISE, 67%PROMISEDRAISE in campus A, at each leg, were drawn from the same population (e.g., whether the 67%SURPRISERAISE sample in leg 1 is drawn from the same population as that in legs 2, 3 and 4 and these samples are no different from the ones drawn for the 67%ANTICIPATEDRAISE and 67%PROMISEDRAISE treatments in the same legs for campus A). The p-values of 0.32, 0.27 and 0.27 in columns (1), (2) and (3), respectively, indicate that we cannot reject that these samples were drawn from the same population. The samples sizes are below the p-values, where N_1, N_2 and N_3 are the sample sizes in the 67%SURPRISERAISE, 67%ANTICIPATEDRAISE and 67%PROMISEDRAISE, respectively, in campus A. Row (2) performs the same analysis for campus B, without its two very high-productivity outliers in the 67%ANTICIPATEDRAISE treatment (two workers inputted twice as many characters than the average worker across the three treatments in this campus). The p-values of 0.21, 0.41 and 0.14 in shifts one through three, documented in columns (1) through (3), respectively, show that we cannot reject that the samples were drawn from the same population. Given that we include these high-productivity outliers in the 67%RAISE condition, as the effort increase in this condition is one of the focuses of the analysis, we show in Row (3), for completeness, how the test for equality of distributions changes when we include these two outliers. As expected, their inclusion changes the results of the test: we can reject the samples are drawn from the same distributions in shift one marginally at the 10% level and in shift three at the 4% level. The outcome measure for this test is the number of characters inputted, which more closely approximates effort, as we argue in Section 4.1.1.

Table A.2: Summary Statistics–Whole sample and Per Campus

	Average across all worker-shifts (1)	Shift One (2)	Shift Two (3)	Shift Three (4)	SD (5)	Min (6)	Max (7)
Panel A: Whole sample							
Control	17,591	16,965	18,286	17,528	6,917	6,298	37,722
N (worker-shifts)	131	46	45	40			
67%Raise	17,361	17,030	17,463	17,597	6,560	4,983	37,828
N (worker-shifts)	207	70	69	68			
50%-100%Raise	17,164	16,816	17,670	16,897	5,587	3,792	30,935
N (worker-shifts)	111	45	43	23			
PieceRate	20,301	18,797	20,893	21,312	8,841	6,455	42,200
N (worker-shifts)	90	31	31	28			
Panel B: Campus A							
Control	21,382	20,593	22,564	20,970	6,295	9,335	37,722
N (worker-shifts)	70	25	24	21			
67%Raise	20,217	19,542	20,344	20,780	6,340	7,800	37,828
N (worker-shifts)	119	40	39	38			
50%-100%Raise	19,343	18,789	20,444	18,397	5,141	11,475	30,935
N (worker-shifts)	73	30	30	15			
PieceRate	26,931	24,197	27,306	29,878	6,972	15,622	42,200
N (worker-shifts)	42	15	15	12			
Panel C: Campus B							
Control	13,240	12,644	13,398	13,723	4,700	6,298	26,183
N (worker-shifts)	61	21	21	19			
67%Raise	13,498	13,682	13,489	13,318	4,594	4,983	28,955
N (worker-shifts)	88	30	29	29			
50%-100%Raise	12,977	12,869	12,494	14,084	3,741	3,792	19,239
N (worker-shifts)	38	15	15	8			
PieceRate	14,500	13,734	14,880	14,887	5,655	6,455	28,526
N (worker-shifts)	48	16	16	16			

Notes: Panel A, column (1) shows the average characters inputted across by all workers in all shifts in each condition, for the whole sample, with the number of worker observations below. For example, workers in the CONTROL inputted an average of 17,591 characters across their total of 131 shifts. Column (2) shows the average number of characters inputted for shift one, with the number of workers per shift below. For example, the average characters inputted by the observed 46 workers in the CONTROL in the shift one was 16,965. Columns (3) and (4) depict the same information but for shifts two and three, respectively. Column (5) shows the standard deviation of characters across all worker-shifts per condition. For example, 6,917 was the standard deviation in the number of characters inputted by workers in the CONTROL across the 131 worker-shifts. Columns (6) and (7) represent the minimum and maximum of characters across all worker-shifts. For example, the 6,298 and 37,722 are the minimum and maximum number of characters entered in any shift across the 131 worker-shifts. The 131 worker-shifts in the CONTROL result from 46 workers in shift one (the character-recording software did not record the characters for one worker), 45 in shift two (as two workers attrited in shift two) and 40 in shift three (as an additional 5 workers attrited in shift three). The 207 worker-observations in the 67%RAISE result from 70, 69 and 68 workers in shifts one, two and three, as one worker attrited in shift two and an additional worker attrited in shift three. The 111 worker-shift observations for the 50%-100%RAISE treatment results from 45 and 43 workers in shifts one and two, respectively (two attritted in shift two) and the additional subsample of 23 workers receiving the additional raise up 100%. The 90 worker-shift observations in the PIECERATE treatment result from 31 workers in shift one, 31 in shift two (as the character-recording software did not capture one worker's characters in shift one and another worker's characters in shift two) and 28 in shift three, (as four workers attrited in shift three). Panels B and C depict the same information as above, but for campuses A and B, respectively.

Table A.3: Side-by-Side Comparison of Differences Between the Treatments and the CONTROL on Characters, Records and Percent of Correct Words Inputted

Dependent Variable	ln(charaters inputted per subject) Within campus X leg X shift			ln(records inputted per subject) Within campus X leg X shift			ln(correct words inputted per subject) Within campus X leg X shift		
	Shifts			Shifts			Shifts		
	One	Two	Three	One	Two	Three	One	Two	Three
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Results using specification (1)									
Difference vs. Control									
67%Raise	0.04	-0.02	-0.04	0.02	-0.05	-0.08	0.06	0.00	-0.07
	(0.07)	(0.07)	(0.07)	(0.08)	(0.08)	(0.07)	(0.09)	(0.09)	(0.09)
50-100%Raise	0.02	-0.04	-0.03	0.00	-0.03	-0.04	-0.01	-0.04	-0.02
	(0.07)	(0.08)	(0.08)	(0.07)	(0.08)	(0.08)	(0.08)	(0.09)	(0.10)
PieceRate	0.06	0.10	0.18	0.03	0.07	0.18	0.02	0.04	0.11
	(0.09)	(0.10)	(0.12)*	(0.10)	(0.11)	(0.13)*	(0.10)	(0.12)	(0.11)
Constant	9.74	9.70	9.69	4.37	4.35	4.35	7.58	7.56	7.54
	(0.04)***	(0.03)***	(0.04)***	(0.04)***	(0.04)***	(0.04)***	(0.074)***	(0.037)***	(0.039)***
CampusXlegXshift fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared		0.38			0.48			0.28	
Number of subjects	192	188	159	192	188	159	175	170	144
Number of subjectXsession observations		539			539			489	
Panel B: Unadjusted Control Averages									
Raw Average number of inputted by workers in the Control	16,965	18,286	17,526	75	86	85	1,943	2,098	2,066

Notes: Panel A, columns (1)-(3) replicate the prior analysis in Table 6, columns (7)-(9), from specification 1, documenting the percentage difference between the treatments and the CONTROL on the number of characters inputted, controlling for campus, leg and shift unobservables. Columns (4)-(6) replicate this analysis but where the outcome variable is the natural log of the number of records inputted by a subject, in a given treatment, campus, leg and shift instead of ln(characters). This analysis yields similar though sometimes smaller and slightly less precise estimates than that with ln(characters) as the dependent variable, as records are a noisier measure of effort (two records may differ substantially in the amount of characters required to enter them, e.g., the title may be longer and the article have more co-authors). Columns (7)-(9) replicate the analysis in specification 1, but where the outcome variable is the natural log of correct words inputted per subject, to assess the quality of the output. The number of observations for the proportion of correct words is slightly smaller at 489 as these data was unavailable across all conditions in one of our seven legs. Panel B contains the raw average of the characters, records and correct words for reference. Standard errors clustered by worker. *Significant at the 10% level, **Significant at the 5% level, ***Significant at the 1% level. All tests are one-tailed.

Table A.4: Distribution of Prosocial Types Among Field Experiment Workers in Wage Raise Treatments

	Workers in Field Experiment	Workers who Responded to Survey		Breakdown Of Respondents by Prosocial Type						
				Prosocial					Selfish	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)		(8)	(9)
Wage Raise Treatments	Total	Total Respondents	Response Rate	Altruists	Conditional Cooperators	Total	Prop. of Respondents		Total	Prop. of Respondents
(1) 67% Raise	70	47	0.67	2	20	22	0.47		14	0.30
(2) 50%-100%Raise	45	35	0.78	2	10	12	0.34		14	0.40

Notes: Column (1) documents the distribution of workers across the two treatments. Column (2) documents the number of workers who responded to the survey in total and by treatment. Column (3) documents the response rates to the survey per treatment. It is thus the ratio of column (2) over column (1). Column (4) documents the number of workers who behaved as *altruists* in the SPD games, by always cooperating, by treatment. Column (5) documents the number of workers who behaved as *conditional cooperators* in the SPD games, by always cooperating except when the first mover defected, by treatment. Column (6) documents the number of workers who behaved as *Prosocial* by treatment. Column (7) documents the proportion of prosocial types among all the respondents, by treatment. It is thus the ratio of column (6) over (2). Column (8) documents the number of workers who behaved as selfish players in the SPD games, by always defecting, by treatment. Column (9) documents the proportion of *Selfish* types among all the respondents by treatment. It is thus the ratio of column (8) over (2). The remainder workers' were at neither end of these spectrum.

Table A.5: Sample Breakdown Per Condition and Shift for Table 8

	Total Sample			Prosocial in Treatments					
				Prosocial Workers in Treatments Versus Full Control			Prosocial Workers in Treatments Versus Survey Respondents in Control		
	Shifts			Shifts			Shifts		
	One	Two	Three	One	Two	Three	One	Two	Three
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Total Observations	539			251			218		
Control	46	45	40	46	45	40	34	34	30
Combined wage raise treatments	115	112	91	34	33	24	34	33	24
67%Raise	70	69	68	22	21	20	22	21	20
50%-100%Raise	45	43	23	12	12	4	12	12	4
PieceRate	31	31	28	10	9	10	10	9	10

Notes: Columns (1)-(3) document the sample breakdown per condition and shift for the analysis in columns (1)-(3) in Table 8. Columns (4)-(6) breakdown the sample for the analysis in columns (4)-(6) in Table 8. One worker observation was missing in shift one for the CONTROL resulting in 46 workers observations (instead of 47) as the character-recording software did not record the characters for one worker in the CONTROL. Similarly the 9 workers in the PIECERATE in the second shift, instead of 10, result from the character-recording software not recording the characters for one worker in this treatment in the second shift. The 4 observations for the 50%-100%RAISE treatment result from only 4 workers among the 23-worker subsample in this treatments in shift three behaving as *Prosocial*. Columns (7)-(9) breakdown the sample for the analysis in columns (7)-(9) in Table 8. The sample for the CONTROL is the sample of those who were in the CONTROL and responded to the survey.

B   Online Appendix - Power Calculations

This section outlines our ex-ante power calculations ensuring that our sample would be large enough to reject the null of no gift exchange, at the 5% level, in favor of the one-sided alternative that a 67% fixed wage raise would elicit 20% extra effort. Based on the study most similar to ours—Gneezy and List (2006), which documented an increase in effort by 27% for the first 1.5 hours of their task, statistically significant at the 5% level in a one-tailed test—we show, through a simple calibration that we achieved 98% power to detect a 20% increase in effort in response to a 67% raise. Thus the reason we cannot document effort responses to fixed wage raises is not due to larger standard errors than those in this prior study—they were, in fact, smaller—but rather due smaller effect sizes ranging from -4% to 4%.

B.1   Ex-Ante Power Calculations Based on the Most Similar Study

The goal was to estimate the sample size necessary to achieve at least 80% power to reject, at the 5% level, the null of no effort increase in favor of the one-sided alternative of an effort increase of 20% following a 67% fixed wage raise. We used the conventional methodology of selecting an effect size and variance estimates of the closest study to ours, which is the data-entry study in Gneezy and List (2006). Our study uses an analogous task to their book digitization task: entering data on academic articles (title, authors etc.). It also has a similar sample (U.S. undergraduates), going market wage ($12 per hour), and wage increase (of $8 per hour to $20 per hour).

We thus use the effect size and variances for the first 1.5 hours in Gneezy and List (2006) to calibrate the sample size required for each two-hour shift. They found a 27% productivity increase for the first 1.5 hours which was statistically significant at the 5% level in a one-sided test. We calibrate our sample conservatively to find a 20% effort increase per two-hour shift, which is statistically significant at the 5% level in a one-sided test. In order to convert the productivity increases into elasticities, we convert the productivity (number of records) data in their study to natural logs.

Mean and Variance Estimates. The parameters in the Gneezy and List (2006) data-entry study were as follows (where the subscripts t and c indicate the treatment and control groups, respectively):

- For the Control group: a sample size $N_c = 10$ and a sample variance $\sigma_c^2 = (0.23)^2$.
- For the Gift treatment (an $8 hourly wage raise to $20 per hour): a sample size $N_t = 9$ and a sample variance $\sigma_t^2 = (0.34)^2$.
- Alternative hypothesis. Let μ_s^a designate the population mean, where the superscript a indicates the alternative hypothesis for $s = t, c$, the treatment and control, respectively. Thus, the effect

48

size—the increase in effort under the alternative hypothesis—corresponds to:[24]

$$H_a: \quad \mu_t^a - \mu_c^a = 0.2$$

where the alternative hypothesis is naturally one-sided following gift exchange's theoretical prediction that above-market wages increase, but never decrease effort. A two-sided alternative hypothesis would not only have been theoretically inaccurate, but would also have required different power calculations resulting in a larger ex-ante sample size.

Null Hypothesis. The null hypothesis of no effort responses to fixed wage raises corresponds to

$$H_0 : \mu_t - \mu_c = 0$$

Power Calculations. We can now compute the sample size to achieve at least 80% power to reject the null of no effort increase in favor of the one-sided alternative of a 20% increase, at the 5% level.

Given our large sample size, the estimator for the difference in population means $\bar{X}_t - \bar{X}_c$ is asymptotically normally distributed through a straightforward application of the Central Limit Theorem. Therefore, we compute:

$$
\begin{aligned}
0.8 &= P(\text{reject } H_0 \text{ at the 5\% level } | H_a \text{ is true}) \\
&= P(\bar{X}_t - \bar{X}_c \text{ is in the rejection region } | H_a \text{ is true}) \\
&= P\left(\frac{\bar{X}_t - \bar{X}_c}{\sqrt{\frac{\hat{\sigma}_t^2}{N_t} + \frac{\hat{\sigma}_c^2}{N_c}}} > 1.65 | H_a \text{ is true} \right) \\
&= P\left(\bar{X}_t - \bar{X}_c > 1.65\sqrt{\frac{\hat{\sigma}_2^t}{N_t} + \frac{\hat{\sigma}_c^2}{N_c}} | H_a \text{ is true} \right) \\
&= P\left(\frac{\bar{X}_t - \bar{X}_c - (\mu_t^a - \mu_c^a)}{\sqrt{\frac{\hat{\sigma}_t^2}{N_t} + \frac{\hat{\sigma}_c^2}{N_c}}} > \frac{1.65\sqrt{\frac{\hat{\sigma}_t^2}{N_t} + \frac{\hat{\sigma}_c^2}{N_c}} - (\mu_t^a - \mu_c^a)}{\sqrt{\frac{\hat{\sigma}_2^t}{N_t} + \frac{\hat{\sigma}_c^2}{N_c}}} \right) \\
&= P\left(Z > \frac{1.65\sqrt{\frac{\hat{\sigma}_t^2}{N_t} + \frac{\hat{\sigma}_c^2}{N_c}} - (\mu_t^a - \mu_c^a)}{\sqrt{\frac{\hat{\sigma}_t^2}{N_t} + \frac{\hat{\sigma}_c^2}{N_c}}} \right) \\
&= P\left(Z > \frac{1.65\sqrt{\frac{\hat{\sigma}_t^2}{N_t} + \frac{\hat{\sigma}_c^2}{N_c}} - 0.2}{\sqrt{\frac{\hat{\sigma}_t^2}{N_t} + \frac{\hat{\sigma}_c^2}{N_c}}} \right)
\end{aligned}
$$

After plugging the parameters shown above, power depends on the sample sizes for the control and

---

[24]The effect size is defined as a specific non-zero value in the population for our alternative hypothesis. This effect size definition was, for example, popularized by Cohen (1988), page 10.

treatment groups according to the following equation :

$$0.8 = P\left(Z > \frac{1.65\sqrt{\frac{0.23^2}{N_t} + \frac{0.34^2}{N_c}} - 0.2}{\sqrt{\frac{0.23^2}{N_t} + \frac{0.34^2}{N_c}}}\right)$$

For example, a sample size of 26 subjects for the treatment and control groups ($N_c = N_t = 26$), would achieve approximately 80% power in rejecting the null that wage raises do not increase effort in favor of the one-sided alternative that they increase it by 20%.

In our specific case, we exceed the sample size of 26 for both our control and treatment groups. Our control contains 47 subjects whereas the 67%RAISE treatment contains 70, achieving 98% power. The standard error in our 67%RAISE, in any two-hour shift, is lower, at 0.07, as shown in Table 6, columns (7)-(9), compared to the benchmark standard error of 0.13 in Gneezy and List (2006), resulting from the samples of 10 and 9 subjects in the control and treatment groups, respectively.[25] This suggests that the lack of detection of effort in response to fixed wage raises is no due larger standard errors in the estimates from the field experiment.

C   Online Appendix - Attrition Analysis

This section documents the results of a linear regression model ascertaining whether attrition in the second or third shifts differs between the treatments and the CONTROL. We estimate the parameters in the specifications that follow using a linear probability model.

Empirical method. We estimate whether subject i, in t_1 (CONTROL), t_2 (67%RAISE), t_3 (50-100%Raise), t_4 (PIECERATE) in campus c, leg l, and shift s attritted as follows:

$$\text{attrit}_{i,t,s,c,l} = \theta_1 + \theta_2 t_2 + \theta_3 t_3 + \theta_4 t_4 + \psi_c \times \psi_l \times \psi_s + \epsilon_{i,s,t,c,l} \tag{2}$$

The variable *attrit* is binary, taking the value one if the worker attrited in a given shift and zero otherwise. The interaction $\psi_c \times \psi_l \times \psi_s$, controls, as discussed previously in the context of specification 1, for unobserved time-invariant campus, leg and shift determinants of attrition. These unobservables can affect the difference in attrition between the treatments and the CONTROL within a given campus, leg and shift. For example, a given leg, in a given campus, shift three may occur closer to finals, leading to less attrition in the wage raise or piece rate treatments, where subjects receive additional payments beyond the contract wage, and the CONTROL, where they do not. Further, this interaction

---

[25]To compute this standard error we just plug 10 and 9 into N_c and N_t respectively, in $\sqrt{\frac{\hat{\sigma}_t^2}{N_t} + \frac{\hat{\sigma}_c^2}{N_c}}$.

allows us to estimate differences between the treatments and the control, within campus, leg and shift, in line with the previously described random assignment scheme, and then to pool them.

The causal parameters of interest are the θ_i. They pool the percentage differences in attrition between the treatments and the CONTROL within a campus, leg and shift, for shifts two and three. For example, θ_2 identifies percentage difference in attrition between treatment two (67%RAISE) in shifts two and three and the CONTROL in shifts two and three, by pooling all these differences within each campus, leg and shift. The parameter θ_1 estimates the average percentage of attriters the baseline category—the outcome for the CONTROL in both shifts two and three—which cannot be separately identified from the fixed effects, as usual.

Due to serial correlation in the attrition of each worker across shifts (serial correlation in $\epsilon_{i,s,t,c,l}$), we cluster the standard errors at the subject level (Bertrand, Duflo, and Mullainathan (2004)).

Results. Table 6, column (1) documents the rate of attrition without controlling for unobserved time-invariant campus, leg and shift factors. It thus transforms into percentages the attrition rates in Table A.2, with the summary statistics for the whole sample and per campus, in columns (2)-(4). For example, the estimate of 10% for the constant in Table 6, column (1), row (4), documents the baseline attrition for the CONTROL. It corresponds exactly to attrition documented Table A.2, columns (2)-(4) for the CONTROL: of the potential 94 worker-shift observations across shifts two and three (from the 47 workers in shift one, we should have had 47 workers in shift two and 47 in shift three), 2 workers missed the second and third shifts (4 missed shifts) and an additional 5 workers missed the third shift (and additional 5 missed shifts). These 9 missed shifts represent the documented rate of attrition of 10% (9/94).

Column (1), rows (1), (2) and (3) document that the unadjusted differences in attrition between the treatments and the CONTROL are not statistically significant, except for the 67%RAISE treatment. Attrition in the 67%RAISE is 7% smaller than that in the CONTROL, though only marginally significant at the 10% level. As before, this 7% estimate also accords with the attrition documented in Table A.2, columns (2)-(4). Of the 140 potential worker-shifts in the 67%RAISE in the second and third shifts (from the 70 workers in shift one, we should have had 70 workers in shifts two and 70 in shift three), one worker missed the second and third shift and an additional worker missed the third shift, resulting in 3 missed shifts. This amounts to an attrition rate of 2.1% (3/140), which is approximately 7% lower than that in the CONTROL, as documented. However, none of these estimates control for time-invariant campus, leg an shift determinants of attrition.

Column (2), with estimates including campus, leg and shift fixed effects, documents that there are

no statistically significant differences in attrition between the treatment and the CONTROL, at the 5% level, when correctly conducting the analysis within campus and leg, following the random assignment scheme. Namely, attrition in the 67%RAISE becomes positive but statistically insignificant and attrition in the PIECERATE though negative, is only marginally significant.

The reason the lower attrition in the 67%RAISE becomes positive and not statistically significant with the campus, leg and shift fixed effects, when it was negative and marginally significant when comparing the raw means in column (1), is that using campus, leg and shift fixed effects answers the question "When the 67%RAISE treatment was run within a given campus and leg, was the attrition in shifts two and three in this condition higher or lower than that in the CONTROL, in that same campus, leg and shifts?". In contrast, the raw means estimation in column (1) compares attrition in the 67%RAISE to that in the CONTROL, which includes students from other campuses and legs where the 67%RAISE was not run (e.g., legs 1 and 2 in campus B and legs 5 through 7 on both campuses A and B, as in Figure A.1.). Thus, for the legs and campuses where we ran the 67%RAISE condition, attrition was 3% higher than that in the CONTROL in those same campuses and legs, though not statistically significant.

In contrast, in the campus and legs where we ran the PIECERATE, attrition was 11% lower than that in the CONTROL in those campuses and legs, though only marginally significant.

Table 6: Differential Attrition Between the CONTROL and the Treatments

Dependent variable: =1 if Subject Attrited in Shift Two or Three		
Sample:	Full Worker Sample	
	Unadjusted	Within campusXlegXshift
	(1)	(2)
Diff. vs. Control group		
(1) 67%Raise	-0.07	0.03
	(0.04)*	(0.02)
(2) 50%-100%Raise	-0.07	0.03
	(0.04)	(0.02)
(3) PieceRate	-0.02	-0.11
	(0.05)	(0.07)*
(4) Constant	0.10	(0.06)
	(0.04)***	(0.02)***
Campus X leg X shift fixed effects	-	Yes
R-squared	0.01	0.18
Number of subjectsXsession observations	366	366

Notes: Number of observations: 194 observations in shift two and 172 observations in shift three, totaling 366 worker-shift observations. The sample declines by 22 observations from shift two to three (from 194 to 172) as we only implemented the 100% wage raise on the subsample of 23 workers in the 50%-100%RAISE treatment for the third shift, instead of the full sample of 45 workers, which naturally reduced the sample size by 22 workers. Standard errors clustered by individual. **Significant at the 5% level, *Significant at the 10% level. All tests are two-tailed, as we did not have a specific hypothesis for whether workers should attrit more or less in the treatments than in the CONTROL.

D Sources for the Overview of the Evidence on Gift Exchange in the Workplace

This Appendix presents the sources and computations for Tables 1 and 2. All studies considered use changes in effort in response to different wage raises. To compare effort responses across studies, we calculate the associated elasticities by dividing the % wage variation by the % effort increase. Effort responses are significant estimates at the 5% level and using two-tailed tests, unless otherwise stated.

PANEL I: Field Studies

(1) **Gneezy and List (2006)**

1.1) *Sample sizes* for the "Gift" treatment and the control ("noGift") are in Table I, column participant number, page 1371 (data-entry task) and in Table V, column participant number, page 1376 (fundraising task).

1.2) *Wage increases*

- *Data-entry task.* Raise from $12 to $20 dollars in Section 2.A, second paragraph in page 1368. It represents a 8*100/12=67% increase.

- *Fundraising task.* Raise from $10 to $20 dollars in Section 2.B, first paragraph in page 1370. It represents a 10*100/10=100% increase.

1.3) *Effort responses* to wage increases and their significance levels for each task were calculated as follows.

- *Data-entry task.* Productivity is measured by the number of books logged. Productivity differences between the "Gift" treatment and the control for the 90, 180, 270 and 360 minutes are presented in Table I, page 1371,

  * The overall effort response over the six hours is reported in Table I, column 360 minutes, row Average, page 1371. The difference between the average number of records imputed by the 9 subjects in the "Gift" treatment over the 6 hours interval (40.3 records) and the same number for the 10 subjects in the control (39.6 records) leads to a difference of 40.3-39.6=0.7 records. This represents a (0.7*100)/39.6=2% increase, which is not statistically significant using a one-sided Wilcoxon test (see page 1372).

  * The effort response for the first 90-minute interval is reported in Table I, column 90 minutes, row Average, page 1371. The difference between the "Gift" treatment and

the control corresponds to 51.7-40.7=11, which represents an statistically significant increase at the 5% level of (11*100)/40.7=27% using a one-sided Wilcoxon test and a t-test (see pages 1370-1372).

* The effort response for the first three hours is reported in Table I, column 180 minutes, row Average, page 1371. The difference between the average number of records imputed by the 9 subjects in the "Gift" treatment over the first three hours (44.9 records) and the same number for the 10 subjects in the control (40.5 records) leads to a difference of 44.9-40.5=4.4 records. This represents a (4.4*100)/40.5=11% increase, which is not statistically significant using a one-sided Wilcoxon test (see second paragraph in page 1372).

1.4) *Elasticities.* The overall elasticity for the whole six hours corresponds to 2%/67%=0.03. For the first 90 minutes it corresponds to 27%/67%=0.40. For the first three hours the elasticity corresponds to 11%/67%=0.16.

- *Fundraising task.* Productivity is measured by the earnings raised. Productivity differences between the "Gift" treatment and the control for the overall six hours and by three-hour intervals are displayed in Table III, page 1374 with their significance levels. Averages by 90-minute intervals are not reported.

  * The overall effect corresponds to the difference between the average earnings by the 13 subjects in the "Gift" treatment over the 6 hours interval ($9.013) and the same number for the 10 subjects in the control ($6.516 dollars) (see Table III, column Gift and NoGift, respectively, row Entire day per hour). The difference 9.013-6.516=2.496 represents a (2.496*100)/6.516=38% increase, which is statistically significant at the 10% level using a one-sided Wilcoxon test (see Table III, column Difference, row Entire day per hour).

  * The effect for the first three-hour window corresponds to the difference between the earnings by the 13 subjects in the "Gift" treatment over the first three hours ($11.00) and the same number for the 10 subjects in the control ($6.40) (see Table III, column Gift and NoGift, respectively, row Pre Lunch per hour). The difference 11.00-6.40=4.6 represents a (4.6*100)/6.40=72% increase, which is statistically significant using a one-sided Wilcoxon test (see Table III, column Difference, row Pre Lunch per hour). For the second-three hours the difference between treatment and control of 7.026-6.633=0.392 represents a (0.392*100)/6.633=6% increase, which is not statistically

55

significant using a one-sided Wilcoxon test (see Table III, column Difference, row Post Lunch per hour).

1.4) *Elasticities.* The overall elasticity corresponds to 38%/100%=0.38; for the first and second hours it corresponds to 72%/100%=0.72 and 6%/100%=0.06, respectively.

(2) **Bellemare and Shearer (2009)**

2.1) *Sample size* of 18 workers is in first paragraph of Section 3, page 253.

2.2) *Wage increase* of $80 dollars in the second day of work (in addition to the $0.20 piece rate for all seven days) is described in Section 3, second paragraph in page 235. The total $215 average daily earnings using the piece rate is described in Section 4, page 236, end of the last paragraph. The gift thus corresponds to an average 80*100/215=37% increase.

2.3) *Effort responses* to the wage increase measured by the increase in the daily average number of trees planted, are presented in Table 2, page 238 with their significance levels. Effort responses change according to whether only data on the experimental block is considered (block for which the workers received the gift) or if productivity of the same workers in neighboring blocks is also included (Table 2, columns I and II, respectively). Fixed effects by planter and block are used in both cases.

* Table 1, column I shows the estimates using daily productivity for the seven-day window considering only productivity in the experimental block. Workers increase productivity by 118 trees on average after receiving the gift, which is statistically at the 1% level. Given that the average baseline effort of all workers pre-gift is unknown (pre-gift effort and worker fixed effects cannot be separately identified), we use the average productivity for the experimental block of 1075.59 tree as a proxy for pre-gift average worker effort when computing the percentage increase in productivity with the gift. Given that raw average of 1075.59 trees incorporates productivity both with and without the gift, the increase in 118 trees lead to an estimate of the lower bound on the average productivity increase, which is 118.31*100/1075.59=11% increase. On the other hand, the upper bound on the increase in productivity arising from the 118 tree increase is 118/(1075.59-118.31)=12%

* Table 1, column II, shows the estimates using daily productivity for the seven-day window considering productivity in the experimental and non-experimental blocks. Workers increased productivity by 132.271 trees on average after receiving the gift,

which is statistically at the 1% level. Again, we use the average productivity for the experimental block of 1075.59 tree as a proxy for pre-gift average worker effort when computing the percentage increase in productivity. The average of 1075.59 trees not only incorporates productivity both with and without the gift but it is also higher than that in the non-experimental blocks 971.55 trees. Thus the increase of 132 trees is an estimate of the lower bound on the average productivity increase, which is 132.27*100/1075.59=13% increase. On the other hand, using the experimental block as the baseline, the upper bound on the increase in productivity arising from the 132.27 tree increase is 132.27/(1075.59-132.27)=14%.

2.4) *Elasticities.* The resulting elasticities range from to 11%/37%=0.30 to 14%/37%=0.38.

(3) **Hennig-Schmidt, Sadrieh, and Rockenbach (2010)**

3.1) *Sample sizes* for the control ("F0"), "F10" and "F40 peer" treatments of 24, 25 and 23 subjects respectively, are in Table 1, columns F0, F10 and F40 peer, respectively, row Number of typist, in page 821.

3.2) *Wage increases* of $2 Deutsche Marks (DM) for the "F10" treatment and of $8 DM for the "F40 peer" treatment above the baseline of $20 DM per hour are in Table 1, columns F0, F10 and F40 peer, respectively, row Wage 2nd hour in page 821. Because the wage raise only applies to the second hour, the percentage wage raise corresponds to 2*100/20=10% for the "F10" treatment and 8*100/20=40% for the "F40 peer" treatment. In the "F40 peer" treatment, the wage raise is accompanied by information about the wage raise of the "F10" treatment.

3.3) *Effort responses* to the wage increase, measured as the number of correctly typed words per minute, are presented in Table 2, page 822. The change from period one to two in the number of correct imputed words per minute in the "F10" treatment corresponds to 0.152, while the control increased by 0.634 (see Table 2 column F10 and F0, row 2nd minus 1st hour). The difference 0.152-0.634=-0.48, which corresponds to a -0.48*100/0.634=-76% increase is not statistically significant (see Table A.1, column F0 vs. F10, row Output ratio-usable difference in Appendix A, page 832). The change from period one to two in the number of correct imputed words per minute in the "F40 peer" treatment corresponds to 0.459 (see Table 2, column F40 peer, row 2nd minus 1st hour). The difference with the control 0.459-0.634=-0.18 corresponds to a -0.18*100/0.634=-28% increase, which is not significant (formal test is not reported in Table A.1; significance only reported verbally in

Section 2.3, page 823, last paragraph).

3.4) *Elasticities.* The elasticity in the "F10" treatment corresponds to -76%10%=-7.6. The elasticity for the "F40peer" corresponds to -28%40%=-0.69.

(4) **Kube, Maréchal, and Puppe (2012)**

4.1) *Sample size* of 34 student workers in the "Money" treatment and 35 in the control ("Baseline") stated at the end of Section I, page 1648, second paragraph.

4.2) *Wage increase* of a total of 7 euros from the 12 euros per-hour baseline in Section I, page 1646, first and fourth paragraphs, respectively. See also Appendix Table A4 in page 1659. Because this is a three-hour task, the gift corresponds to a 7*100/(3*12)=19% increase.

4.3) *Effort responses* to the wage increase, measured by the number of charters entered, are presented in Table I, page 1649 in percentages and with their significance levels. Subjects in the "Money" treatment increased their productivity by 5.2%, which is not significant (see Table I, column baseline, row Money; In Table 1 we round this estimate to 5% to keep all effort estimates without decimals). For reference, productivity levels for the treatment and the control are presented in Appendix Table A2, page 1658. The average productivity of the control, which lumps the productivity of the "Baseline I" and "Baseline II" treatments, corresponds to 7,983.5+8,622.1=16,605.6 (see Table A2, column Characters, rows Average for Baseline I and Baseline II treatments). The average productivity of the "Money" treatment, which lumps the productivity of the "Money" and the "MoneyUpfront" treatments, correspond to 8,462.3+8,989.9=17,452.2 (see Table A2, column Characters, rows Average for Money and MoneyUpfront treatments).

4.4) *Elasticity* corresponds to 5%/19%=0.26.

(5) **Kube, Maréchal, and Puppe (2013)**

5.1) *Sample size* of 22 student workers in the "PayRaise" treatment and 25 in the control ("Baseline") stated at the end of Section 2, page 858, third paragraph.

5.2) *Wage increase* of 5 euros per-hour from 15 euros per-hour baseline is at the end of Section 2, page 857, second paragraph. The gift corresponds thus to a 5*100/15=33% increase.

5.3) *Effort responses* to the wage increase, measured by the number of books entered, are presented in percentages in Table 1, page 859 by 90-minutes intervals and in the overall six hours with their significance levels.

* The overall effort response of subjects in the "PayRaise" treatment corresponds to a

productivity decrease of by -0.3% (see Table 1, column PayRaise-Baseline, row All quarters), which is not significant (see Table 1, column p>|z|, row All quarters). For reference, productivity levels for the treatment and control are presented in the last paragraph of page 859. The average productivity of the control corresponds to 219.4 books entered. The average productivity of the PayRaise treatment corresponds to 218.6 books entered.

* The effort response of the "PayRaise" treatment by 90-minutes intervals correspond to -9.5%, 1%, 0.2% and 6.5% (see Table 1, column PayRaise-Baseline, rows Quarter I, Quarter II, Quarter III and Quarter V, respectively; In Table 1 we round these estimates to -10%, 1%, 0.2% and 7% to keep all effort estimates above one without decimals). None is statistically significant (see Table 1, column p>|z|, rows Quarter I, Quarter II, Quarter III and Quarter IV, respectively).

5.4) *Elasticity* for the overall six hours corresponds to -0.3%/33%=-0.01. For each 90-minute intervals, the elasticities correspond to -10%/33%=-0.33, 1%/33%=0.03, 0.2%/33%=0.01 and 7%/33%=0.21 for the first, second, third and fourth intervals respectively.

(6) **Cohn, Fehr, and Goette (forthcoming)**

Refereces to page numbers are omitted since only the online version of the paper is currently available.

6.1) *Sample size* of 196 workers of a promotion agency hired to distribute the newly launched newspaper of a publishing company is in the first paragraph in section 2.4.

6.2) *Wage increase* of 5 Swiss Francs (CHF) per hour from the 22 CHF per-hour baseline is in the first paragraph in section 2.2. The gift corresponds thus to a 5*100/22=23% increase.

6.3) *Effort responses* to the wage increase were measured by the hourly number of newspaper copies distributed.

* The effort response for the full sample of workers comes from Table 6 displaying the coefficient estimates of regressing the logarithm of hourly number of copies distributed on a treatment dummy variable (1 if received a wage raise; 0 otherwise) plus location and day fixed effects. Table 6, column (1), row CHF27, shows that the parameter associated with the treatment dummy is 0.037, which is significant. Column (2) shows that when adding worker fixed effects this coefficient estimate corresponds to 0.030, which is also significant. Table 1 presents this estimate. Thus, the estimated increase

in effort is 3%.

    * The effort response for reciprocal workers who felt underpaid at the baseline wage is displayed in column (1) Table 10, where workers were classified as reciprocal and non-reciprocal using a Moonlighting game. The coefficient estimates of regressing the logarithm of hourly number of copies distributed on a treatment dummy variable (1 if received a wage raise; 0 otherwise) and the interaction between the treatment dummy and the difference between the wage a worker considered to be fair and the base wage correspond to 0.000 (not significant) and 0.028 (significant), respectively (See Table 10, column (1), rows Intercept, CHF27 and CHF27×Δ_i). Therefore, the total effort increase for reciprocal workers who felt underpaid by 5 CHF corresponds to 0.000+5*0.028=0.14 or 14%. For reference, neither the treatment dummy nor the interaction term is significant for the non-reciprocal workers.

6.4) *Elasticity* for the overall sample it corresponds to 3%/23%=0.13. The elasticity for the reciprocal subjects who felt 5 CHF underpaid at the base wage corresponds to 14%/23%=0.61%

(7) **Gilchrist, Luca, and Malhotra (2014)**

Because this paper had not been published at the moment this review was conducted, reported estimates are from the last version available online: May 2014. Consequently, no pages are cited.

7.1) *Sample size* of 58 oDesk workers in the "3+1" treatment and 110 in the control ("3") are shown in Table 1, fourth and fifth columns, fourth row.

7.2) *Wage increase* of $1 dollar per-hour from baseline $3 dollars per-hour is presented in Figure 2 on the experimental design. The gift corresponds to an 1*100/3=33% increase.

7.3) *Effort response* to wage increases and their significance levels were calculated as follows. Productivity is measured by the number of completed and correct CAPTCHAs entered in the 4-hour task.

    * The effort response for the full sample of workers comes from Table 2, where treatment "3+1" corresponds to the baseline category. Column (1) shows the difference between the control and the "3+1" treatment using the number of completed and correct CAPTCHAs as the dependent variable. The baseline productivity of the "3+1" treatment is captured by the constant term amounting to 938.5, while the coefficient of the dummy for the control (1 if wage is 3; 0 otherwise) corresponds to -146.8,

which is significant using robust standard errors (see Table 2, Column (1), rows Constant and Wage=3+1). The effort increase for the full sample thus corresponds to 146.4*100/(938.5-146.4)=18%.

* The effort response for the experienced oDesk workers comes from Table 3, where treatment "3+1" corresponds to the baseline category. Column (1) shows the difference between the control and the 3+1 treatment for experienced and unexperienced oDesk workers using the number of completed and correct CAPTCHAs as the dependent variable. The productivity in the "3+1" treatment for experienced workers is 861.8+111.2=973 records (see Table 3, column (1)), whereas it is 973-199.9=773.1 for the control. This difference in 199.9 records is statistically significant at the 5% level. The effort increase for the experienced sample corresponds thus to 199.9*100/773.1=26%.

* The effort response for the unexperienced oDesk workers comes from Table 3 and mimics the computation for experienced workers. In this case, however, the difference between the baseline "3+1" and the control for unexperienced workers is the dummy for the control (1 if wage is 3; 0 otherwise) alone, which corresponds to -27.34, which is not significant (see Table 3, Column (1) rows Wage=3). The effort increase for the unexperienced sample corresponds thus to 27.34*100/(861.8-27.34)=3%.

7.4) *Elasticities.* The elasticity for the full sample corresponds to 18%/33%=0.55. For the experienced workers it corresponds to 30%/33%=0.91 and for the unexperienced workers it corresponds to 3%/33%=0.09.

PANEL II: Most Cited Laboratory Studies

(1) Fehr, Kirchsteiger, and Riedl (1993)

1.1) *Sample size* in Section II, second paragraph in page 440.

1.2) *Average Wage offer* of 72 experimental units above the market wage of 30 units is in the first paragraph of Section V, page 446. Market wage is in Section III, page 443, second paragraph. This represents a (72-30)*100/30= 140% increase.

1.3) *Average effort response* of 0.4 units is in the first paragraph Section V, page 446 (effort range is 0.1, 0,2, ..., 1). The competitive effort level, which corresponds to the minimum effort of 0.1, is in Section III, page 443, second paragraph. This represents a (0.4-0.1)*100/0.1= 300% increase.

1.4) *Elasticity* corresponds to 300%/140%=2.14.

(2) **Fehr, Kirchsteiger, and Riedl (1998)[26]**

    2.1) *Sample size* in first paragraph of Section V, pages 9-10.

    2.2) *Average Wage offer* of 74 experimental units above the market wage of 30 units is in the first paragraph of Section VI, page 11. Market wage is in Section V, first paragraph in page 11 (denoted by f). This represents a (74-30)*100/30= 147% increase.

    2.3) *Average effort response* is not reported. We estimate it using the experimental data provided in Tables 5 to Table 8 in Appendix B, pages 24 to 32. These tables display the observed wage, the effort, the cost of effort and the id numbers for workers and firms for all shifts and for all periods for the control and reciprocity treatments. The observed wage and effort are displayed in columns "p" and "q", respectively. There is no row or column indicating to which treatment each observation corresponds to, but since in the control condition effort was exogenously determined by the experimenter, for this condition effort is filled with a dash in the "q" column. The average effort for the reciprocity treatment corresponds to the raw average of the 213 effort observations in the "q" columns, which are not dashed, across Tables 5 to 8. See section 6, first paragraph in page 11 for a quote that 213 is right number of effort observations for this treatment. This raw average, therefore, pools effort across all employer-employee matches, all rounds and all shifts. From this calculation the average effort response corresponds to 0.35 units (effort range is 0.1, 0,2, ..., 1). The competitive effort level (denoted by q_0), which corresponds to the minimum effort of 0.1, is in Section V, first paragraph in page 11. This represents a (0.35-0.1)*100/0.1= 250% increase.

    2.4) *Elasticity* corresponds to 250%/147%=1.70.

(3) **Fehr, Kirchler, Weichbold, and Gächter (1998)**

    3.1) *Sample sizes.* Number of shifts is in the first paragraph of Section III, page 333. Number of subjects per shift is in the fist paragraph of Section II.A, page 329 ("Bilateral GE treatment") and in Section II.C, page 331 ("GE Market treatment").

    3.2) *Average Wage offers* by treatment are not reported. We estimate them using Figure 2a in page 340.

---

[26]Fehr, Kirchsteiger, and Riedl (1998) frame the experiment in terms of prices offered by buyers and quality offered by sellers, but argue this framing applies to labor markets, where buyers are employers and sellers are workers. Hence the wages and effort terminology we use here.

Table 7: Calculation of the Average Effort for the Bilateral GE Treatment in Fehr, Kirchler, Weichbold, and Gächter (1998)

Wage Interval (Experimental Currency)	Percentage of Trades Per Wage Interval	Approximated Average Effort Per Wage Interval	Weighted Average Effort Per Wage Interval
(1)	(2)	(3)	(4)
21-30	9	0.16	1.44
31-40	7	0.21	1.47
41-50	17	0.29	4.93
51-60	20	0.35	7
61-70	21	0.43	9.03
71-80	15	0.44	6.6
+ 80	11	0.53	5.83
Total	100	2.41	36.3
Total Weighted Average Effort			0.36

Notes: Columns (1) shows the wage intervals as shown in the x-axis of Figure 1. Column (2) shows the percentage of trades (employer-employee matches realised) as shown in Figure 1. Column (3) corresponds to an approximation of the average effort of each wage interval, which was estimated visually from Figure 1. Column (4) corresponds to the multiplication of columns (2) and (3). "Total Weighted Average Effort" corresponds to the summation of column (4) divided by 100.

- *Bilateral GE treatment.* The approximated average wage offer is 63 experimental units above the market wage of 20 units. Market wage is in Table 1, first column, last row, page 329. This represents a (63-20)*100/20= 215% increase.

- *GE Market treatment.* The approximated average wage offer is 59 experimental units above the market wage of 20 units. Market wage is in Table 1, second column, last row, page 329. This represents a (59-20)*100/20= 195% increase.

3.3) *Average effort responses* are not reported. We estimate them using Figure 1 in page 334. Figure 1 shows the average effort by wage intervals. Intervals correspond to 21 to 30 wage units, 31 to 40, 41 to 50, ... , 71 to 80 and more than 80 wage units. Figure 1 also reports the percentage of employer-employee matches in each wage interval. The average effort response is calculated as the weighted average of the average efforts by wage intervals.

- *Bilateral GE treatment.* The approximated average effort response corresponds to 0.36 units (effort range is 0.1, 0,2, ..., 1). Table 7 shows the exact calculation. The competitive effort level, which corresponds to the minimum effort of 0.1, is in Table 1, first column, last row, page 329. This represents a (0.36-0.1)*100/0.1= 260% increase.

- *GE Market treatment.* The approximated average effort response corresponds to 0.4 units (effort range is 0.1, 0,2, ..., 1). Table 8 shows the exact calculation. The competitive effort level, which corresponds to the minimum effort of 0.1, is in

Table 8: Calculation of the Average Effort for the GE Market Treatment in Fehr, Kirchler, Weichbold, and Gächter (1998)

Wage Interval (Experimental Currency)	Percentage of Trades Per Wage Interval	Approximated Average Effort Per Wage Interval	Weighted Average Effort Per Wage Interval
(1)	(2)	(3)	(4)
21-30	3	0.1	0.3
31-40	7	0.15	1.05
41-50	18	0.3	5.4
51-60	33	0.42	13.86
61-70	26	0.48	12.48
71-80	11	0.5	5.5
+ 80	2	0.55	1.1
Total	100	2.5	39.69
Total Weighted Average Effort			0.40

Notes: Columns (1) shows the wage intervals as shown in the x-axis of Figure 1. Column (2) shows the percentage of trades (employer-employee matches realised) as shown in Figure 1. Column (3) corresponds to an approximation of the average effort of each wage interval, which was estimated visually from Figure 1. Column (4) corresponds to the multiplication of columns (2) and (3). "Total Weighted Average Effort" corresponds to the summation of column (4) divided by 100.

Table 1, second column, last row, page 329. This represents a $(0.4-0.1)*100/0.1=$ 300% increase.

3.4) *Elasticities.* The elasticities correspond to 260%/215%=1.21 ("Bilateral GE treatment") and 300%/195%=1.54 ("GE Market treatment").

3.5) *Effort response and elasticity to a 67% wage increase in the "Bilateral GE treatment".* An increase in 67% in wages corresponds to an increase from the market-clearing wage of 20 to 33. Figure 1 in page 334, documents that for a wage of 33, the effort response raises from 0.1 to 0.15-0.20. This corresponds to a 50% to 100% increase, and a pay-effort elasticity between 50%/67%=0.75 and 100%/67%=1.5.

(4) **Gächter and Falk (2002)**

4.1) *Sample size.* Number of shifts is in the first paragraph of Section IV, page 7. Number of subjects per shift is in Appendix: Instructions, page 22.

4.2) *Average Wage offer* is not reported. We approximate it using the reported average payoff of the firm, (120-W)*e, which corresponds to 19.4 (see Section IV, first paragraph in page 8). Using the average effort of 0.41 (see below), we have that (120-W)*0.41=19.4, which means that the average wage offer is approximately W=120-(19.4/0.41)=73. The market wage of 21 is in Section IV, first paragraph in page 8 (denoted as w^*). This represents a (73-21)*100/21= 248% increase.[27]

---

[27]Figure 1, in page 7, however, shows that the average wage, per shift for the "OS" treatment ("One-Shot" treatment)

4.3) *Average effort response* of 0.41 units (effort range is 0.1, 0,2, ..., 1) is in Section IV, first paragraph in page 9. The competitive effort level of 0.1 is in Section IV, first paragraph in page 8 (denoted as e^*). This represents a (0.41-0.1)*100/0.1= 310% increase.

4.4) *Elasticity* corresponds to 310%/248%=1.25.

(5) **Brown, Falk, and Fehr (2004)**

5.1) *Sample size.* Number of shifts is in second paragraph in page 755. Number of subjects per shift is second paragraph in Section 4, page 759.

5.2) *Average Wage offer* is not reported. We estimate it using Figure 3, page 763, which shows the average wage offer by period. From this calculation, the average wage offer corresponds to approximately 24 units. The market wage of 5 is in the first paragraph of Section 3, page 755. This represents a (24-5)*100/5= 380% increase.

5.3) *Average effort response* is not reported. We estimate it using Figure 5, page 767, which shows the average effort by period. From this calculation, the average wage offer corresponds to approximately 3.3 units. The minimum effort of 1 is in the first paragraph of Section 3, page 755. This represents a (3.3-1)*100/1= 230% increase.

5.4) *Elasticity* corresponds to 230%/380%=0.61.

PANEL III: Companion Real-Effort Laboratory Experiments

(1) **Hennig-Schmidt, Sadrieh, and Rockenbach (2010)**

3.1) *Sample sizes* for the control ("L0"), "L10" and "L10 surplus" treatments of 10, 10 and 19 subjects respectively, are in Table 3, columns L0, L10 and L10 surplus, respectively, row Number of typists in page 825.

3.2) *Wage increase* of 0.25 euros for both the "L10" and "L10 surplus" treatments above the baseline of 2.5 euros per each 15-minutes shift are in Table 3, columns L10 and L10surplus, respectively, 2nd work unit wage (15 mins) row in page 825. Because the wage raise only applies to the second hour, the percentage wage raise corresponds to 0.25*100/2.5=10% for both treatments. In the "L10 surplus" treatment, the wage raise is accompanied by information about the employer's surplus as a result of work effort.

---

hovers around 61 units, which is below the average 73 units implied by the average payoff for the firm of 19.4 stated in the text. In the case of an average wage of 61, the percentage wage increase is (61-21)/21*100=190%, resulting in an pay-effort elasticity of 310%/190%=1.63. We report the most conservative elasticity of 1.25, calculated below, in the table.

3.3) *Effort responses* to the wage increase, measured as the number of filled envelopes, are presented in Table 5, page 826. The change from period one to two in the number of filled envelopes in the "L10" treatment corresponds to 50.4-40.5=9.90, while the control "L0" increased output by 41.1-31.1=10 (see Table 5, column L0 and L10, row Output quantity work unit 2 minus row Output quantity work unit 1). The difference 9.90-10=-0.10 corresponds to a -0.10*100/10=-1% increase. This difference is not statistically significant (see Section 3.2, third paragraph in page 827). The change from period one to two in the number of filled envelopes in the "L10 surplus" treatment corresponds to 56.2-43.3=12.9. The difference with the control "L0" corresponds to 12.9-10=2.9, which is a 2.9*100/10=29% increase. The significance of this estimate is not reported.

3.4) *Elasticities.* The elasticity in the "L10" treatment corresponds to -1%10%=-0.1. The elasticity for the "L10 surplus" corresponds to 29%10%=2.9.

E   Online Appendix - Protocols

E.1   Protocol and Wording for Treatments

Students who answer the campus fliers contact the recruiting assistant by phone or email. The recruiting assistant gathers their contact information and availability. Since this is a natural field experiment, no consent forms are signed to enter the employment relationship. Workers coordinate with the project manager (our research assistant) the time and place to perform the job. Each subject works in a different room, in isolation, where rooms are spread across campus to avoid any contamination.

The description and wording for each treatment is as follows:

1) CONTROL

 – Subjects are hired at $12 per hour for the duration of the task—the six hours of work—as advertised. On the first day, subjects are briefly instructed on the very simple bibliographic software before they start working. Immediately after, they execute the agreed two hours of work. On the two subsequent days, subjects work two hours each day as agreed upon recruitment. Finally, subjects are paid the recruiting $12 per hour ($72 total) when they submit their time sheets at the end of the six hours.

2) 67%SURPRISERAISE

– Subjects are hired at \$12 per hour for the duration of the task—the six hours of work—as advertised. On the first day of work, they are briefly instructed on the very simple bibliographic software, but before they start working the agreed two hours, subjects are offered the envelope with the raise raise of \$8 dollars per hour for the first day (\$16 dollars total). They are further told that there will be similar gifts for the next two shifts. On the two subsequent days of work subjects are given the raise before they start their two-hour shift. Finally, subjects are paid the recruiting \$12 per hour (\$72 total) when they submit their time sheets at the end of the six hours.

– Wording:

* At the beginning of day 1: "We have a thank you gift, in the amount of \$8 per hour in addition to the \$12 per hour pay. We will give this gift for the hours you work today and we will also give you the same gift on each of the next two shifts."

* At the beginning of days 2 and 3: "As promised, here is the gift for today".

* At the end of day 3, upon receiving the time sheets. "Thank you for your work. Here is the \$72 payment."

3) 50%-100%RAISE

– This treatment is exactly like 67%SURPRISERAISE, except that the surprising permanent raise amounts to \$6 per hour (instead of \$8 and thus they receive \$12 total instead of \$16) and in the last day (third day) of work, before subjects start the task, they are given a second raiseraise (on top of the first one) in the amount of \$6 per hour of work. In this last shift, the research assistant gives the subject two envelopes: one with \$12 and another with \$12 before they start the task.

– Wording:

* At the beginning of days 1 and 2: same wording as in 67%SURPRISERAISE treatment, but changing the size of the gift.

* At the beginning of day 3: "We have a further thank-you gift in the amount of \$6 per hour in addition to the \$12 per hour pay and the gift of \$6 per hour in previous shifts. Here are the gifts."

* At the end of day 3, upon receiving the time sheets. "Thank you for your work. Here is the \$72 payment."

5) 67%AnticipatedRaise

– This treatment is exactly like 67%SurpriseRaise, except that the research assistant meets subjects exactly one week in advance of the start of work to give them instructions on the program and shows them the time sheets used to pay them the agreed hiring wage of $72.[28] Further, during this extra shift the research assistant shows workers the envelope with the $8 per hour raise (the $16 in the envelope).[29]

– Wording:

  ∗ For the extra initial shift: "We have a thank you gift, in the amount of $8 per hour in addition to the $12 per hour pay. We will give this gift at the beginning of your shift of next week and we will give the same gift on each of the next two shifts. (The research assistant only shows the gift, does not give it to subjects. He is instructed to make it natural and to this end he shows the envelope with the gift on top of the time sheets).

  ∗ At the beginning of days 1, 2 and 3: "As promised, here is the gift for today" (offered again right before the students start the shift).

  ∗ At the end of day 3, upon receiving the time sheets. "Thank you for your work. Here is the $72 payment."

6) 67%PromisedRaise

– This treatment is exactly like SurpriseRaise, except that instead of handing in the gifts immediately before subjects work on the task, the gifts are announced in the first shift, but they are only received at the end of the last shift. That is, subjects receive the raise at the end of the third shift ($48) together with the hiring pay of $72.

– Wording:

  ∗ At the beginning of day 1: "We have a gift in the amount of $8 per hour in addition to the $12 per hour pay. We will give this gift for the hours you work today and we will give the same gift on each of the next two shifts. You will receive it at the end of the last shift."

  ∗ For days 2 and 3: Nothing is said.

---

[28]If the meeting was not possible exactly one week in advance, it was scheduled week and 1 day in advance.

[29]The workers in all there treatment also receive time sheets used to pay them the agreed hiring wage of $72, but on the first day they report to work.

∗ At the end of day 3, upon receiving the time sheets. "Thank you for your work. Here is the $48 gift and the $72 payment."

7) PIECERATE

– Subjects are hired at $12 per hour for the duration of the task—the six hours of work—as advertised. On the first day of work, before they start working the agreed two hours, subjects are informed that, in addition to the $12 per hour agreed upon hiring, they will receive a piece rate for each record they enter.

– The piece rate corresponds to $0 \times x$ if $x < 70$; $0.05 \times x$ if $70 \leqslant x \leqslant 110$; $0.10 \times x$ if $110 < x \leqslant 140$; and $0.20 \times x$ if $x > 140$; where x is the number of records entered on the shift. The piece rate is paid in cash at the end of each shift. Subjects are paid the recruiting $12 per hour when they submit their time sheets at the end of the six hours, as it is customary at the host university where experiments were conducted.

– Wording:

∗ Wording for the communication with subjects right before the start of the first shift, when they are informed about the piece rate: "In addition to the agreed $12 per hour, you will receive a piece rate in each of the three shifts. The piece rate is as follows (research assistant walks subjects through Table 9 below). The payment for the piece rate will be given to you in cash at the end of each shift".

∗ For days 1, 2, 3, when handing in payment for the piece rate at the end of each shift: "You logged XX records during your two-hour shift. This implies that you receive $YY. Here is your payment".

Table 9: Piece Rate Table Shown to Subjects in PIECERATE Treatment

In addition to the $12 per hour in each shift, you will receive an amount for each record imputed, per shift, as follows below.

Number of records	Extra payment per record	Total compensation per two-hour shift
69 or less	$0	$24 for the two-hours of work
Between 70 and 110	$0.05 per record inputted	For example: • If input 70 records, receive an extra 70x0.05=$3.5. (So total compensation per shift is $24+$3.5=$27.5) • If input 110 records, receive an extra 110x0.05=$5.5. (So total compensation per shift is $24+$5.5=$29.5)
Between 111 and 140	$0.1 per record inputted	For example: • If input 111 records, receive an extra 111x0.1=$11.1 (So total compensation per shift is $24+$11.1=$35.1) • If input 140 records, receive an extra 140x0.1=$14 (So total compensation per shift is $24+$14=$38)
141 or more	$0.20 per record inputted	For example: • If input 141 records, receive an extra 141x0.20=$28.2 (So total compensation per shift is $24+$28.2 =$52.2)

E.2 Protocol for the Post-Field Experiment Survey

Next we present the exact protocol used in the survey. Notes to the reader are in corresponding footnotes, which were not part of the protocol.[30]

"Instructions

We will ask you to make decisions on two related situations ("games"). In these games, in addition to your participation fee of \$10, you can earn up to \$15. To this end you will be (anonymously) paired with another undergraduate student from your university who will be the "other player" (or "partner") in these games.[31]

We will start with a brief training period for you to familiarize yourself with these simple games. You will play exactly the same games you will face in the actual decision period, except that you will not receive the payment corresponding to the outcome of the games during this practice period. After you practice playing each game, you will be asked whether you want to ask the research assistant clarifying questions, play the practice game again or whether you want to play the actual games.[32]

Remember you can contact a research assistant to answer any questions you may have about the survey and the payments anytime between 9am and 5pm. The research assistant is available by (email), (Skype) or (phone).[33]

(New screen)

PRACTICE GAMES

Your will not be paid for the outcomes of these two games

Game 1

First, you have to choose between action A and B. The other player, having observed your choice, will also choose between A and B.

---

[30]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.

[31]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.

[32]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.

[33]The existence of a research assistant who would support subjects in the survey was communicated to the subject in the email that invited them to participate in the survey.

The payment you will receive from your choice depends on your choice **and** on your partner's choice, and it is represented in the following diagram[34]:

	Other player chooses A	Other player chooses B
You choose A	You get $4 Other player gets $4	You get $0 Other player gets $7.5
You choose B	You get $7.5 Other player gets $0	You get $1 Other player gets $1

Please examine the payments on the diagram carefully and choose between actions A or B by clicking below:

(button) I choose A
(button) I choose B

Message on screen: "Thank you for your choice"

Message on the next screen (This screen also displays the payment diagram):

"You chose X *(A or B)*.

(1) Suppose your partner plays A

- What would be you payment in this case?

(If payment is correct, display "Yes, that is correct". If subject is incorrect, display "That is not correct, please try again" and display the same question again)

- What would be your partner's payment in this case?

(If payment is correct, display "Yes, that is correct". If subject is incorrect, display "That is not correct, please try again" and display the same question again)

(2) Suppose your partner plays B

- What would be you payment in this case?

(If payment is correct, display "Yes, that is correct". If subject is incorrect, display "That is not correct, please try again" and display the same question again)

---

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

– What would be your partner's payment in this case?

(If payment is correct, display"Yes, that is correct". If subject is incorrect, display"That is not correct, please try again" and display the same question again)

(New screen) Now suppose that you had chosen Y *(Y is A if X was B and Y is B if X was A)*

(3) Suppose your partner plays A

– What would be you payment in this case?

(If payment is correct, display"Yes, that is correct". If subject is incorrect, display"That is not correct, please try again" and display the same question again)

– What would be your partner's payment in this case?

(If payment is correct, display"Yes, that is correct". If subject is incorrect, display"That is not correct, please try again" and display the same question again)

(4) Suppose your partner plays B

– What would be you payment in this case?

(If payment is correct, display"Yes, that is correct". If subject is incorrect, display"That is not correct, please try again" and display the same question again)

– What would be your partner's payment in this case?

(If payment is correct, display"Yes, that is correct". If subject is incorrect, display"That is not correct, please try again" and display the same question again)

(New screen) What would you like to do?

  a. Ask the research assistant clarifying questions
  b. Play this practice game again
  c. Continue to the second game of the practice shift

If subject selects a) "You can contact a research assistant (Name) by calling (Phone) or by calling the Skype id [···]".
If subject selects b), repeat the game.
If subject selects c) proceed to Game 2 below.

Game 2

Suppose now that instead of choosing first, you will choose second. That is, your partner will choose between A or B and having observed her/his choice, you will choose between A and B. The payment you will receive is represented in the following diagram, which is the same diagram, with the same amounts, as that in the previous games.

	Other player chooses A	Other player chooses B
You choose A	You get $4 Other player gets $4	You get $0 Other player gets $7.5
You choose B	You get $7.5 Other player gets $0	You get $1 Other player gets $1

Please examine the payments on the diagram carefully and choose between actions A or B by clicking below:

If my partner chooses A:

(button) I choose A
(button) I choose B

If my partner chooses B:

(button) I choose A
(button) I choose B

Message on the next screen (This screen also displays the payment diagram):

(1) "You chose X *(A or B)* if your partner chooses A.

– What would be you payment in this case?

(If payment is correct, display "Yes, that is correct". If subject is incorrect, display "That is not correct, please try again" and display the same question again)

– What would be your partner's payment in this case?

(If payment is correct, display "Yes, that is correct". If subject is incorrect, display "That is not correct, please try again" and display the same question again)

(2) "You chose X *(A or B)* if your partner chooses B.

– What would be you payment in this case?

(If payment is correct, display "Yes, that is correct". If subject is incorrect, display "That is not correct, please try again" and display the same question again)

    – What would be your partner's payment in this case?

(If payment is correct, display "Yes, that is correct". If subject is incorrect, display "That is not correct, please try again" and display the same question again)

(3) "Suppose you had chosen Y if your partner had chosen A. *(Y is A if X was B or Y is B if X was A)*

    – What would be you payment in this case?

(If payment is correct, display "Yes, that is correct". If subject is incorrect, display "That is not correct, please try again" and display the same question again)

    – What would be your partner's payment in this case?

(If payment is correct, display "Yes, that is correct". If subject is incorrect, display "That is not correct, please try again" and display the same question again)

(4) "Suppose you had chosen Y if your partner had chosen B. *(Y is A if X was B or Y is B if X was A)*

    – What would be you payment in this case?

(If payment is correct, display "Yes, that is correct". If subject is incorrect, display "That is not correct, please try again" and display the same question again)

    – What would be your partner's payment in this case?

(If payment is correct, display "Yes, that is correct". If subject is incorrect, display "That is not correct, please try again" and display the same question again)

(New screen) What would you like to do?

    a. Ask the research assistant clarifying questions
    b. Play this practice game again
    c. Continue to the second game of the practice session

If subject selects a) "You can contact a research assistant (Name) by calling (Phone) or by calling the Skype id [⋯]".

If subject selects b), return to the initial screen "PRACTICE GAMES"

If the subject selects c), go to new screen with the "ACTUAL GAMES".

(New screen)
ACTUAL GAMES

Now you will make your actual decisions. The resulting monetary outcomes of the games and gambles will be added to your Amazon gift card with your participation fee. To play the game you have now been anonymously paired with another undergraduate student from your university.

Game 1

First, you have to choose between action A and B. The other player, having observed your choice, will also choose between A and B.

The payment you will receive from your choice depends on your choice **and** on your partner's choice, and it is represented in the following diagram:

	Other player chooses A	Other player chooses B
You choose A	You get $4 Other player gets $4	You get $0 Other player gets $7.5
You choose B	You get $7.5 Other player gets $0	You get $1 Other player gets $1

Please examine the payments on the diagram carefully and choose between actions A or B by clicking below:

(button) I choose A
(button) I choose B

Message on screen: "Thank you for your choice.You will know the outcome of the game once you play the second game."

Game 2

Suppose now that instead of choosing first, you will choose second. That is, your partner will choose between A or B and having observed her/his choice, you will choose between A and B.

The payment you will receive is represented in the following diagram, which is the same diagram, with the same amounts, as that in the previous games.

	Other player chooses A	Other player chooses B
You choose A	You get $4 Other player gets $4	You get $0 Other player gets $7.5
You choose B	You get $7.5 Other player gets $0	You get $1 Other player gets $1

Please examine the payments on the diagram carefully and choose between actions A or B by clicking below:

If my partner chooses A:

(button) I choose A
(button) I choose B

If my partner chooses B:

(button) I choose A
(button) I choose B

(New screen)

OUTCOME OF GAMES AND PAYMENTS

Outcome of Game 1:

You chose X
Your partner chose XX
Therefore, you won XX

Outcome of Game 2:

You chose X
Your partner chose XX
Therefore, you won XX

Closing Window: Farewell Message

"Thank you for participating in this survey. You will receive a payment of $XX in addition to your participation fee. The payment is being processed now. You will receive an email within the next hour with an electronic Amazon gift card. Please contact the research assistant (Name) by calling (Phone) or by calling the Skype id [⋯] if you have any concerns."