A Additional Figures and Tables

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

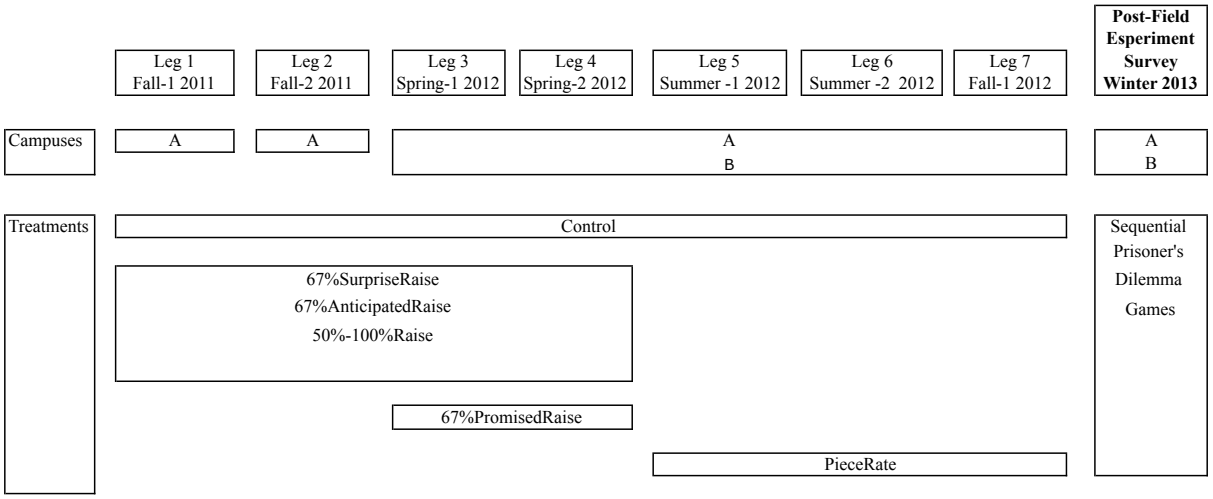

Table A.1: Kruskall-Wallis Test for Equality of Distributions in the 67%SurpriseRaise, 67%AnticipatedRaise and 67%PromisedRaise Treatments

	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=11, N_2=5, N_3=12)	0.41 (N_1=11, N_2=5, N_3=11)	0.14 (N_1=11, 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=11, N_2=7, N_3=12)	0.17 (N_1=11, N_2=7, N_3=11)	0.04 (N_1=11, 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.

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

	Average across all worker shifts	Shift One	Shift Two	Shift Three	SD	Min	Max
	(1)	(2)	(3)	(4)	(5)	(6)	(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	40	39			
50%-100%Raise	19,343	18,789	20,444	18,397	5,141	11,475	30,935
N (worker-shifts)	73	30	28	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: Differences Between the Treatments and the CONTROL on Characters, Records and Correct Words Inputted

Dependent Variable	ln(charaters inputted per subject)			ln(records inputted per subject)			ln(correct words inputted per subject)		
	Within campus X leg X shift			Within campus X leg X shift			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)
	[0.26/0.43]	[0.62/0.59]	[0.71/0.57]	[0.37/0.49]	[0.73/0.55]	[0.84/0.70]	[0.26/0.44]	[0.50/0.49]	[0.78/0.64]
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)
	[0.37/0.49]	[0.70/0.53]	[0.67/0.62]	[0.49/0.49]	[0.65/0.55]	[0.68/0.64]	[0.56/0.53]	[0.65/0.52]	[0.58/0.60]
PieceRate	0.07	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)
	[0.22/0.38]	[0.17/0.30]	[0.08/0.10]	[0.38/0.48]	[0.27/0.43]	[0.08/0.11]	[0.44/0.50]	[0.36/0.48]	[0.15/0.27]
Constant	9.79	9.76	9.74	4.37	4.35	4.35	7.54	7.56	7.54
	(0.07)***	(0.03)***	(0.04)***	(0.04)***	(0.04)***	(0.04)***	(0.07)***	(0.04)***	(0.04)***
CampusXlegXshift fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared		0.38			0.48			0.40	
Number of subjects	192	188	159	194	189	159	175	168	154
Number of subjectXsession observations		539			542			497	
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	2,064

Notes: Panel A, columns (1)-(3) replicate the prior analysis in Table 6, columns (7)-(9), from specification 1. Columns (4)-(6) replicate this analysis but where the outcome variable is the natural log of the number of records inputted by a subject. 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., their titles' length may differ). The sample for the records analysis has three more observations than the one for characters (542 versus 539) because the software in three instances recorded the number of records entered but not the characters. Columns (7)-(9) replicate the analysis in specification 1, but where the outcome variable is the natural log of correct words inputted per subject. The number of observations for this measure is slightly smaller at 497 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 from the regression are in parentheses and clustered by worker (*Significant at the 10% level, **Significant at the 5% level, ***Significant at the 1% level). All tests are one-tailed. The square brackets contain the [p-values under regression/p-values under MHT]. The MHT procedure in this table corrects the p-values not only for the test of multiple conditions (fixed-wage-raise and PIECERATE conditions) but also multiple outcomes (ln(characters), ln(records) and ln(correct words)).

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	34	0.76	2	9	11	0.32	14	0.41

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

				Prosocial in Treatments					
	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)
Total Observations	539			250			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%.

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, we also used the PieceRate condition to help us assess the power of our gift exchange test. As documented in the main text, the PieceRate detected statistically significant effort increases despite its having a smaller sample than wage-raise treatments, which 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. We now describe the power calculations.

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 size—the increase in effort under the alternative hypothesis—corresponds to:[30]

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

  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:

[30] 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.

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

After plugging the parameters shown above, power depends on the sample sizes for the control and 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.[31] This suggests that the lack of detection

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

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 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 percent-

ages 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)**

Gneezy and List (2006) contains two task. In the first, students were hired 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.). In the second, 23 students were 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.

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 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)**

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)).

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)**

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 raise of 40% and information that peers were only receiving a 10% raise.In this paper. The 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))

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)**

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.

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)**

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.

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 en-

tered. 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 (2014)**

In the within-subject test in Cohn, Fehr, and Goette (2014) 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. References 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 (2015)**

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.[32] Only the online version of this paper is available so no pages are cited.

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

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

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 1/3=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 "Wage=3" corresponds to the baseline category. Column (1) shows the difference between this baseline and the "Wage=3+1" treatment using the number of completed and correct CAPTCHAs as the dependent variable. The baseline productivity of the "Wage=3" treatment is captured by the constant term amounting to 792.1, while the coefficient for "Wage=3+1" corresponds to 146.8, which is significant using robust standard errors (see Table 2, Column (1). The effort increase for the full sample thus corresponds to 146.4*100/(792.1)=18%.
* The effort response for the experienced oDesk workers comes from Table 3, Panel A. The productivity in the "Wage=3+1" treatment for experienced workers is 973 records whereas it is 773.1 for the baseline "Wage=3". 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, panel A. The effort increase for the unexperienced sample corresponds thus to (861.8-834.4)*100/(834.4)=3%.

7.4) *Elasticities.* The elasticity for the full sample corresponds to 18%/33%=0.55. For the experienced workers it corresponds to 26%/33%=0.79 and for the unexperienced workers it corresponds to 3%/33%=0.09.

PANEL II: Most Cited Laboratory Studies

The most cited laboratory studies on gift exchange rely on laboratory games with 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. We now detail each study.

(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)**[33]

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

[33]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.

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.

- *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.

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.

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 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.

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.

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.[34]

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.

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

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)** 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.[35]

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 L10 surplus, 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.

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

[35]The statistical significance of this result was not reported.

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%Raise

In this treatment, workers were offered a 67% wage raise versus the contract wage of $12 per hour: they received a raise to $20 per hour (i.e., an additional $8 per hour) for the duration of the contract.

The 67%Raise treatment aggregates three subtreatments varying the timing of the information about the raise (immediately before or one week before the first shift)

and when the raise was paid (at the start or at the end of each shift): 67%SurpriseRaise, 67%AnticipatedRaise and 67%PromisedRaise. Given that the distribution of outcomes for these three subtreatments was not statistically different we aggregated them into a single condition 67%Raise.

2.A) 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."

2.B) 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.[36] Further, during this extra shift the research assistant shows workers the envelope with the $8 per hour raise (the $16 in the envelope).[37]
- Wording:

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

[37] 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.

* 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."

2.C) 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.
  * 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."

3) 50%-100%Raise

- In this treatment subjects receive a raise that mounts to $6 per hour in the first and second days of work (i.e., shifts one and two, respectively) to $18 per hour. In the third day before subjects start the task, they are given a further additional raise of $6 per hour, to $24 per hour.
- 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."

4) 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.[38]

"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.[39]

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.[40]

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).[41]

(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.

[38]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.

[39]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.

[40]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.

[41]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[42]:

	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?

[42]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).

(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) 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."