Triggering reciprocity in a principal-agent game

Debing Ni Xiaoyuan Wang* Kaiming Zheng

School of Management and Economics

University of Electronic Science and Technology of China

Preliminary

Abstract

This paper considers a modified principal-agent environment, where the agent makes costly work effort in exchange for wage, and the principal chooses a combination of fixed rate and piece rate transfer to the agent in hope for higher effort. Because the principal in our environment is liability constrained and because the production is subjected to great uncertainty, she suffers from significant efficiency loss with self-regarding agents. However, the agent may be potentially reciprocal. In our specific theoretic setting, by triggering some agents' reciprocity preference, the principal may achieve a better outcome. Is it possible that principals reward agents' trustworthiness and "trustworthy" agents improve labor market efficiency? We test the modified principal-agent model using a lab experiment. We find that, compared with a market with self-regarding agents, the market witnesses significant higher offers of fixed rate wage. Estimations on agents' effort choice confirms the effects of both positive and negative reciprocity. It reveals that negative reciprocity has a greater impact on efforts and thus on principals' wage offers. The estimated reciprocity reference point decreases over the experiment in the social information sessions, but not in the individual information sessions. Only at later rounds, principals' offer are correlated with the trustworthiness level of the agents.

Keywords: Reciprocity, Moral hazard, Trustworthiness, Contract

*Corresponding author. Please contact at wangxy@uestc.edu.cn.

1 Introduction

It is well documented that real world labor contracts often use much weaker monetary incentives than conventional principal agent theory dictates (Englmaier and Leider, 2012, Lemieux et al., 2009). In the standard theory, each principal only observes the noisy production output as a proxy for an agent's effort, and it is essential for them to tie compensation to output, in order to avoid moral hazard issues. In reality, alternative sources of incentives may coexist (Fehr and Schmidt, 1999, Rabin, 1993) and too strong financial incentives have been reported to "crowd out" intrinsic ones in specific settings (Ariely et al., 2009, Gneezy and Rustichini, 2000). For example, education is sometimes viewed as a "moral" work, where performance is easy to manipulate and difficult to contract on. Financial incentives can lead to lower exam standards, inflated test scores and credit hours. In contrast, alternative incentives, if appropriately designed, may lead to more efficient outcomes. In this paper, we focus on agents' reciprocity preference–a topic that has received great attention by both theorists (Dufwenberg and Kirchsteiger, 2004, Englmaier and Leider, 2012, Falk and Fischbacher, 2006, Rabin, 1993) and experimentalists(Anderhub et al., 2002, Fehr and Gächter, 2000, Fehr et al., 1993). We experimentally investigate an economic setting, where principals are liability constrained, so that financial incentives cannot achieve the social optimal outcome. However, principals have good chance of triggering agents' reciprocity preference, because they have access to the information of "trustworthiness" measure of the agents, and can customize a flexible offer.

In our specific theoretical setting, principals are allowed to select a combination of both fixed rate and piece rate. Without reciprocal preference, principals' optimal strategy is to use piece rate and only the minimum level of fixed rate. However, they may use a contract with a high fixed rate to trigger agents' reciprocity preference and achieve a better economic return. Our first research question is about agents' preference. We directly model agents' reciprocity preference, estimate the key parameters governing the reciprocity preference, and compare with the benchmark model lacking such preference component. Our second research question concerns whether principals make use of the trustworthiness information elicited before the contract game. We investigate if there is any aggregate difference, compared with the benchmark model with self-regarding agents. In addition, we test if principals make personalized offers, given individual level information on trustworthiness.

Our theoretical setting reflects features in real organizational relations. Anecdotally, firms use more and more screening methods, which are not all related to candidates' abilities[1]. The costly resource firms spent reflect the fact that two equally competent candidates may have different willingness to contribute to the firm. Such willingness to focusing on principals' payoff should also be priced in when considering an optimal contract. Therefore, it becomes reasonable for firms to form certain beliefs of the subjects' reciprocity preference and how to trigger such preference.

To investigate such triggering mechanism, we adopt a two stage experimental design. At the first stage, each agent's trustworthiness is elicited using the Trust game (Berg et al., 1995). The percentage returned in the Trust game has been well regarded as a practical measure of trustworthiness (Glaeser et al., 2000). We simultaneously elicit the principal's transfer and the agent's all contingent choices when they face possible transfers. [2] No feedback is provided at this stage and participants do not know the detailed second stage instruction until the beginning of the second stage. At the beginning of the second stage, principals are provided with agents' trustworthiness information so that they form appro-

[1] A recent research report by Glassdoor (Chamberlain, 2015) found both "explained" and "unexplained" increase in average hiring time for US job seekers. Interview methods that has became more popular include background checks (25% in 2010 and 42% in 2014) and personality tests (12% in 2010 and 18% in 2014)–all contribute significantly to the hiring length.

[2] The design is commonly referred as the strategy method (Selten, 1967). See also Casari and Cason (2009), Solnick (2007) for trust game experiments involving the strategy method.

priate beliefs on agents' reciprocity preference. In half of the sessions, each principal can access the trustworthiness measure as the group average ("social information sessions"). In the other half of the sessions, each principal can observe her paired agent's individual decisions in the first stage ("individual information sessions"). In the second stage, each principal and each agent play 15 rounds of randomly re-matched contract game, where the principal chooses a combination of a fixed rate and a piece rate, and the agent chooses an effort level.

We have conducted four social information sessions and four individual information sessions. Our results show that agents responded positively to the fixed wage, and a significant potion of principals use higher-than-minimum fixed rate. However, agents' negative reciprocity has a stronger impact, compared with the positive one in both information conditions. Principals, under such preference, use high fixed rate to reduce the impact of negative reciprocity, instead of rewarding the other. The estimated reciprocity point parameter declines over the experiment in the social information sessions, so that the triggering of negative (positive) reciprocity becomes harder (easier). However it is not reduced in the individual information sessions. The information on trustworthiness—originally designed to proxy positive reciprocity—is only correlated with principals' offers at later rounds.

2 Literature

The trust game introduced by Berg et al. (1995) is one of the most frequently replicated measure of trust and trustworthiness.[3] The game involves a two-person-two-step game between a trustor, who makes an investment, and a trustee, who decides how to split a

[3]For example, Johnson and Mislin (2011) collect 162 replications of the trust game, involving 23,000 participants.

return from such investment. In such a game, to trust is risky, because trustee's behavior is not contractually enforced. to reciprocate also must go against the trustee's self-interest, because she becomes a dictator at the second stage (where she chooses to transfer back any amount from the return) of the game. The typical measure of trust and trustworthiness are based on the percentage transferred and returned in the game. Camerer (2003) summarizes that trustees behavior is motivated by moral obligations, while also offers additional evidence showing that altruism (Cox, 2004, Dufwenberg and Gneezy, 2000) may play a more important role. Di Bartolomeo and Papa (2016) extend the later argument and find initial inequality affects agents' motivations. In our contract game, we provide each principal with the paired agent's contingent choices in the trust game as a measure of trustworthiness of the agent and allow them to update their beliefs on the other party's altruistic and reciprocating behavior.

Our second stage game is related to the gift exchange experiments (Fehr and Gächter, 2000, Fehr et al., 1993). In the typical gift exchange game, each principal (firm) offers a fixed rate offer (sometimes with expected effort) and then the agent (worker) who accepts it chooses (costly) effort. The principal makes a risky choice of the fixed wage, because once hired, she cannot control the agent's effort. Fehr and Gächter (2000) demonstrate a positive correlation between wage rents and the actual effort chosen; yet, in another treatment, with explicit disincentives for shirking, agents exert less effort and their responses do not reciprocally correlated with job rents. The comparison between treatments shows that incentives may undermine reciprocity. [4] Gift exchange experiments show that reciprocity seems to work well in the lab, however, in those games, the generosity of the principals (firms) can be easily seen as intentional. In a field setting, such link may be difficult

[4]In the field, financial incentives may not work well for the similar reason. For example, Gneezy and Rustichini (2000), both in the lab and in the field, demonstrate a U-shaped relationship between piece rate and effort.

to establish. For example, Gneezy and List (2006) suggest that work effort in the first few hours of the gift exchange is higher; after that, however, there is little difference in effort. Furthermore, the total yield in the gift treatment is less than that using market clearing wage. Negative reciprocity seems to be more effective in a field setting. Kube et al. (2006) find that wage reduction has a significant and lasting negative (asymmetric) impact on efforts. Krueger and Mas (2004) show labor strife as the main cause of the defective tires manufactured by Firestone in the 1990s. Bewley (1999) document evidence that managers expect potential retributions from workers. However, robust evidence of social preferences in real workplace remains scarce (Charness and Kuhn, 2011). Jun et al. (2015) argued that non-monetary incentives play a significant role in a fixed rate contract, and it can be socially efficient when piece rate contract is too costly to implement.

One way to look at the lab-field discrepancy in generating reciprocity is to introduce more realistic field features. For example, by endogenizing both piece rate with an incentive device in a more standard principal-agent setting, Anderhub et al (2002) test a principal-agent model where the principal can choose a combination of a piece rate wage and a return share. They find a high degree of incentive-compatible behavior. In addition, a small proportion of both positive and negative reciprocity coexist with the incentive device. In our theoretical setting, the principals choose a combination of fixed and piece rate compensation, however, the wage is constrained to be higher than a positive minimum amount so that principals cannot properly incentivizing agents. Principals in theory can form belief on agents' reciprocity preference and send an even higher fixed wage. Our specification of the agent utility is based on Englmaier and Leider (2012); we consider a specific case where fixed wage can trigger both positive and negative reciprocity, governed by an unobserved reference point. This reference point of "moral obligation" is crucial when investigating the generation of reciprocity under different information conditions.

Finally, in order to incorporate the real life feature of production shock and to give the triggering process a better chance, we assume the production is uncertain. Rubin and Sheremeta (2015) consider a gift exchange setting in the lab, and find that the introduction of shocks reduces wages and effort. In our game, uncertainty serves another purpose: principals can use fixed rate to transfer the risk to their side as a way to triggering positive reciprocity. We conjecture that, with uncertainty considered, triggering agents' positive reciprocity preference may become an appealing option for both side.

3 The Experimental Design

3.1 The game

The experiment consists of a one-shot trust game (with strategy method) and a finitely repeated contract game between two players. In the first round of the game (Stage One), a principal (Role A) and an agent (Role B) are randomly paired and agents' contingent choices for all possible transfers in the trust game are recorded. In the next fifteen rounds (Stage Two), principals and agents are randomly rematched with fixed roles. Each Principal has the opportunity to provide a production task with linear wage schedule and upon acceptance, each agent can choose her costly effort and receive payment correspondingly. In this game, principals are liability constrained and can only use a fixed rate higher than a minimum and a non-negative piece rate. However, information on agents' performance in the first round Trust game is available and may potentially affect the decisions of the game. The detail of the two stage games are as follows:

The Trust game: The principal is initially endowed with 20 tokens, while the agent has zero tokens. The principal decides the number of tokens transferred to the agent and the agent decides, for all scenarios, her preferred return. They simultaneously decide on their own strategies and no feedback will be provided until the end of the game. The asymmetric initial endowment is chosen to reflect potential principal-agent interactions in modern organizations.

The Contract game: At the beginning of each repeated contract game round, the principal first learns the agent's performance in the first round Trust game. In the social information sessions, all agents' group average of the contingent choices are reported; in the individual information sessions, each randomly paired agent's performance in the trust game is reported. If the principal agrees to provide a work contract in each of the 15 rounds, she chooses a fixed rate higher than a minimum amount($\omega >= \omega_0$) and a piece rate (β) for the work contract for each round. Then, the agent chooses to reject or accept the contract offer and upon acceptance, she selects a work effort level e.

The agent's work effort e and a random shock jointly determine production outcome. In half of the cases, production is determined by $q(e, low) = 5 \times e$; in the rest cases, production is determined by $q(e, high) = 15 \times e$. For each principal, the marginal benefit for production satisfies $\alpha = 80$. Therefore, the principal's monetary payoff can be written as,

$$u^P(e; \omega(q)) = \alpha * q(e, \epsilon) - (\omega + \beta q(e, \epsilon))$$

$$q(e, \epsilon) = e * \epsilon, \ \epsilon = 5 \text{ or } 15 \text{ with } 50\% \text{ each}$$

For agents, the total cost of work effort is $c(e) = 50 \times e^2$. Therefore, the agent's monetary payoff can be written as,

$$u^A(e; w(q)) = \omega + \beta q(e, \epsilon) - c(e)$$

The choices of the parameters guarantee that agents are not binded by the participation constraint.

To ensure that the numbers are easy for the agents to digest, in the experiment, we use discretized strategy space for principals and agents. The choices of the fixed rate (ω) and the piece rate (β) offers obey the following restrictions:

$$\omega \in \{100, 200, ..., 1000\}$$

$$\beta \in \{0, 10, ..., 100\}$$

The agents choose 10 levels of effort: $e \in \{1, 2, ..., 10\}$.

3.2 Game-theoretic solution

When a risk neutral principal and a risk neutral agent perfectly align their goals, the social optimal of the organization can be given by the following maximization problem:

$$Max_e \; \pi^{social} = \alpha E q(e, \epsilon) - c(e)$$

Therefore, the social optimal effort level is 8 and the total expected surplus created is $80 * 80 - 50 * 8 * 8 = 3200$. However, with self-regarding and risk neutral agents, her best response must satisfy:

$$e \in \underset{e}{\operatorname{argmax}} \, E_\epsilon[u^A(e; \omega, \beta)]$$

$$\text{or in this case } e* = \beta/10$$

The above condition shows that self-regarding agents only respond to piece rate wage and are not affected by the fixed rate. They will always prefer to take the job and the participation constraint is never binding. For example, at the boundary, with minimum wage of 100 fixed rate and 0 piece rate wage, the agent is better off by choosing effort level one, and receive a positive payoff of $100 - 50 = 50$. In this case, principals' expected payoff is 700. With the maximum piece rate of 100 and another 100 fixed rate, the agent's optimal effort level is 10 and receive a expected payoff of $100 + 100 * 100 - 50 * 10^2 = 5100$. However, principals should never choose a piece rate higher than the marginal benefit of production ($\alpha = 80$).

Each principal takes account of the above "incentive compatibility constraint", as well as the fact that it is not practical to impose a fixed rate lower than ω_0. she solves the following problem:

$$\begin{aligned} Max_{\omega,\beta}\ Eu^P &= \alpha * Eq(e,\epsilon) - (\omega + \beta Eq(e,\epsilon)) \\ & e \in \underset{e}{\operatorname{argmax}} E_\epsilon[u^A(e;\omega,\beta)] \\ & E_\epsilon[u^A(e;\omega,\beta)] \geq 0 \\ & \omega >= \omega_0 = 100;\ \beta >= 0 \end{aligned}$$

The result suggests that it is optimal to use the lowest possible fixed rate, $\omega_0 = 100$, and a piece rate of 40, which induces an effort level 4 from the agents and generates a surplus of $80 * 40 - 50 * 4 * 4 = 2400$. Ideally, the principal and the agent can achieve a better outcome. For example, the principal can make a risky transfer through the fixed rate, and trust that the agent, who is triggered by the kind gesture, would reciprocate the kindness. In order to incorporate this component in to the model, we modify the benchmark case, so that after the agent's reciprocity preference is triggered, she will also care about the principal's utility.

For the reciprocal agent case, assume each agent's payoff has the following form:

$$u^{AO} = \omega + \beta Eq(e,\epsilon) + \lambda(\omega - K) \cdot \pi^P - 50e^2 \tag{1}$$

The component $\lambda(\omega - K)$ governs the triggering process: with a positive λ, it is assumed that if the fixed rate is greater (less) than a reference value K, the agent will positively (negatively) reciprocate the principal's utility. [5] When $\lambda = 0$, the utility function degenerates to the self-regarding case. The specification flexibly captures both positive and negative reciprocity with only two additional parameters. Under such preference, agents' effort depends on both piece rate and fixed rate:

$$e_r^* = \lambda(K - \omega) \cdot \alpha/10 + (1 - \lambda(K - \omega)) \cdot \beta/10$$

The principal who expects such reciprocal behavior will take account of this in her optimization process. The fixed wage now plays an additional role of triggering positive and negative reciprocity and thus the principal may find it's optimal to set a much higher fixed rate. For example, if only considering the interior solution and $\lambda^{-1/2}$ is less than α, the principal's optimal strategy can be written as:

$$\omega^* = K + (\lambda^{-1} - \alpha/2\lambda^{-1/2}) \tag{2}$$

$$\beta^* = \alpha - \lambda^{-1/2} \tag{3}$$

The optimal choice depends on both unobserved parameters K and λ for each individual. Even if the agent is potentially not self-regarding, it can be costly to trigger positive reciprocity, with a significantly higher reservation value, K.

In the above game, there is another source of uncertainty besides agents' level of trustworthiness: the production shock. With the existence of production shock, principals can choose to transfer such risk to their own side and demonstrate "kindness" toward

[5] The positive value of λ represents the relative weights agents put on the principal. The whole term $\lambda(\omega - K)$ is likely to be small, because a value greater than one would indicate the agent treat the principal's payoff as more important than her own. In the later estimation part, we impose no restriction on both parameters and allow the data to provide us further information on agents' behavior.

agents. However, agents' risk aversion itself may change the response of both sides and confound with the reciprocal behavior. In order to rule out this possibility, we calculate equilibrium strategies for both risk neutral and risk averse agents. Table 1 reports the best responses for both principal and agent. The upper panel reports the benchmark case: in the top-left table, given principals' fixed wage (row) and piece rate (column), agents' choice of effort level only depends on piece rate. The top-right table lists the expected payoff of the principal, when facing such agents. The principal's best response is to choose the lowest fixed rate, 100, and a piece rate of 40. The payoff is relatively symmetric around the optimal response. The lower panel lists the best responses with a highly risk-averse agent[6]. The new equilibrium under the risk-averse agent has moved to the minimum fixed rate and a piece rate of 30. Even with the extremely risk-averse agents, the optimal fixed rate is lower than or equal to 300. In contrast, a reciprocity-based transfer is likely to be a lot higher, because it has to provide substitute for the piece rate. [7]

3.3 Discussions of the design

Making sure that all participants understand the game is crucial. In our game, agents' decisions involve a discrete choice among 10 effort levels, while principals strategy space is 10 fixed rates $\times$ 11 piece rates. Backward induction for each strategy creates significant calculation burden. In order to make them more focused on the task, we provide principals with 10 reference tables. Each table lists the principal's expected payoff for each wage strategy, conditional on the agents 10 levels of effort. Appendix Table A.1 illustrates one of the ten tables for example; a principal may choose a fixed rate of 300 and a piece rate of 30, believing that the agent facing the wage will provide an effort level of 4. Agents

[6]The result of the table is based on the standard loss aversion utility function (Tversky and Kahneman, 1992), namely $v(x) = x^{\sigma}$, $x >= 0$ and $v(x) = -\theta(-x)^{\sigma}$, $x < 0$, where σ is set to 0.7 and θ is set to 2

[7]In Table 6, we return to this topic and provide simulated equilibrium strategies for reciprocal agents.

Table 1: **Simulated Responses for Risk Neutral and Risk Averse Agents** Each row represents the fixed rate level and Each column represents the piece rate level. The left tables list agents' optimal effort levels and the right tables list principals' expected payoffs correspondingly. The equilibrium effort and payoff are in bold.

Best Response with Risk Neutral and Self-regarding Agents

Agents' Optimal Effort Level

	0	10	20	30	40	50	60	70	80
100	1	1	2	3	**4**	5	6	7	8
200	1	1	2	3	4	5	6	7	8
300	1	1	2	3	4	5	6	7	8
400	1	1	2	3	4	5	6	7	8
500	1	1	2	3	4	5	6	7	8
600	1	1	2	3	4	5	6	7	8
700	1	1	2	3	4	5	6	7	8
800	1	1	2	3	4	5	6	7	8
900	1	1	2	3	4	5	6	7	8
1000	1	1	2	3	4	5	6	7	8

Principals' Expected Payoff

	0	10	20	30	40	50	60	70	80
100	700	600	1100	1400	**1500**	1400	1100	600	-100
200	600	500	1000	1300	1400	1300	1000	500	-200
300	500	400	900	1200	1300	1200	900	400	-300
400	400	300	800	1100	1200	1100	800	300	-400
500	300	200	700	1000	1100	1000	700	200	-500
600	200	100	600	900	1000	900	600	100	-600
700	100	0	500	800	900	800	500	0	-700
800	0	-100	400	700	800	700	400	-100	-800
900	-100	-200	300	600	700	600	300	-200	-900
1000	-200	-300	200	500	600	500	200	-300	-1000

Best Response with Risk Averse and Self-regarding Agents

Agents' Optimal Effort Level

	0	10	20	30	40	50	60	70	80
100	1	1	2	**3**	3	4	5	6	7
200	1	1	2	3	3	4	5	6	7
300	1	1	2	3	4	4	5	6	7
400	1	1	2	3	4	4	5	6	7
500	1	1	2	3	4	4	5	6	7
600	1	1	2	3	4	4	5	6	7
700	1	1	2	3	4	5	5	6	7
800	1	1	2	3	4	5	5	6	7
900	1	1	2	3	4	5	5	6	7
1000	1	1	2	3	4	5	5	6	7

Principals' Expected Payoff

	0	10	20	30	40	50	60	70	80
100	700	600	1100	**1400**	1100	1100	900	500	-100
200	600	500	1000	1300	1000	1000	800	400	-200
300	500	400	900	1200	1300	900	700	300	-300
400	400	300	800	1100	1200	800	600	200	-400
500	300	200	700	1000	1100	700	500	100	-500
600	200	100	600	900	1000	600	400	0	-600
700	100	0	500	800	900	800	300	-100	-700
800	0	-100	400	700	800	700	200	-200	-800
900	-100	-200	300	600	700	600	100	-300	-900
1000	-200	-300	200	500	600	500	0	-400	-1000

will also receive a table of the production cost and possible productions conditional on each effort level. The documents are distributed when participants read the second stage instructions.

Appendix Figure A.1 and A.2 demonstrate the screen shots of the game. In Figure A.1, Role A players are reminded of the summarized instruction of the game on the left panel; they are also informed about the social aggregate reciprocity level or the individual reciprocity level of their own pair on the right panel. All contingent choices elicited in the first round are reported. In Figure A.2, Role B players are reminded of their summarized instruction as well as the paired Role A players' strategies. After each Role B player's choice, the production and both role's earnings will be reported on each one's screen.

After the 15 rounds of the Stage Two had elapsed, subjects filled out a questionnaire consisting of demographics, a risk-preference test (Eckel and Grossman, 2008) and a Cognitive Reflection Test (Frederick, 2005). They were also requested to describe the rationale behind their strategy choices. Participants were paid for both stages of the game. Stage One and one randomly selected round in Stage Two were used for each one's payment. The conversion rate was 1.5 Experimental Currency Units (ECU):1 Chinese yuan and 150 ECU:1 Chinese yuan for the corresponding Stage One and Two.

The experiment was conducted at the University of Electronic Science and Technology of China. All subjects were UESTC undergraduate or master students recruited using a Wechat-based subject recruitment system(Keyan Assistant[8]). 106 subjects from eight sessions have participated in the experiments; 52 of them (from four sessions) are under the social information condition.[9] The average earning of the experiments were 36 Chinese Yuan and the typical 10 or 12 subject session took about one hour and a half. The

[8]https://www.ancademy.org/

[9]There is one larger session of 18 or 20 subjects for each condition. The rest six sessions(three in each condition) contains 10 or 12 subjects.

experiment was programmed and conducted with the software Z-tree (Fischbacher, 2007).

4 Results

4.1 The Trust game

In the Trust game, we have 53 subjects' return decisions conditional on all (integer) transfers. Among them, 77.4% have non-decreasing absolute return, indicating potential reciprocal preference. The average amount transferred in the trust game is 11.4 (out of 20) and the average percentage returned is 30.7%. Both the average amount transferred and the average percentage returned are not statistically different between treatments (P-value $\approx$ 0.200 and 0.177 correspondingly [10]). We approximate trustworthiness using the simple average over all transfer scenarios for each person. Based on the measures of trustworthiness, we investigate the contract game in the subsequent sub-sections. There are 761 valid observations recorded in the contract game, which represents 95.7%[11] of the total observations (53 subjects $\times$ 15 rounds).

4.2 Behavior of Principals

For describing principals' behavior, we categorize a fixed rate offer higher or equal than 300 as evidence against our theoretical benchmark and report the summary statistics in Table 2. In the social information sessions, the most frequently chosen strategy is the lowest fixed rate with a piece rate of 40, which is the optimal strategy under risk neutrality. 40.3% of the time, principals choose to use the lowest fixed rate. However, 38.4% of the time, principals use a fixed rate too high to be explained by risk aversion. This deviation can be either due to a noisy play or a tendency to rewarding trustworthiness. On the

[10]Without further notification, the std. err. of each statistic was all clustered at the individual level

[11]In few occasions, principals refused to provide a wage contract and/ or agents refused to accept one.

aggregate level, principals also use a lower piece rate, compared with the risk neutral benchmark (T - test, mean ≈ 34.9, p - value ≈ 0.0645). Overtime, the use of piece rate is significantly increasing with a rate of 0.382 (T - test, p - value ≈ 0.013). The changing rate of the fixed rate over time is negative and insignificant (T - test, p - value ≈ 0.190).

Table 2: **Principals' Offer Choices**

Social Info; total obs: 380 (% is listed below)				Ind Info; total obs: 381 (% is listed below)			
	100	200, 300	>300		100	200, 300	>300
0	17 (4.47)	5 (1.32)	4 (1.05)	0	10 (2.62)	1 (0.26)	0 (0)
10	12 (3.16)	8 (2.11)	16 (4.21)	10	12 (3.15)	12 (3.15)	8 (2.10)
20	2 (0.53)	23 (6.05)	15 (3.95)	20	8 (2.10)	16 (4.20)	26 (6.82)
30	22 (5.79)	20 (5.26)	17 (4.47)	30	**42** **(11.02)**	21 (5.51)	43 (11.29)
40	**54** **(14.21)**	40 (10.53)	15 (3.95)	40	19 (4.99)	19 (4.99)	55 (14.44)
50	24 (6.32)	19 (5.00)	24 (6.32)	50	14 (3.67)	26 (6.82)	20 (5.25)
>60	22 (5.79)	9 (2.37)	12 (3.16)	>60	12 (3.15)	9 (2.36)	8 (2.10)
Total	153 (40.26)	124 (32.63)	103 (27.11)		117 (30.71)	104 (27.30)	160 (41.99)

In the individual sessions, the most frequently chosen strategy is the lowest fixed rate with a piece rate of 30; meanwhile we see a significantly more use of higher fixed wage higher than or equal to 400. The average use of piece rate is 34.0, which is significantly lower than the theory benchmark of 40 (T - test, p-value ≈ 0.005). Although average use of fixed rate is 330, not significantly different from that in the social information sessions (T- test, p-value ≈ 0.164), the distribution of the fixed rate are distinctively different across sessions (Wilcoxon rank-sum test, P - value ≈ 0.006). The distribution of the piece rate is similar(Wilcoxon rank-sum test, P - value ≈ 0.606). It appears that principals use individual level information in making the fixed wage, not the piece rate.

4.3 Behavior of Agents

After observing the contract detail, agents choose work effort accordingly. In theory the optimal decision for a risk neutral agent is to choose an effort level of $\beta/10$, where β is the piece rate. Other factors may also affect the agents' choices, such as agents' risk attitude, cognitive abilities as well as reciprocity preference. We first compare each agent's effort choice and her optimal strategy. Out of 761 observations, there are 236 cases of optimal decisions and 280 cases with under-provision of effort–both negative reciprocity and risk aversion can explain some of those choices. The rest 245 cases are cases with over-provision of effort. Besides decision errors, positive reciprocity can explain the over-provision of effort. If the assumption of reciprocal agent is appropriate, we expect that fixed rate will have a positive impact on agents' effort choice after controllingng for decision errors. This conjecture can be verified using the panel data regression model below for each agent j:

$$effort_{jt} = \left[CRT_j, Risk_j, Gender_j, fixed_{jt}, piece_{jt}, TW_{jt}\right]\beta^{agents} + \{session\} + \{round\} + \phi_j^{agents} + \epsilon_{jt}^{agents},$$

where CRT_j, $Risk_j$ and $Gender_j$ corresponds to the agent j's number of corrected answers in the Cognitive Reflection Test, the multiple choice test on risk attitude, and whether the agent is female; TW_j is the trustworthiness measure. After controlling for the session dummy variables, the round dummy variables and the random effects (ϕ_j^{agents}), we expect both the fixed rate and the trustworthiness measure to be positively correlated with effort provision.

Table 3 lists the estimation results for both information conditions.[12] The fixed rate

[12]For better display of the coefficients, we rescale the fixed rate by 1/100, because the fixed rate increases in

Table 3: **The effect of trustworthiness on effort choices**

	social info sessions			individual info sessions		
	first five rounds	second five rounds	third five rounds	first five rounds	second five rounds	third five rounds
fixed	0.145***	0.0456	0.315***	0.107**	0.0749	0.216***
	(0.0432)	(0.0636)	(0.0674)	(0.0336)	(0.0581)	(0.0653)
piece	0.749***	0.545***	0.723***	0.626***	0.479***	0.751***
	(0.0749)	(0.0814)	(0.0782)	(0.0948)	(0.117)	(0.104)
crt	-0.0681	0.0440	-0.0323	0.250+	0.444*	0.211
	(0.180)	(0.244)	(0.255)	(0.146)	(0.211)	(0.143)
risk	0.00275	0.264+	0.0235	0.310+	0.447*	0.248
	(0.127)	(0.151)	(0.172)	(0.164)	(0.175)	(0.210)
gender	-0.129	-0.394	-0.142	-0.432+	-0.785**	-0.151
	(0.339)	(0.437)	(0.437)	(0.262)	(0.298)	(0.488)
TW	2.807*	4.786**	2.348	0.397	-0.647	0.903
	(1.357)	(1.690)	(1.813)	(0.628)	(1.007)	(0.579)
constant	0.605	-0.127	-0.914	0.289	0.327	-1.351
	(0.829)	(1.050)	(0.996)	(0.697)	(0.821)	(0.899)
round dummies	Yes	Yes	Yes	Yes	Yes	Yes
session dummies	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	26	26	26	27	27	27
N	380	124	127	381	124	130

Standard errors in parentheses
$^{+}$ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

has a growing impact on agents' effort provision in both treatments. Increasing the fixed rate by 300–500 ECU from the minimum fixed wage will lead to additional level of effort, holding all else constant. As expected, the piece rate has a significant effect over all rounds. Other individual characteristics seems to only affect initial rounds.

4.4 Reciprocity preference of Agents

The above analysis shows that agents' preference is beyond self-regarding, otherwise a fixed rate should not increase their effort. However, regressions typically cannot inform us further about the unobserved structural parameters. Therefore, we model agents' choice over effort levels directly, in hope to explore both λ, the weight the agent has on the reciprocity part of her preference, and K, the reference point for triggering reciprocity in either direction (positive or negative). Following our theoretical setting, agent js'

the unit of 100 ECU; we also rescale the piece rate by 1/10 for the same reason.

(expected) utility of effort level k can be written as

$$U_{jkt} = \xi \left[(\omega_{jt} + 10 \cdot \beta_{jt} \cdot k - k^2) + \lambda(\omega_{jt} - K) \cdot (10 \cdot (\alpha - \beta_{jt}) \cdot k - \omega_{jt}) \right] + \epsilon_{jkt}, \quad (4)$$

where k represents agents' effort level, ξ affects the precision of the agents' choice, $\alpha = 80$ is the marginal benefit from production, and ϵ_{jkt} is the standardized (logit) random error for each subject j's effort choice k at round t. Given the principal's wage offer ω and β, we can estimate the set of parameters that maximize the conditional probabilities of the realized effort choices. Table 4 reports results from the conditional logit estimations. Because $1/\xi$ captures the nosiness in the play, we observe agents make improving decisions over time. The estimates are also consistent across different information conditions: a 100 increase in Experimental Currency Unit (ECU) leads to a 0.3 utility point increase. In the social information sessions, the average reference point for reciprocity preference is 404, significantly higher than the minimum wage. When such preference is triggered, the weight an agent assign to the principal's utility depends both on λ and K. For example, with a fixed rate given at 300 ECU, 7% of the principal's payoff would enter the agent's utility in a negative way. The higher the effort is, the larger impact the negative reciprocity will cast on the agent's utility. Therefore, principals' fixed rate offer can be positively correlated with effort input because of negative reciprocity. With similar argument, positive reciprocity can also create such correlation.

In the benchmark model, the effects of positive and negative reciprocity are symmetric around K. However, if we view K as a reference point, agents may be subjected to greater dis-utility in the loss domain (Tversky and Kahneman, 1992). We add more flexibility to λ, allowing different values at both sides around the reference point K. Table 5 provides the estimation results for the alternative specification. λ_+ is the weighting parameter when a principal's fixed rate is higher than the reference point K; jointly with $(\omega - K)$

Table 4: **Maximum Likelihood Estimation on Agents' Effort Choice with Symmetric Effects of Reciprocity** The standard errors are calculated numerically using the quasi-newton algorithm provided by Matlab Optimization Toolbox (fminunc)

	social information sessions			individual information sessions		
	all rounds	first five rounds	third five rounds	all rounds	first five rounds	third five rounds
ξ	0.298	0.172	0.426	0.297	0.198	0.456
	(0.031)	(0.044)	(0.067)	(0.039)	(0.048)	(0.082)
λ	0.065	0.058	0.097	0.051	0.034	0.069
	(0.012)	(0.031)	(0.020)	(0.036)	(0.051)	(0.027)
K	404.117	475.602	345.134	691.227	759.462	571.749
	(49.861)	(180.005)	(43.217)	(239.765)	(481.320)	(106.040)
Log Likelihood	663.383	234.541	206.384	642.316	230.817	197.035

it captures the effect of positive reciprocity. Similarly, $-\lambda_{-}(\omega - K)$ captures the effect of negative reciprocity. The estimation confirms the asymmetric effects of reciprocity preference. In the social information session, the effect of negative reciprocity is twice as big as the positive reciprocity, which is consistent with previous estimates on loss aversion Tversky and Kahneman (1992). In the individual sessions, although the effect of negative reciprocity remains at the similar level, the effect of positive reciprocity is essentially zero– at least for the first several rounds. Over time, the average level of λ_{+}s are improving, however the asymmetry remains large. In the last five rounds, the estimate of the reference point becomes lower at 263 ECU for the social information sessions, yet it remains high in the individual information sessions at 519 ECU. It appears that the triggering of positive reciprocity under individual information sessions is more difficult, compared with that under social information sessions, while the effects of negative reciprocity are consistent across information conditions and remain dominant over the effects of positive reciprocity.

4.5 Revisiting Principals' Strategies

With the estimated agent response, it is useful to first think about the principal's best strategy. In Table 6, we assume two simulated agents based on the estimation results in

Table 5: **Maximum Likelihood Estimation on Agents' Effort Choice with Asymmetric Effects of Reciprocity** Because the utility function is not differentiable, we use the differential evolution (DE) method to estimate the global optimum and report the leave-two-out jacknife resampling standard deviation as the estimate of standard error of each coefficient below.

	social information sessions			individual information sessions		
	all rounds	first five rounds	last five rounds	all rounds	first five rounds	last five rounds
ξ	0.296	0.171	0.429	0.302	0.203	0.457
	(0.026)	(0.015)	(0.060)	(0.022)	(0.020)	(0.035)
λ_+	0.044	0.037	0.068	0.000	0.000	0.041
	(0.006)	(0.048)	(0.019)	(0.001)	(0.000)	(0.020)
λ_-	0.105	0.082	0.165	0.111	0.114	0.082
	(0.034)	(0.030)	(0.040)	(0.011)	(0.026)	(0.015)
K	323.784	400.000	263.300	446.888	400.000	519.414
	(49.891)	(94.051)	(42.646)	(26.887)	(36.217)	(55.033)
Likelihood	662.158	234.398	206.064	637.146	228.886	196.935

Table 5 and calculate the empirical best response for the agents and the principal. The agent in the bottom panel satisfy $\lambda_+ = 0.04$, $\lambda_- = 0.1$, and K=324, while the positive reciprocity is removed for the agent in the top panel. Compared with the self-regarding agent, now the fixed rate has a significant influence on the agent's effort choice. In the top-left panel, negative reciprocity causes the agent to use effort that is one level lower, when the fixed rate is low and the piece rate is not high enough. However it cannot explain the over-provision of effort. In the bottom-left panel, when positive reciprocity is allowed, the agent increases her effort given sufficient high fixed wage offers. The agent's response alters the principal's optimal choices: with only negative reciprocal agents, her best strategy is to offer a fixed rate of 300 and a piece rate of 40 (the top-right panel of Table 6); when positive reciprocity is considered, another best response emerges (the bottom-right panel of Table 6), in which she uses a much higher fixed wage of 700 and a lower piece rate of 30 to trigger the agent's positive reciprocity. Compared with the self-regarding case in Table 1, the principal's expected payoff is lower, indicating that the positive reciprocity cannot complement the surplus loss caused by the negative reciprocity

Table 6: **Simulated Responses for Negative and/or Positive Reciprocal Agents** Each row represents the fixed rate level and Each column represents the piece rate level. The left tables list agents' optimal effort levels and the right tables list principals' expected payoffs correspondingly. The equilibrium effort and payoff are in bold.

Agents with Only Negative Reciprocity Component

Agents' Optimal Effort Level										Principals' Expected Payoff									
	0	10	20	30	40	50	60	70	80		0	10	20	30	40	50	60	70	80
100	0	1	1	1	3	4	5	7	8	100	-100	600	500	400	1100	1100	900	600	-100
200	1	1	1	2	3	4	6	7	8	200	600	500	400	800	1000	1000	1000	500	-200
300	1	1	1	2	**4**	5	6	7	8	300	500	400	300	700	**1300**	1200	900	400	-300
400	1	1	2	3	4	5	6	7	8	400	400	300	800	1100	1200	1100	800	300	-400
500	1	1	2	3	4	5	6	7	8	500	300	200	700	1000	1100	1000	700	200	-500
600	1	1	2	3	4	5	6	7	8	600	200	100	600	900	1000	900	600	100	-600
700	1	1	2	3	4	5	6	7	8	700	100	0	500	800	900	800	500	0	-700
800	1	1	2	3	4	5	6	7	8	800	0	-100	400	700	800	700	400	-100	-800
900	1	1	2	3	4	5	6	7	8	900	-100	-200	300	600	700	600	300	-200	-900
1000	1	1	2	3	4	5	6	7	8	1000	-200	-300	200	500	600	500	200	-300	-1000

Agents with both Positive and Negative Reciprocity Component

Agents' Optimal Effort Level										Principals' Expected Payoff									
	0	10	20	30	40	50	60	70	80		0	10	20	30	40	50	60	70	80
100	0	1	1	1	3	4	5	7	8	100	-100	600	500	400	1100	1100	900	600	-100
200	1	1	1	2	3	4	6	7	8	200	600	500	400	800	1000	1000	1000	500	-200
300	1	1	1	2	**4**	5	6	7	8	300	500	400	300	700	**1300**	1200	900	400	-300
400	1	1	2	3	4	5	6	7	8	400	400	300	800	1100	1200	1100	800	300	-400
500	1	1	2	3	4	5	6	7	8	500	300	200	700	1000	1100	1000	700	200	-500
600	1	2	2	3	4	5	6	7	8	600	200	800	600	900	1000	900	600	100	-600
700	1	2	3	**4**	4	5	6	7	8	700	100	700	1100	**1300**	900	800	500	0	-700
800	1	2	3	4	5	5	6	7	8	800	0	600	1000	1200	1200	700	400	-100	-800
900	2	2	3	4	5	6	6	7	8	900	700	500	900	1100	1100	900	300	-200	-900
1000	2	3	3	4	5	6	6	7	8	1000	600	1100	800	1000	1000	800	200	-300	-1000

under the simulation's parameter constellation.

The above example gives us some ideas about principals' reaction based on the two simulated agents who reciprocate. How do principals generally perform in our contract game? We elicite individual level measures of trustworthiness, which reflects each agent's tendency to positively reciprocate principals, we expect that principals would use the information when making (fixed rate) offers. Estimations from panel data regression models suggest this is true in the later rounds. The following specification is considered:

$$y_{it} = \beta_0 + [CRT_i, Risk_i, Gender_i, TW_i]\,\beta + \{session\} + \{round\} + \phi_i + \epsilon_{it},$$

where y_{it} is the fixed rate[13] used by principal i at round t); CRT_i represents the number of corrected answers in the Cognitive Reflection Test; $Risk_i$ is based on the Eckel & Grossman measure; TW_i is the paired agent's trustworthiness measure; "{session}", "{round}" are the corresponding session and round dummy variables and ϕ_i is the random effect. Table 7 lists the estimation results. Principals start to reward trustworthiness at later rounds. For example, agents who has returned more than one third of the transfer in the first stage will receive about 87 more fixed rate on average, compared with those who has returned nothing. The finding is generally consistent with our estimation on agents' preference structure: in the early rounds of the game, the agents do not illustrate positive reciprocity preference.

By examining the realized expected payoff for each strategy combination, we do see that reciprocity-based wage offer yield higher expected payoff. For example, in the individual sessions, the wage offers with the top three highest expected profit are featured by much higher fixed rates of 600 ECU (with 40 piece rate), 400 ECU (with 30 piece rate) and 700 ECU (with 30 piece rate). Those offers on average generates 1533 ECU, 1523 ECU and 1400 ECU correspondingly[14]—higher than the benchmark case (100 ECU fixed rate and 40 ECU piece rate), which generates only 1173 ECU on average (T-test, P-value $\approx$ 0.180), and significantly higher than the most frequently chosen offer (100 ECU fixed rate and 30 ECU piece rate), which generates only 1055 ECU on average (T-test, P-value $\approx$ 0.002). The standard deviation of the expected profit (700 ECU) of the possible reciprocity driven offers is not far from the self-regarding case (921 ECU). The use of reciprocity based offers can also be observed in the social information sessions. In those sessions, principals have less information, face more uncertainty and their group average offers are lower. While free riders are still likely to take full advantage against a high fixed rate, the

[13] For comparison, the appendix Table A.2 reports the same regression for the use of the piece rate offer.
[14] There are 6, 13, and 5 occasions, in which those offers are used.

reciprocity reference point may be adjusted accordingly, making those relatively scattered high fixed rate more “worthy” for a high repay from trustworthy agents. Our estimation results confirm a decreasing trend of reciprocity reference point in the social information sessions, and thus positive reciprocity can be relatively easier to trigger.

In the individual information sessions, the rewarding to the relatively “trustworthy” agents (positively reciprocal agents who returned more than one third of the investment) is not associated with lower expected average return than the less “trustworthy” counterpart. In contrast, principals have higher expected return (T-test, P-value ≈ 0.170). The two types of agents also have similar payoffs (T-test, P-value ≈ 0.736), although there is also an additional non-monetary component associated with the “trustworthy” agents. Therefore, our lab experiment shows that reciprocity based offers perform at least about equal to the standard offer, on the principal side. The overall performance of the reciprocity based wage offer seems to also depend on the (endogenously determined) effect of positive reciprocity and the reciprocity reference point among agents. When the market is lack of trust, a fixed rate offer is likely to earn a high repay by “trustworthy” agents; in contrast, if the market has sufficient high fixed rate offers, it will be harder to trigger such preference. While our paper features a one-shot game, we leave this as an intriguing topic for future research.

5 Conclusion

In this paper, we provide an experimental test of a modified principal-agent model, in which the agent can be reciprocal. The theory tells us that principals with potential individual level trustworthiness information have an incentive to trigger such preference, since a positively reciprocal agent care more about the (weighted) total efficiency. Therefore, principals can benefit by providing higher fixed rate to those trustworthy agents.

Table 7: **The effect of trustworthiness on fixed rate choices**

	individual information sessions		
	all rounds	first five rounds	third five rounds
crt	0.281	0.0508	0.493
	(0.446)	(0.808)	(0.451)
risk	0.0584	0.270	0.0391
	(0.212)	(0.332)	(0.220)
gender	-0.291	-0.854	0.525
	(0.768)	(1.181)	(0.621)
$TW_{individual}$	1.336^{+}	1.130	2.614^{**}
	(0.683)	(1.037)	(0.890)
constant	2.935	3.632	0.0854
	(1.866)	(2.979)	(1.332)
round dummies	Yes	Yes	Yes
session dummies	Yes	Yes	Yes
Cluster	27	27	27
N	381	127	130

Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

And agents should also respond to wage offers in a way predicted by the theory. Our design reflects an environment where reciprocity contract is powerful: because principals cannot use a negative fixed rate to incentivize agents, traditional wage offer based on self-regarding agents leads to efficiency loss, while reciprocity contract may mitigate such loss. In our individual information sessions, principals also have information to screen positively reciprocal agents. We expect that this environment can be used to test agents' reciprocating behavior and correspondingly principals' offer strategies.

Our estimation results show that agents respond positively to the fixed rate and principals reward trustworthiness in later rounds of the individual information sessions. To examine agents' preference more closely, we directly model agents' preference using discrete choice models. The reciprocity component follows the theoretical setting and is assumed to be governed by a preference weighting parameter, as well as a reciprocity reference point. Our structural model shows that agents are subjected to both positive and negative reciprocity preferences; the later seems to have a strong impact on agents' preference. The estimated reciprocity reference point seems to be decreasing over time in the social information sessions, where principals only receive the aggregate information on agents' trustworthiness. This indicates that agents are less affected by the negative reciprocity and are more likely to positively reciprocate. In the individual information sessions, the reference point is not significantly moved.

Together, our experiment provides evidence supporting assumptions on reciprocal behavior made in recent theoretical contributions. Our results also suggest that loss aversion needs to be considered when applying the theory to the field: negative and positive reciprocity do not affect preference in a symmetric way. Negative reciprocity is generated more easily and it is considerably stronger than the positive reciprocity. Moreover, repeated interactions of the game in the social information sessions leads to significant changes in the reciprocity reference point. The fact that reciprocity reference point can be changed show a possible way for the principals or policy makers to alter agents' (reciprocity) preference.

References

Vital Anderhub, Simon Gächter, and Manfred Königstein. Efficient contracting and fair play in a simple principal-agent experiment. *Experimental Economics*, 5(1):5–27, 2002.

Dan Ariely, Uri Gneezy, George Loewenstein, and Nina Mazar. Large stakes and big mistakes. *The Review of Economic Studies*, 76(2):451–469, 2009.

Joyce Berg, John Dickhaut, and Kevin McCabe. Trust, reciprocity, and social history. *Games and Economic Behavior*, 10(1):122 – 142, 1995.

Truman F Bewley. *Why wages don't fall during a recession*. Harvard University Press, 1999.

Colin Camerer. *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press, 2003.

Marco Casari and Timothy N. Cason. The strategy method lowers measured trustworthy behavior. *Economics Letters*, 103(3):157–159, June 2009. ISSN 0165-1765. doi: 10.1016/j.econlet.2009.03.012.

Andrew Chamberlain. Why is hiring taking longer? new insights from glassdoor data, 2015. URL `https://www.glassdoor.com/research/studies/time-to-hire-study/`.

Gary Charness and Peter Kuhn. Lab labor: What can labor economists learn from the lab? *Handbook of labor economics*, 4:229–330, 2011.

James C Cox. How to identify trust and reciprocity. *Games and economic behavior*, 46(2): 260–281, 2004.

Giovanni Di Bartolomeo and Stefano Papa. Trust and reciprocity: extensions and robustness of triadic design. *Experimental Economics*, 19(1):100–115, 2016.

Martin Dufwenberg and Uri Gneezy. Measuring beliefs in an experimental lost wallet game. *Games and economic Behavior*, 30(2):163–182, 2000.

Martin Dufwenberg and Georg Kirchsteiger. A theory of sequential reciprocity. *Games and Economic Behavior*, 47(2):268–298, 2004.

Catherine C. Eckel and Philip J. Grossman. Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization*, 68(1):1 – 17, 2008.

Florian Englmaier and Stephen Leider. Contractual and organizational structure with reciprocal agents. *American Economic Journal: Microeconomics*, 4(2):146–183, 2012.

Armin Falk and Urs Fischbacher. A theory of reciprocity. *Games and Economic Behavior*, 54 (2):293–315, 2006.

Ernst Fehr and Simon Gächter. Fairness and retaliation: The economics of reciprocity. *The Journal of Economic Perspectives*, 14(3):159–181, 2000.

Ernst Fehr and Klaus M. Schmidt. A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868, 1999.

Ernst Fehr, Georg Kirchsteiger, and Arno Riedl. Does fairness prevent market clearing? an experimental investigation. *The quarterly journal of economics*, pages 437–459, 1993.

Urs Fischbacher. z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178, 2007.

Shane Frederick. Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4):25–42, 2005.

Edward L Glaeser, David I Laibson, Jose A Scheinkman, and Christine L Soutter. Measuring trust. *The Quarterly Journal of Economics*, 115(3):811–846, 2000.

Uri Gneezy and John A List. Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. *Econometrica*, 74(5):1365–1384, 2006.

Uri Gneezy and Aldo Rustichini. Pay enough or don't pay at all. *Quarterly journal of economics*, pages 791–810, 2000.

Noel D Johnson and Alexandra A Mislin. Trust games: A meta-analysis. *Journal of Economic Psychology*, 32(5):865–889, 2011.

Goto Jun, Sawada Yasuyuki, Aida Takeshi, and Aoyagi Keitaro. Incentives and social preferences: Experimental evidence from a seemingly inefficient traditional labor contract. Technical report, Center for Economic Institutions, Institute of Economic Research, Hitotsubashi University, 2015.

Alan B Krueger and Alexandre Mas. Strikes, scabs, and tread separations: Labor strife and the production of defective bridgestone/firestone tires. *Journal of Political Economy*, 112(2):253, 2004.

Sebastian Kube, Michel André Maréchal, and Clemens Puppe. Putting reciprocity to work-positive versus negative responses in the field. *University of St. Gallen Economics Discussion Paper*, (2006-27), 2006.

Thomas Lemieux, W. Bentley MacLeod, and Daniel Parent. Performance pay and wage inequality. *The Quarterly Journal of Economics*, 124(1):1–49, 2009.

Matthew Rabin. Incorporating fairness into game theory and economics. *The American Economic Review*, pages 1281–1302, 1993.

Jared Rubin and Roman Sheremeta. Principal–agent settings with random shocks. *Management Science*, 62(4):985–999, 2015.

R. Selten. Die strategiemethode zur erforschung des eingeschrnkt rationalen verhaltens im rahmen eines oligopolexperiments in Sauerman, H.(ed.) Beitrge zur Experimentellen Wirtschaftsforschung, Tbingen: JCB Mohr (Paul Siebeck). 1967.

Sara J. Solnick. Cash and alternate methods of accounting in an experimental game. *Journal of Economic Behavior & Organization*, 62(2):316–321, 2007.

Amos Tversky and Daniel Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5(4):297–323, 1992.

Appendices: Not Intended for Publication

Table A.1: Principals' Expected Payoff when Effort Level = 4

	0	10	20	30	40	50	60	70	80	90	100
100	3100	2700	2300	1900	1500	1100	700	300	-100	-500	-900
200	3000	2600	2200	1800	1400	1000	600	200	-200	-600	-1000
300	2900	2500	2100	1700	1300	900	500	100	-300	-700	-1100
400	2800	2400	2000	1600	1200	800	400	0	-400	-800	-1200
500	2700	2300	1900	1500	1100	700	300	-100	-500	-900	-1300
600	2600	2200	1800	1400	1000	600	200	-200	-600	-1000	-1400
700	2500	2100	1700	1300	900	500	100	-300	-700	-1100	-1500
800	2400	2000	1600	1200	800	400	0	-400	-800	-1200	-1600
900	2300	1900	1500	1100	700	300	-100	-500	-900	-1300	-1700
1000	2200	1800	1400	1000	600	200	-200	-600	-1000	-1400	-1800

Figure A.1: Experimental Interface: Role A

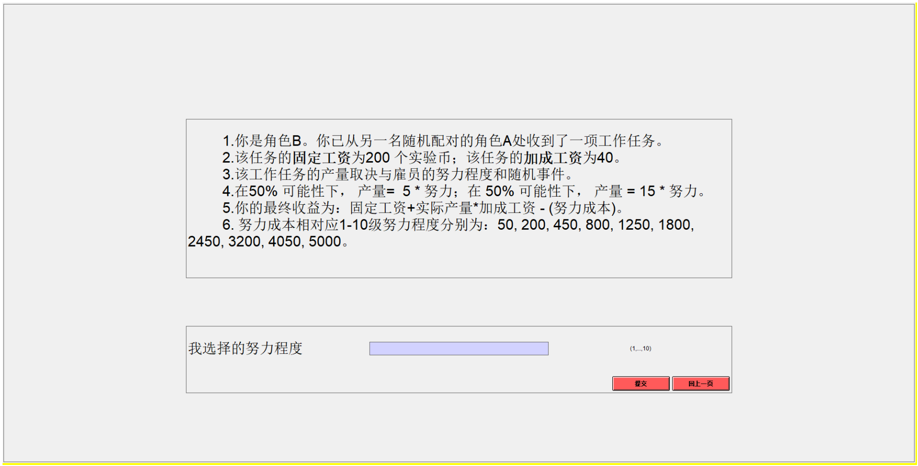

Figure A.2: Experimental Interface: Role B

Table A.2: The effect of trustworthiness on piece rate choices

	individual information sessions		
	all rounds	first five rounds	third five rounds
crt	0.259	0.101	0.393
	(0.279)	(0.337)	(0.346)
risk	0.221	0.0753	0.341
	(0.207)	(0.200)	(0.249)
gender	0.593	0.106	1.107*
	(0.452)	(0.478)	(0.560)
$TW_{individual}$	-0.448	0.141	-0.290
	(0.395)	(0.757)	(0.583)
constant	1.708	2.431	1.551
	(1.132)	(1.600)	(1.250)
round dummies	Yes	Yes	Yes
session dummies	Yes	Yes	Yes
Cluster	27	27	27
N	380	127	130

Standard errors in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$