Forthcoming: Econometrica

Reciprocity as a Contract Enforcement Device Experimental Evidence

Ernst Fehr*, Simon Gächter*, Georg Kirchsteiger**

Abstract: Numerous experimental studies indicate that people tend to reciprocate favors and punish unfair behavior. It is hypothesized that these behavioral responses contribute to the enforcement of contracts and, hence, increase gains from trade. It turns out that if only one side of the market has opportunities for reciprocal responses, the impact of reciprocity on contract enforcement depends on the details of the pecuniary incentive system. If both sides of the market have opportunities for reciprocal responses robust and powerful reciprocity effects occur. In particular, reciprocal behavior causes a substantial increase in the set of enforceable actions and, hence, large efficiency gains.

Revised Version: September 1996

JEL-Classification No.: J33, J41, J64, C91, C92

Keywords: Contract Enforcement, Reciprocity, Moral Hazard, Principal Agent Theory

* Institute for Empirical Research in Economics - University of Zurich
Blümlisalpstrasse 10
CH-8006 Zurich
Switzerland
Fax: +41-1-3640366
fehriew@iew.unizh.ch
gaechter@iew.unizh.ch

** University of Vienna
Department of Economics
Hohenstaufengasse 9
A-1010 Vienna
Georg.Kirchsteiger@univie.ac.at

This paper is part of a research project on the impact of social norms on wage formation which is financed by the Swiss National Science Foundation under the project no. 12-43590.95. We would like to thank Douglas Gale, three anonymous referees, Klemens Binswanger, Jordi Brandts, Josef Falkinger, Rebecca Morton, Dieter Pfaff, Jan Potters, Reinhard Selten and participants at meetings of the Econometric Society, the Economic Science Association, the Amsterdam Workshop on Experimental Economics, the Verein für Socialpolitik and seminar participants at the Universities of Berlin, Linz, Vienna, and Zurich for encouraging comments. Research assistance by Martin Brown, Armin Falk, Urs Fischbacher, Jean-Robert Tyran and Paolo Vanini is gratefully acknowledged.

1. Introduction

Contracts are a core element of market economies. Without voluntary agreements, i. e. contracts, there would be no market and without the enforcement of agreements parties would have no reason to conclude them. The problem of contract enforcement is, therefore, a central issue of the functioning of market economies. During the last two decades economic theory has made much progress in the understanding of the *endogenous* enforcement of contracts. It has been shown that informational asymmetries and the absence of third parties who enforce contracts exogenously have important economic consequences. They give rise to incentive compatibility requirements that impose limits on the set of enforceable contracts. These limits will, in general, move the economy away from first best allocations and often even do not allow the achievement of constrained Pareto optima (Laffont and Maskin 1982, Grossman and Hart 1983, Hart and Holmström 1987, Milgrom and Roberts 1992).

The standard approach to the enforcement of contracts derives incentive compatibility constraints under the assumption of fully rational and selfish individuals. In this paper we argue that the exclusive reliance on selfishness and, in particular, the neglect of reciprocity motives may lead to wrong predictions and to wrong normative inferences. We argue that reciprocal behavior may cause an increase in the set of enforceable contracts and may thus allow the achievement of non-negligible efficiency gains.

Our starting point is a recently developed model of reciprocal behavior (Rabin 1993). Rabin shows that the interactions of reciprocally motivated agents may produce outcomes that differ significantly from the predictions of a model that is based on purely selfish behavior. Reciprocity motives may, for example, generate a cooperative outcome in a one shot prisoners' dilemma. If the set of enforceable contracts is limited by incentive compatibility constraints the contracting parties are in a situation that is similar to a prisoners' dilemma. In principle they could agree on a Pareto-superior contract that is not incentive compatible. Yet, since such an agreement violates the incentive compatibility constraints it is not in the interests of the parties to meet their obligations.

Whether reciprocity motives are sufficiently strong to overcome contract enforcement problems is an empirical question. To examine this question we have developed an experimental design which allows for the isolation of reciprocity effects on contract enforcement. We conducted a series of market experiments in which reciprocal motivations and interactions could potentially ease incentive compatibility constraints. That reciprocity motives may be an *empirically* relevant factor in the enforcement of contracts is suggested by two types of observations. **First**, in Fehr, Kirchsteiger and Riedl (1993, 1996), Fehr, Kirchler, Weichbold and Gächter (1994) and in Berg, Dickhaut and McCabe (1995) it has been shown that generous behavior often induces reciprocal responses. Recipients of a gift frequently respond by being generous to those who give the gift. We hypothesized that in the context of contract enforcement reciprocal responses might increase the set of enforceable

a verifiable contract violation. This constraint on the maximum feasible fine generates a constraint on the enforceable effort level. We hypothesized that by offering generous contract terms firms in the WRT could be able to induce workers to choose effort levels above the level that is enforceable by the maximum fine. Likewise, in the SRT firms' reciprocity and its anticipation by the workers may generate even higher effort levels than in the WRT.

The results of the WRT experiments indicate that firms' contract offers are affected by reciprocity considerations. The number of generous offers in the WRT is considerably higher than in the NRT. In addition, the generosity of employment offers increases with the effort level firms would like to enforce. However, in the WRT the strength of workers' reciprocal responses is strongly affected by the details of the pecuniary incentive environment. While in our main experimental condition we observe only weak reciprocal responses there are other conditions in which workers exhibit strong reciprocity. In contrast, in the SRT we observe a very strong impact of reciprocity on contract enforcement irrespective of the details. There is strong evidence that firms behave reciprocally, that is, they punish shirking workers and reward those who fulfill the contract. This provides incentives for those workers who are not or only weakly motivated by reciprocity considerations. Our data indeed show that workers anticipate firms' reciprocity and shirk much less than in the WRT. Furthermore, firms demand and enforce much higher effort levels than in the WRT. As a result, both, workers and firms, are better off in the SRT compared to the WRT. *Therefore, the data suggest that if both parties in a trade have the opportunity to reciprocate, reciprocal motivations have a robust and very powerful impact on the enforcement of contracts.*

The rest of the paper is organized as follows: In the next section we present a simple labor market model which provides the basis for our treatment conditions. In Section 3 we discuss the implications of reciprocity in the WRT and the SRT. In Section 4 we present our experimental procedures. Section 5 shows the regularities in the data. In Section 6 additional evidence regarding the robustness of our results is reported. The final section provides a summary and concludes with some remarks about the implications of our results for principal agent theory.

2. A simple labor market with moral hazard

2.1 The two stage design

In this section we present a simple one-period labor market model in which workers can underprovide effort relative to firms' desired effort level. Firms have the opportunity to fine workers who underperform. The implementation of this model in the laboratory constitutes the WRT. We assume that there are L identical risk neutral workers and $N < L$ risk neutral firms in the labor market. Each firm can employ at most one worker and each worker can accept at most one employment offer. The

if she shirks. For any offer $(w, \tilde{e}, f)$ a rational employee will never perform $e > \tilde{e}$ because of $c'(e) > 0$. On the other hand, if the worker prefers to shirk she will always shirk fully $(e = 0)$ because whether she has to pay the fine f does not depend on the amount of shirking. She will exactly perform $\tilde{e}$ if $u^{ns} \geq u^{s}$ holds. This yields

(2) $$sf \geq c(\tilde{e}) .$$

The firm's profit from trading with a nonshirking worker is given by

(3) $$\pi = (q - w)\tilde{e}.$$

q is an exogenously given redemption value. Note that by offering $w \leq q$ firms can always rule out losses irrespective of whether the worker shirks. If we had instead implemented a more common profit equation, for example $\pi = qe - w$, firms would have suffered losses in case of shirking. In our main experiment we ruled out losses to prevent that subjects' behavior is affected by loss aversion[3]. This allows us to study pure reciprocity effects. In subsequent experiments we allowed for the interaction of loss aversion and reciprocity effects because firms could make losses. These experiments are described in more detail in Section 6.

If there are no restrictions on the firms' choices of f they can enforce any effort level in $[0,1]$. Since we are interested in the question to what extent reciprocity provides a mean for the efficiency enhancing enforcement of contracts we have to create an (experimental) environment in which firms face an enforcement problem. This is done by restricting f by an exogenous upper bound f^{0}.[4] The existence of f^{0} implies a maximum enforceable effort level of e^{0}. e^{0} obeys the equation $sf^{0} = c(e^{0})$. We assume that f^{0} is sufficiently low to ensure $e^{0} < 1$ and that $e > e^{0}$ is more profitable than $e \leq e^{0}$.

What are the terms of the equilibrium contract in case that $f \leq f^{0}$ is binding? Notice that it is never profitable to provoke shirking because in that case $\pi = 0$. Hence, each firm demands the maximum enforceable effort level e^{0} and pays the reservation wage $w^{r} = c(e^{0})$ that corresponds to e^{0}. The equilibrium offer is thus given by

(4) $$w^{*} = c(e^{0}),\ \tilde{e}^{*} = e^{0} = c^{-1}(sf^{0}),\ f^{*} = f^{0}.$$

---

3 A number of experiments indicate that subjects behave differently when losses can or will occur (see e.g. Kahneman and Tversky 1979 and Tversky and Kahneman 1992).

4 It is perhaps worthwhile to emphasize that restrictions on f are not just an experimental tool for the isolation of reciprocity effects. The real world is frequently characterized by constraints on firms' sanction opportunities. Such constraints may be imposed by law or by collective bargaining agreements. They may even arise endogenously because monitoring technologies may not allow the measurement of effort without error or because of problems of firms' moral hazard. In our experiments a firm did not have the opportunity to claim $(e < \tilde{e})$ although the worker met the effort requirement. Yet, in reality this is of course possible.

willingness to pay for *responding* fairly (unfairly) to a behavior that is perceived as fair (unfair). Whether an action is perceived as fair or unfair depends on the distributional consequences of the action relative to a neutral reference action.

In the following we derive the *qualitative*[5] implications of reciprocity for our two stage labor market. Since there is an excess supply of workers in this market firms are in a strong position because they need not pay positive job rents to workers. Therefore, the voluntary payment of job rents signals a certain generosity. We hypothesize that the higher the job rent the higher will be, on average, the perceived generosity of a contract offer. This means that workers who are motivated by reciprocity will derive a nonpecuniary utility gain if they provide effort above the level that is dictated by their pecuniary interest. As a consequence, by paying higher rents firms are able to elicit higher effort levels from reciprocating workers.

Of course, workers may differ in their preferences for reciprocal actions. For example, they may have different reference standards against which the generosity of a particular rent is compared. In case of heterogeneous reference standards not all workers will in general be willing to provide a given $\tilde{e} > e^0$ for a given positive rent. However, the fraction of workers who is willing to choose $\tilde{e} > e^0$ will in general increase if r is increased. Thus, in the presence of reciprocal preferences we should observe that the probability of nonshirking is positively related to r.

The pecuniary incentive to shirk is the larger the larger $\tilde{e}$. In the presence of reciprocal preferences firms can compensate for this increased incentive to shirk by raising r. Therefore, if firms anticipate a sufficiently steep positive relation between the probability of nonshirking and r they have a reason to pay higher job rents if they demand higher effort levels. When firms are fine tuning their $\tilde{e}$ -choices we should thus observe a positive relation between $\tilde{e}$ and r. Notice that this contrasts sharply with the prediction of the standard model that job rents are zero irrespective of the level of $\tilde{e}$.

If firms anticipate reciprocal effort responses in the WRT we should also observe higher rents in the WRT than in the NRT because in the NRT effort is exogenously fixed. In particular, we should observe that those job rent levels which successfully trigger higher effort levels are more frequently chosen in the WRT. A final prediction concerns the relation between firms' profits and their contract

---

5 At present there does not exist a general theory that allows to precisely locate reference standards. Nor do there exist empirical methods for the exact determination of reference points. This makes precise *quantitative* predictions of behavior that depends on reference standards difficult. Rabin (1993 p. 1286), for example, admittedly introduces "a crude reference point against which to measure how generous" an action is. Yet, as long as one accepts that an action is perceived the more generous the more resources a person gives up in favor of another person, it is possible to make qualitative predictions. Notice that reference points are parameters in the utility function of a person with reciprocity motives. Therefore, our problem is not different from the general problem of deriving *exact* quantitative predictions in the absence of exact knowledge of people's preferences. For this reason predictions in applied economics are to a large extent predictions of the sign of comparative static results, i.e. qualitative predictions.

4. Parameters and experimental procedures

In total we conducted eighteen experimental sessions.[7] Four sessions (S1 - S4) implemented the NRT, in six sessions (S5 - S10) we conducted the WRT and in two sessions (S11 - S12) the SRT was implemented. To investigate the sensitivity of behavior with regard to several design features we conducted additional six sessions. In S13 - S16 we changed firms' payoff function such that losses could occur. In sessions S17 - S18 we analyzed subjects' ability to perform backward induction in the presence of reciprocity motives. The results of S1 - S12 are presented in Section 5 while the results of S13 - S18 will be reported in Section 6.

In the WRT a trading period consisted of the two stages of our model of Section 2.1 and a session lasted for 16 periods. The NRT sessions lasted for 16 periods, too. In the SRT a trading period consisted of the three stages described in Section 2.2. Time constraints forced us to conduct only 12 trading periods in the SRT. In all sessions there was one trial period which allowed subjects to become familiar with the trading institution. The participants were student volunteers mainly from the universities of technology in Vienna and Zürich (computer scientists, engineers, etc.). They were recruited with the announcement that, depending on their decisions, they could earn a considerable amount of money during the experiment. In general we had 8 workers and 6 firms.[8]

Since reciprocal behavior is triggered by the distributional impact of actions, our experimental design allowed both parties of a given trade to compute the monetary gains of their trading partner. This information condition was implemented by rendering the parameters of the experiments common knowledge. Thus, q, s, N, L, the cost function $c(e)$, and the exogenously determined and enforced value of $\tilde{e}$ in the NRT were common knowledge. To ensure that subjects were able to compute the monetary gains, they had to compute their own gains and the gains of their partner in three hypothetical examples before the experiment started. All subjects solved these exercises correctly.

To rule out the possibility of reputation formation and of rewarding or punishing a subject's previous behavior, the identities of the trading partners were not revealed; exchange took place between anonymous agents. To ensure the anonymity of the trading partners, firms and workers were located in two separate rooms and the messages between these rooms were transmitted via telephone. To exclude any kind of group pressure other workers or other firms were not informed about a worker's effort choice; only the worker's firm was informed about e.

In the two-stage experiments we chose the following parameters: $q=120$, $f^0 = 10$ and an effort cost schedule according to Table 1a. Each subject received an initial endowment of 70 Austrian Schillings

---

7 A highly compressed version of the experimental instructions is presented in the appendix. A full set of instructions is available upon request.

8 Unfortunately some subjects who had signed up for the experiment did not show up in S5 and S9. In S5 the worker-firm relation was 6:4; in S9 it was 7:5.

5. Experimental results

In this section the results of our main sessions (S1 - S12) are presented.[11] In our four NRTs (S1 - S4) there were 384 potential trades while in our six WRTs (S5 - S10) there were 540 potential trades. The number of actual trades amounted to 353 in the NRT and to 509 in the WRT. In our two SRT sessions (S11 - S12) there were 144 potential trades all of which have been realized. Subjects' average earnings (net of show up fees) were 173 ATS in the NRT, 148 ATS in the WRT while in the SRT they earned 440 ATS on average. NRT-sessions lasted on average 1.5 hours, WRT-sessions lasted 2.5 hours and SRT-sessions lasted on average 3.5 hours.[12] In the following we present first the results of the WRT and compare them with the data pattern in the NRT. After that we compare the SRT with the WRT.

5.1 Regularities in the Weak Reciprocity Treatment

Our first result concerns the effort demanded by firms in the WRT.

R1: Firms persistently tried to induce effort levels above the risk neutral subgame perfect equilibrium level of $e^0 = 0.1$. Yet, due to high shirking rates the actual average effort is close to e^0.

To provide evidence for R1 we depicted the evolution of the average effort demanded in each period in Figure 1 . It is obvious from Figure 1 that in all periods firms demanded, on average, effort levels above $e^0 = 0.1$. In particular, at the beginning and towards the end of a session firms tried to induce relatively high effort levels which - in case of risk neutral workers - were only incentive compatible if the maximum fine f^0 were twice as high as in our experiments. The actual average effort is, however, much lower than $\tilde{e}$. During the first twelve periods it is slightly below, during the last four periods it is slightly above e^0.

INSERT FIGURE 1 HERE

The fact that firms demand $\tilde{e} > e^0$ is in itself no unambiguous evidence for the impact of reciprocity on contract terms. $\tilde{e} > e^0$ may also be caused if firms expected workers to be risk averse because risk averse workers are willing to perform above e^0. Table 2, however, casts doubt on the assumption of risk aversion. Among other things the table shows the number of trades that occurred at each level of $\tilde{e}$ together with the percentage of trades in which workers shirked. If workers are risk averse we should observe no shirking[13] at levels of $\tilde{e}$ that are at or slightly above $e^0 = 0.1$. Yet, at $\tilde{e} = 0.2$ the shirking rate is 64 percent and at $\tilde{e} = 0.1$ it is 55 percent. At $\tilde{e} = 0.064$ the shirking rate is still 22

---

11 Our data are available upon request.

12 This includes the time spent on reading and understanding the instructions.

13 This argument assumes that firms impose the maximum fine. In more than 95 percent of all WRT-trades this was the case.

Notice that Figure 2 understates workers' true reciprocity because there is a positive relation between r and $\tilde{e}$. This means that a rise in r is associated with an increase in the pecuniary incentive to shirk. To examine the ceteris paribus impact of r on the probability of nonshirking, we ran the following probit regressions:

$$\theta_{ti} = \beta_1 + \beta_2\left(u_{ti}^{ns} - u_{ti}^{s}\right) + \beta_3 r_{ti} + \beta_4 \sum_{j=1}^{t-1} u_{ji} + \varepsilon_{ti} \tag{6}$$

where $\theta_{ti} = 1$ ($\theta_{ti} = 0$) in case that worker i does not (does) shirk in period t. u^{ns} - u^{s} measures the pecuniary incentive to perform $\tilde{e}$ while r measures the nonpecuniary incentive to provide $\tilde{e}$ that arises from reciprocity motives. To control for wealth effects we also included the sum of worker i's earnings up to period t as a regressor. In Table 4 the results of regression (6) are presented. As Table 4 shows, β_2 and β_3 have the expected sign. In particular, β_3 is always positive. It is significant at the ten percent level in four sessions and below the 0.1 percent level for the data of all sessions. The results of regressions (6) and Figure (3) indicate that on average workers respond reciprocally.

INSERT FIGURE 2 HERE
INSERT TABLE 4 HERE

On the basis of regression (6) we can compute firms' expected profits $E\pi$ for any combination of r and $\tilde{e}$. $E\pi$ is given by $E\pi = F(\theta)[q - r - c(\tilde{e})]\tilde{e}$ where F is the standard normal distribution[15]. If we plug the result of regression (6) for S5-S10 into $E\pi$ we can examine whether workers' reciprocation is sufficient to render the payment of higher rents in case of higher $\tilde{e}$ a rational strategy. Computing the $E\pi$-maximizing level of r for different levels of $\tilde{e}$ shows that it is indeed a rational strategy to pay a higher r if one demands a higher $\tilde{e}$. In the light of this fact R2 suggests that in the process of fine tuning their $\tilde{e}$-choices firms took advantage of workers reciprocity by adjusting their rents accordingly.

R2 and R3 suggest that if firms demanded high effort levels they appealed to workers' reciprocity by paying high rents. Yet, it does not inform us about the frequency and the strength of firms' overall appeal to workers' reciprocity. To shed more light on the overall importance of reciprocity we have to compare the WRT with the NRT. The fact that reciprocity cannot play a role in the NRT leads to the conjecture that on average less generous offers will be observed in the NRT. Moreover, since Figure 2 suggests that the biggest impact of rents on the average effort occurs for rent levels between 20 and 40 we expect that the relative frequency of rent offers above 20 is larger in the WRT than in the NRT:

R4: On average firms pay higher rents in the WRT compared to the NRT. Moreover, the percentage of rents above 20 is larger in the WRT.

---

15 Notice that u^{ns} - $u^{s} = sf - c(\tilde{e})$ while w can be written as $w = r + c(\tilde{e})$. Thus if $f = f^0$ $E\pi$ can be considered as a function of only $\tilde{e}$ and r.

maximizing levels of $\tilde{e}$ it turns out that for rents below 20 $\tilde{e}$ = 0.2 is the best strategy while for rents between 20 and 40 $\tilde{e}$ =0.3 and 0.4 is the best strategy. Moreover the latter strategy generates a higher $E\pi$ than the former.[17] In contrast, $E\pi$ is never maximized at $e^0 \leq 0.1$ irrespective of the level of r.

When interpreting the results of these computations one should keep in mind that individual firms have much less information about workers' behavior than the econometrician who estimates regression (6). Therefore, it is not surprising that we observe a wide variety of $\tilde{e}$ and r levels in the experiment that do not maximize $E\pi$. However, the fact that during the last periods of the WRT we observe a shift towards rents above 20 and towards higher $\tilde{e}$ levels indicates that firms moved in the direction of more profitable strategies.

The regularities of the WRT suggest that workers' exhibited enough reciprocity to render a high effort - high rent strategy profitable and that firms' contract offers were affected by the anticipation of workers' reciprocity. However, although strategies that appealed to workers' reciprocity were on average more profitable the impact of reciprocity on the average effort is not overwhelmingly strong. This raises the question whether two-sided reciprocity generates a larger increase in the set of enforceable effort levels.

5.2 Regularities in the Strong Reciprocity Treatment

The main result of the SRT is

R 6: Firms demand and enforce significantly higher effort levels in the SRT compared to the WRT.

Figure 1 makes these differences between SRT and WRT transparent. It is obvious that firms demanded considerably higher effort levels in the SRT. While in the SRT the average $\tilde{e}$ always was above 0.8 and converged towards the maximum effort of $e = 1$, it was - except for the first three periods - always below 0.4 in the WRT. These behavioral differences across treatments were also present at the level of individual firms. We have computed the average effort demanded for each firm in each session. It turned out that the *highest* average $\tilde{e}$ among all firms in the WRT was well *below* the *lowest* average $\tilde{e}$ of all firms in the SRT. With regard to the actual average effort there also was a strikingly large difference between the two treatments (see Figure 1). In the SRT the actual average effort per period was almost always above 0.7 while in the WRT it almost never exceeded 0.1. As in the case of $\tilde{e}$ these differences across treatments could also be observed at the level of individual firms. The *highest* average effort received by WRT-firms was well *below* the *lowest* average effort received by a SRT-firm. Thus, there is unambiguous support for R6.

---

17 On the basis of regression (6) $\tilde{e}$ = 0.3 is in general slightly better than $\tilde{e}$ = 0.4. The maximum expected profit of $E\pi$ = 8.7 is attained at r = 49 and $\tilde{e}$ = 0.3. However, all rent levels above 33 (at $\tilde{e}$ = 0.3) yield $E\pi > 8$.

underprovision of effort $[(\tilde{e} - e)/\tilde{e}]$ is lower in the SRT. In the SRT we observe also more frequently that workers provide excess effort.

INSERT TABLE 6 HERE

What are the welfare implications of the introduction of a third stage? The enforcement of higher effort levels in the SRT implies that the total cake that is available for distribution is larger than in the WRT. This raises the question whether the introduction of the third stage leads to a Pareto improvement. The answer is given by

R10: In the SRT both workers and firms are, on average, better off compared to the WRT.

To substantiate R10 we computed workers' and firms' actual average gains from trade. In the WRT workers earned on average 17 ATS from a trade while in the SRT they earned 24 ATS. The firms' increase in the gains from a trade was even larger. In the WRT they reaped 6 ATS on average compared to 42 ATS in the SRT. The statistical significance of these differences is confirmed by a robust rank order test (p-value < 0.009 for the workers, p-value < 0.0001 for firms). The data indicate, however, not only a Pareto improvement if one takes the average over all observations of the WRT and the SRT. Even if one compares the average gains by period there is a clear pattern. Figure 4 shows that except in period 9 workers' gains from trade are higher in the SRT than in the corresponding period of the WRT. Even more impressive is the firms' increase in gains from trade. In each period of the SRT firms earned between three and seven times more than in the corresponding period of the WRT. Therefore, our data provide rather strong evidence in favor of a Pareto improvement.

INSERT FIGURE 4 HERE

6. The robustness of reciprocity effects

In this section[19] we present the results of S13 - S18. S13 - S16 deal with the issue of losses and risk. S17 - S18 investigate the issue of backward induction in the presence of reciprocity considerations. In S1 - S12 we have ruled out the possibility that firms can make losses to prevent the interaction of loss aversion and reciprocity effects. Our solution had the disadvantage that for $e < 1$ wages did no longer represent pure transfer payments and, hence wage increases led to a rise in the sum of payoffs[20]. In accordance with most principal-agent models we also implemented a random effort verification

19 Space limitations prevent us from a more detailed presentation of the evidence on robustness. Upon request we send interested readers an earlier version of this paper in which we deal with the robustness issue in more detail.

20 The sum of u^{ns} and π is given by $q\bar{e} - c(\bar{e}) + w(1-\bar{e})$

the lowest effort level in case that they shirk whereas in the modified WRT *partial* shirking is very frequent.

INSERT FIGURE 5 HERE

For space limitations we omit a detailed presentation of the other regularities of our modified WRT and SRT. Qualitatively, both, the modified and the original design produce very similar patterns: There is a strong positive correlation between $\tilde{e}$ and r as well as between e and r. This correlation is present at the aggregate and at the individual level. Firms punish shirking and reward $e \geq \tilde{e}$ and workers anticipate firms' reciprocity. In addition, the shirking rate is much lower in the three-stage design. For this reason the actual average effort is significantly higher in the modified SRT compared to the modified WRT. Finally, in both modified treatments gains from trade are much larger than predicted by the standard model. In our view these results provide strong evidence that the reciprocity effects we detected in our main experiments are also present in our modified design. In the modified WRT the reciprocity effects are even stronger while in the modified SRT they are equally strong. This suggests that in the WRT reciprocity is more easily affected by the details of the pecuniary incentive system.

In sessions S17 and S18 we conducted three stage experiments to examine subjects' capability to perform backward induction. Remember that our reciprocity predictions rest on the assumption that subjects are able to correctly perform the required backward induction. While this may be easy in the WRT it requires quite a bit of sophistication in the SRT. A sceptic might thus argue that deviations from the standard prediction are not due to reciprocity but to the lack of backward induction. To shed some light on the validity of this argument we varied the costs $k(p)$ in S17 and S18: In the first six periods of S17 we increased the costs $k(p)$ by a factor of five while in periods 7-12 we implemented the $k(p)$-schedule as given in Table 1b. To control for order effects we reversed the order of high and low $k(p)$-costs in S18. If subjects are indeed not capable to perform the required backward induction the change in $k(p)$-costs should have no systematic effects on behavior. Yet, if subjects correctly anticipate a positive, but limited, willingness to pay for reciprocal acts the cost change should affect behavior systematically in the following way: (i) Firms should punish and reward less at stage three. (ii) Therefore, workers should shirk more at stage two. (iii) As a consequence, firms who perform the backward induction should lower $\tilde{e}$.

All three predictions are met by the data. Firms propensity to reciprocate is lower in the high cost condition. Workers anticipate less reciprocity and, hence, shirk more and firms lower their $\tilde{e}$ levels. What is interesting, however, is that not all three behavioral changes occur in the first period after the cost change. While the reduction in firms' reciprocity and the increase in shirking takes place immediately, the lowering of $\tilde{e}$ takes a few periods. This suggests that workers immediately understand that there will be less punishment and rewarding in the high cost condition while firms need some feedback to learn the required backward induction.

References

Berg, J., J. Dickhaut and K. McCabe (1995): "Trust, Reciprocity and Social History", *Games and Economic Behavior* 10, 122-142.

Camerer, C. (1988): "Gifts as Economic Signals and Social Symbols", *American Journal of Sociology* 94, S180 - S214.

Camerer, C. and R. Thaler (1995): "Ultimatum Games", *Journal of Economic Perspectives* 9, 209 - 220.

Cameron, L. (1995): "Raising the Stakes in the Ultimatum Game: Experimental Evidence from Indonesia", Discussion Paper, Dept. of Economics, Princeton University.

Carmichael, L. and B. MacLeod (1995): "Gift Giving and the Evolution of Cooperation", Discussion Paper, Queen's University.

Fehr, E., E. Kirchler, A. Weichbold and S. Gächter (1994): "When Social Norms Overpower Competition - Gift Exchange in Experimental Labour Markets", Discussion Paper, University of Zurich.

Fehr, E., G. Kirchsteiger and A. Riedl (1993): "Does Fairness prevent Market Clearing? An Experimental Investigation", *Quarterly Journal of Economics* 108, 437-460.

Fehr, E. and E. Tougareva (1995): "Do Competitive Markets with High Stakes Remove Reciprocal Fairness? - Evidence from Russia", Discussion Paper, University of Zürich

Fehr, E., G. Kirchsteiger and A. Riedl (1996): "Gift Exchange and Reciprocity in Competitive Experimental Markets", forthcoming: *European Economic Review*.

Greene, W. (1993): *Econometric Analysis*, 2nd ed., Prentice Hall.

Grossman, S. and O. Hart (1983): "An Analysis of the Principal-Agent Problem", *Econometrica* 51, 7 - 45.

Güth, W. and R. Tietz (1990): "Ultimatum Bargaining Behavior - A Survey and Comparison of Experimental Results", *Journal of Economic Psychology* 11, 417-449.

Güth, W. and M. Yaari (1992): "An evolutionary approach to explain reciprocal behavior in a simple strategic game" in *Explaining Process and Change - Approaches to Evolutionary Economics*, ed. by U. Witt, Ann Arbor 1992.

Güth, W. (1995): "An Evolutionary Approach to Explaining Cooperative Behavior by Reciprocal Incentives", *International Journal of Game Theory* 24, 323 - 344.

Hart, O. and B. Holmström (1987): "The Theory of Contracts", in *Advances in Economic Theory*, 5th World Congress of the Econometric Society, Cambridge University Press.

Hoffman, E., K. McCabe and V. Smith (1995): "The Behavioral Foundations of Reciprocity: Experimental Economics and Evolutionary Psychology", Discussion Paper, University of Arizona.

Hoffman, E., K. McCabe and V. Smith (1994): "On Expectations and Monetary Stakes in Ultimatum Games, Discussion Paper, University of Arizona.

Appendix

For space limitations we present only a shortened version of the instructions for the sellers (workers) in the three-stage experiment[22]. The sellers' questionnaire, the documentation sheet and the instruction summary for the sellers are not presented here. We avoided, of course, value loaden terms like effort, fine, penalty or reward. Instead we used terms like quality, price, commission, transformation factor, etc. In our instructions the difference between the price a seller (worker) gets in case that the desired quality (effort) is delivered and the price paid in case of verifiable underprovision of quality (effort) constitutes the fine f. The sum of the commission plus the price (in case that the desired quality is delivered) is tantamount with the wage w in the paper.

INSTRUCTIONS FOR THE THREE-STAGE EXPERIMENT

General Information (for both market sides):

We now give you a short outline of the experiment. Below you will get an exact description. The experimental subjects are either buyers or sellers. The experiment consists of a trial-trading day and 12 further trading days. In the trial-trading day you cannot earn money. This trial trading day allows you the get some experience for trading days in which you can earn money, so it is in your interest to take it seriously. Every trading day consists of three stages. At the **first stage** every buyer makes a bid which stipulates the conditions at which he is prepared to buy the experimental good from the seller. Such a bid consists of a price, a desired quality and a quality-independent commission. There are fourteen possible quality levels. The commission can be positive or negative. At **the second stage** a random mechanism determines the order in which the sellers can choose among the offers made. No seller is forced to accept an offer, and no buyer is forced to state an offer. This procedure ends if either all offers have been accepted or if all sellers have had the opportunity to choose an offer. After all offers have been accepted or every seller has had the opportunity to choose, every seller who accepted an offer must decide whether he delivers the desired quality or not. If a seller delivers the desired quality or a higher one, he gets the accepted price plus the commission. If a seller delivers a lower quality as desired this can be verified with a probability of 50 %. In case that too low a quality is verified the seller only gets a **"fixed price"**, which is determined by us, plus the commission. If it cannot be verified that too low a quality has been chosen the seller will get the price and the commission stated in the accepted offer. **Whether a quality below the desired quality can be verified, will be told to the seller after he has made his quality decision.** Right after the seller has determined his quality level, the respective buyer will be informed. At the **third stage** the buyer determines a transformation factor, which affects the actual gains at this trading day. Then the trading day ends and the next one will start.

There are more sellers than buyers and everybody knows this. Every seller (buyer) can only sell (buy) one commodity per trading day. A detailed description of each stage, i.e. which choice opportunities will be available, and how the gains are calculated, will be given below. At the end of this instructions you will find a control questionnaire, Sheet 2, which serves documentation purposes, and an Instruction Summary, which summarizes important information.

price in any case. **After** you have determined your actual quality, and noted it on Sheet 2, we will told you - by ticking the respective box on Sheet 2 - whether the actual quality can be verified or not. **This entry is only important for you if your actual quality is below the desired quality!** If you have chosen the desired quality, or a higher one, you will get the accepted price in any case. Your accepted bid will now be deleted from the blackboard and the next seller is free to choose among the remaining bids.

If you do not sell a good on a trading day, you will get ATS 10,- from us. All sellers have the same cost conditions.

Your acceptance and your actually chosen quality will be transmitted by us via telephone to the buyers' room. Buyers are not told who has chosen which quality, nor which seller has accepted a certain bid.

If "your" buyer has been informed about which quality you have actually chosen, the **third stage** of the experiment begins. The gains you have made in the first two stages are measured in units of experimental money. By his choice of a **transformation factor (TF)** "your" buyer now decides how many Austrian Schillings you will get for a unit of experimental money. **Which TF can be chosen by the buyer is indicated on your Instruction Summary!** The choice of a TF causes the buyer costs which are also indicated on the Instruction Summary. At the beginning of the third stage, the buyer gets from us ATS 10,-, which he can use for covering the costs of TF.

If "your" buyer, for example, chooses a TF of 0.3, this costs him ATS 7,- and one unit of experimental money is worth ATS 0.3 for you. If "your" buyer chooses, for example a TF of 1.5 he has costs of ATS 5,- and one unit of experimental money is worth ATS 1.5.

If your actual quality is higher or equal to the desired quality, "your" buyer is only allowed to choose a TF above or equal to 1. If your actual quality falls short of the desired quality, "your" buyer is allowed to choose a TF smaller or equal to 1.

At the second stage of the experiment, where you have to choose your "actual quality", you also have to write down which TF you will expect realistically from "your" buyer. We ask you to insert this "expected TF" in the box "expected TF" on Sheet 2. **Nobody will be informed about your expected TF. For the calculation of your gains ONLY the actual TF of "your" buyer is relevant!**

The TF chosen by "your" buyer will be communicated to you only. Now you can calculate your gain as well as the gain of "your" buyer. This ends a trading day and the next one will start.

A further important remark: **all the information which you document on Sheet 2 is only for your private use.You are not allowed to communicate this information to other sellers!**

At the end of a trading day there are the following possibilities:

1. In case that you have not accepted a bid or you did not have the opportunity, your **gain** on this trading day is **ATS 10,-**.

Table 1a

Effort cost schedule

e	0.001	0.008	0.027	0.064	0.1	0.2	0.3
c(e)	0	1	2	3	5	7	9

e	0.4	0.5	0.6	0.7	0.8	0.9	1.0
c(e)	11	12.5	14	15.5	17	18.5	20

Table 1b

Firms' cost of punishing and rewarding

p	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
k(p)	10	9	8	7	6	5	4	3	2	1

p	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2
k(p)	0	1	2	3	4	5	6	7	8	9	10

Table 3: *OLS-Regression of job rents r on effort demanded* $\tilde{e}$

$$r_{ti} = \alpha_1 + \alpha_2 \tilde{e}_{ti} + \varepsilon_{ti}$$

Session #	N	α_1	α_2	$\overline{R}^2$
S5	64	13.77 (0.0001)	10.15 (0.075)	0.03
S6	89	- 0.47 (0.569)	22.81 (0.000)	0.53
S7	92	5.36 (0.036)	38.50 (0.000)	0.25
S8	95	4.57 (0.015)	33.84 (0.000)	0.43
S9	78	4.45 (0.004)	12.73 (0.066)	0.04
S10	91	2.45 (0.221)	37.71 (0.000)	0.43
S5-S10	509	3.95 (0.000)	29.34 (0.000)	0.29

N: Number of observations
p-values (marginal significance levels) are in parentheses

Table 5(a): *Firms' punishment/reward decision at stage three, given workers' effort decision*

actual punishment/reward:	**shirking** $e < \tilde{e}$ 30 trades	**no shirking** $e = \tilde{e}$ 104 trades	**excess effort** $e > \tilde{e}$ 10 trades
$p < 1$	18 (0.19)	not possible	not possible
$p = 1$	12	52	6
$p > 1$	not possible	52 (1.62)	4 (1.53)

Note: The number in parentheses shows the average level of p

Table 5(b): *Workers' expectation formation: Do they anticipate firms' reciprocity?*

expected punishment/reward:	**shirking** $e < \tilde{e}$ 30 trades	**no shirking** $e = \tilde{e}$ 104 trades	**excess effort** $e > \tilde{e}$ 10 trades
$p^e < 1$	18 (0.59)	not possible	not possible
$p^e = 1$	12	29	0
$p^e > 1$	not possible	75 (1.51)	10 (1.61)

Note: The number in parentheses shows the average level of p^e

Table 6: *Effort behavior in the WRT and the SRT*

treatment	**No. trades**	**shirking** $e < \tilde{e}$		**no shirking** $e = \tilde{e}$	**excess effort** $e > \tilde{e}$	
		% of trades with $e < \tilde{e}$	average amount of $(\tilde{e}-e)/\tilde{e}$	% of trades with $e = \tilde{e}$	% of trades with $e > \tilde{e}$	average amount of $(e-\tilde{e})/(1-\tilde{e})$
WRT	509	65.42	0.97	33.01	1.57	0.20
SRT	144	20.83	0.82	72.22	6.94	0.83

Figure 2: Actual average effort for given rents in the WRT

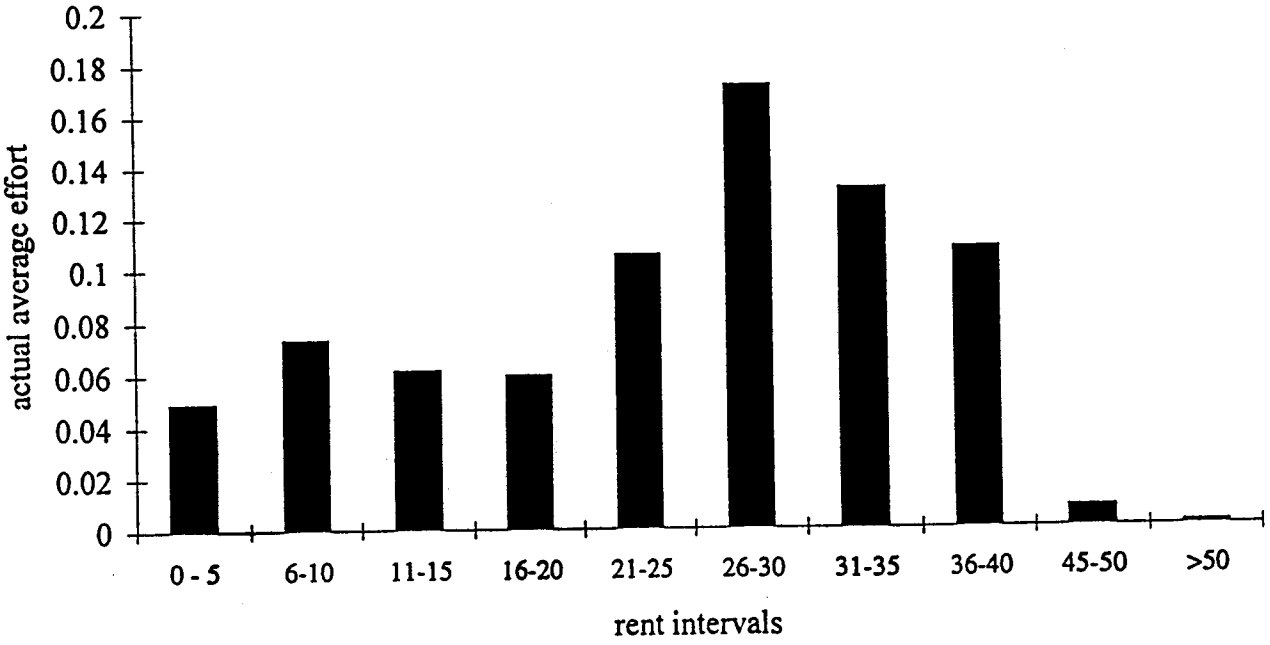

Figure 4: Firms' and workers' gains per trade in the SRT and the WRT