ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Fehr, Ernst; Gaechter, Simon

Working Paper

Fairness and Retaliation: The Economics of Reciprocity

CESifo Working Paper, No. 336

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Fehr, Ernst; Gaechter, Simon (2000) : Fairness and Retaliation: The Economics of Reciprocity, CESifo Working Paper, No. 336, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:
https://hdl.handle.net/10419/75602

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

CESifo *Working Paper Series*

FAIRNESS AND RETALIATION: THE ECONOMICS OF RECIPROCITY

Ernst Fehr
Simon Gächter*

Working Paper No. 336

September 2000

CESifo
Poschingerstr. 5
81679 Munich
Germany
Phone: +49 (89) 9224-1410/1425
Fax: +49 (89) 9224-1409
http://www.CESifo.de

---

* This paper is part of the MacArthur Foundation Network on Economic Environments and the Evolution of Individual Preferences and Social Norms. Some research reported in this paper has also been funded by the EU-TMR research network ENDEAR (FMRX-CT98-0238). We are very grateful for helpful comments by Alan Krueger, Timothy Taylor and Bradford De Long, as well as by Ken Binmore, Iris Bohnet, Terence Burnham, Colin Camerer, Gary Charness, Jim Cox, Vince Crawford, Armni Falk, Urs Fischbacher, Diego Gambetta, Robert Gibbons, Herbert Gintis, Felix Oberholzer, Larry Samuelson, Rajiv Sethi, Herbert Simon, Vernon Smith and Frans van Winden. We regret that – in view of the very large number of comments we received – we could not do justice to all of them.

CESifo *Working Paper No. 336*
September 2000

FAIRNESS AND RETALIATION: THE ECONOMICS OF RECIPROCITY

Abstract

This paper shows that reciprocity has powerful implications for many economic domains. It is an important determinant in the enforcement of contracts and social norms and enhances the possibilities of collective action greatly. Reciprocity may render the provision of explicit incentive inefficient because the incentives may crowd out voluntary co-operation. It strongly limits the effects to competition in markets with incomplete contracts and gives rise to noncompetitive wage differences. Finally, reciprocity it is also a strong force contributing to the existence of incomplete contracts.

Ernst Fehr
University of Zurich
Institute for Empirical Research in Economics
Blümlisalpstr. 10
8006 Zurich
Switzerland
e-mail: efehr@iew.unizh.ch

Simon Gächter
University of Zurich
Institute for Empirical Research in Economics
Blümlisalpstr. 10
8006 Zurich
Switzerland
e-mail: gaechter@iew.unizh.ch

A long-standing tradition in economics views human beings as exclusively self-interested. In most economic accounts of individual behavior and aggregate social phenomena, the "vast forces of greed" (Arrow, 1980) are put at the center of the explanation. In economic models human actors are typically portrayed as "self-interest seeking with guile (which) includes .... more blatant forms, such as lying, stealing, and cheating .... (but) more often involves subtle forms of deceit" (Williamson, 1985).

However, as we will document below, many people deviate from purely self-interested behavior in a reciprocal manner. Reciprocity means that in response to friendly actions, people are frequently much nicer and much more cooperative than predicted by the self-interest model; conversely, in response to hostile actions they are frequently much more nasty and even brutal. The *Edda*, a 13th century collection of Norse epic verses, gives a succinct description of reciprocity: "A man ought to be a friend to his friend and repay gift with gift. People should meet smiles with smiles and lies with treachery." There is considerable evidence that a substantial fraction of people behave according to this dictum: People repay gifts and take revenge even in interactions with complete strangers and even if it is costly for them and *yields neither present nor future material* rewards. Our notion of reciprocity is thus very different from kind or hostile responses in repeated interactions that are solely motivated by future material gains.

We term the cooperative reciprocal tendencies "positive reciprocity" while the retaliatory aspects are called "negative reciprocity." In this paper, we first offer a brief overview of the evidence for reciprocal actions in relatively abstract one-shot games. Then, we show that reciprocity has powerful implications for many important economic domains. George Stigler (1981) wrote that when "self-interest and ethical values with wide verbal allegiance are in conflict, much of the time, most of the time in fact, self-interest-theory ... will win." Our evidence indicates that Stigler's position is often not correct. When the world is made up of self-interested types and reciprocal types, interacting with each other, the reciprocal types dominate the aggregate outcome in certain circumstances, while the self-interested types will dominate the aggregate outcome in other circumstances. We will provide evidence that there are important conditions in which the self-interest theory is unambiguously refuted. For example, in competitive markets with *incomplete contracts* the reciprocal types dominate the aggregate results. Similarly, when people face strong material incentives to free ride, the self-interest model predicts no cooperation at all. However, if individuals have opportunities to punish others in this situation, then the reciprocal types vigorously punish free riders even when the punishment is costly for the punisher. As a consequence of the punishing behavior of the reciprocal types, a very high level of cooperation can in fact be achieved. Indeed, the power to enhance collective actions and to enforce social norms is probably one of the

most important consequences of reciprocity.

This line of thought brings out another important implication of the presence of reciprocal types: Details of the institutional environment, the presence of incomplete contracts or of costly individual punishment opportunities determine whether the reciprocal or the selfish types are pivotal. Institutional features like this may thus have a tremendous impact on patterns of aggregate behavior that is neglected by the self-interest model. As a consequence, economic predictions regarding the impact of different institutions will be questionable if they do not take into account the presence of reciprocal types. Moreover, it turns out that the existence of reciprocal types may actually give rise to a world of incomplete contracts, so that reciprocity helps to generate those conditions under which it can flourish.

Positive and Negative Reciprocity: Some Evidence

Reciprocity is fundamentally different from "cooperative" or "retaliatory" behavior in repeated interactions. These behaviors arise because actors expect future material benefits from their actions; in the case of reciprocity, the actor is *responding* to friendly or hostile actions even if no material gains can be expected. Reciprocity is also fundamentally different from altruism. Altruism is a form of *un*conditional kindness; that is, altruism given does not emerge as a response to altruism received. Again, reciprocity is an in-kind response to beneficial or harmful acts.

Examples for retaliatory behavior abound. Many wars and gang crimes fit well into this category. A vivid example is provided by the recent events in Kosovo when many Albanian refugees took bloody revenge after the victory of NATO over Serbian forces. Other examples are given by the rise in employees' theft rates after firms have cut employees' wages (Giacalone and Greenberg, 1997) or by the social ostracism exercised by coworkers against strike breakers during and after industrial disputes.

Likewise, positive reciprocity is deeply embedded in many social interactions. Psychological studies show, for example, that smiling waitresses get tipped much more than the less friendly ones (Tidd and Lochard, 1978). Calls for contributions to charities are often accompanied by small gifts. Apparently, charities believe that this raises the propensity to donate. Uninvited favors, in general, are likely to create feelings of indebtedness obliging many people to repay the psychological debt. A particularly powerful example of this is the use of free samples as a sales technique (Cialdini 1993). In supermarkets customers are frequently given small amounts of a certain product for free. For many people it seems to be very difficult to accept

samples from a smiling attendant without actually buying anything. Some people even buy the product although they do not like it very much. The normative power of reciprocity is also likely to have an important impact on social policy issues (Bowles and Gintis, 1998). Social policies are much less likely to be endorsed by public opinion when they reward people independent of whether and how much they contribute to society.

Since in real world interactions, it is very difficult to rule out with certainty that an actor derives a future material benefit from a reciprocal response, we provide in the following evidence on reciprocity from controlled laboratory experiments. In these experiments, real subjects interact anonymously and face real, and sometimes rather high, material costs of reciprocal actions, in a context where it can be precluded that reciprocal responses will lead to future material rewards.

Perhaps the most vivid game to demonstrate negatively reciprocal behavior is the ultimatum bargaining experiment. In this game, two subjects have to agree on the division of a fixed sum of money. Person A, the Proposer, can make exactly one proposal of how to divide the amount. Person B, the Responder, can accept or reject the proposed division. In the case of rejection, both receive nothing; in the case of acceptance, the proposal is implemented. A robust result in this experiment, across hundreds of trials, is proposals that give the Responder less than 30 percent of the available sum are rejected with a very high probability (for example, see Güth, Schmittberger and Schwarze, 1982; Camerer and Thaler, 1995; Roth, 1995, and the references therein). Apparently, Responders do not behave in a self-interest maximizing manner. In general, the motive indicated for the rejection of positive, yet "low", offers is that subjects view them as unfair.

Negative reciprocity in an ultimatum game has been observed in many countries, including Indonesia, Israel, Japan, many European countries, Russia and the United States (for example, see Roth, Prasnikar, Okuno-Fujiwara and Zamir, 1991).[1] Moreover, rather high monetary stakes do not change or have only a minor impact on these experimental results. In the study of Cameron (1999) the amount to be divided represented the income of three months for the subjects. Other studies with relatively high stakes have involved college students dividing amounts of $100 or more (Hoffman, McCabe and Smith, 1995; Henrich (forthcoming); Slonim and Roth, 1998).

Positive reciprocity has been documented in many trust or gift exchange games (for example, Fehr, Kirchsteiger and Riedl, 1993; Berg, Dickhaut and McCabe, 1995; McCabe, Rassenti and Smith 1996). In a trust game, for example, a Proposer receives an amount of money x from the experimenter, and then can send between zero and x to the Responder. The experimenter then

[1] The only exception is the study of Henrich (forthcoming) among the Machiguenga in the Peruvian Amazon. Machiguengas exhibit very low rejection rates.

triples the amount sent, which we term *y*, so that the Responder has *3y*. The Responder is then free to return anything between zero and 3*y* to the Proposer. It turns out that many Proposers send money and that many Responders give back some money. Moreover, there is frequently a positive correlation between *y* and the amount sent back at the individual as well as at the aggregate level. Again, positive reciprocity does not appear to diminish even if the monetary stake size is rather high; for example, Fehr and Tougareva (1995) found strong positive reciprocity in experiments conducted in Moscow, where their subjects earned on average the monetary income of ten weeks in an experiment that lasted for two hours.

The fraction of subjects who show a concern for fairness and behave reciprocally in one-shot situations is relatively high. Many studies have carried out detailed analyses of individual decisions and found that the fraction of subjects exhibiting reciprocal choices is between 40 and 66 percent (Gächter and Falk; 1999; Berg et al., 1995; Fehr and Falk, 1999; Abbink, Irlenbusch and Renner, 2000). However, these same studies also find that between 20 and 30 percent of the subjects do not reciprocate and behave completely selfishly. Thus, a nontrivial minority of subjects exhibits selfish behavior. Burnham (1999) found that male behavior in the ultimatum game is systematically linked with testosterone levels. Males who reject unfair offers have higher testosterone levels than males who accept unfair offers. This is interesting because testosterone levels are thought to be important mediators of male willingness to engage in aggressive behavior.

There is now little disagreement among experimental researchers about the facts indicating reciprocal *behavior*. There also seems to be an emerging consensus that the propensity to punish harmful behavior is stronger than the propensity to reward friendly behavior (Offerman, 1999; Charness and Rabin, 2000). There is, however, disagreement regarding the main sources of reciprocal behavior. Some believe that the desire to maintain equity is most important (Bolton and Ockenfels, forthcoming). Others emphasize that the desire to punish hostile intentions and to reward kind intentions is also important (Rabin, 1993; Blount, 1995; Dufwenberg and Kirchsteiger, 1998; Falk and Fischbacher, 1999). A third possibility is that people do not respond to the intention but to the *type* of person they face (Levine, 1998). A fourth group of researchers, in contrast, views the reciprocal actions in laboratory experiments as a form of boundedly rational behavior (Gale, Binmore, and Samuelson, 1995; Roth and Erev, 1995). However, differences in interpretation notwithstanding, many researchers now agree that reciprocity is a rather stable behavioral response by a non-negligible fraction of the people that can be reliably elicited under appropriate circumstances.[2]

[2] This stability of reciprocal behavior suggests that it has deep evolutionary roots. For explanations of the evolutionary

In our view this stability and reliability renders reciprocity important for economics and raises exciting questions: How do reciprocal types change the nature of collective action problems that permeate people's interactions in firms, public bureaucracies, markets and the political sphere? To what extent can reciprocal people constrain the opportunistic tendencies of selfish people? Which institutions render the reciprocal types decisive in shaping aggregate social phenomena and when are the selfish types pivotal? How does the presence of reciprocal types change organizational outcomes, contractual and institutional choices, and the interactions in competitive markets? How do explicit economic incentives affect the propensity for voluntary cooperation among the reciprocal people? Do explicit incentives crowd out or enhance voluntary cooperation? In the rest of this paper, we offer answers to these questions.

Economic Applications

Public Goods

Many societies face the problem of how to provide public goods. For a group of self-interested agents, of course, public goods present the difficulty that since all agents will want to be free riders on the efforts of others, no agent will contribute willingly to the public good.

To take a specific example of this situation, consider the basic structure of a public good experiment run by Fehr and Gächter (forthcoming). In this experiment, there are four group members who are each given 20 tokens. All four subjects decide simultaneously how many tokens to keep for themselves and how many tokens to invest in a common public good project. For each token that is privately kept by a subject, that subject earns exactly one token. For each token a subject invests into the project *each* of the four subjects, whether they have invested in the public good or not, earns .4 tokens. Thus, the private return for investing one additional token into the public good is .4 tokens while the social return is 1.6 tokens. Since the cost of investing one token is exactly one token while the private return is only .4 tokens, it is always in the material self-interest of a subject to keep all tokens. Yet, if all group members keep all tokens privately, each subject earns only 20 tokens, while if all invest their total endowment in the public good, each subject earns 32 tokens. Thus, in this simple example, the highest level of social welfare would be achieved if everyone contributed all of their assets to the public good, but it is in the self-interest of

---

emergence of reciprocity see, e.g., Güth and Yaari (1992); Bowles and Gintis (1999); and Sethi and Somanathan (2000).

each individual to free ride, regardless of what others contribute, and to contribute nothing.[3]

To what extent can reciprocity provide the basis for agents deciding to make at least some contribution to the public good? Positive reciprocity implies that subjects are willing to contribute something to the public good if others are also willing to contribute, because a contribution to the public good represents a kind action, which induces reciprocally motivated people to contribute, too (Sugden, 1984; Keser and van Winden, forthcoming). However, to sustain contributing to the public good as a stable behavioral regularity, a sufficiently high proportion of the agents in the game have to be reciprocally motivated. Since we know that a non-negligible minority of subjects is motivated by pure self-interest, not reciprocity, it is unlikely that a positive level of contributions to the public good can be sustained as an equilibrium.

Up to this point, negative reciprocity has not played a role, because in the game as described there are no opportunities for *direct* retaliation in response to observed free riding. However, negative reciprocity can play the role that if subjects expect that others free ride, and if they interpret that as a hostile act, then they can "punish" others by free riding, too. The result is likely to be that self-interested types choose to free ride because they are self-interested, and reciprocal types free-ride because they observe others free-riding. Although the motivation to free ride is different for the reciprocal type, in the end the behavior of the selfish and the reciprocal type is indistinguishable. This public good game provides, therefore, an example where selfish types can induce reciprocal types to make "selfish" choices.[4]

However, the impact of negative reciprocity changes radically if subjects are given the opportunity to observe the contributions of others, and to punish those who do not contribute. Suppose, for example, that each subject in a group has the opportunity to *reduce* the income of each other subject in the group. Suppose further, that a reduction of the income of one other group member by x tokens costs the punisher $(1/3)*x$ tokens. It is important that punishment be costly to the agent who imposes it. After all, if punishing is costly for the punisher, selfish subjects will never punish. Hence, if all subjects were purely self-interested, contribution decisions would be unaffected by the punishment opportunity. However, negatively reciprocal subjects, who are willing to pay a price in order to act reciprocally, will use the costly punishment opportunity to punish free riders. This, in turn, will induce self-interested subjects to contribute to avoid the punishment. The public good game with *direct* punishment opportunities provides, therefore, an example where the reciprocal types can induce the selfish types to make “cooperative” choices. Fehr and Schmidt

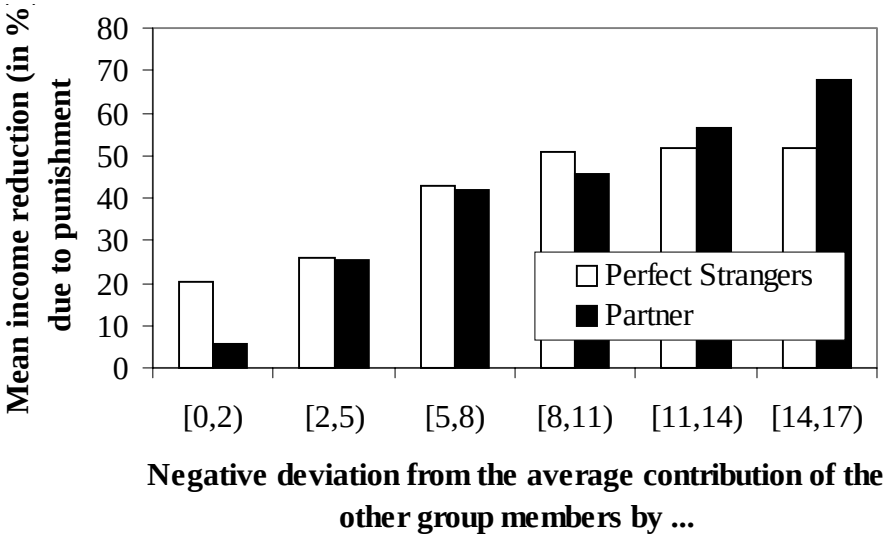
[3] For a survey of the experimental literature on public goods up to 1995 see Ledyard (1995).

[4] In Proposition 4, Fehr and Schmidt (1999) provide a formal proof of these arguments. They show that even a small minority of purely selfish subjects can induce the reciprocal subjects to behave “selfishly” in this game.

(1999, Proposition 5) show theoretically that even a *minority* of reciprocal subjects is capable of inducing a *majority* of selfish subjects to cooperate in these circumstances.

Fehr and Gächter (forthcoming) have conducted a variety of public good experiments with and without punishment opportunities, using the basic structure of the four-person, 20-token public good game just described.[5] The experiment was conducted in two versions: a "Perfect Stranger" version and a "Partner" version. In an experimental session of the Perfect Stranger version 24 subjects formed six groups with four members in each group. The public good games were repeated for six periods and in each period completely new groups were formed so that nobody met another group member more than once. The Perfect Stranger version ensures that the actions in a particular period have no rewards in future periods. In contrast, in the Partner version the *same* four members played exactly ten times.[6] In this version there are possible strategic spillovers across periods so that present actions can have future returns. However, as in the Perfect Stranger version, all subjects knew the total number of periods in advance.

Figure 1: Mean income reduction for a given negative deviation from the mean contribution of other group members



Source: Fehr and Gächter (forthcoming).

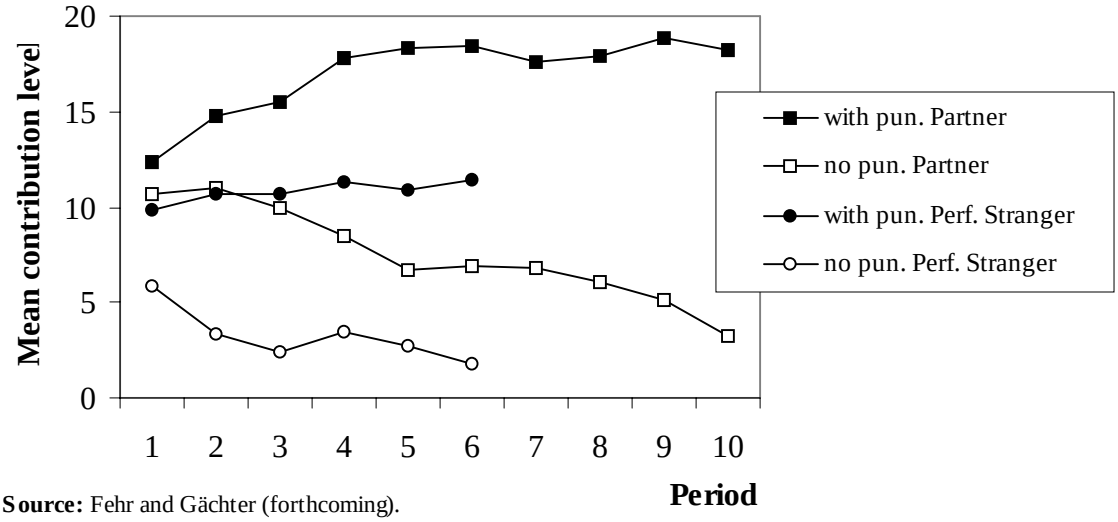
[5] For an exciting experiment with punishment opportunities in a common pool resource context, see Ostrom, Walker and Gardner (1992).

[6] In the Partner and the Perfect Stranger version experiments with and without the punishment opportunity were conducted and all interactions were completely anonymous. In the presence of the punishment opportunity subjects could punish in each period after they observed others contributions in this period.

Figure 1 shows how much a subject is punished for a given negative deviation from the average contribution of other members in the group. The punishment is measured by the average percentage reduction in the incomes of the punished subject. It turns out that the negative deviation from others' average contributions to the public good is a strong determinant of punishment. The more a subject free rides *relative* to the others the more it gets punished. Moreover, this pattern is almost the same in the two versions of the game: Free riders are punished irrespective of whether there are future rewards for the punisher. Questionnaire evidence that elicits subjects' motives and emotions indicates that the deviation from the norm of cooperation causes resentment and the impulse to punish.[7]

The heavy punishment of free riders, in turn, has a large disciplining effect on subjects' cooperation behavior, as indicated in Figure 2. This figure compares the time paths of the average contributions in the two versions of the public good game.

Figure 2: Evolution of average contributions with and without the punishment option in the Partner and the Perfect Stranger condition



Source: Fehr and Gächter (forthcoming).

[7] On the role of emotions in similar contexts see also the experimental study by Bosman and van Winden (1999).

The first observation is that in the absence of a punishment opportunity average cooperation converges to very low levels in the later periods. For instance, in period six of the Perfect Stranger version, 79 percent of the subjects free ride completely and the rest contributes little. This high defection rate stands in sharp contrast to the contribution behavior in the games with a punishment opportunity: When subjects are perfect strangers they can at least stabilize contributions at relatively high levels. In the Partner version they almost converge to the maximum level of contributions. It is particularly remarkable that in the *final* period of the Partner version subjects still contribute 90 percent of the endowment (of 20 tokens) indicating the disciplining force of the punishment opportunity.

From Public Goods to Social Norms

The problem of public goods may seem a rather limited economic application, and it may seem far-fetched to link the experiment here to government spending on basic research and development or national defense.

While we believe that such links can be made, we readily concede that the most important applications of this line of thought are not found in government budget decisions. Instead, we believe that the analytical structure of the public good problem is a good approximation to the question of how social norms are established and maintained.

At this point, it is useful to define a social norm more precisely. It is: 1) a behavioral regularity; that is 2) based on a socially shared belief how one ought to behave; which triggers 3) the enforcement of the prescribed behavior by informal social sanctions. Thus, a social norm can be thought of as a sort of behavioral public good, in which everybody should make a positive contribution -- that is, follow the social norm -- and also where individuals must be willing to enforce the social norm with informal social sanctions, even at some immediate cost to themselves.

Casual evidence and daily experience suggests that social norms are pervasive in social and economic life. The large majority of interactions in people's lives take place in the family, in residential neighborhoods, in formal or informal clubs and at people's workplaces. Typically, these interactions are not regulated by explicit contracts but by informal social norms.

For example, it has been observed in many studies that social norms influence work morale and behavior against "rate busters" (Roethlisberger and Dickson, 1947). Social sanctions by peer members are probably a very important determinant of effort behavior in work relations. It has often been observed that consumption and savings decisions are to a large degree affected by social norms that determine what others regard as "appropriate" consumption. Norm-governed attitudes,

social interactions and conformism among peers, relatives, and in neighborhoods may have important consequences for human capital decisions, for the decision to take part in elections, and for criminal activities. Social norms also often regulate the use of common pool resources (Ostrom, 1998) and the ways landowners settle disputes (Ellickson, 1994). There is a huge literature that argues that in collective action problems and in the provision of public goods, social norms play a decisive role (Elster, 1989; Ostrom, 1998). There is also reason to believe that social norms are relevant for the amount of tax evasion and the abuse of welfare payments, and for attitudes towards the welfare state in general (Lindbeck, Nyberg and Weibull, 1999). Social norms also constitute perhaps one of the most important elements of what recently has been termed "social capital" -- the informal cooperative infrastructure of our societies. Finally, there are powerful arguments that social norms also have a decisive impact on the functioning of markets. Solow (1990), for instance, has argued that they can lead to involuntary unemployment. The above examples also indicate that social norms are not necessarily beneficial for society. Dependent on the specific context of the norm they may deter or encourage socially beneficial behavior.

In our views there can, thus, be little doubt that human behavior is shaped by social norms. They constitute constraints on individual behavior beyond the legal, information and budget constraints usually considered by economists. In view of the fact that most social relations in neighborhoods, families and work places are not governed by explicit agreements but by social norms, the role of reciprocity as a norm enforcement device is perhaps its most important function.

Reciprocity as a Contract Enforcement Device

Real-world contracts are often highly incomplete, which gives rise to strong incentives to shirk (Williamson, 1985). Economic historians like North (1990) have argued that differences in societies' contract enforcement capabilities are probably a major reason for differences in economic growth and human welfare.

The employment relationship, in particular, is characterized by incomplete contracts. Labor contracts often take the form of a fixed wage contract without explicit performance incentives and in which workers have a considerable degree of discretion over effort levels. In such a situation, a worker's general job attitude, loyalty (Simon, 1991), or what Williamson (1985) called "consummate cooperation," which is "an affirmative job attitude whereby gaps are filled, initiative is taken, and judgment is exercised in an instrumental way," become important. Under a complete labor contract, of course, a generally cooperative job attitude would be superfluous, because all relevant actions would be unambiguously described and enforceable. But how can any explicit

contract unambiguously describe, assess, and enforce terms like "initiative," "good judgment" and "potentially arising gaps"?

The requirement of a generally cooperative job attitude renders reciprocal motivations potentially very important in the labor process. If a substantial fraction of the work force is motivated by reciprocity considerations, employers can affect the degree of "cooperativeness" of workers by varying the generosity of the compensation package -- even without offering explicit performance incentives.

The conjecture that reciprocity plays a role in the choice of effort has been investigated in several tightly controlled laboratory experiments. For example, in Fehr, Gächter and Kirchsteiger (1997), experimental employers could offer a wage contract that stipulated a binding wage w and a desired effort level $\hat{e}$. If an experimental worker accepted this offer, the worker was free to choose the actual effort level e between a minimum and a maximum level. The employer always had to pay the offered wage irrespective of the actual effort level. In this experiment effort is represented by a number e between 1 and 10. Higher numbers represent higher effort levels and, hence, a higher profit π for the employer and higher effort costs $c(e)$ for the worker. The effort cost for $e = 1$ was zero. The profit π from the employment of a worker was given by $\pi = 10*e - w$ and the monetary payoff for the experimental worker was $u = w - c(e)$. In each experimental session there were eight workers and six employers, who could employ at most one worker. All participants knew the excess supply of workers. It ensured that a worker's reservation wage, if he is purely selfish, was zero so that employers could, in principle, enforce very low wages. The crucial point in this experiment is that selfish workers have no incentives to provide effort above the minimum level $e = 1$. The question, therefore, is to what extent experimental employers do appeal to workers' reciprocity by offering generous compensation packages and to what extent workers honor generous offers.

It turns out that many employers indeed make quite generous offers. On average, the offered contracts stipulate a desired effort of $\hat{e} = 7$ and the offered wage implied that the worker receives 44 percent of the total surplus $u + \pi$. Interestingly, many workers honor this generosity somewhat but not fully. The actual average effort is given by $e = 4.4$ – substantially above the selfish choice of $e = 1$. However, only in 14 percent of all cases workers abide by the terms of the contract while in 83 percent of all cases they shirk. In 74 percent of all instances of shirking they do not shirk fully. Thus, although shirking is still quite prevalent in this situation the evidence suggests that in response to generous job offers, people are on average willing to put forward extra effort above what is implied by purely pecuniary considerations.

A large interview study conducted by Bewley (1995, 2000) provides field evidence supporting this view. The managers who were interviewed stress "that workers have so many

opportunities to take advantage of employers that it is not wise to depend on coercion and financial incentives alone as motivators. ... Employers believe that other motivators are necessary, which are best thought of as having to do with generosity" (Bewley, 1995, p. 252).

In the situation described above only workers can react reciprocally while employers cannot. Employers can only try to elicit reciprocal effort choices from the workers. Yet, what happens if employers can also respond reciprocally by rewarding or punishing workers after they observe actual effort choices? This question is examined as well in Fehr, Gächter and Kirchsteiger (1997). In this experiment everything is kept identical to the previous experiment except that employers could now reward or punish the workers *ex post*. For every token spent on rewards they could raise the worker's monetary income by 2.5 tokens (reflecting the possible higher marginal utility of income for workers in reality). Likewise, for every token spent on punishment they could reduce the worker's income by 2.5 tokens. Since rewarding and punishing is costly a selfish employer will never reward or punish. Hence, the reward and punishment opportunity is irrelevant according to the self-interest model.

If workers shirked in the experiments, however, employers punished in 68 percent of these cases. If there was overprovision, employers rewarded in 70 percent of these cases. If workers exactly met the desired effort, employers still rewarded in 41 percent of the cases. As a result of these expectations, workers chose much higher effort levels when employers have a reward/punishment opportunity. Indeed, although in these experiments the average *desired* effort level is slightly higher than in the previous experiment, the shirking rate declined from 83 percent to 26 percent. In 38 percent of the cases workers even provided a higher effort than requested. An important consequence of this increase in average effort was that the aggregate monetary payoff increased by 40 percent -- even if one takes the payoff reductions that result from actual punishments into account.

This evidence strongly suggests that reciprocity substantially contributes to the enforcement of contracts. The power of reciprocity derives from the fact that it provides incentives for the potential cheaters to behave cooperatively or to limit at least their degree of non-cooperation. In the above experiments, for example, even purely selfish employers have an incentive to make a generous job offer, if they expect sufficiently many workers to behave in a reciprocal manner. Similarly, even purely selfish workers have an incentive to provide a high effort in case of a reward/punishment opportunity, if they expect that sufficiently many employers respond reciprocally to their effort choices.

Work motivation and performance incentives

The previous experiments focus on fairness and reciprocity as a means to enforce contracts. In reality, material incentives are, of course, also used to mitigate the enforcement problem. The question, therefore, arises, how explicit material incentives to abide by the terms of the contract interact with motivations of fairness and reciprocity. One possibility is that reciprocity gives rise to extra effort on top of what is enforced by material incentives alone. However, it is also possible that explicit incentives may cause a hostile atmosphere of threat and distrust, which reduces any reciprocity-based extra effort. Bewley (1995, p. 252), for example, reports that many managers stress that explicit "punishment should be rarely used as a way to obtain co-operation" because of the negative effects on work atmosphere.

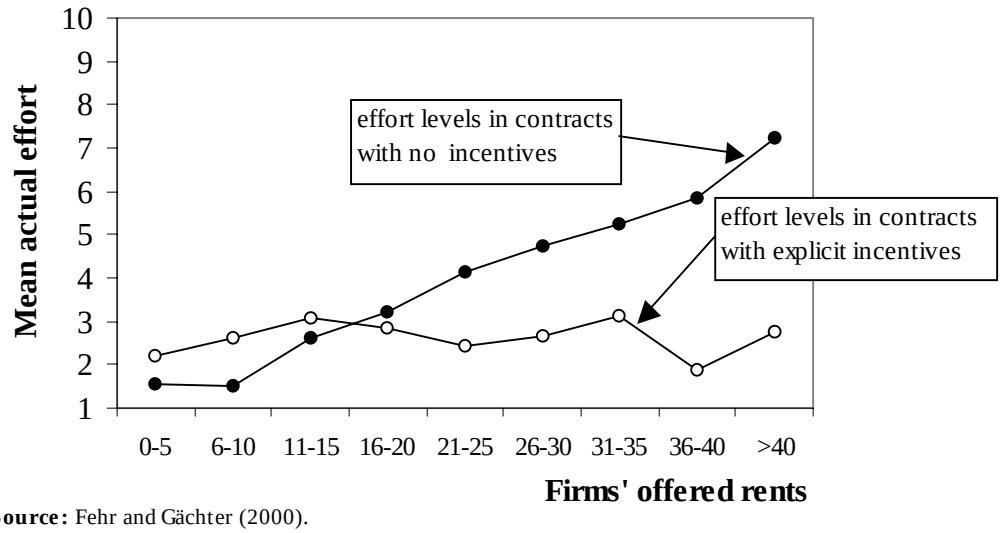
In a new series of experiments Fehr and Gächter (2000) examine this possibility. They implement a control treatment that is identical to the previous contract enforcement experiment *without* reward and punishment opportunities. Remember that in this treatment there are no material incentives. In addition, they also implement a treatment with explicit performance incentives. This treatment keeps everything constant relative to the control treatment except that employers now have the possibility to stipulate a fine, to be paid by the worker to the employer in case of verified shirking. The probability of verification is given by 1/3 and the fine is restricted to an interval between zero and a maximal fine. The maximal fine is fixed at a level such that a selfish risk-neutral worker will choose an effort level of 4 when faced with this fine.[8]

The line with the black dots in Figure 3 shows workers' effort behavior in the control treatment. It depicts the average effort on the vertical axis as a function of the rent offered to the workers. The offered rent is implied by the original contract offer; it is defined as the wage minus the cost of providing the desired effort level. Due to the presence of many reciprocal workers the average effort level is strongly increasing in the offered rent and rises far above the selfish level of $e = 1$. The line with the white dots in Figure 3 shows the relationship of rent to effort in the presence of the explicit performance incentive. Except at the low rent levels, the average effort is *lower* in the presence of the explicit incentives! This result suggests that reciprocity-based effort elicitation and explicit performance incentives may indeed be in conflict with each other. In particular, explicit incentives may "crowd" out reciprocal effort choices. In the experiments of Fehr and Gächter (2000) the average effort taken over all trades and, hence, the aggregate monetary surplus, is lower in the incentive treatment than in the control treatment. However, employers' profits are higher

[8] To prevent that hostility is introduced merely by the use of value laden terms we avoided terms like "fine", "performance", etc. Instead we used a rather neutral language like, for example, "price deduction".

because in the incentive treatment they rely much less on the "carrot" of generous wage offers. Instead they threaten the maximal fine in most cases. For the employers the savings in wage costs more than offset the reductions in revenues that are caused by the lower effort in the incentive treatment. However, while the wage savings merely represent a transfer from the workers to the firms, the reduction in effort levels reduces the aggregate surplus. This shows that in the presence of reciprocal types efficiency questions and questions of distribution are inseparable. Since the perceived fairness of the distribution of the gains from trade affects the effort behavior of the reciprocal types different distributions are associated with different levels of the aggregate gains. Thus, lump-sum transfers between trading parties have allocative consequences.

Figure 3: Actual effort-rent relation in the absence and presence of explicit performance incentives



Source: Fehr and Gächter (2000).

Our "crowding out" result may seem counterintuitive, since it is almost axiomatic to some economists that material incentives should produce a better outcome. However, this position neglects the existence of reciprocity-based voluntary cooperation. Similar problems with explicit incentives are obtained in the experiments of Gneezy and Rustichini (forthcoming) and Bohnet, Frey and Huck (1999) -- explicit material incentives may have counterproductive effects. These results, of course, do not provide a general case against the use of explicit incentives.[9] In some

[9] There exists a large psychological literature on the crowding out of intrinsic motivation by explicit rewards (Deci and

cases, there is evidence that explicit incentives can leave voluntary cooperation intact (e.g., Güth, Klose, Königstein and Schwalbach, 1998). In particular, notice that the incentive devices discussed here involved punishments and it may well be that reward-based explicit incentives do not destroy reciprocal inclinations, or may even strengthen them. However, the results do indicate that in the presence of reciprocity-based voluntary co-operation, the task of providing explicit incentives is considerably more complicated than envisaged by standard principal-agent theory.

Wage Rigidity, Rent-Sharing and Competition

In a seminal paper, Akerlof (1982) argued that labor markets are characterized by considerations of "gift exchange", by which he meant that employers offer a gift of pay which is more than labor's opportunity cost and employees offer a gift of more than minimal effort. This exchange may explain why employers are reluctant to cut wages in a recession, as found by many researchers (see, e.g., Bewley, 2000, and the references therein). The reason is that wage cuts may decrease productivity. In addition, the gift exchange notion implies that, ceteris paribus, more profitable firms pay on average higher wages. Higher profitability is likely to be associated with a higher marginal product of effort. Therefore, the return of a given effort increase is higher and employers have an incentive to pay higher wages.

The fact that the presence of reciprocal types in the labor market gives rise to downward wage rigidity has been demonstrated in a number of experiments. In the following we draw on Fehr and Falk (1999), because they confirmed the existence of downward wage rigidity in a version of the most competitive environment -- the competitive double auction. In this environment, both experimental firms and workers can make wage bids. A large body of research has shown the striking competitive properties of experimental double auctions. In hundreds of such experiments, prices and quantities quickly converged to the competitive equilibrium predicted by standard self-interest theory (see Holt, 1995, for a survey of important results). Therefore, showing that reciprocity causes wage rigidity in a double auction provides a strong piece of evidence in favor of the importance of reciprocity in markets.

Fehr and Falk (1999) carried out a series of double auction experiments set in the context of a labor market. Both experimental firms and experimental workers could make wage bids. If a bid

---

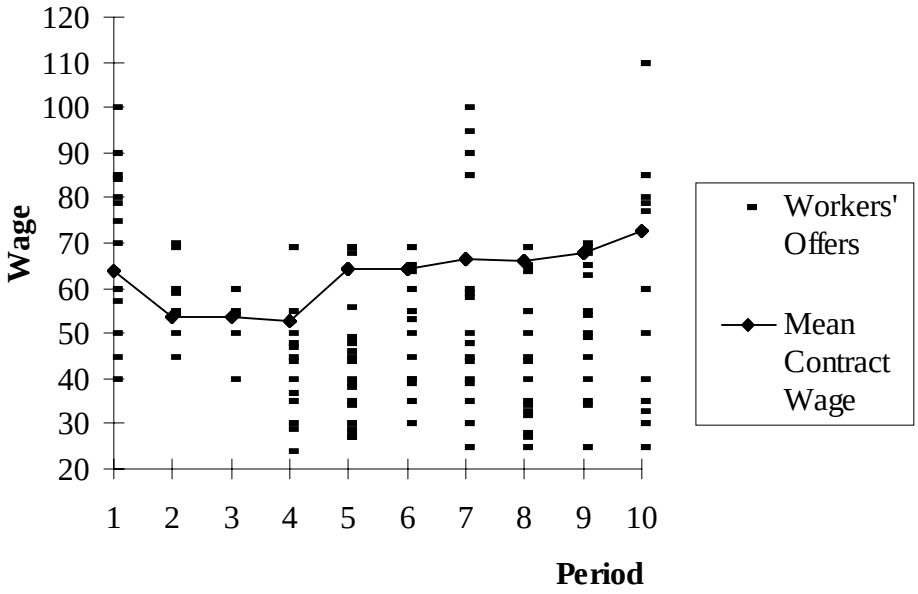
Ryan 1985). For applications of intrinsic motivation theory to economics, see Frey (1997). There are, however, considerable differences between the literature discussed above and the psychological studies on crowding out (see Fehr and Gächter, 2000).

was accepted, a labor contract was concluded. There were 8 firms and 12 workers and each firm could employ at most one worker. A worker who concluded a contract had costs of 20. Therefore, due to the excess supply of labor, the competitive wage level was 20. Within this broader context, Fehr and Falk (1999) considered two treatment conditions: one condition in which the labor contract was complete because the experimenter enforced a given effort level; and one where the labor contract was incomplete because employees could choose an effort level between a minimum and maximum after the wage contract was concluded, and the employers could neither stipulate nor enforce an effort level above the minimum level.[10]

The time path of the average wage in a typical double auction with incomplete contracts is shown in Figure 4a. Figure 4b shows average wages in a typical double auction with complete contracts. In addition, both figures show workers' wage bids. Clearly, wage levels are radically different in the two conditions. In the market with complete contracts, employers take full advantage of the low wage offers made by the workers and, as a consequence, wages are close to the competitive level in this market. In contrast, in the market with incomplete contracts employers are very reluctant to accept workers' underbidding of prevailing wages. From period 4 onwards, wages move even further away from the competitive level – despite fierce competition among workers for scarce jobs. The data analysis in Fehr and Falk (1999) shows that employers' high wage policy in the market with incomplete contracts was quite rational, because in this way they could sustain higher effort levels and increase profits relative to a low wage policy.

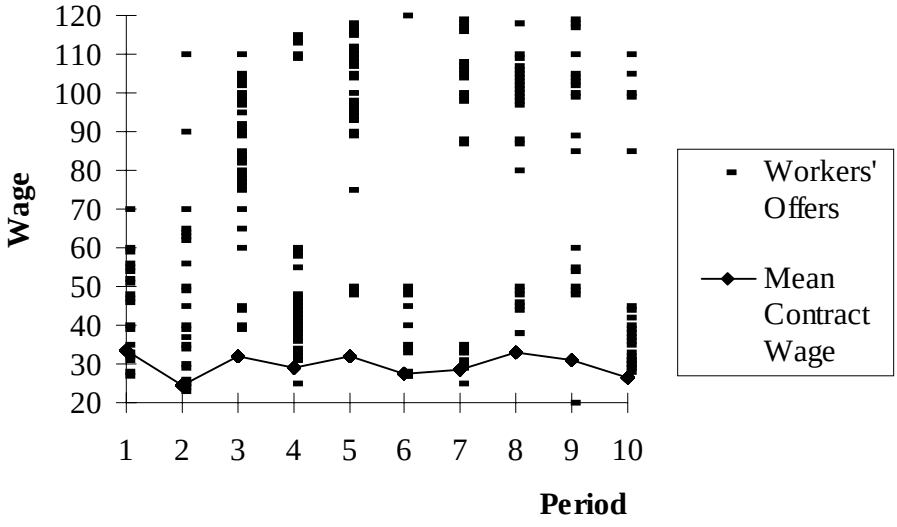
[10] One double auction lasted for ten periods and a period lasted for three minutes. In each period the same stationary situation was implemented, i.e., there were 12 workers, 8 firms, and each worker's reservation wage was 20. In a given period employers and workers could make as many wage bids as they liked, as long as they had not yet concluded a contract. In the condition with incomplete contracts workers had to choose an effort between a minimum and a maximum level after they had concluded a contract.

Figure 4a: Workers' offers and mean contract wages in the double auction market with *incomplete* contracts



Source: Fehr and Falk (1999).

Figure 4b: Workers' offers and mean contract wages in the double auction market with *complete* contracts



Source: Fehr and Falk (1999).

The big difference in the impact of reciprocity on wage formation in markets with complete and incomplete contracts illustrates again the importance of institutional details. In the incomplete contracts condition, a reciprocal worker can punish the firm by choosing a low effort level after the labor contract has been concluded. Since firms anticipate this possibility, they have a reason to pay generous wages. In contrast, in the complete contracts condition, the only method for a worker to punish a firm who offers a low wage is to reject such an offer. However, due to the presence of a certain proportion of purely self-interested workers, the reciprocal worker knows that others will accept low wage offers. Thus, reciprocal workers have, in fact, no possibility to punish firms -- which induces them to accept low wage offers, too. Since firms anticipate or notice this willingness to accept low offers, they have no reason to offer generous wages. Thus, the ability to punish is an institutional detail that means that reciprocity will have a very different impact on wage formation in the two conditions. Fehr and Schmidt (1999) provide a rigorous derivation of this argument.

A variety of studies have found that one major reason why managers are reluctant to cut wages in a recession is the fear that wage cuts may inhibit work performance. Among others, Bewley (2000) reports that managers are afraid that pay cuts "express hostility to the work force" and will be "interpreted as an insult."

A comparison of wage levels in Figure 4a and 4b shows that workers earn rents in the market with incomplete contracts. The existence of rents is also indicated by the many wage bids below the prevailing wage level in Figure 4a. This raises two questions: 1) Does reciprocity also cause a reduction in employment if employers can hire more than one worker? 2) Do the rents vary systematically with firms' profitability? In a recent paper Falk and Fehr (2000) addressed the first question. They show that firms indeed reduce employment in response to workers' reciprocity. The second question is examined in Fehr, Gächter and Kirchsteiger (1996) who conducted competitive market experiments in which experimental firms differed according to their profit opportunities. They found a clear positive correlation between firms' profit opportunities and the rents paid to workers.

This result is compatible with empirical evidence on rent sharing in real companies. For example, Blanchflower, Oswald and Sanfey (1996) show that there is a positive relation between long-run wages and the profitability of non-unionized companies or non-unionized industries, respectively. Also, Krueger and Summers (1987), for instance, have shown that estimated industry wage differentials are positively correlated in a cross-section with industry profitability. Such findings are not consistent with competitive theories of the labor market, because in those theories, firms should pay no more the opportunity cost of wages, and there is no reason that the profit opportunities of a certain firm should affect those market-determined opportunity costs. However,

this finding is predicted by rent-sharing theories of the labor market based on the presence of reciprocal agents in the market.

The combination of the findings of laboratory studies on rent sharing and the field evidence on rent sharing suggests that rent-sharing theories have explanatory power. The laboratory results have the advantage that they unambiguously show the existence of profit-related job rents, due to their ability to control fully for other factors like unobservable heterogeneity in working conditions or skill levels. Such factors can create havoc in interpreting the results of real-world studies of wage differentials (e.g., Gibbons and Katz, 1992). In addition, the laboratory approach allows for the isolation of the gift-exchange mechanism as a cause for non-compensating wage differentials. The disadvantage of the laboratory data is that further assumptions are necessary to render the results relevant for real labor markets. This comparison illustrates that field and laboratory studies should be viewed as complementary methods of economic exploration.

Foundations of Incomplete Contracts

Standard principal-agent models predict that contracts should be made contingent on all verifiable measures that are informative with regard to the agent's effort. But in reality, we often observe highly incomplete contracts. For example, as noted earlier, wages are often paid without explicit performance incentives. To this point, the discussion has focused on demonstrating that reciprocity has powerful economic effects in situations where contracts are incomplete.

This section seeks to explore the underlying causes for the prevalence of incomplete contracts in the first place. One common explanation for the absence of explicit incentives is that when employees are expected to carry out multiple tasks or when the measures of effort and performance are distorted in some way, providing powerful incentives will induce agents to focus too much on what is being measured and not enough on the other dimensions and tasks of the job (Holmström and Milgrom, 1991). This line of explanation certainly has some truth in it. However, our aim here is to show that the presence of reciprocal types is an independent source of the absence of explicit incentives.

To study the impact of reciprocity on contractual choices, Fehr, Klein and Schmidt (2000) conducted an experiment in which principals had the choice between an explicit contract and an implicit, less complete, contract. In a typical session of this experiment there are 12 principals and 12 agents who play for ten periods. In each of the 10 periods an employer faces a different principal, which ensures that all matches are one-shot. A period consists of three stages. At stage one of a period, the principal has to decide whether to offer the agent an implicit or an explicit contract. The

implicit contract specifies a fixed wage and a desired effort level (where effort choices can range from 1 to 10). In addition, the principal can promise a bonus that may be paid after actual effort has been observed. In the implicit contract, there is no contractual obligation to pay the announced bonus, nor is the agent obliged to choose the desired effort level. The principal is, however, committed to pay the wage. An explicit contract also specifies a binding fixed wage and a desired effort level between 1 and 10. Here, however, the principal can impose a fine on the agent that has to be paid to the principal in case of verifiable shirking. Except for one detail the explicit contract is identical to the performance contract discussed above in the context of "crowding out" of reciprocity. The difference concerns the fact that the choice of the explicit contract involves fixed verification cost. This reflects the fact that the verification of effort is, in general, costly. Note that the implicit contract does not require third-party verification of effort. It is only necessary that effort is observable by the principal.[11] The explicit contract is more complete than the implicit contract, because in the explicit contract the employer conditions the fine on the actual effort level in a credible manner, while in the implicit contract, no such credible commitments are possible.

At stage two, the agent observes which contract has been offered and decides whether to accept or reject the offer. If the agent rejects the offer, the game ends and both parties get a payoff of zero. If the agent accepts, the next step is for the agent to choose the actual level of effort.

At stage three, the principal observes the actual effort level. If the principal has offered an implicit contract, the next decision is whether to award the bonus payment to the agent. If the principal offered an explicit contract and if the agent's effort falls short of the agreed effort, a random draw decides with probability 1/3 whether shirking is verifiable, in which case the agent has to pay the fine.

If all players have purely selfish preferences, the analysis of this game is straightforward. A selfish principal would never pay a bonus. Anticipating this, there is no incentive for the agent to spend more than the minimum effort. If the principal chooses the explicit contract, the principal should go for the maximum punishment because this is the best deterrence for potential shirkers. The parameters of the experiment are chosen such that a risk neutral and selfish agent maximizes expected utility by choosing an effort level of 4 if faced with the maximum fine. Since the enforceable effort level is 4 under the explicit contract while it is only 1 under an implicit contract, the self-interest model predicts that principals prefer the explicit contract.

The experimental evidence is completely at odds with these predictions. In total, the

[11] Employers are, in general, not free to cut a worker's wage for shirking while they have little legal problems when they refuse to pay a promised bonus.

implicit contract was chosen in 88 percent of the cases. In view of the relative profitability of the different contracts, the popularity of the implicit contract is not surprising. Those principals who chose the explicit contract made an average loss of 9 tokens per contract, while those who chose the implicit contract made an average profit of 26 tokens per contract. Since the fixed verification cost in the explicit contract was 10 tokens, the explicit contract would have been much less profitable even in the absence of these costs. For both contracts the average income of the agents was roughly 18 tokens. Implicit contracts were more profitable because -– contrary to the standard prediction -– they induced much higher effort levels. The effort level in the implicit contract was 5.2 on average (on a scale of 1 to 10), while the effort level in the explicit contract was 2.1 on average.

How did implicit contracts induce effort levels so much higher than expected? A major reason is that in the presence of reciprocal principals, the promised bonus does not merely represent cheap talk, because reciprocal principals can -- and actually do -- condition the bonus payment on the effort level. The average data clearly reflect this impact of the reciprocal types because the actual average bonus rises steeply with the actual effort level. The principal's capability to commit to paying a conditional bonus is based on their reciprocal inclinations. Conditional bonus payments, in turn, provide a strong pecuniary incentive for the agents to perform as desired by the principals. Why did explicit contracts induce effort levels lower than expected? A likely reason is that these contracts crowd out positive reciprocity, and perhaps even induce negative reciprocity, as shown in the section on work motivation above.

One also might conjecture that the preference for implicit contracts in this particular experiment is caused by the fact that the explicit contract involves a punishment while the implicit contract involves a reward. Further experiments in Fehr et al. (2000) cast, however, doubt on this explanation. If the above-described implicit contract competes with a piece rate contract the vast majority of principals still prefer the implicit contract.

The endogenous formation of incomplete contracts through reciprocal choices shows that reciprocity may not only cause substantial changes in the functioning of given economic institutions but that it also may have a powerful impact on the selection and formation of institutions. To provide a further example: The present theory of property rights (Hart 1995) predicts that joint ownership will in general severely inhibit relations-specific investments so that it emerges only under very restrictive conditions. This may no longer be true in the presence of reciprocal actors who are willing to cooperate if they expect the trading partner to cooperate as well, and who are willing to punish even at a cost to themselves.

Concluding Remarks

The assumption that economic agents make their decisions based on pure self-interest has served economists well in many areas. In situations where contracts are reasonably complete, the underlying assumption of self-interest should continue to be especially important. However, the self-interest model has also failed to give satisfactory explanations for a wide variety of questions of interest to economists, including questions about labor market interactions, public goods, and social norms. We believe that for important questions in these areas progress will not come from additional tweaking of a pure self-interest model, but rather from recognizing that a sizeable proportion of economic actors act on considerations of reciprocity.

In view of the powerful implications of reciprocity it is also important to know why a sizeable fraction of the people has reciprocal inclinations. Which factors in the evolution of the human species have contributed to this? Which social and economic conditions produce the propensity to reciprocate? There are now several evolutionary models (see section 2) that provide answers to this question. At the empirical level, however, little is known.

References

Abbink, Klaus, Irlenbusch, Bernd and Renner, Elke (2000). "The Moonlighting Game. An Experimental Study on Reciprocity and Retribution." *Journal of Economic Behavior and Organization* 42(2), 265-277.

Akerlof, George (1982): "Labor Contracts as Partial Gift Exchange", *Quarterly Journal of Economics*, vol. 97, 543-69.

Arrow, Kenneth (1980): "Discrimination in the Labour Market", in: J.E. King (ed.): *Readings in Labour Economics.* Oxford: Oxford University Press.

Berg, Joyce and John Dickhaut, and Kevin McCabe (1995): "Trust Reciprocity and Social History", *Games and Economic Behavior*, Vol. 10, 122-142.

Bewley, Truman (1995): "A Depressed Labor Market as Explained by Participants", *American Economic Review, Papers and Proceedings*, vol. 85, 250-254.

Bewley, Truman (2000): Why Wages Don't Fall During a Recession. Cambridge: Harvard University Press.

Blanchflower, David, Andrew Oswald, and Peter Sanfey (1996): "Wages, Profits, and Rent-Sharing", *Quarterly Journal of Economics*, Vol. 111, No. 2, 227-252.

Blount, Sally (1995): "When Social Outcomes aren't Fair: The Effect of Causal Attributions on Preferences", *Organizational Behavior & Human Decision Processes*, 63, 131-144.

Bohnet, Iris, Bruno Frey and Steffen Huck (1999): "More order with less law: On contract enforcement, trust, and crowding", *mimeo*, Harvard University.

Bolton, Gary and Axel Ockenfels (forthcoming): "ERC - A Theory of Equity, Reciprocity, and Competition", *American Economic Review.*

Bosman, Ronald and Frans van Winden (1999): "The Behavioral Impact of Emotions in a Power-to-Take Game: An Experimental Study", CREED/Tinbergen Institute working paper, Amsterdam.

Bowles, Samuel and Herbert Gintis (1998): "Is Equality Passé?" *Boston Review*, 23(6), December/January 1998-99.

Bowles, Samuel and Herbert Gintis (1999): "The Evolution of Strong Reciprocity", *Mimeo,* University of Massachusetts.

Burnham, Terence (1999): "Testosterone and Negotiations", mimeo, John F. Kennedy School of Government, Harvard University.

Camerer, Colin and Richard Thaler (1995): "Anomalies: Ultimatums, Dictators and Manners", *Journal of Economic Perspectives* Vol. 9(2), 209-219.

Cameron, Lisa (1999): "Raising the Stakes in the Ultimatum Game: Experimental Evidence from Indonesia", *Economic Inquiry*, Vol. 37(1), 47-59.

Charness, Gary and Matthew Rabin (2000): "Social Preferences: Some Simple Tests and a New Model", *Mimeo*, University of Berkeley.

Cialdini, Robert (1993): Influence – The Psychology of Persuasion. New York: Quill William Morrow.

Deci, Edward and Richard Ryan (1985): Intrinsic Motivation and Self-Determination in Human Behavior. New York and London: Plenum Press.

Dufwenberg, Martin and Georg Kirchsteiger (1998): "A Theory of Sequential Reciprocity", *mimeo,* CentER Tilburg.

Ellickson Robert (1994): Order without Law – How Neighbors Settle Disputes, Cambridge, Massachusetts: Harvard University Press.

Elster, Jon (1989): The Cement of Society - A Study of Social Order, Cambridge: Cambridge University Press.

Falk, Armin and Ernst Fehr (2000): "Fair Wages and Unemployment", *mimeo,* University of Zürich.

Falk, Armin and Urs Fischbacher (1999): "A Theory of Reciprocity", *Working paper* No. 6, Institute for Empirical Research in Economics, University of Zurich.

Fehr, Ernst and Armin Falk (1999): "Wage Rigidity in a Competitive Incomplete Contract Market", *Journal of Political Economy* 107, 106-134.

Fehr, Ernst and Simon Gächter (forthcoming): "Cooperation and Punishment in Public Goods Experiments", *American Economic Review.*

Fehr, Ernst and Klaus Schmidt (1999): "A Theory of Fairness, Competition, and Cooperation", *Quarterly Journal of Economics,* vol. 114, pp. 817-868.

Fehr, Ernst and Elena Tougareva (1995): "Do High Stakes Remove Reciprocal Fairness - Evidence from Russia", Discussion paper, Univ. of Zurich.

Fehr, Ernst, Simon Gächter and Georg Kirchsteiger (1996): "Reciprocal Fairness and Noncompensating Wage Differentials", *Journal of Institutional and Theoretical Economics,* Vol. 152, No. 4, 608-640.

Fehr, Ernst, Simon Gächter, and Georg Kirchsteiger (1997): "Reciprocity as a Contract Enforcement Device", *Econometrica,* Vol. 65, No. 4, 833-860.

Fehr, Ernst, Georg Kirchsteiger, and Arno Riedl (1993): "Does Fairness prevent Market Clearing? An Experimental Investigation", *Quarterly Journal of Economics,* Vol. 108, Issue 2, 437-460.

Fehr, Ernst, Alexander Klein, and Klaus Schmidt (2000): "Endogenous Incomplete Contracts", *mimeo,* University of Munich.

Frey, Bruno (1997): Not Just For the Money - An Economic Theory of Personal Motivation, Cheltenham: Edward Elgar Publishing.

Gächter, Simon and Armin Falk (1999): "Reputation or Reciprocity?", *Working paper No.* 19, Institute for Empirical Research in Economics, University of Zurich.

Gale, John, Ken Binmore and Larry Samuelson (1995): "Learning to Be Imperfect: The Ultimatum Game", *Games and Economic Behavior* 8, 56-90.

Giacalone, Robert and Jerald Greenberg (1997): Antisocial Behavior in Organizations. Thousand Oaks: Sage Publications.

Gibbons, Robert and Lawrence Katz (1992): "Does Unmeasured Ability Explain Inter-Industry Wage Differentials?" *Review of Economic Studies,* Vol. 59, 515-535.

Gneezy, Uri and Aldo Rustichini (forthcoming): "A Fine is a Price", *Journal of Legal Studies.*

Güth, Werner and Menachem Yaari (1992): “Explaining reciprocal behavior in simple strategic games: An evolutionary approach”, in: Ulrich Witt (ed.), *Explaining process and change: Approaches to evolutionary economics*. Michigan University Press.

Güth, Werner and Rolf Schmittberger, and Schwarze, Bernd (1982): "An Experimental Analysis of Ultimatum Bargaining", *Journal of Economic Behavior and Organization*, vol. 3, no. 3, 367-88.

Güth, Werner, Wolfgang Klose, Manfred Königstein, and Joachim Schwalbach (1998): "An Experimental Study of a Dynamic Principal-Agent Relationship", *Managerial and Decision Economics* 19, 327-341.

Hart, Oliver (1995): Firms, Contracts, and Financial Structure. Oxford: Clarendon Press.

Henrich, Joe (forthcoming): "Does culture matter in economic behavior? Ultimatum Game Bargaining among the Machiguenga of the Peruvian Amazon, *American Economic Review.*

Hoffman, Elizabeth, Kevin McCabe, and Vernon Smith (1996): "On Expectations and Monetary Stakes in Ultimatum Games", *International Journal of Game Theory*, Vol. 25, 289-301.

Holmström, Bengt and Paul Milgrom (1991): "Multi-Task Principal-Agent Analyses: Incentive Contracts, Asset Ownership and Job-Design", *Journal of Law, Economics and Organization*, vol. 7, 24-52.

Holt, Charles (1995): "Industrial Organization: A Survey of Laboratory Research", in: Alvin Roth and John Kagel (eds.), Handbook of Experimental Economics. Princeton: Princeton University Press.

Keser, Claudia and Frans van Winden (forthcoming): "Conditional Cooperation and Voluntary Contributions to Public Goods", *Scandinavian Journal of Economics.*

Krueger, Alan and Lawrence Summers (1987): "Reflections on the Inter-Industry Wage Structure", in: Kevin Lang and Jonathan Leonhard (eds.), *Unemployment and the Structure of Labor Markets*, New York: Basil Blackwell.

Ledyard, John (1995): "Public Goods: A Survey of Experimental Research", in: Alvin Roth and John Kagel (eds.), Handbook of Experimental Economics. Princeton: Princeton University Press.

Levine, David (1998): "Modeling Altruism and Spitefulness in Experiments", *Review of Economic Dynamics*, 1, 593-622.

Lindbeck, Assar, Sten Nyberg and Jörgen Weibull (1999): "Social Norms and Economic Incentives in the Welfare State", *Quarterly Journal of Economics*, Vol. 116, Issue 1, 1-35.

McCabe, Kevin, Stephen Rassenti and Vernon Smith (1996): “Game Theory and Reciprocity in some Extensive Form Experimental Games”, *Proceedings National Academy of Science*, 93, 13421-13428.

North, Douglass (1990): Institutions, Institutional Change and Economic Performance. Cambridge: Cambridge University Press.

Offerman, Theo (1999): "Hurting hurts more than helping helps: the role of the self-serving bias", mimeo, CREED, University of Amsterdam.

Ostrom, Elinor, James Walker and Roy Gardner (1992): “Covenants with and without a Sword: Self-Governance is Possible”, *American Political Science Review*, Vol. 86, 404-417.

Ostrom, Elinor (1998): "A Behavioral Approach to the Rational Choice Theory of Collective Action", *American Political Science Review*, Vol. 92, 1-22.

Rabin, Matthew (1993): "Incorporating Fairness into Game Theory and Economics", *American Economic Review*, 83(5), 1281-1302.

Roethlisberger, F. J. and Dickson, W. J. (1947): Management and the Worker: An Account of a Research Program Conducted by the Western Electric Company, Hawthorne Works, Chicago, Cambridge, MA: Harvard University Press.

Roth, Alvin (1995): "Bargaining Experiments", in: Alvin Roth and John Kagel (eds.), Handbook of Experimental Economics. Princeton: Princeton University Press.

Roth, Alvin and Ido Erev (1995): "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term", *Games and Economic Behavior*, 8, 164-212.

Roth, Alvin and Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir (1991): "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study", *American Economic Review*, vol. 81, 1068-95.

Sethi, Rajiv and E. Somanathan (2000): “Preference Evolution and Reciprocity”, *mimeo*, Columbia University.

Simon, Herbert (1991): “Organizations and Markets”, *Journal of Economic Perspectives*, 5, 25-44.

Slonim, Robert and Alvin Roth (1998); "Financial Incentives and Learning in Ultimatum and Market Games: An Experiment in the Slovak Republic", *Econometrica* 66, 569-596.

Solow, Robert (1990): *The Labor Market as a Social Institution.* Oxford: Basil Blackwell.

Stigler, George (1981): "Economics or ethics?" in: S. McMurrin (ed.), *Tanner Lectures on Human Values*, Cambridge: Cambridge University Press.

Sugden, Robert (1984): “Reciprocity: The supply of public goods through voluntary contributions”, *Economic Journal*, 84, 772-787.

Tidd, K.L and J.S. Lochard (1978): "Monetary significance of the affiliative smile: a case for reciprocal altruism", *Bull. Psychonom. Soc.* 11, 344-346.

Williamson, Oliver (1985). The Economic Institutions of Capitalism. New York: Free Press.