Econometrica, Vol. 65, No. 4 (July, 1997), 833–860

RECIPROCITY AS A CONTRACT ENFORCEMENT DEVICE: EXPERIMENTAL EVIDENCE

BY ERNST FEHR, SIMON GÄCHTER, AND GEORG KIRCHSTEIGER[1]

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.

KEYWORDS: Contract enforcement, reciprocity, moral hazard, principal agent theory.

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 nonnegligible efficiency gains.

[1] 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 a co-editor, 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.

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 contracts. For example, by making a generous employment offer, a firm might induce a worker to provide effort beyond the level that is enforceable by incentive compatible means. The output increase produced by the higher effort level may ultimately render the firm and the worker better off. The *second* type of observation comes from numerous ultimatum game experiments (Güth and Tietz (1990), Roth (1995), Camerer and Thaler (1995)). The results of these experiments suggest that people are frequently willing to forego some money in order to punish unfair behavior. In our context this means that somebody who offered a generous contract which the trading partner subsequently violated might be willing to punish the trading partner. Yet, if the trading partner anticipates this willingness to punish, she has a reason not to violate the contract in the first instance. Punishing unfair behavior can also be considered as a form of reciprocal behavior. While the first type of observation suggests that people are kind to those who are kind to them, the second type of observation suggests that people hurt those who hurt them.[2]

[2] It is also worthwhile to stress that the reciprocal behavior observed in the above mentioned experiments cannot be considered as an artifact of low monetary incentives. Reciprocity can be observed even if up to three months' income is at stake (Hoffman, McCabe, and Smith (1996); Cameron (1995); Fehr and Tougareva (1995)). Recently several papers (Güth and Yaari (1992); Güth (1995); Hoffman, McCabe, and Smith (1995)) have provided evolutionary foundations for the existence of reciprocal preferences. The evolution of gift giving, a phenomenon that is closely related to reciprocity, has been studied in Carmichael and MacLeod (1995). For an explanation of gift giving as a signaling device, see Camerer (1988).

To study the impact of reciprocity on contract terms and their enforcement we implemented three experimental conditions: a no-reciprocity-treatment (NRT) in which contract terms are exogenously enforced so that reciprocity cannot contribute to contract enforcement; a weak-reciprocity-treatment (WRT) in which only one side of the market can respond reciprocally to the previous action of the trading partner; and a strong-reciprocity-treatment (SRT) in which both sides of the market can *respond* reciprocally to previous actions. In the WRT firms post employment contracts in a competitive market. Once a contract is accepted the worker has to decide whether to supply the effort that is demanded by the contract or whether to shirk. The NRT is identical to the WRT except that the effort level is exogenously fixed by the experimenter. The SRT is identical to the WRT except that after workers' effort decisions firms have the opportunity to punish or to reward the worker. Both reactions are costly for the firm.

In all three experimental conditions we implemented a competitive market with an exogenous excess supply of workers. This puts firms in a strong position in which they have the power to enforce rather unfair contracts, i.e. contracts with zero rents for the workers. We hypothesize that the actual rent offered to a worker serves as an indicator for the generosity of an employment offer. The NRT is mainly a control treatment for the WRT. Since workers in the NRT cannot respond reciprocally whereas in the WRT reciprocal responses are possible, the difference in the rents offered in the NRT and the WRT measures the extent to which firms in the WRT want to elicit workers' reciprocity. The WRT in turn serves as a control treatment for the SRT. In the WRT and the SRT we impose identical limits on the enforcement of contracts by constraining the maximum fine that has to be paid in case of 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 would 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 labor market consists of two stages (see Table I). At the first stage employment contracts are concluded. At the second stage the effort decisions and payments are made. At the first stage firms simultaneously offer employment contracts which stipulate a wage w, an effort demanded $\tilde{e}$, and a fine f.[3] Then all workers are informed about firms' offers and can choose among the available contracts according to a randomly determined order. At the second stage those workers who accepted an offer have to choose an effort level e. Once e is chosen a random mechanism determines with probability s $(0 < s < 1)$ whether shirking $(e < \tilde{e})$ is verifiable by a third party. In case of verifiability the worker has to pay f to the firm.

If the worker performs $e \geq \tilde{e}$ she receives w; if she shirks she gets paid w with probability $(1 - s)$ and $(w - f)$ with probability s. Without loss of generality effort is restricted to the interval $[0,1]$. $c(e)$ denotes effort costs which are strictly increasing in e with $c(0) = 0$. To derive the equilibrium we have to determine how workers choose e for any given contract $(w, \tilde{e}, f)$. Then we have to examine how firms choose their contracts taking into account workers' effort responses. The (expected) utility of an employment offer $(w, \tilde{e}, f)$ is given by

$$(1a) \qquad u^{ns} = w - c(e)$$

[3] Notice that in a one-period model dismissal imposes no costs on the shirking worker. Therefore we assume that firms can directly fine shirking workers.

TABLE I

SEQUENCE OF EVENTS

• Firms simultaneously post employment contracts $(w, \tilde{e}, f)$. • Workers observe all contracts and choose among the available offers in a randomly determined order.	Stage 1
• Workers who accepted an offer choose $e \gtreqless \tilde{e}$. • Random device determines whether shirking $(e < \tilde{e})$ is verifiable. • Firms are informed about the effort choice of their worker.	Stage 2

if a worker does not shirk $(e \geq \tilde{e})$, and by

$$u^s = (1-s)[w - c(e)] + s[w - f - c(e)] \tag{1b}$$

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

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

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

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

where 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.[4] 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.[5] The existence of f^0 implies a maximum enforceable effort level of

[4] 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 (1991)).

[5] 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.

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 if $f \leq f^0$ 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

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

The relations in (4) imply that the job rent as defined by $r \equiv w - w^r = w - c(\tilde{e})$ is zero. Notice that this prediction (i.e. $r^* = 0$) is also valid if firms demand $e \neq e^0$. For any given $\tilde{e}$ profit maximizing firms have no reason to pay more than $c(\tilde{e})$.

2.2. *The Three Stage Design*

While the two stage design constitutes our weak reciprocity treatment (WRT), the three stage design represents the strong reciprocity treatment (SRT). At the first and the second stage the WRT and the SRT are identical. Yet, in the SRT there is a *third stage* in which firms can punish or reward their workers in the following way: After firms observe their worker's effort level they have to choose a number $p \in [0, 2]$. A worker's gain from the first two stages is then multiplied by p. Both $p < 1$ (a penalty) and $p > 1$ (a reward) causes costs $k(p)$. $k(p)$ is decreasing for $p \in (0, 1)$ and increasing for $p \in (1, 2)$. In addition $k(0) = 0$ holds. In the SRT workers' payoffs are given by

$$(1a') \qquad u^{ns} = [(w - c(\tilde{e}))]p$$

and

$$(1b') \qquad u^s = [(1 - s)(w - c(e)) + s(w - f - c(e))]p,$$

while the payoff of firms is given by

$$(3') \qquad \pi = (q - w)e - k(p).$$

The implications of the third stage for equilibrium behavior are straightforward. If firms and workers are payoff maximizers and if this is common knowledge, the third stage will have no impact at all. Firms will never punish or reward at stage three because it is costly. Rational workers will anticipate that $p = 1$, i.e. they will choose the same effort as in the WRT for any given contract. Therefore, rational firms will offer the same contracts in the WRT and the SRT.

3. THE IMPLICATIONS OF RECIPROCITY

3.1. *Reciprocity in the Two Stage Design*

To what extent are the above predictions about $\tilde{e}$ and r changed if reciprocity motives are taken into account? To discuss this question we have to define the

notion of reciprocity in more detail. Rabin (1993) has developed a model of reciprocity which provides the basis for the following exposition. According to Rabin a person has reciprocity motives if she is willing (i) to sacrifice resources to be kind to those who are being kind (= positive reciprocity) and (ii) to sacrifice resources to punish those who are being unkind (= negative reciprocity). The essential feature of reciprocity motives is thus a 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*[6] 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 are 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

[6] At present there does not exist a general theory that allows precise location of 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.

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 choices. At least over some interval above e^0 there should be a positive relation between firms' profits and the levels of $\tilde{e}$ and r, respectively.

3.2. *Reciprocity in the Three Stage Design*

In the presence of reciprocal effort choices even profit maximizing firms have an incentive to pay generous wages. In our view the power of reciprocity to affect people's behavior derives to a large extent also from the fact that it changes the incentives for those who are *not* motivated by reciprocity. Previous experimental work[7] strongly indicates that not all subjects do behave reciprocally and that those who do, exhibit different levels of reciprocity. This suggests that in our contract enforcement experiments a fraction of subjects will only be motivated by pecuniary incentives, too.

Since the acceptance of a contract offer by a worker renders $\tilde{e}$ a natural reference standard, it seems rather likely that firms will consider shirking as unfair while $e \geq \tilde{e}$ will be considered as fair. Therefore, in the presence of reciprocal firms there will be a positive probability of being punished for $e < \tilde{e}$ and of being rewarded for $e \geq \tilde{e}$. In case that firms indeed behave reciprocally even those workers who are purely selfish, or are only weakly motivated by reciprocity, now have some incentive to provide effort levels above e^0. If firms take into account that workers anticipate firms' reciprocity, they have an incentive to demand higher effort levels in the SRT compared to the WRT. Thus, we predict that firms demand and enforce higher effort levels in the SRT.

Before we describe the experimental procedures in more detail, we want to stress that we do not interpret the third stage as a literal mapping of real world phenomena. Yet, long-term relationships like the employment relation are usually only incompletely regulated by binding contracts and in the course of such relationships it seems very likely that for both parties reward and punishment possibilities arise. We conjecture that people utilize these possibilities not only for purely pecuniary reasons (e.g. investment in reputation formation) but that reciprocity considerations will also play an important role. The addition of the third stage should, therefore, be viewed as a useful experimental tool to analyze the effects of two-sided reciprocity without running into the problems of multiple equilibria that arise in repeated games (Fudenberg and Maskin (1986)).

[7] See Fehr, Kirchsteiger, and Riedl (1993, 1996), Fehr, Kirchler, Weichbold, and Gächter (1994), and Fehr and Tougareva (1995). Although there is always a clear majority of 60–75 percent of the subjects who do behave reciprocally, between 15 and 25 percent of subjects make purely selfish choices. The remaining subjects make choices that are neither reciprocal nor purely selfish.

4. PARAMETERS AND EXPERIMENTAL PROCEDURES

In total we conducted eighteen experimental sessions.[8] 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 six additional 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 Zurich (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.[9]

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 $\bar{e}$ 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 behavior in previous periods, 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 IIa. Each subject

[8] A highly compressed version of the experimental instructions is presented in the Appendix. A full set of instructions is available upon request.

[9] 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.

TABLE IIa

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

received an initial endowment of 70 Austrian Schillings (70 ATS $\approx$ 7\$) for participation in the experiment. When determining wages, firms were bound by the constraint $w^r \leq w \leq q$.

In the following we derive the numerical predictions for the risk neutral equilibrium as a benchmark case. This equilibrium is given by (4). Notice that $f \leq f^0 = 10$ implies that risk neutral nonreciprocal workers can at most be induced to bear effort costs of $c(e) = 5$ which implies a maximum enforceable effort of $e = 0.1$. Since in equilibrium firms pay no rents, the prediction is $(r^*, \tilde{e}^*, f^*) = (0, 0.1, 10)$. In the presence of risk neutral, yet reciprocally motivated workers, firms may be able to induce workers to perform $e > 0.1$. Since we expected some positive reciprocity in the WRT and to maximize the comparability between the WRT and the NRT we fixed the exogenously enforced effort level in the NRT at $e = 0.2$. All other parameters and the experimental procedures in the NRT were the same as in the WRT. If workers and firms are fully rational and their preferences are given by (1a) and (3), the prediction for the NRT is straightforward. All firms offer contracts with zero rents and these offers are accepted by the workers.

The procedural details of the first two stages of the SRT did not differ from those of the WRT. Since the sessions of the SRT lasted longer, we paid all participants an initial endowment of 100 ATS ($\approx$ 10\$) to render recruitment easier. The costs of punishing ($p < 1$) and rewarding ($p > 1$) workers were common knowledge; for all firms they were given by the schedule in Table IIb. When workers had chosen their effort at stage two they had to write down the level of p they expected to get from "their" firm. No other subject in the market was informed about a worker's expectation. The actual level of p was only known to the firm and the worker that were involved in the trade. To avoid that

TABLE IIb

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

idiosyncratic subjects could create confusion,[10] firms were not allowed to punish ($p < 1$) nonshirking workers or to reward ($p > 1$) shirking workers. This constraint was known to the workers. Note that if firms are profit maximizers this constraint is never binding because they always prefer $p = 1$.

The parameters in the three stage game were chosen such that in the absence of reciprocal behavior the highest enforceable effort level is the same as in our WRT.[11] Therefore, all predictions regarding the first two stages are also the same: Firms pay no job rents, demand $\tilde{e} = 0.1$, and impose the largest feasible fine $f^0 = 10$. Yet, if firms' reciprocity is behaviorally relevant, their chances to enforce $\tilde{e} > e^0$ are now much better.

5. EXPERIMENTAL RESULTS

In this section the results of our main sessions (S1–S12) are presented.[12] 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.[13] 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

[10] Probably every experimenter has once encountered such subjects. They can easily contaminate a whole experimental session.

[11] The same $c(e)$ schedule and the same values of f^0 and s have been implemented.

[12] Our data are available upon request.

[13] This includes the time spent on reading and understanding the instructions.

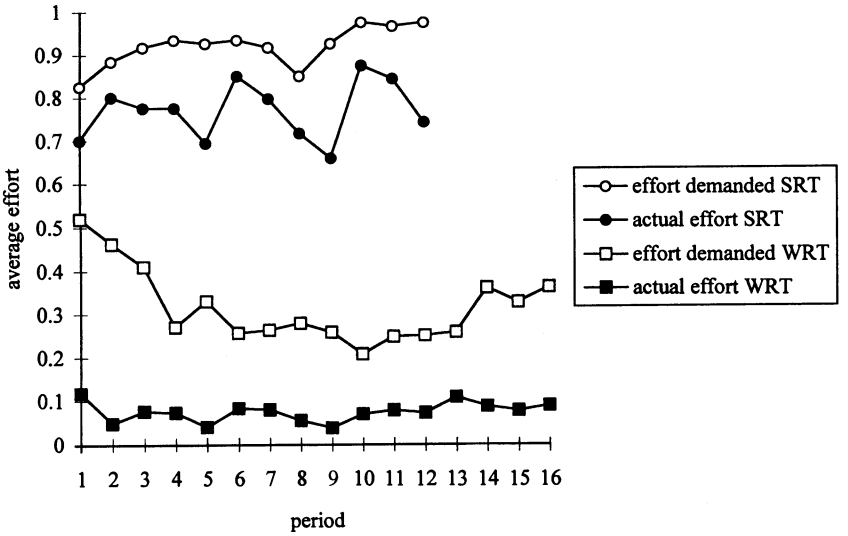

FIGURE 1.—Average effort demanded and actual average effort in the SRT and the WRT.

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.

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 III, however, casts doubt on the assumption of risk aversion. Among other things the table shows the number of trades that

TABLE III

BEHAVIOR IN THE WEAK RECIPROCITY TREATMENT

Effort Demanded	No. of Trades	Percentage of Shirking	Actual Average Effort	Average Rent Offered	Average Profit of Firms
0.001	5	0	0.001	0.80	0.25
0.008	4	0	0.008	8.50	0.76
0.027	11	18.18	0.022	2.55	2.24
0.064	58	22.41	0.050	3.12	4.93
0.1	78	55.13	0.045	5.91	4.71
0.2	102	63.73	0.078	9.46	6.72
0.3	96	76.04	0.084	15.12	6.37
0.4	32	84.38	0.084	16.84	4.28
0.5	29	86.21	0.098	18.91	4.61
0.6	32	90.63	0.095	25.25	12.08
0.7	14	85.71	0.101	32.43	8.68
0.8	17	100	0.001	28.29	0.10
0.9	8	62.50	0.339	29.13	19.79
1	23	100	0.072	24.04	1.09

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[14] 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 percent. These data indicate that the majority of effort choices at $\tilde{e} = 0.1$ and at $\tilde{e} = 0.2$ are not compatible with risk aversion and that a substantial fraction of choices at $\tilde{e} = 0.064$ exhibit risk seeking behavior.

To examine whether the anticipation of workers' reciprocity has been responsible for firms' $\tilde{e}$ choices we have to look first at the relation between r and $\tilde{e}$. Remember that if there is a sufficiently steep positive relation between the probability of nonshirking and r, it is profitable to compensate the higher pecuniary incentive to shirk that arises from an increase in $\tilde{e}$ by a rise in r. Therefore, if firms anticipate sufficiently strong reciprocal responses, the relation between r and $\tilde{e}$ should be positive.

R2: *Firms pay higher job rents if they demand higher effort levels.*

This result is supported by Table III and the following OLS regressions:

$$r_{ti} = \alpha_1 + \alpha_2 \tilde{e}_{ti} + \varepsilon_{ti}. \tag{5}$$

A comparison of column 1 ('effort demanded') and column 5 ('average rent offered') of Table III shows that the average rent increases if $\tilde{e}$ increases. In addition, Table IV shows that α_2 is positive below the 1 percent level in four sessions and below the 10 percent level in two sessions. It is highly significant if we use the data of all sessions. The positive relation between r and e also shows up at the level of individual firms. To document this we computed the Spearman rank correlation between r and $\tilde{e}$ for each firm. For 29 of the 33 firms this correlation is positive and for 20 firms the positive correlation is significant at the five percent level. Thus, at the individual as well as at the aggregate level, R2 is strongly supported by our data.

R2 indicates that firms anticipated workers' reciprocity. Our next result concerns the question whether workers behaved indeed reciprocally.

R3: *Workers in the WRT behave on average reciprocally; i.e., by raising r firms can increase the probability of nonshirking.*

Figure 2 provides a first indication for workers' reciprocity. It shows that for $r \in [0, 20]$ the actual average effort is lower than for $r \in [21, 40]$. A Wilcoxon signed rank test with average effort levels of the 45 workers as observations confirms that this difference is significant below the 1 percent level.[15] Notice

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

[15] Moreover, the Spearman rank correlation between actual average effort and rents is significantly positive [$\rho(e, r) = 0.23$, p value < 0.07].

TABLE IV

OLS-REGRESSION OF JOB RENTS r ON EFFORT DEMANDED $\tilde{e}$

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

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

[a] N: Number of observations.
[b] p values (marginal significance levels) are in parentheses.

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:

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

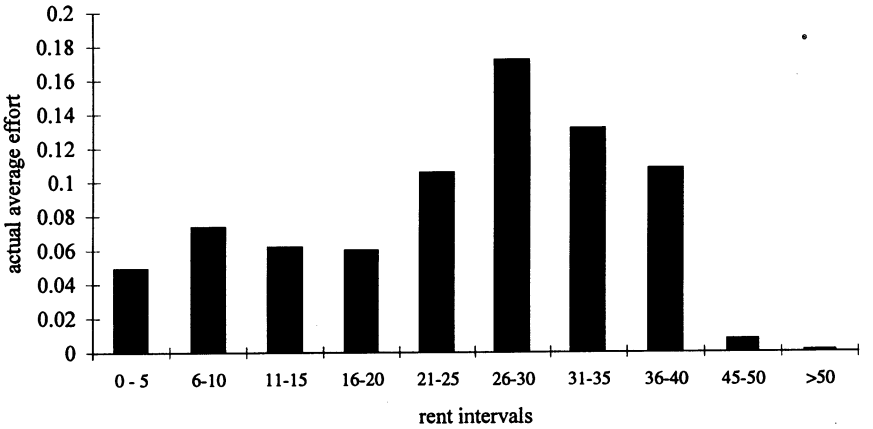

FIGURE 2.—Actual average effort for given rents in the WRT.

TABLE V

DETERMINANTS OF THE PROBABILITY OF NON-SHIRKING

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

Session #	N[a]	β_1[b]	β_2[b]	β_3[b]	β_4[b]	LRI[c]
S5	64	1.066	0.188	0.038	−0.005	0.26
		(0.027)	(0.0001)	(0.029)	(0.067)	
S6	89	0.480	0.377	0.076	−0.005	0.31
		(0.292)	(0.000)	(0.052)	(0.223)	
S7	92	0.493	0.318	0.005	−0.001	0.31
		(0.200)	(0.0001)	(0.688)	(0.666)	
S8	95	0.441	0.161	0.033	−0.005	0.15
		(0.173)	(0.001)	(0.087)	(0.030)	
S9	78	−0.273	0.504	0.065	−0.006	0.35
		(0.370)	(0.0001)	(0.007)	(0.283)	
S10	91	0.580	0.134	0.008	−0.008	0.23
		(0.091)	(0.002)	(0.576)	(0.009)	
S5–S10	509	0.147	0.169	0.020	−0.002	0.17
		(0.202)	(0.000)	(0.000)	(0.079)	

[a] N: Number of observations.
[b] p values (marginal significance levels) are in parentheses.
[c] LRI: As a measure of the goodness of fit we used the likelihood ratio index (LRI) as defined in Greene (1993, p. 651).

where $\theta_{ti} = 1$ ($\theta_{ti} = 0$) in case that worker i does not (does) shirk in period t. The pecuniary incentive to perform $\tilde{e}$ is measured by $u^{ns} - u^{s}$, 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 V the results of regression (6) are presented. As Table V 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 regression (6) and Figure 2 indicate that on average workers respond reciprocally.

On the basis of regression (6) we can compute firms' expected profits $E\pi$ for any combination of r and e. $E\pi$ is given by $E\pi = F(\theta)[q - r - c(\tilde{e})]\tilde{e}$ where F is the standard normal distribution.[16] 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

[16] 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.

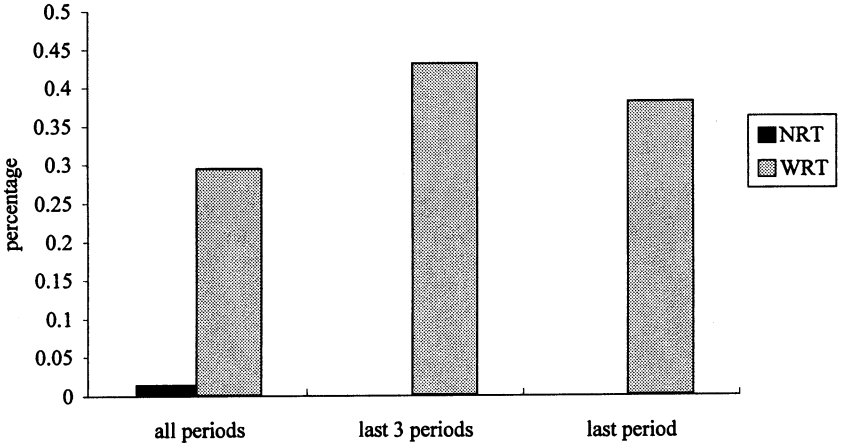

FIGURE 3.—Percentage of rent offers above 20 in the No Reciprocity Treatment and the Weak Reciprocity Treatment.

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.*

This result is also strongly supported by the data. Irrespective of taking the data of all periods or only those of the last or the last three periods, the average rent in the WRT is significantly larger than in the NRT. For the last three periods the average rent in the WRT is 17 compared to 10 in the NRT. This difference is significant at all conventional significance levels.[17] Similar results hold for the percentage of offers with rents above 20 (see Figure 3).

While in the NRT there are virtually no offers above 20, in the WRT we observe on average about 30 percent of offers above 20. The figure also shows that the higher percentage of generous offers is not a temporary phenomenon that vanishes over time. During the last three periods the difference in the relative frequency of generous offers even increases. While there are no(!) offers with rents above 20 in the NRT, they constitute more than 40 percent in the WRT. Together with the previous results we take this as rather convincing evidence that reciprocity affected firms' contract offers. It remains to check

[17] Rent differences between the NRT and the WRT are confirmed by a t test as well as by a one-sided Kolmogorov-Smirnov test (p values < 0.01).

whether this appeal to reciprocity was profitable for the firms. In view of the fact that the most frequently demanded effort levels are 0.2 and 0.3 (see Table III), it is particularly important to know whether they yielded higher profits in comparison to $\tilde{e} \leq 0.1$.

R5: *There exist high-effort/high-rent strategies that yield on average higher profits than low-effort/low-rent strategies.*

To distinguish between high $(\tilde{e}, r)$- and low $(\tilde{e}, r)$-strategies we take the prediction of the standard model that only effort demands $\tilde{e} \leq e^0$ are profitable as a benchmark. In contrast to this prediction the last column in Table III shows that by increasing $\tilde{e}$ to 0.2 and 0.3 firms increased their average profits relative to $\tilde{e} \leq e^0$. Notice that this profit increase occurs despite the fact that firms pay much higher job rents at $\tilde{e} = 0.2$ and $\tilde{e} = 0.3$ compared to $\tilde{e} \leq e^0$. Therefore, if higher rents would not increase the probability of nonshirking, profits would have decreased. That high-effort/high-rent strategies were more profitable is also confirmed by formal statistical tests. First a chi-square test reveals that firms who offered above average rents earned above average profits and vice versa (p value < 0.05). The same holds for firms who demanded above average effort levels (p value < 0.04). Furthermore, the Spearman rank correlation $\rho(\pi,r)$ between actual profits and r, is significantly positive ($\rho(\pi, r) = 0.38$, p value < 0.015). The same holds true for the rank correlation between profits and $\tilde{e}$ ($\rho(\pi, \tilde{e}) = 0.43$, p value < 0.007).

A final piece of evidence for the superior profitability of high-effort/high-rent strategies comes from the computation of firms' expected profits on the basis of regression (6). If we compute the $E\pi$-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.[18] 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

[18] 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 $E\pi > 8$.

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 R6.

R6: *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.

R6 indicates that the firms' opportunity to punish or reward workers' behavior at stage 2 has strong effects. Did these effects arise because of firms' reciprocal behavior at stage three? To check this we examine firms' p choices in more detail.

R7: *In case workers shirked, firms chose on average* $p < 1$, *while in case of* $e \geq \tilde{e}$ *they rewarded* ($p > 1$) *workers.*

R7 is supported by Table VIa. If workers shirked, firms punished in 18 out of 30 cases with an average choice of $p = 0.19$. Therefore, the expected value of p in case of shirking is given by[19] $Ep = 0.51$. If workers did not shirk, firms rewarded in 52 out of 104 cases and Ep is given by $Ep = 1.31$. If workers provided excess effort firms also rewarded on average workers' behavior. Punishing and rewarding is present at the level of individual firms, too. Eight of twelve firms punished in case of shirking and rewarded in case of nonshirking. Only one firm always chose $p = 1$ while the other three firms either only punished or

[19] $Ep = 0.19 * 0.6 + 1 * 0.4 = 0.51$.

only rewarded. Thus, Table VIa and individual firm data show that punishing and rewarding was quite frequent. Yet, did workers anticipate firms' behavior correctly?

R8: *On average workers expected to be punished in case of $e < \tilde{e}$ and to be rewarded for $e \geq \tilde{e}$. However, they expected less severe punishments and more frequent rewards than actually occurred.*

Table VIb provides workers' expectation data. They correctly anticipated the frequency of punishments but they underestimated the severity of the punishments. While firms chose $p = 0.19$ if they punished, workers estimated $p = 0.59$ if they expected a punishment. A comparison of columns three and four of Tables VIa and VIb also reveals that in case of $e \geq \tilde{e}$ workers overestimated the frequency of being rewarded.

In our view Table VIb indicates that workers expected firms to behave reciprocally. Yet quantitatively their expectations were not quite correct. To what extent did workers' expectation affect their shirking behavior?

R9: *Although firms demanded higher effort levels in the SRT than in the WRT, the rate of shirking is lower in the SRT.*

TABLE VIa

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 VIb

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 VII

EFFORT BEHAVIOR IN THE WRT AND THE SRT

		Shirking $e < \tilde{e}$		No Shirking $e = \tilde{e}$	Excess Effort $e > \tilde{e}$	
Treatment	No. Trades	% 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

The evidence in Table VII provides strong support for R9. The shirking rate declined from 65 percent to 21 percent when a third stage was added. Moreover, in case workers shirked, the relative underprovision of effort $[(\tilde{e} - e)/\tilde{e}]$ is lower in the SRT. In the SRT we also observe more frequently that workers provide excess effort.

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.

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.

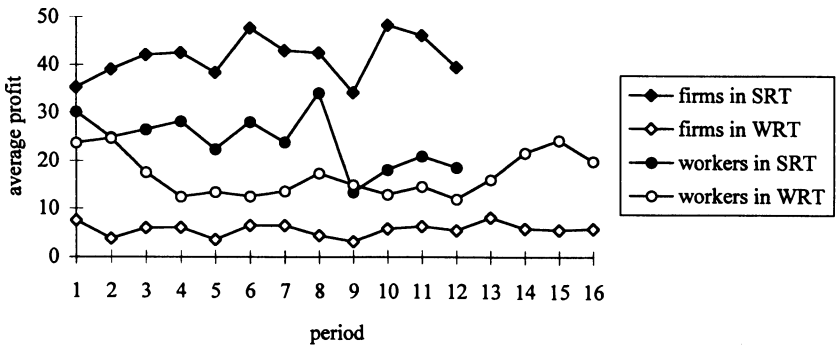

FIGURE 4.—Firms' and workers' gains per trade in the SRT and the WRT.

6. THE ROBUSTNESS OF RECIPROCITY EFFECTS

In this section[20] 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.[21] In accordance with most principal-agent models we also implemented a random effort verification procedure in our WRT and SRT. This causes, however, the problem of controlling for workers' risk preferences.

In S13–S16 we have conducted two additional two and three stage sessions, respectively, in which (i) firms could make losses, (ii) wages represented a pure lump-sum transfer, and (iii) workers' risk preferences were irrelevant for their effort choices. The purpose of these experiments was to check the robustness of the reciprocity effects on $\tilde{e}$ and e. To remove the impact of workers' risk preferences on effort choices we eliminated the random verification procedure and the fines for shirking. Therefore, a contract consisted only of $(w, \tilde{e})$. To achieve (i) and (ii) we implemented the following payoff functions:

$$u = \begin{cases} w - c(e) & \text{in the two stage design,} \\ w - c(e) + bp & \text{in the three stage design,} \end{cases} \tag{7}$$

$$\pi = \begin{cases} qe - w & \text{in the two stage design,} \\ qe - w - k(p) & \text{in the three stage design.} \end{cases} \tag{8}$$

[20] Space limitations prevent us from a more detailed presentation of the evidence on robustness. Upon request we will send interested readers an earlier version of this paper in which we deal with the robustness issue in more detail. Experiments on the robustness of reciprocity effects are also discussed in more detail in Fehr and Gächter (1997).

[21] The sum of u^{ns} and π is given by $q\tilde{e} - c(\tilde{e}) + w(1 - \tilde{e})$.

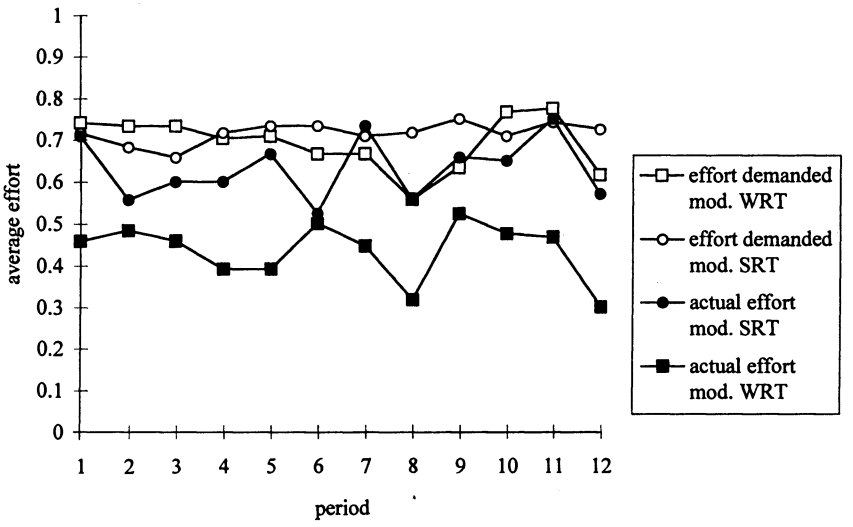

FIGURE 5.—Average effort demanded and actual average effort in the modified SRT and WRT.

Notice that if firms pay high wages and get low effort levels they will experience losses. In addition, wages constitute a pure transfer payment and our parameterization ensured that the welfare maximizing levels of e and p are well defined. The parameters are given by $q = 100$, $b = 25$, $e \in [0.1, 1]$ while $c(e)$ represents a discrete approximation of $c = (10e - 0.1)^{1.3}$; p had to be in the interval $[-1, 1]$ and $k(p)$ was given by $10p$ for $p \in [0, 1]$ and $-10p$ for $p \in [-1, 0]$.[22]

In the absence of reciprocity it is easy to see that rational subjects will choose $p = 0$, $e = 0.1$, and $r = 0$ while the level of $\tilde{e}$ is behaviorally irrelevant. In the presence of reciprocity workers in the WRT will provide higher effort in response to higher rents. As a consequence, firms who want to elicit reciprocal responses will pay higher rents if they demand higher effort levels. Therefore, we should observe a positive relation between r and $\tilde{e}$ as well as between r and e. In addition r is predicted to be significantly larger than zero and e will exceed 0.1 significantly. For the modified SRT we predict that firms punish shirking and reward nonshirking. As a consequence workers will shirk less and firms can enforce higher effort levels compared to the modified WRT.

The effort results of S13–S16 are depicted in Figure 5. Interestingly, firms demand approximately the same effort in the modified WRT as in the modified SRT. In both treatment conditions, $\tilde{e}$ is far above the enforceable effort level $e = 0.1$. As in our main experiments (S5–S12) the actual average effort in the modified WRT is lower than in the modified SRT. Yet, in contrast to our original WRT the actual average effort is now well above 0.1. In the original

[22] The parameters ensure that the effort and job rent predictions do not change relative to our original WRT and SRT. We chose $b = 25$ because, on average, workers actual earnings in the original SRT after stage 2 were roughly 25.

WRT workers almost always choose the lowest effort level if they shirk, whereas in the modified WRT *partial* shirking is very frequent.

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 IIb. 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.

7. SUMMARY AND CONCLUDING REMARKS

This paper argues that reciprocal motivations have important implications for the enforcement of contracts. We designed experiments in which these implications can be observed if subjects are motivated by reciprocity. To isolate the impact of reciprocity on contract enforcement we designed two major treatments: the weak reciprocity treatment (WRT) and the strong reciprocity treatment (SRT). In the WRT only workers could respond reciprocally to firms' actions while in the SRT workers *and* firms could respond reciprocally to the other party's actions.

The regularities of the WRT suggest that firms' employment offers are strongly affected by the appeal to workers' reciprocity. Yet, the strength of workers' reciprocal responses depends on the details of the pecuniary incentive system. In our main WRT workers exhibit only weak reciprocity while in our modified WRT their reciprocal responses are rather strong. In contrast, in the SRT reciprocity is a powerful device for the enforcement of contracts irrespective of the differences between the original and the modified design. The SRT-data show (i) that firms reciprocate, (ii) that workers anticipate firms' reciprocation and, hence, shirk less than in the WRT, and (iii) that firms demand and enforce effort levels far above the incentive compatible level.

To our knowledge modern principal-agent theory has so far not been concerned with the impact of reciprocity on contract terms and their enforcement. Our results indicate, however, that the neglect of reciprocity may render principal agent models seriously incomplete. As a consequence it may limit their predictive power. Moreover, the normative conclusions that follow from models that neglect reciprocity may not be correct. This is indicated by the large efficiency gains in our SRT. Both workers and firms were considerably better off in the SRT relative to the WRT and relative to an equilibrium without reciprocity. In view of the powerful behavioral impact and the efficiency consequences of reciprocity in our SRT, it seems doubtful that one can design optimal incentive contracts on the basis of a neglect of reciprocal motivations.

Institute for Empirical Research in Economics, University of Zürich, Blümlisalpstrasse 10, CH-8006 Zürich, Switzerland,

and

Dept. of Economics, University of Vienna, Hohenstaufengasse 9, A-1010 Wien, Austria.

Manuscript received September, 1995; final revision received September, 1996.

APPENDIX

For space limitations we present only a shortened version of the instructions for the sellers (workers) in the three-stage experiment.[23] The sellers' questionnaire, the documentation sheet, and the instruction summary for the sellers are not presented here. We avoided, of course, value laden

[23] A full set of the instructions is available on request.

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 to 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 than 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 these instructions you will find a control questionnaire, Sheet 2, which serves documentation purposes, and an Instruction Summary, which summarizes important information.

Detailed Information for Sellers

In the market a certain commodity will be traded, and every seller sells the same good. Each seller can sell to each buyer and each buyer can buy from each seller. Every buyer gets on every trading day 105 units of experimental money, which he can use for buying one good. Every seller and every buyer knows this. The market is organized as follows. We open the market for a trading day. Then buyers (without having the possibility to communicate with other buyers) can make a bid. *A bid consists of a price, a quality-independent commission; and a desired quality.* Whether you receive the price stated in the bid depends on your delivered quality and whether—if your actual quality falls short of the desired quality—this can be verified. The commission, however, must be paid in any case, independently of your quality choice. There are *fourteen feasible quality levels* among which the buyer, or you as a seller, can choose. The lowest quality is 1/1000 and the highest 1000/1000. The effect of the delivered quality on the monetary gains will be explained in more detail below. In general, however, a high quality increases your costs and the gains of the buyer. In your *Instruction Summary* you will find a schedule with all feasible quality levels and the associated quality costs. The price bid can be at most ATS 20,-.

The commission can be positive or negative. If the proposed commission is *positive*, for example ATS + 10,-, the seller who accepts the bid, receives the accepted price *plus* ATS 10,-. If, however, the proposed commission is *negative*, for example ATS −10,-, the seller gets the price *less* ATS 10,-.

When all buyers have had the opportunity to make a bid, the proposed bids will be transmitted by phone and written on the blackboard in a random order. You are *not* told who has made which bid. Now an order is determined at random in which you as a seller can choose among the bids. For this purpose you have to draw a numbered card (numbered from 1–8, but without seeing the numbers, of course) out of a bundle of cards. The seller who draws the card with number 1 is the first who is allowed to choose a bid, the seller who draws number 2, the second, and so on. The choice of a bid takes place as follows: When it is your turn to make a choice, you have to announce your *Seller No.* and your chosen bid. On each trading day you can choose only one bid. You are not forced to accept a bid. *Buyers do not learn which bid you have accepted; they only know whether their bid is accepted or not.*

If you have accepted a bid, please enter price, desired quality, and commission in the box "*accepted bid*" on *Sheet 2*. You now have to decide whether you deliver the desired quality or not. As already mentioned, the quality choice is associated with *quality costs* for you as a seller. *The schedule with the feasible quality levels and the associated quality costs is on your Instruction Summary! This schedule is known to every buyer and seller.*

We ask you to insert your actually chosen quality into the box "*actual quality*" (Sheet 2). Other sellers are not told your quality choice. *Do not announce your "actual quality" in the classroom.* Every buyer is only told the "actual quality" "his" seller has chosen. They also do *not know* the identity of "their" sellers. The anonymity of your quality choice is completely guaranteed, therefore.

In the meantime every buyer throws a six-sided die. If the numbers 1, 2, or 3 show up, the actual quality can be verified, with die numbers 4, 5, or 6, however, this is not possible. If your actual quality is below the desired quality and if this can be verified (die numbers 1, 2, 3) you do not receive the accepted price, but instead a "fixed price," which is determined by us. *This "fixed price" is ATS* 10,-. If your actual quality is at least as high as the desired quality, or higher, you will get the accepted price in any case. *After* you have determined your actual quality, and noted it on Sheet 2, we will tell 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,-.

2. You have accepted a bid, which means you have sold a good, and your "actual quality" corresponds to the desired quality, or it is higher, then your *gain* (in ATS) will be calculated as follows:

Your gain in ATS = (*price* + *commission* − *quality costs*) × "*actual TF*".

The *gain of your buyer* in ATS is:

Buyer's gain in ATS = (105 − price − commission) × "actual quality" + (10 − costs of TF).

3. You have accepted a bid and your "actual quality" is lower than desired.

(a) Your "actual quality" can be verified. Your monetary gain in ATS is then calculated as follows:

Your gain in ATS = (*fixed price of ATS* 10 + *commission* − *quality costs*) × "*actual TF*".

The *gain of your buyer* in ATS is:

Buyer's gain in ATS = (105 − fixed price of ATS 10 − commission) × "actual quality" + (10 − costs of TF).

(b) Your "actual quality" cannot be verified. In this case your gain is

Your gain in ATS = (*price* + *commission* − *quality cost*) × "*actual TF*".

The *gain of your buyer* in ATS is:

Buyer's gain in ATS = (105 − price − commission) × "actual quality" + (10 − costs of TF).

Every seller and every buyer knows the details of this calculation of monetary gains. You are, therefore, able to calculate the gain (in ATS) of "your" buyer, and "your" buyer can calculate your gain.

Do you have further questions?

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., AND S. GÄCHTER (1997): "How Effective are Trust- and Reciprocity-Based Incentives?" in *Economics, Values, and Organization*, ed. by A. Ben-Ner and L. Putterman. Cambridge: Cambridge University Press.

FEHR, E., E. KIRCHLER, A. WEICHBOLD, AND S. GÄCHTER (1994): "When Social Norms Overpower Competition—Gift Exchange in Experimental Labour Markets," forthcoming, *Journal of Labor Economics*.

FEHR, E., G. KIRCHSTEIGER, AND A. RIEDL (1993): "Does Fairness Prevent Market Clearing? An Experimental Investigation," *Quarterly Journal of Economics*, 108, 437–460.

——— (1996): "Gift Exchange and Reciprocity in Competitive Experimental Markets," forthcoming, *European Economic Review*.

FEHR, E., AND E. TOUGAREVA (1995): "Do Competitive Markets with High Stakes Remove Reciprocal Fairness?—Evidence from Russia," Discussion Paper, University of Zürich.

FUDENBERG, D., AND E. MASKIN (1986): "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information," *Econometrica*, 54, 533–554.

GREENE, W. (1993): *Econometric Analysis*, 2nd ed. Englewood Cliffs, NJ: Prentice Hall.

GROSSMAN, S., AND O. HART (1983): "An Analysis of the Principal-Agent Problem," *Econometrica*, 51, 7–45.

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

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: University of Michigan Press.

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

HOFFMAN, E., K. MCCABE, AND V. SMITH (1996): "On Expectations and Monetary Stakes in Ultimatum Games," *International Journal of Game Theory*, 25, 289–301.

——— (1995): "The Behavioral Foundations of Reciprocity: Experimental Economics and Evolutionary Psychology," Discussion Paper, University of Arizona.

KAHNEMAN, D., AND A. TVERSKY (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47, 263–291.

LAFFONT, J., AND E. MASKIN (1982): "The Theory of Incentives: An Overview," in *Advances in Economic Theory*, ed. by W. Hildenbrand. Cambridge: Cambridge University Press.

MILGROM, P., AND J. ROBERTS (1992): *Economics, Organization and Management*. Englewood Cliffs: Prentice Hall.

RABIN, M. (1993): "Incorporating Fairness into Game Theory and Economics," *American Economic Review*, 83, 1281–1302.

ROTH, A. E. (1995): "Bargaining Experiments," in *Handbook of Experimental Economics*, ed. by J. Kagel and A. E. Roth. Princeton, NJ: Princeton University Press.

TVERSKY, A., AND D. KAHNEMAN (1991): "Loss Aversion in Riskless Choice: A Reference Dependent Model," *Quarterly Journal of Economics*, 106, 1039–1062.