Towards an understanding of the relative strengths of positive and negative reciprocity

Omar Al-Ubaydli*
George Mason University

Uri Gneezy
University of California at San Diego

Min Sok Lee
The Kenneth and Anne Griffin Foundation

John A. List
University of Chicago and NBER

Abstract

Scholars in economics and psychology have created a large literature studying reward, punishment and reciprocity. Labor markets constitute a popular application of this body of work, with particular emphasis on how reciprocity helps regulate workplace relationships where managers are unable to perfectly monitor workers.

We study how idiosyncratic features of the labor market (compared to most scenarios in which reciprocity applies) affect the nature of worker reciprocity. In particular, we show how having an excess supply of workers (simulating unemployment) and managers who can observe the reciprocal behavior of workers and hire/fire them on that basis (simulating the reputational concerns inherent in labor market transactions) profoundly alters worker reciprocity. In the absence of reputational concerns, workers tend to reward kind behavior and punish unkind behavior by managers in approximately equal measure. In the presence of reputational concerns, workers exhibit a marked increase (decrease) in the propensity to reward kind (punish unkind) behavior by managers. We demonstrate how this is a consequence of workers and managers responding to changes in the strategic incentives to reward and punish.

Keywords: reciprocity, reputation, reward, punishment, gift-exchange.

1 Introduction

The principle of retaliation is as old as mankind. As far back as the Hammarabian code some 3000 years ago, retaliation of some form has served to organize behavior in both market and non-market situations. Perhaps illustrating the importance of revenge most succinctly is the Biblical injunction of Exodus 21:23–25: "Life for life, eye for eye, tooth for tooth… bruise for bruise". For their part, scholars have explored the importance of negative actions alongside their seemingly more benign cousins, positive actions.

One of the key insights that can be taken from the decades of research within the social sciences is that reciprocity in general is important, and that negative actions toward an individual induce a greater behavioral response than comparable positive actions.[1] This stylized fact is perhaps best illustrated in the words of Baumeister et al. (2001), who provide a broad survey of several areas of study examining positive and negative reciprocity, and conclude that (p. 354–355, italics added): "The breadth and convergence of evidence, however, across different areas were striking, which forms the most important evidence. In no area were we able to find a consistent reversal, such that one could draw a firm conclusion that good is stronger than bad. *This failure to find any substantial contrary patterns occurred despite our own wishes and efforts.… Hence, we must conclude that bad is stronger than good at a pervasive, general level.*"

*We wish to thank Jon Baron, Ali Hortacsu and two anonymous referees for helpful comments, Jonathan Alevy for programming help and Nicholas Curott for excellent research assistance. Address: Omar Al-Ubaydli, Department of Economics, George Mason University, MSN 3G4, 4400 University Dr., Fairfax VA, 22030. Email: omar@omar.ec.

[1] A particularly simple exposition involves no more than asking subjects to list emotions within a time limit (Van Goozen and Frijda, 1993). The number of negative emotions listed almost always exceeds positive ones. Along similar lines, Oehman et al. (2001) found that people identified threatening faces more quickly and accurately than happy faces. Such negative visual stimuli also induce larger amplitude brain responses than positive ones (Ito et al., 1998). Generally, the negative domain commands affect and cognition more than the positive. In their survey, Baumeister et al. (2001) somewhat playfully draw our attention to Fiedler's (1982) finding that nobody has ever written a successful novel about a happy marriage; there is something about negative events that seizes our attention. Similar to scholars in other social sciences, economists have found that negative events also call forth greater responses than their positive counterparts (see, e.g., Offerman, 2002; Pereira et al., 2006; and Al-Ubaydli & Lee, 2009). Also see the review by Rozin and Royzman (2001). For applications of positive vs. negative reciprocity to the labor market, see Charness (2004) and the review in Charness and Kuhn (2005).

https://doi.org/10.1017/S1930297500001698 Published online by Cambridge University Press

Within economics, such results have served as the classic example of loss aversion — that people are more sensitive to negative realizations than to positive realizations of uncertainty (Tversky & Kahneman, 1991) — have played an important role in policymaking (see List, 2003), and have informed mechanism design. In terms of the latter, the principal is confronted with an interesting decision problem if framing of the incentive scheme matters to agent behavior or the number of instruments available to the principal is constrained. In this manner, choosing between carrots and sticks, for example, plays an important role in the outcome (see Andreoni et al., 2003). More generally, scholars have frequently remarked that loss aversion represents one of the most robust general behavioral patterns in the social sciences (see the citations in Baumeister et al., 2001).

In this study, we explore a general, labor-market setting wherein economic theory provides predictions that positive reciprocity should be stronger than negative reciprocity. The two key features are that the agent is on the short end of a market that includes reputational considerations and that being out of the market provides less utility than being a participant. Under this design, a worker that respects her initial affective reaction and punishes the employer will find herself unemployed. Alternatively, a worker who is nice to the employer will be more likely to be employed in the next period. Since being employed dominates unemployment, we predict that the worker will restrain herself and will not follow the initial affective reaction. On the other hand, if the employer is nice, the worker will reciprocate strongly since in this situation not only is she employed, but also by a nice employer. Thus, in this situation, positive reciprocity will be stronger than negative reciprocity.

To test our theory, we design a simple controlled laboratory experiment, which yields several insights. First, consonant with the literature, agents reciprocate. And, when the interactions are anonymous, negative reciprocity is slightly more important than positive reciprocity, but not significantly so. Also consonant with the literature is the fact that agents become emotionally charged when treated poorly. Yet, this emotional charge does not readily transfer to actions when realistic institutional features of labor markets are in place. For example, when agents can form reputations, they respond much more acutely to positive than to negative stimuli. Second, the data suggest that the source of the behavioral differences observed is strategic, rather than a change in the social norm of reciprocation in a gift exchange setting.

The remainder of our study proceeds as follows. Section 2 contains the experimental design. Section 3 summarizes the experimental results. Section 4 concludes.

2 Experimental design

To provide insights into measuring the strength of positive and negative reciprocity, we designed a simple two treatment experiment to test three conjectures. First, in one-shot environments, negative reciprocity is stronger than positive reciprocity. Second, in a repeated environment with a threat of exclusion, positive reciprocity is stronger than negative reciprocity. Finally, any observed difference in behavior across the two environments is not driven by affective reactions. In other words, when we test for differences in reciprocity conditional on affective reactions, behavioral differences should remain.

The game played by subjects is a discrete version of the trust game (Berg et al., 1995) where the agent (worker) has the opportunity to punish as well as reward. It is also repeated with an opportunity for principals (managers) to choose among the workers in each period. We begin by describing the simple stage game.

2.1 Stage game

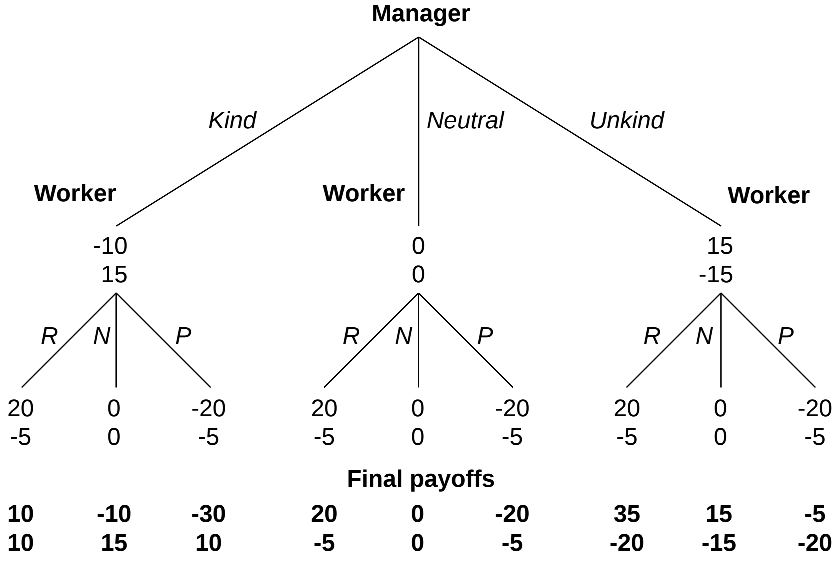
There are two players: a manager and a worker who play the stage game in Figure 1. In each column vector of Figure 1, the top (bottom) number is the manager's (worker's) payoff. The manager has three actions: *kind*, *neutral* and *unkind*. After the manager makes her choice, the worker observes the manager's choice and has three available actions: *reward*, *neutral* and *punish*.[2] Similar to Offerman (2002), the payoff consequences of actions are separable across the choices of each player. If the manager plays *neutral*, then both players' payoffs are unchanged. If the manager plays *kind*, then she loses 10 points and the worker gains 15 points. If the manager plays *unkind*, then she gains 15 points and the worker loses 15 points. Note that *kind* and *unkind* have symmetric effects on the worker's payoff.

At all nodes, if the worker plays *neutral*, both players' payoffs are unchanged. If the worker plays *reward*, she loses 5 points and the manager gains 20 points. If the worker plays *punish*, she loses 5 points and the manager loses 20 points. Thus *reward* and *punish* are equally costly to the worker and have symmetric effects on the manager's payoff (as in Offerman, 2002). Similar to most versions of the trust game, the unique subgame perfect equilibrium (assuming selfish preferences) is for the worker to play *neutral* at all nodes and the manager to therefore play *unkind*. The unique symmetric efficient outcome is (*kind*,*reward if kind*).

[2] In the experiment, roles and strategies were given neutral names. See below for more details. There is a large literature on reward and punishment in public goods games; see, e.g., Abbink et al. (2000), Dickinson (2001), Sefton et al. (2007), though this literature typically focuses on the ability of reward and/or punishment to improve the efficiency of outcomes compared to the absence of reward/punishment.

https://doi.org/10.1017/S1930297500001698 Published online by Cambridge University Press

Figure 1: Stage game. In each column vector, top number is manager payoff, bottom number is worker payoff. Final payoffs are obtained by summing the each of the two vectors implied by the strategy. *R* denotes reward, *N* denotes neutral and *P* denotes punishment.



2.2 Experimental game and treatments

The full game has 8 players: 3 managers and 5 workers. It is repeated for an uncertain number of periods. Throughout all versions of the game, the 3 managers have fixed IDs. In the key "reputation" treatment, workers also have fixed IDs, which are carried from period to period. In the one-shot, baseline treatment, the same fixed selection of IDs is randomly reassigned to each of the workers every round. For example, worker 3 might be a different person in round 1 to worker 3 in round 2. These features are common knowledge.[3]

Play proceeds in the following manner.

1. Nature decides the order in which each manager gets to select the worker that will be her partner (uniform distribution).

2. The first manager selects her partner; the next manager selects from the remaining workers, and so on.

   The 2 workers who are unselected for the round lose 25 points. This is 10 points less than the worst guaranteed payoff when selected. In other words, being employed by a nasty manager is better than being unemployed.

3. Each manager chooses between *kind*, *neutral* and *unkind*.

4. Each worker finds out which manager selected her and what choice the manager made. The workers who were not selected are informed of these choices.

5. Each paired worker chooses between *reward*, *neutral* and *punish*.

6. All players see their own payoffs for the round. They also see the choices made by all manager-worker pairs in that round and they see the history of choices by all pairs.

   Recall that worker IDs are fixed only in the reputation treatment. In the one-shot treatment, it is common knowledge that, say, worker 5's choice in round 3 may not have been made by the same player as worker 5's choice in round 4.

The game is repeated for a total of 11 rounds, though at the start of the experiment to prevent end-game effects subjects are told only that the experiment will continue for a "number of rounds". Moreover, the decision-making stages of all sessions end at least 25 minutes in advance of the 90 minutes for which subjects sign up (it takes about 10 minutes to calculate earnings and pay subjects).

The reputation treatment is designed to capture a situation where reputational concerns — allied with an exclusion threat — will make positive reciprocity dominate negative reciprocity.[4] The one-shot treatment is designed

[3] Our one-shot vs. reputation design mimics that of Brown et al. (2004), though their study focuses on third-party enforcement vs. reputation as mechanisms for eliciting efficient behavior.

[4] Fehr et al. (1998) also look at the effect of competition with an excess supply of workers, though their study focuses on the effect of competition on the ability of social norms to avoid inefficient outcomes.

https://doi.org/10.1017/S1930297500001698 Published online by Cambridge University Press

as a control where the elimination of reputational concerns and strategic exclusion will move the balance back towards the stylized fact, i.e., dominance of negative reciprocity.

To explore affective reactions, in some sessions we asked paired workers to declare privately how they feel about their respective managers' choices: *very unhappy*, *somewhat unhappy*, *neutral*, *somewhat happy*, *very happy*.[5] This permits an exploration of whether affective reactions have a greater influence on reward/punishment decisions in the one-shot treatment. Worker behavior was not different across the sessions in which this information was elicited, so we pool the data below.

2.3 Procedure

In total, we ran 9 sessions at George Mason University during spring 2009. Subjects were recruited from a database of students who had declared an interest in participating in economics experiments. Each session had 11 periods and 16 subjects divided into two groups of 8 (3 managers, 5 workers). Roles were assigned randomly and all interactions were anonymous. Subjects' roles were fixed and they interacted exclusively with members of their own group. The total number of worker observations collected is therefore 591.[6]

The experiment was computer-based and used z-Tree (Fischbacher, 2007). Instructions were on-screen, though a hard copy was given to each subject. All subjects were in the same room and within earshot and eyeshot of each other. This is potentially important because the monitor read the instructions aloud to ensure common knowledge. The game was presented in a neutral frame, i.e., managers chose between *left*, *middle* and *right*, and workers chose between *add 20*, *nothing* and *subtract 20*. Managers were called *Reds* and workers were called *Blues*. After completing the experiment, subjects were paid privately.[7]

3 Empirical results

Table 1 presents a summary of the experimental results. In this summary, to provide a first glimpse of behavioral patterns we have ignored data dependencies and pooled individual play over all 11 periods of the game. As we discuss each result below, we supplement these raw data patterns with conditional analysis. We begin with a first result.

Result 1: In both treatments, workers reciprocate manager choices: they frequently reward *kind* and punish *unkind*.

Evidence to support this result can be seen in Table 1, where it is shown that workers reward *kind* actions in 51% of the cases in the one-shot treatments and in 84% of the cases in the reputation treatment.[8] Likewise, they punish *unkind* actions in 53% of the instances they occur in the one-shot treatment and in 34% of the instances they occur in the reputation treatments.[9] These figures are both significantly higher than the propensities to react positively or negatively to other manager actions (i.e., the percentage of reward play in response to a manager choosing *neutral*), and are all significantly different from zero using conventional parametric statistical tests at the p < .01 level. Similar results are found when using nonparametric tests that have a null hypothesis of no treatment effect, or that the two samples are derived from identical populations.

These unconditional tests ignore statistical dependence between observations. As a robustness check, we estimate conditional parametric models of the following form, where i denotes worker and t denotes period:

$$Y_{it} = \alpha + \beta_K K_{it} + \beta_U U_{it} + \sum_{s=2}^{11} \tau^s T_{it}^s + \sum_{g=2}^{J} \gamma^s G_{it}^g + \varepsilon_{it}$$

Y_{it} is a dummy variable that takes the value 1 if and only if the worker plays reward (or punish, where appropriate).[10] K_{it} is a dummy variable taking the value 1 if and only if the worker's manager played kind. U_{it} is a dummy variable taking the value 1 if and only if the worker's manager played unkind. To allow for dependence across workers, within time periods, T_{it}^s is a dummy variable taking the value 1 if and only if $s = t$, i.e., it is a period s time effect. To allow for dependence across workers and time periods, and within groups, G_{it}^g is a dummy variable taking the value 1 if and only if the worker is in group g, i.e., it is a group effect. (Recall that each session is

[5] The only part of the game that was not common knowledge was asking about emotions; it was not read out at the start (see procedure below) and appeared only on the workers' screens. This was done to minimize priming of the subjects to think in terms of affective reactions. As noted, for robustness we also ran sessions in which we did not ask the workers to declare their emotions. None of the results were affected.

[6] The treatment breakdown was 396 from one-shot and 195 from reputation. We lost the data from the last period of one group in one session (3 observations).

[7] To compensate for the fact that some subjects had a negative payoff, there was a large show-up fee, though this was announced only after the start of the experiment, i.e., subject recruitment utilized the usual GMU show-up fee. Average earnings were approximately $18.

[8] For full statistical comparisons of cells in Table 1, see Table A1 in the Appendix.

[9] All of our inference (for all the results) excludes the data corresponding to managers selecting *neutral* since it is not relevant to comparisons of positive and negative reciprocity. For the interested reader, here are a few features of the excluded data. Mangers play *neutral* 8% of the time. In both treatments, worker emotions in response to *neutral* are insignificantly different from 0 (on a scale of -2 to +2) using a t-test (p > .30).

[10] For ease of interpreting estimated coefficients, we estimate linear regressions, i.e., linear probability models. For robustness, we also estimate probits with the same explanatory variables.

https://doi.org/10.1017/S1930297500001698 Published online by Cambridge University Press

Table 1: Summary statistics by treatment. The percentage in a cell should be interpreted as the frequency with which a worker plays the *row* strategy given that the manager played the *column* strategy, e.g., in the one-shot, when the manager plays *unkind*, the worker plays *punish* 53% of the time. Emotions are scaled as follows: -2 = very unhappy, -1 = somewhat unhappy, 0 = neutral, +1 = somewhat happy, +2 = very happy. There are less observations for emotions because we did not elicit emotions in all sessions.

		Manager choice		
		Unkind	Neutral	Kind
Worker response: One-shot (6 sessions = 12 groups)	Observations (total)	216	36	144
	Punish	53%	17%	3%
	Neutral	37%	75%	46%
	Reward	10%	8%	51%
	Observations (emotions)	82	15	35
	Emotion	−1.60	−0.20	1.70
	Standard deviation	0.83	0.86	0.84
Worker response: Reputation (3 sessions = 6 groups)	Observations (total)	99	13	83
	Punish	34%	16%	1%
	Neutral	51%	62%	14%
	Reward	15%	23%	84%
	Observations (emotions)	64	13	52
	Emotion	−1.40	−0.31	1.80
	Standard deviation	1.00	1.10	0.55

composed of two groups of eight individuals who operate independently for the entire session.) Finally, ε_{it} is an error term that allows for dependence across time periods and within an individual worker: $\varepsilon_{it} = u_i + e_{it}$ where $Cov(e_{it}, e_{js}) = 0$ for $i \neq j$ or $s \neq t$, and $Cov(u_i, u_j) = 0$ for $i \neq j$, also known as clustering at the individual level.

The results from estimating the above model echo the unconditional results. For parsimony, we relegate them to the appendix. This result is not surprising, as scores of studies have found reciprocal behavior — from student subjects to CEOs (see, e.g., Offerman, 2002; Andreoni et al. 2003, Fehr and List, 2004). Examining the data at a slightly deeper level, we observe another result that is in line with the literature.

Result 2: In the one-shot treatment, negative reciprocity is slightly stronger than positive reciprocity, though the difference is statistically insignificant.

Evidence to support this result can be seen in Tables 1 and 2. In Table 1, for example, we find that in the one-shot treatment, workers reward *kind* actions in 51% of cases, while they punish *unkind* at the slightly higher rate of 53% of cases, though this difference is insignificant using Mann-Whitney (n = 360, p = .73) and t-tests (n = 360, p = .73).

To allow for the likely sources of dependence in the data, we estimate the following model:

$$R_{it} = \alpha + \beta_K K_{it} + \sum_{s=2}^{11} \tau^s T^s_{it} + \sum_{g=2}^{J} \gamma^s G^g_{it} + \varepsilon_{it}$$

R_{it} is a dummy variable that takes the value 1 if and only if the worker reciprocates the manager's choice, i.e., responds to *kind* with *reward* or to *unkind* with *punish*. In addition to controlling for period and group effects, we allow for a particularly refined form of correlation in the error term ε_{it}: we allow for correlation within worker given his/her decision node (*kind* vs. *unkind*), i.e., we use two clusters per worker.[11]

In this model, the control group are the observations where the manager plays *unkind* and let the treatment group are the observations where the manager plays *kind* (both in the one-shot sessions). In this sense, the treatment group is the "positive reciprocity" group.

The results can be seen in Table 2. The estimated coefficient on "positive reciprocity" should be read as how

[11] Using a common cluster affects instead none of our results. In fact using a common cluster shrinks the standard errors — as one would expect – though not by enough to alter any result.

https://doi.org/10.1017/S1930297500001698 Published online by Cambridge University Press

Table 2a: Conditional results. The dummy variable “Reciprocate” takes the value 1 when the worker reciprocates *kind* with *reward* or *unkind* with *punish*. “Positive reciprocity” is a dummy variable that takes the value 0 when the manager plays unkind and 1 when the manager plays kind. All models contain clusters at the individual level. In probits, the reported figure is the estimated marginal effect. All models exclude the 49 observations corresponding to the manager playing neutral. Estimated period/group fixed effects are omitted for parsimony. Asterices denote statistical significance (* = .10, ** = .05, *** = .01).

Model		1	2	3	4
Data included		One-shot	One-shot	Reputation	Reputation
Estimation method		Regression	Probit	Regression	Probit
Dependent variable		Reciprocate	Reciprocate	Reciprocate	Reciprocate
Explanatory variable:	Positive reciprocity	−0.09	−0.09	0.51***	0.54***
	P-value from z-test	.31	.32	< .01	< .01
R^2 / Pseudo R^2		0.11	0.09	0.32	0.26
Observations		108	108	47	47
Degrees of freedom		360	360	182	182

much more likely a worker is to reward *kind* than she is to punish *unkind*. Using both a linear probability model (model 1) and a probit model (model 2), punishing *unkind* is 9% more likely than rewarding *kind*, though this difference is again insignificant (p = .31). This is directionally consistent with the existing literature’s finding that in one-shot environments, the negative reciprocity is stronger than positive reciprocity (Offerman, 2002; Al-Ubaydli & Lee 2009).[12]

Result 3: In the reputation treatment, positive reciprocity is stronger than negative reciprocity.

Evidence to support this result can be seen in Tables 1 and 2a. Looking at Table 1, we see that in the reputation treatment, unconditionally, a worker is 50% more likely to reward *kind* than she is to punish *unkind*. Both Mann-Whitney and t-tests are significant at conventional levels (n = 182, p < .01).

To allow for the likely sources of dependence in the data, we estimate the model from Result 2 for data from the reputation sessions (see Table 2a; model 3 is a linear probability model and model 4 is a probit). We find that the estimated treatment effect is over 50% and statistically significant (p < .01).

An even more conservative approach is to treat each group (there are two independent groups of eight participants per session) as yielding only two data points: the relative frequency of rewarding *kind* (treatment) and the relative frequency of punishing *unkind* (control), both obtained by averaging across all players and rounds within a group. This method implies 12 total data points. Both a paired value t-test and a Wilcoxon signed-rank test reject the null hypothesis of equality (p < .05; 5 df).

Results 4-to-6 attempt to shed light on the underpinnings for this result.

Result 4: The difference in reciprocity between one-shot and reputation sessions is driven primarily by a large increase in positive reciprocity when going from one-shot to reputation sessions.

Empirical evidence to support this result can be seen in Table 1, where positive reciprocity increases by 34% when moving from the one-shot to the reputation treatment, whereas punishment falls by 19%. Testing this formally requires a conditional parametric specification. In models 5 and 6 in Table 2b, we pool the data from our one-shot and reputation treatments and include a reputation session dummy (D_{it}^{REP}) variable and an interaction term between positive reciprocity and the reputation session dummy $(D_{it}^{REP} K_{it})$:

$$R_{it} = \alpha + \beta_{REP} D_{it}^{REP} + \beta_K K_{it} + \beta_{K,REP} D_{it}^{REP} K_{it} + \sum_{s=2}^{11} \tau^s T_{it}^s + \sum_{g=2}^{J} \gamma^s G_{it}^g + \varepsilon_{it}$$

The reputation by session dummy coefficient $\beta_{K,REP}$ tells us how much more likely workers are to punish *unkind* in the reputation treatment than in the one-shot treatment. The point estimates, of roughly 20%, suggest that workers are substantially less likely to punish in the reputation treatment, though this is not statistically significant at conventional levels. Thus negative reciprocity is at most slightly smaller in the reputation treatment than in the one-shot treatment.

Given this result, the large (greater than 46%) and significant (p < .01) coefficient of the interaction of the rep-

[12] Essentially all our results are robust to using only data from periods 6-to-11, reinforcing the design’s attempts at avoiding any “end-of-session” effects. The only slight exception is Result 4. See below.

https://doi.org/10.1017/S1930297500001698 Published online by Cambridge University Press

Table 2b: Conditional results. In addition to the description below Table 2a: emotions are scaled as follows: −2 = very unhappy, −1 = somewhat unhappy, 0 = neutral, +1 = somewhat happy, +2 = very happy. 'Reputation session' is a dummy that takes the value 1 in reputation sessions.

Model		5	6	7
Data included		Pooled	Pooled	Pooled
Estimation method		Regression	Probit	Regression
Dependent variable		Reciprocate	Reciprocate	Emotions
Explanatory variables	Positive reciprocity	−0.08	−0.08	3.26***
	P-value from z-test	.36	.37	< .01
	Reputation session	−0.21	−0.22	0.09
	P-value from z-test	30%	15%	76%
	Pos. recip. x Rep. sess.	0.56***	0.47***	−0.09
	P-value from z-test	< .01	< .01	.76
R^2 / Pseudo R^2		0.16	0.13	0.80
Observations		542	542	233
Degrees of freedom		156	156	65

utation and positive reciprocity is primarily the result of a substantial increase in positive reciprocity when going from one-shot to reputation.[13] Our next result concerns the underpinnings of the reversal.

Result 5: The difference in the balance of positive and negative reciprocity across one-shot and reputation sessions is not the result of differences in the affective reactions to *kind* and *unkind* across one-shot and reputation sessions.

Evidence to support this result can be seen in Tables 1 and 2b. Recall that the scale for worker's declared emotion after seeing the manager's choice is: –2 = very unhappy, –1 = somewhat unhappy, 0 = neutral, +1 = somewhat happy, +2 = very happy.

In Table 1, the mean emotion in response to *kind* is +1.7 in the one-shot treatment and +1.8 in the reputation treatment. This difference is insignificant using three unconditional tests (n = 87; Mann-Whitney: p = .19, t-test: p = .30; Kolmogorov-Smirnov: p = .96). The mean emotion in response to *unkind* is –1.6 in the one-shot treatment and –1.4 in the reputation treatment. This difference is marginally significant for insignificant depending on the unconditional test employed (n = 146; Mann-Whitney: p = .10, t-test: p = .12; Kolmogorov-Smirnov: p = .66).

In model 7 in Table 2, we estimate a regression model of emotions (E_{it}):

$$E_{it} = \alpha + \beta_{REP} D_{it}^{REP} + \beta_K K_{it} + \beta_{K,REP} D_{it}^{REP} K_{it} + \sum_{s=2}^{11} \tau^s T_{it}^s + \sum_{g=2}^{J} \gamma^s G_{it}^g + \varepsilon_{it}$$

In this model, the estimated coefficients on the reputation sessions dummy (β_{REP}) and on the interaction between positive reciprocity and reputation sessions dummies $(\beta_{K,REP})$ are statistically insignificant. In other words, worker emotive responses to manager choices do not depend upon being in one-shot vs. reputation treatments. Thus, the marginal significance levels obtained in a couple of the unconditional tests are the result of failing to deal with the statistical dependence between observations.[14]

Perhaps the most compelling evidence that the results are not driven by differences in affective reactions is that in the one-shot sessions, when the worker was very unhappy at the manager playing *unkind*, she rewarded the manager 3% of the time. The corresponding figure for the reputation sessions was 17% (n = 103, p < .05 using a t-test and a MW-test). Clearly several workers were willing to reciprocate due to the threat of strategic exclusion in the reputation sessions.

[13] When we use data only from periods 6-to-11, in addition to the large increase in reward (73%), there is a large (but still smaller) decrease in punishment (52%). This does not affect the paper's main argument.

[14] Minor results that we omit for parsimony are that in both treatments, first, manager moves predict emotions in the expected way, i.e., the kinder the manager's action, the happier the worker. Second, emotions predict reward and punishment in the expected way, i.e., subjects who report more positive emotions are more likely to reward and less likely to punish.

https://doi.org/10.1017/S1930297500001698 Published online by Cambridge University Press

Table 2c: Conditional results. In addition to the description below Table 2a: "Reselect" is a dummy variable that takes the value 1 when a worker is reselected by (any) manager in that period given selection in the previous period.

Model		8	9	10	11
Data included		Reputation	Reputation	Reputation	Reputation
Estimation method		Regression	Probit	Regression	Probit
Dependent variable		Reselect	Reselect	Reselect	Reselect
Explanatory variables	Rewarded *kind* last period	0.36**	0.40***	–	–
	P-value from z-test	.01	< .01	–	–
	Punished *unkind* last period	–	–	−0.25**	−0.30***
	P-value from z-test	–	–	.02	< .01
R^2 / Pseudo R^2		0.26	0.27	0.25	0.20
Observations		77	77	87	87
Degrees of freedom		29	29	45	45

This suggests that the difference in the balance of positive and negative reciprocity across one-shot and reputation sessions is not the result of differences in the affective reactions, leading to our next result.

Result 6: The difference in the balance of positive and negative reciprocity across one-shot and reputation sessions is the result of strategic differences in the environment, specifically the threat of systematic exclusion.

Evidence to support this result can be seen in Tables 2c and 3. The dimensionality of the history space is too large for sophisticated structural modeling (the data demands are not met either), however result 6 can still be derived from a more modest structural approach. The basic hypothesis is that in the reputation sessions, managers seek workers who have rewarded kind or who did not punish unkind.

In Table 3, for example, we see that in the reputation treatment, if a worker rewards a play of *kind*, then this increases her chances of being reselected by a manager in the subsequent round by 43% compared to not rewarding (significant at p < .03 using Mann-Whitney and t-tests; n = 77). Similarly, if a worker punishes a play if *unkind*, then this decreases her chances of being reselected by a manger in the subsequent round by 24% compared to not punishing (significant at p < .03 using Mann-Whitney and t-tests; n = 87). In the one-shot treatments, statistically speaking, reselection chances are unaffected by past play. This is comforting since the worker IDs were scrambled every round and so the managers could not perform any systematic exclusion.[15]

Table 2c reveals that allowing for dependence across observations yield consistents results. We estimate the following model:

$$Z_{it} = \alpha + \beta_V V_{it} + \sum_{s=2}^{11} \tau^s T_{it}^s + \sum_{g=2}^{J} \gamma^s G_{it}^g + \varepsilon_{it}$$

Z_{it} is a dummy variable that takes the value 1 if and only if a player who was selected in the previous round was reselected in the current round. V_{it} is a dummy variable that takes the value 1 if and only if a worker responds to *kind* with *reward* or *unkind* with *punish*. We estimate this model for observations where the manager played *kind* last period (models 8 and 9) and where the manager played *unkind* last period (models 10 and 11).

The models confirm that rewarding *kind* increases a worker's probability of being selected in the next round by over 35%. On the negative reciprocity side, models 10 and 11 confirm that punishing *unkind* diminishes a worker's probability of being reselected by over 24%. Again, re-estimating models 8–11 using data from the one-shot sessions (omitted for parsimony), we find that all coefficients are statistically insignificant (all have a p-value greater than .50) and have very small magnitudes (smaller than 5%).

While it is clear that managers account for worker actions in their partnerships choices, another manner in which managers' behavior potentially changes from the treatment itself. Our data reveal that workers punish much less often but reward more frequently in the repeated game than in the one shot game. An interesting question is whether managers use this information effectively in their choices.

[15] Since all subjects can see the entire history of play every round, in principle, it is possible to examine how behavior is affected by player earlier than the immediately preceding period. However when behavior in earlier periods differs from behavior in the immediately preceding period, the dimensionality of differences is too large (given the amount of data) to permit a cogent statistical analysis.

Table 3: Reselection. Relative frequency of a worker being reselected by (any) manager given selection in the previous round and given manager/worker choice in the previous round.

		Manager choice		
		Unkind	Neutral	Kind
Observations		87	13	77
Worker choice	Punish	37%	50%	0%
	Neutral	61%	38%	58%
	Kind	61%	67%	91%

Suppose that we treat workers choices as independent and identically distributed draws from the unconditional relative frequencies of reward and punishment behavior. In the one-shot treatment, this means that managers maximize their payoff by playing *unkind*. (Moreover, *kind* yields a higher expected return than *neutral*.) This is loosely reflected in managers' choices (see Table 1): *unkind* (54%), *kind* (36%), *neutral* (9%). The comparatively large incidence of *kind* is consonant with equity considerations.

In the reputation treatment, both the optimal and realized rank-ordering of manager choices is unchanged. While there is a slight increase (7%) in *kind* at the expense of *neutral* and *unkind*, none of the changes are statistically significant using conditional or unconditional tests. To some extent, this is unsurprising since increased reward and decreased punishment renders *kind* and *unkind* simultaneously more lucrative in absolute terms. This exploration leads to our final result.

Result 7: Efficiency is substantially higher in the reputation treatment.

The total realized payoff in the one-shot treatment is 14% of the total potential payoff, while the corresponding Figure for the reputation treatment is 40%. The reason why both are so low is because any deviation from *kind* and *reward* leads to a lower aggregate payoff, and such deviations are very frequent. Naturally, the efficiency improvement is driven by workers moving away from punishment towards reward.

4 Conclusion

As Arrow (1972, p. 357) put forth decades ago when he noted that "Virtually every commercial transaction has within itself an element of trust", most economic and non-economic transactions require a degree of trust. With the element of trust comes the necessary ingredient reciprocity. Scholars as far back as Aristotle (2004) appreciated the importance of negative reciprocity, as he extolled that revenge serves to discourage mistreatment. More recently, scientists have come to the firm conclusion that both negative and positive reciprocity are important, and have studied the factors that determine the balance between the two, especially in the context of labor markets.

This paper revisits this issue by infusing two realistic features — the agent is on the short end of a market that includes reputational considerations and that being out of the market provides less utility than being a participant — into a popular laboratory game. We argue that these additional considerations provide a setting that is representative of many common economic situations, especially labor markets. This alteration permits us to examine the relative strengths of positive and negative reciprocity while simultaneously exploring the underpinnings for reciprocity.[16]

We find that in a baseline without reputational concerns, negative and positive reciprocity are approximately equal in frequency. In repeated environments with a threat of systematic exclusion, positive reciprocity becomes much more frequent than negative reciprocity.[17] This holds because being employed by an exploitative manager dominates unemployment. Rational agents understand that this is the case and act accordingly. Importantly, this reversal is not the consequence of a change in affective reactions. People are as happy about kind behavior in the repeated environment as they are in the one-shot environment, and they are equally riled by unkind behavior across the two environments. Rather, the prominence of positive reciprocity in the repeated environment is driven by strategic concerns: those workers who are cooperative — either by reciprocating kind behavior or refraining from punishing unkind behavior — avoid unemployment by acting appropriately in the environment.

[16] When we use the term "reciprocity", we do not distinguish between reward/punishment behavior that is motivated by social (other-regarding) preferences vis-à-vis reputational concerns.

[17] A related study is Rand et al. (2009), which finds that in repeated a public goods game, reward is superior to punishment in eliciting cooperation. Also, see Kube et al. (2006) for a field comparison of positive and negative reciprocity in a one-shot setting.

https://doi.org/10.1017/S1930297500001698 Published online by Cambridge University Press

We view these results as important in several domains. First, they move us toward a deeper understanding of the relative strengths of positive and negative reciprocity. In this way, the results highlight the importance of the economic and psychological features embedded in any economic environment. Second, in doing so they open new paths of inquiry. For instance, in public policymaking, the general discussion of whether preferences are defined over consumption levels or changes in consumption has moved policymakers to more carefully consider the differences between willingness to pay and willingness to accept in cost benefit analysis. Understanding the mechanisms that underlie these valuation divergences is invaluable. Also, the practitioner interested in mechanism design might regard the results of import when crafting incentive schemes to alter agent behavior.

References

Abbink, K., Irlenbusch, B., & Renner, E. (2000). The moonlighting game: An experimental study on reciprocity and retribution. *Journal of Economic Behavior and Organization, 42*, 265–277.

Al-Ubaydli, O., & Lee, M. (2009). An experimental study of asymmetric reciprocity. *Journal of Economic Behavior and Organization, 72*, 738–749.

Andreoni, J., Harbaugh, W., & Vesterlund, L., (2003). The carrot or the stick: rewards, punishments, and cooperation. *American Economic Review. 93*, 893–902.

Aristotle (2004). *Rhetoric (Book II)*, translated by W. Rhys Roberts. New York: Dover Publications.

Arrow, K. (1972). Gift and exchanges. *Philosophy and Public Affairs, 1*, 343–362.

Baumeister, R., Bratslavsky, E., Finkenauer, C., & Vohs, K. (2001). Bad is stronger than good. *Review of General Psychology, 5*, 323–370.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity and social history. *Games and Economic Behavior, 10*, 122–142.

Brown, M., Falk, A., & Fehr, E. (2004). Relational contracts and the nature of market interactions. *Econometrica, 72*, 747–780.

Charness, G. (2004). Attribution and reciprocity in an experimental labor market. *Journal of Labor Economics, 22*, 665–688.

Charness, G., & Kuhn, P. (2006). Does pay inequality affect worker effort? Experimental evidence. *NBER Working Paper 11786*.

Dickinson, D. (2001). The carrot vs. the stick in work team motivation. *Experimental Economics, 4*, 107–124.

Fehr, E., & List, J. (2004). The hidden costs and returns of incentives – trust and trustworthiness among CEOs. *Journal of European Economic Association, 2*, 743–771.

Fehr, E., Kirchler, E., Weichbold, A., & Gachter, S. (1998). When social norms overpower competition: Gift exchange in experimental labor markets. *Journal of Labor Economics, 16*, 324–351.

Fiedler, L. (1982). *Love and death in the American novel*. New York: Stein & Day,.

Fischbacher, U. (2007). Z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics, 10*, 171–178.

Ito, T. , Cacioppo, J., & Lang, P. (1998). Eliciting affecting using the International Affective Picture System: Bivariate evaluation and ambivalence. *Personality and Social Psychology Bulletin, 24*, 855–879.

Kube, S., Marechal, M., & Puppe, C. (2006). Putting reciprocity to work — positive versus negative responses in the field. Unpublished manuscript, University of Karlsruhe.

List, J. (2003). Does market experience eliminate market anomalies?. *Quarterly Journal of Economics, 118*, 41–71.

Oehman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: A threat advantage with schematic stimuli. *Journal of Personality and Social Psychology, 80*, 81–396.

Offerman, T. (2002). Hurting hurts more than helping helps. *European Economic Review, 46*, 1423–1437.

Pereira, P., Silva, N., & Silva, J. (2006). Positive and negative reciprocity in the labor market. *Journal of Economic Behavior and Organization, 59*, 406–422.

Rand, D., Dreber, A., Ellingsen, T., Fudenberg, D., & Nowak, M. (2009). Positive interactions promote public cooperation. *Science, 325*, 1272–1275.

Rozin, P., & Royzman, E. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review, 5*, 296–320.

Sefton, M., Shupp, R., & Walker, J. (2007). The effect of rewards and sanctions in provision of public goods. *Economic Inquiry, 45*, 671–690.

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: a reference-dependent model. *Quarterly Journal of Economics, 106*, 1039–1061.

Van Goozen, S., & Frijda, N. (1993). Emotion words used in six European countries. *European Journal of Social Psychology, 23*, 89–95.

https://doi.org/10.1017/S1930297500001698 Published online by Cambridge University Press

Appendix

Table A1: Statistical comparisons across cells from Table 1. Relation frequency of a worker being reselected by (any) manager given selection in the previous round and given manager/worker choice in the previous round.

	Cell 1	Cell 2	*d*	P-value	Deg. of freedom
One-shot	Punish / Unkind	Punish / Neutral	$-.36$	$< .01$	58
	Punish / Unkind	Punish / Kind	$-.50$	$< .01$	59
	Punish / Neutral	Punish / Kind	$-.14$	.06	53
	Neutral / Unkind	Neutral / Neutral	.38	$< .01$	58
	Neutral / Unkind	Neutral / Kind	.09	.12	59
	Neutral / Neutral	Neutral / Kind	$-.29$	$< .01$	53
	Reward / Unkind	Reward / Neutral	$-.02$	.31	58
	Reward / Unkind	Reward / Kind	.41	$< .01$	59
	Reward / Neutral	Reward / Kind	.43	$< .01$	23
Reputation	Punish / Unkind	Punish / Neutral	$-.18$	.29	29
	Punish / Unkind	Punish / Kind	$-.33$	$< .01$	29
	Punish / Neutral	Punish / Kind	$-.15$	.26	23
	Neutral / Unkind	Neutral / Neutral	.11	.45	29
	Neutral / Unkind	Neutral / Kind	$-.37$	$< .01$	29
	Neutral / Neutral	Neutral / Kind	$-.48$	$< .01$	23
	Reward / Unkind	Reward / Neutral	.08	.93	29
	Reward / Unkind	Reward / Kind	.69	$< .01$	29
	Reward / Neutral	Reward / Kind	.61	$< .01$	23

https://doi.org/10.1017/S1930297500001698 Published online by Cambridge University Press

Table A2a: Conditional results for Result 1. The dummy variable “Reward” takes the value 1 when the worker plays *reward*. The dummy variable “Kind” takes the value 1 when the manager plays *kind*. The dummy variable “Unkind” takes the value 1 when the manager plays *unkind*. In probits, the reported figure is the estimated marginal effect. All models contain clusters at the individual level. Estimated period/group fixed effects are omitted. All models exclude the 49 observations corresponding to the manager playing neutral. Asterices denote statistical significance (* = .10, ** = .05, *** = .01).

Model		1	2	3	4
Data included		One-shot	One-shot	Reputation	Reputation
Estimation method		Regression	Probit	Regression	Probit
Dependent variable		Reward	Reward	Reward	Reward
Explanatory variables	Kind	0.43***	0.55***	0.64***	0.81
	P-value from z-test	< .01	< .01	< .01	< .01
	Unkind	0.04	0.07	−0.03	0.03
	P-value from z-test	.60	.56	.81	.88
R^2 / Pseudo R^2		0.35	0.36	0.58	0.52
Observations		396	396	195	195
Degrees of freedom		131	131	54	54

Table A2b: Conditional results for Result 1. In addition to the description below Table A1a: the dummy variable “Punish” takes the value 1 when the worker plays *punish*.

Model		1	2	3	4
Data included		One-shot	One-shot	Reputation	Reputation
Estimation method		Regression	Probit	Regression	Probit
Dependent variable		Punish	Punish	Punish	Punish
Explanatory variables	Kind	−0.11	−0.24 **	−0.15*	−0.22**
	P-value from z-test	.23	.03	.10	.01
	Unkind	0.39***	0.32***	0.13	0.08
	P-value from z-test	< .01	< .01	.17	.30
R^2 / Pseudo R^2		0.40	0.29	0.35	0.31
Observations		396	396	195	195
Degrees of freedom		131	131	54	54

https://doi.org/10.1017/S1930297500001698 Published online by Cambridge University Press

Experimental instructions

Welcome to our experiment in decision making.

If you read these instructions carefully and make good decisions, you may earn a considerable amount of money. At the end of the experiment, your earnings will be paid to you, privately and in cash.

At the beginning of the experiment, you will be randomly separated into groups of 8. You will only interact with your group members. Three of you will be randomly assigned the role of Red and five of you the role of Blue. So there are more Blues than Reds. [NEXT]

At the beginning of the experiment, both Reds and Blues get 200 points each as a show-up fee. We will convert the points you earned into dollars at the rate of 10 points = $1. [NEXT]

The experiment has a number of rounds. Each round has two stages.

Stage 1:

- The Reds take turns to individually choose a Blue. The order in which the Reds get to choose a Blue is randomly determined.

- This results in 3 Red-Blue pairs with2 Blues left unmatched.

- The 2 Blues who have not been chosen in this round do nothing and their final earnings for the round are -25 points. [NEXT]

- Each Red chooses LEFT, MIDDLE or RIGHT.

  ○ If a Red chooses LEFT: Red's earnings are -10 points, Blue's earnings are +15 points.

  ○ If a Red chooses MIDDLE: Red's earnings are 0 points, Blue's earnings are 0 points.

  ○ If a Red chooses RIGHT: Red's earnings are +15 points, Blue's earnings are -15 points. [NEXT]

Stage 2:

- Each Blue learns if they are in a pair with a Red.

- If they are in a pair, they will learn the action their Red partner chose among LEFT, MIDDLE and RIGHT, and the corresponding earnings in points.

- The paired Blues will choose 1 of 3 actions:

  ○ To add 20 points to their Red partner at the cost of 5 points to them.

  ○ To subtract 20 points from their Red partner at the cost of 5 points to them.

  ○ Do nothing at zero cost. [NEXT]

- Red's final earnings will be changed by their Blue partner's choice to add or subtract. If Blue chose to subtract, then Red's earnings decrease by 20 points. If Blue chose to add, then Red's earnings increase by 20 points. In both cases Blue's earnings decrease by 5 points.

- If Blue does nothing, then Red's and Blue's final earnings are the points initially decided by the action chosen by the Red partner.

- Blues who have not been chosen in this round do nothing and their final earnings for the round are -25 points. [NEXT]

That's a round. The experiment will last a number of rounds. [NEXT]

Reputation treatment

- Reds and Blues will have IDs (e.g., Red 2 or Blue 4). Both Reds and Blues always keep the same ID.

- After every round, everyone will see the actions chosen by the 3 Red-Blue pairs up to and including that round.

https://doi.org/10.1017/S1930297500001698 Published online by Cambridge University Press

People's action choices are labeled by their ID.

We will begin the experiment now. Your role in the experiment will be decided in the next screen. After your role is decided, I will read the specific instructions for Reds and Blues. Please refrain from asking any questions until I finish reading these instructions. [NEXT]

Your ID for this round is Red ***. Your ID will be the same for all rounds. Specific instructions for Reds:

- The order in which you choose Blues as partners may change every round.
- If you are the first, you can pick any of the 5 Blues.
- If you are not the first, then you only pick from the Blues that were not chosen before.
- The 2 Blues that are not picked will earn -25 points for that round.
- Remember: both Reds' and Blues' IDs are always the same throughout the experiment.
- After choosing a Blue counterpart for the first stage, we ask that you choose LEFT, MIDDLE or RIGHT. The corresponding earnings are shown in the supplementary Figure.
- You will find out your Blue partner's choice. [OK]

Now the Blues. Your ID for this round is Blue ***. Your ID will be the same for all rounds. Specific instructions for Blues:

- If a Red picks you in a given round, you will find out which action your partner chose. The corresponding earnings are shown in the supplementary Figure.
- Then you will choose 1 action among 3 actions
  - ○ To add 20 points to your Red partner at the cost of 5 points.
  - ○ To subtract 20 points from your Red Partner at the cost of 5 points.
  - ○ Do nothing at zero cost.
- If a Red does not choose you, you will do nothing and your final earnings for this round are -25 points.
- Remember: both Reds' and Blues' IDs are always the same throughout the experiment. [OK]

One-shot treatment

- Reds and Blues will have IDs (e.g., Red 2 or Blue 4). Reds always keep the same ID. Blues get a random ID every round. (e.g. Blue 1 in round 1 may or may not be the same person as Blue 1 in round 2.)
- After every round, everyone will see the actions chosen by the 3 Red-Blue pairs up to and including that round. People's action choices are labeled by their ID.

We will begin the experiment now. Your role in the experiment will be decided in the next screen. After your role is decided, I will read the specific instructions for Reds and Blues. Please refrain from asking any questions until I finish reading these instructions. [NEXT]

Your ID for this round is Red ***. Your ID will be the same for all rounds. Specific instructions for Reds:

- The order in which you choose Blues as partners may change every round.
- If you are the first, you can pick any of the 5 Blues.
- If you are not the first, then you only pick from the Blues that were not chosen before.
- The 2 Blues that are not picked will earn -25 points for that round.
- Remember: Blues' IDs may change every round.
- After choosing a Blue counterpart for the first stage, we ask that you choose LEFT, MIDDLE or RIGHT. The corresponding earnings are shown in the supplementary Figure.

https://doi.org/10.1017/S1930297500001698 Published online by Cambridge University Press

- Remember: you keep the same ID throughout the experiment.
- You will find out your Blue partner's choice. [OK]

Now the Blues. Your ID for this round is Blue ***. Your ID may change every round. Specific instructions for Blues:

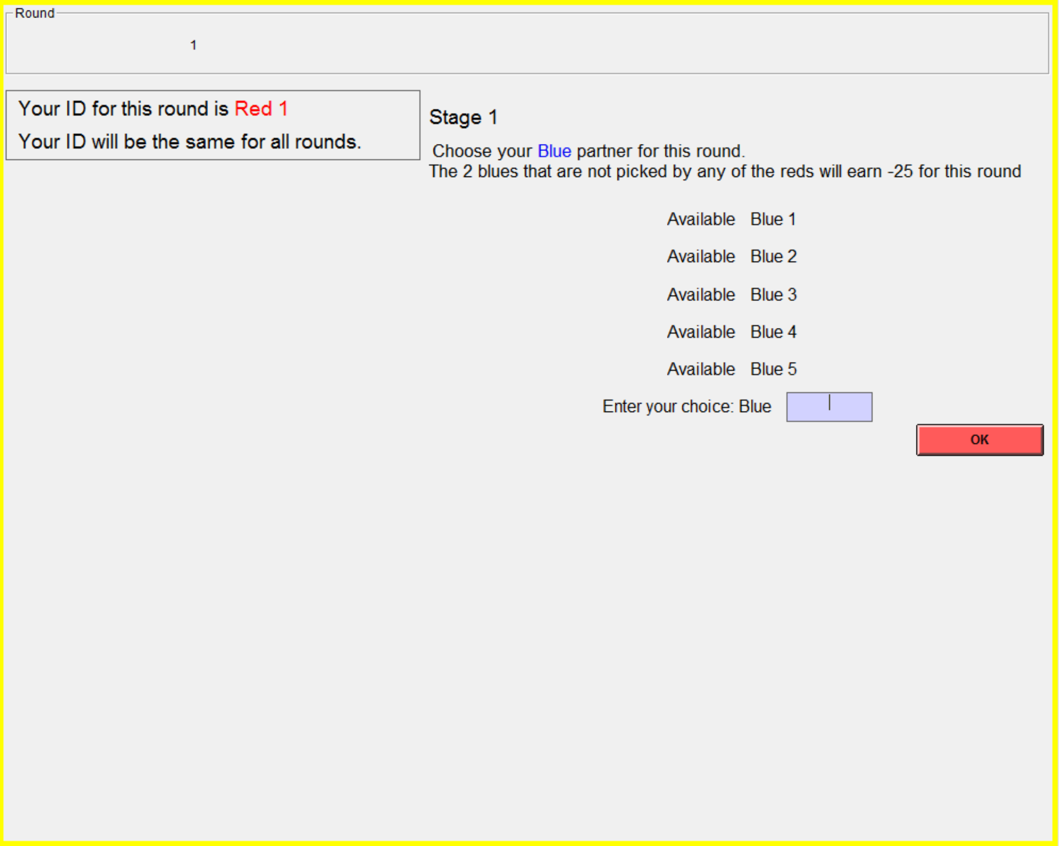
- If a Red picks you in a given round, you will find out which action your partner chose. The corresponding earnings are shown in the supplementary Figure.
- Remember: your ID may change every round.
- Then you will choose 1 action among 3 actions
  - ○ To add 20 points to your Red partner at the cost of 5 points.
  - ○ To subtract 20 points from your Red Partner at the cost of 5 points.
  - ○ Do nothing at zero cost.
- If a Red does not choose you, you will do nothing and your final earnings for this round are -25 points.
- Reds always keep the same ID. [OK]

Check if there are any questions. If not, the game should start. If anyone asks about the number of rounds, the experimenter should simply repeat:

"The experiment will last a number of rounds."

The experiment should go on for 11 rounds or 70 minutes — whatever comes first. That leaves 20 minutes for paying people etc.

Sample screenshot of a manager's choice:



Round

1

Your ID for this round is Red 1

Your ID will be the same for all rounds.

Stage 1

Choose your Blue partner for this round.
The 2 blues that are not picked by any of the reds will earn -25 for this round

Available Blue 1

Available Blue 2

Available Blue 3

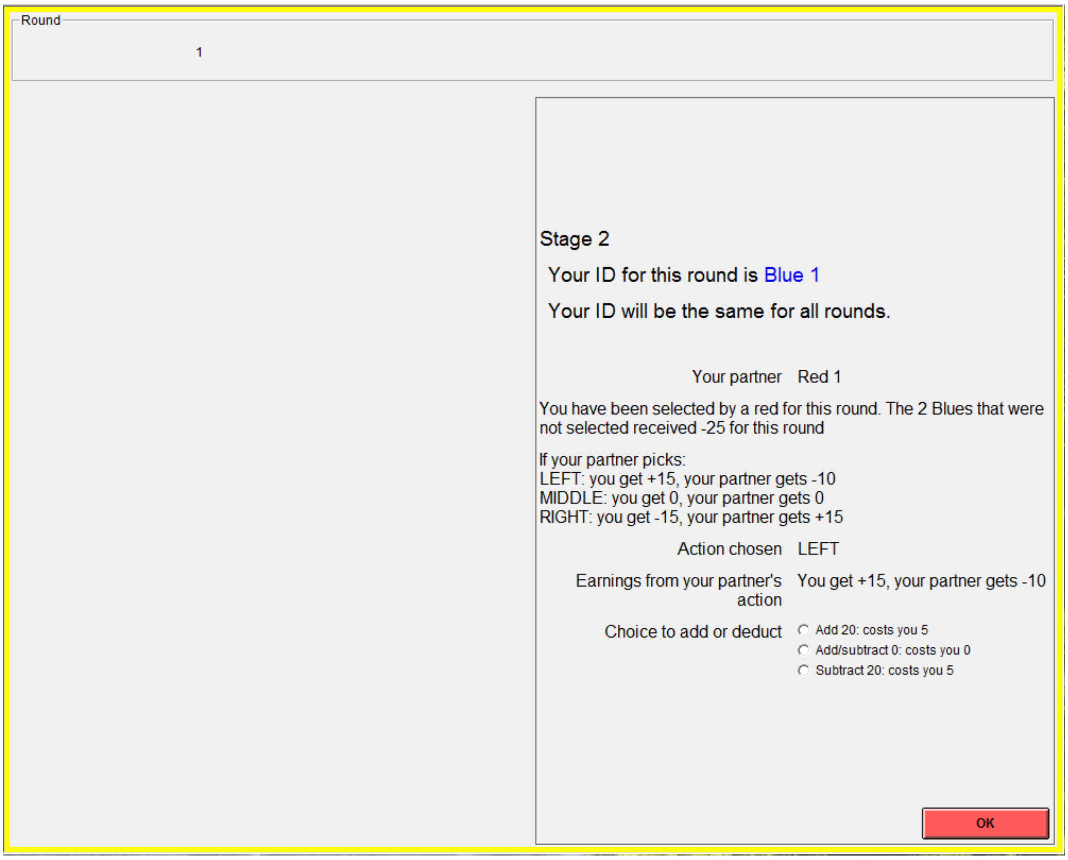
Available Blue 4

Available Blue 5

Enter your choice: Blue

https://doi.org/10.1017/S1930297500001698 Published online by Cambridge University Press

Sample screenshot of a worker's choice:



Stage 2

Your ID for this round is Blue 1

Your ID will be the same for all rounds.

Your partner Red 1

You have been selected by a red for this round. The 2 Blues that were not selected received -25 for this round

If your partner picks:
LEFT: you get +15, your partner gets -10
MIDDLE: you get 0, your partner gets 0
RIGHT: you get -15, your partner gets +15

Action chosen LEFT

Earnings from your partner's action You get +15, your partner gets -10

Choice to add or deduct
○ Add 20: costs you 5
○ Add/subtract 0: costs you 0
○ Subtract 20: costs you 5

https://doi.org/10.1017/S1930297500001698 Published online by Cambridge University Press