Fairness, Incentives and Contractual Incompleteness*

Ernst Fehr[a)]
University of Zurich, CESifo and CEPR

Alexander Klein[b)]
University of Munich

Klaus M. Schmidt[c)]
University of Munich, CESifo and CEPR

This version: December 20, 2002

Abstract: We report on a series of experiments that show that concerns for fairness have dramatic consequences for the optimal provision of incentives in a moral hazard context. Incentive contracts that are optimal when there are only selfish actors become inferior when some agents are concerned about fairness. Conversely, contracts that are doomed to fail when there are only selfish actors provide powerful incentives and become superior when there are also fair-minded players. These results are consistent with the theory of inequity aversion. Furthermore, they suggest that the existence of fair actors may be an important reason why many contracts are left deliberately incomplete.

Keywords: Incentive Contracts, Moral Hazard, Fairness, Reciprocity, Incomplete Contracts

JEL Classification Numbers: C7, C9, J3

---

* We would like to thank Sam Bowles, Colin Camerer, Mathias Erlei, Dan Friedman, Paul Milgrom, Jean Tirole and seminar participants at Bocconi University, the California Institute of Technology, the University of California at Berkeley and at Santa Cruz, the University of Munich, Stanford University, the University of Toulouse, the University of Washington and University College London for helpful comments and suggestions. Susanne Kremhelmer provided excellent research assistance for the experiments discussed in this paper. Part of this research was conducted while the third author visited Stanford University and he would like to thank the Economics Department for its great hospitality. Financial support by Deutsche Forschungsgemeinschaft through grant SCHM-1196/4-1 is gratefully acknowledged. Ernst Fehr also gratefully acknowledges support from the Swiss National Science Foundation (project number 1214-05100.97) and the Network on the Evolution of Preferences and Social Norms of the MacArthur Foundation and the EU-TMR Research Network ENDEAR (FMRX-CTP98-0238).
a) Ernst Fehr, Institute for Empirical Research in Economics, University of Zurich, Bluemlisalpstrasse 10, CH-8006 Zurich, Switzerland, email: efehr@iew.unizh.ch.
b) Alexander Klein, Department of Economics, University of Munich, Ludwigstrasse 28, D-80539 Muenchen, Germany, email: alexander.klein@Lrz.uni-muenchen.de.
c) Klaus M. Schmidt, Department of Economics, University of Munich, Ludwigstrasse 28, D-80539 Muenchen, Germany, email: klaus.schmidt@Lrz.uni-muenchen.de (corresponding author).

I. Introduction

This paper deals with the question how concerns for fairness affect the optimal provision of incentives. We conducted a series of experiments in which principals could choose which type of contract to offer to the agents. The type of contract that is optimal according to standard contract theory turns out to be far less efficient than predicted by this theory, while contracts that are predicted to be very inefficient if all agents are purely self-interested turn out to be superior. The experimental results clearly suggest that this reversal in contract efficiency is due to the existence of fair subjects. The principals in our experiments seem to understand this quite well. A large majority of them chooses a contract that relies on fairness as an enforcement device. Those who choose the contract predicted by standard contract theory do very poorly. In the second part of the paper we show that these results are largely consistent with the theory of inequity aversion by Fehr and Schmidt (1999).

To better understand the nature of our results consider one of our experiments in more detail. Suppose that the principal wants to induce the agent to spend effort which is personally costly to the agent. Effort is observable by both parties, but it can be verified to the courts only if the principal invests in a verification technology. If she makes this investment, she can offer an "incentive contract" to the agent, which fines the agent for unsatisfactory performance. The problem with the incentive contract is that the verification technology is imperfect and that the fine that can be imposed on the agent is bounded, so the highest implementable effort level is positive but falls short of the efficient level of effort. Alternatively, the principal can offer a "bonus contract" which does not rely on effort verification and enforcement by third parties. Instead, the principal announces to make a voluntary bonus payment if the agent's effort is to her satisfaction. The problem with the bonus contract is that the principal cannot be forced to pay the bonus.

Given that each principal interacts with each agent only once in the experiment, a selfish principal would never pay the bonus. If it were common knowledge that all principals are selfish, rational agents would choose the minimum effort level. Thus, according to standard contract theory, the bonus contract is doomed to fail, while the incentive contract is predicted to do much better. Yet, in our experiments the overwhelming majority of principals offered a bonus contract. Even though many principals did not pay the bonus, some of them made quite generous bonus payments which induced the agents to spent much more effort than under an incentive contract. Thus, the bonus

contract turns out to induce more efficient effort choices than the incentive contract and to generate a higher payoff for both, principals and agents.

These results contradict standard contract theory that is based on the assumption that principals and agents are solely interested in their own material payoffs. While this assumption may be an accurate description of the behavior of many people, it is clearly contradicted if applied to *all* people. In fact, there is by now considerable evidence indicating that a substantial fraction of the people also cares about fairness (see e.g. the surveys of Camerer 2000, and Fehr and Schmidt 2000).[1]

In the second part of the paper we use the theory of inequity aversion (Fehr and Schmidt, 1999) to better understand our experimental results. We show theoretically that contracts that are optimal when all actors are selfish may be much less efficient when there is a minority of people who care about fairness. Furthermore, contracts that would be very inefficient if all actors were selfish, may achieve astonishingly high levels of efficiency when there are some fair-minded people. The qualitative and quantitative predictions of the theory of inequity aversion are largely consistent with the experimental results and help to organize the data surprisingly well. Thus, our theoretical results suggest that concerns for fairness can and should be taken into account in the design of optimal incentive schemes.

The theory can also be used to illustrate some of the intricate and often surprising effects that arise when some (but not all) people are fair-minded. A fair principal will pay the bonus if the agent worked sufficiently hard. Thus, if the fraction of fair principals is not too small, it pays off in expected terms for a purely self-interested agent to choose a high effort level. This makes it profitable for a selfish principal to mimic the contract offered by the fair principals in order to benefit from the high effort levels of the selfish agents under a bonus contract without paying, in fact, the bonus. Yet, fair-minded agents strongly dislike being the sucker, i.e., they experience *additional* disutility if they work hard but do not receive the bonus. If the principal cannot credibly signal that she is going to pay the bonus, fair agents are *not* willing to work under a bonus contract. Hence, it may happen that the existence of fair principals induces the selfish agents to perform well while the existence of selfish principals induces the fair agents to provide little effort under a bonus contract. Thus, too many fair agents can be detrimental to the efficiency of a bonus contract.

[1] Reciprocally fair people respond to friendly actions in a friendly manner, but they retaliate in response to hostile actions. These friendly and hostile responses also prevail in one-shot situations and when they are costly for the responder (see Rabin 1993 for this definition of reciprocal fairness).

Our paper can also been seen as a contribution to the literature on incomplete contracts. The simplicity and incompleteness of many contracts is a puzzle for traditional contract theory which predicts that contracts should depend on all verifiable information that contains statistical information about the agent's action or type (see, e.g., Holmström, 1982, Laffont and Tirole, 1993). However, actual contracts frequently specify important obligations of the contracting parties in fairly vague terms, and they do not tie the parties' monetary payoffs to measures of performance that would be available at a relatively small cost.[2] Thus, it seems that many contracts are left deliberately incomplete.

There have been many attempts to explain contractual incompleteness. Some papers, e.g. Aghion, Dewatripont, and Rey (1994), Nöldeke and Schmidt (1995), or Edlin and Reichelstein (1996) argue that in some situations a properly designed incomplete contract that is completed by renegotiation implements the first best already, so there is no need to write a more complete contract. Other papers (e.g. Che and Hausch, 1998, Segal, 1999, and Hart and Moore, 1999) show that, although the best incomplete contract does not implement the first best, a more complete contract cannot improve the situation and is of no value to the contracting parties. Spier (1992) and Allen and Gale (1992) argue that a principal may offer an incomplete contract, because a more complete contract would be interpreted as an unfavorable signal by the other party.

Our paper differs from this literature by pointing out that an incomplete contract may rely on fairness and reciprocity as an enforcement device. The experimental results show that this enforcement mechanism can be very effective, and that the large majority of principals used it quite successfully. The highly incomplete bonus contract is more efficient than the less incomplete incentive contract *because* it is less complete and thus gives more freedom to the parties to reciprocate.

This last insight is closely related to the literature on implicit incentives in repeated relationships. MacLeod and Malcomson (1989) consider a principal agent game similar to the one considered in this paper where the agent is free which effort to spend while the principal is free what bonus to pay. If this game is infinitely repeated, then there exists a subgame perfect equilibrium in which the agent works efficiently and the principal pays a generous bonus in every period. The equilibrium is sustained by the threat that, if any party deviates, the inefficient equilibrium where the agent does not work and the principal does not pay a bonus will be played in all subsequent periods.

[2] For example, a typical contract for a university professor does not make the salary directly contingent on easily

Baker, Gibbons and Murphy (1994), Schmidt and Schnitzer (1995) and Bernheim and Whinston (1998) extend this analysis and show in different setups that it may be beneficial not to contract on some variables that are easily verifiable and contractible in order to give more discretion to the parties to retaliate to a deviation from the implicit contract. However, all of these models are based on repeated game effects. In contrast, our paper shows that implicit contracts can be sustained even in a one-shot principal-agent game if some players are concerned about fairness and reciprocity.

The rest of this paper is organized as follows. Section II sets up the simple principal agent problem that we used in the experiments. Section III describes the experimental design and procedures and Section IV reports the results of the experiments. In Section 5 we discuss the experimental results. There we show that the theory of inequity aversion is largely consistent with the data and can be quite useful to organize and to better understand the data. Section VI summarizes our main results and concludes.

measurable and verifiable measures of performance such as citations, teaching ratings or the placement of Ph.D. students.

II. A Simple Principal-Agent Problem

Consider a principal who hires an agent to carry out production. If the agent exerts effort $e \geq \underline{e}$, he generates a gross profit $v(e)$ that is strictly increasing and concave in e, but he also has to incur a private cost $c(e)$ (measured in monetary terms with $c(\underline{e})=0$, $c'(e) > 0$, and $c''(e) > 0$). Let $e^{FB} > \underline{e}$ denote the unique first best efficient effort level that maximizes $v(e)-c(e)$.

Gross profits and effort costs cannot be contracted upon. Both parties observe the agent's effort level, but in order to contract on effort, it has to be verified to the courts. At date 0, before the agent chooses e, the principal can invest in a verification technology at a fixed cost k that renders effort partially verifiable. To fix ideas we assume that if the principal invested k and required the agent to work at least e^*, then with probability p, $0 < p < 1$, the courts observe whether $e \geq e^*$ or $e < e^*$. The principal can impose a fine f on the agent if shirking $(e \leq e^*)$ has been verified. However, the agent cannot be punished arbitrarily harshly, i.e., the fine f is bounded above by $\overline{f}$. Let $\bar{e}$ denote the highest effort level such that $p \cdot \overline{f} \geq c(\bar{e})$, i.e., $\bar{e}$ is the highest effort level such that it is more profitable for a risk neutral agent to choose this effort level rather than to shirk (choose $e = \underline{e}$) and to incur the expected punishment $p \cdot \overline{f}$. We will call $\bar{e}$ the *highest incentive compatible effort level*. To make the problem interesting, we assume that $\underline{e} < e^{FB}$.

The timing of events is as follows. At date 0 the principal decides whether to incur the verification cost and makes a take-it-or-leave-it contract offer to the agent. If the agent rejects the offer, both parties get their reservation utilities that we normalize to 0. If the agent accepts, he has to choose e at date 1. At date 2, a random draw determines whether the agent's effort is verifiable (in case k has been invested). Then payoffs are realized and payments are made.

Two types of contracts are feasible in this setting. If the principal did not invest in the verification technology, she can only offer a contract with a fixed wage w to the agent. If she invested in the verification technology, she can offer a contract (w, e^*, f) that stipulates a wage w, a demanded effort level e^*, and a fine f, to be paid in case that shirking ($e<e^*$) has be verified. Such a contract, that relies on effort verification and explicit, enforceable incentives, will be called *Incentive Contract (IC)*. An IC implies the following (expected) monetary payoffs:

$$M^A = \begin{cases} w - c(e) & if \quad e \geq e^* \\ w - c(e) - pf & if \quad e < e^* \end{cases} \quad and \quad M^P = \begin{cases} 10e - w - 10 & if \quad e \geq e^* \\ 10e - w + pf - 10 & if \quad e < e^* \end{cases}$$

From the point of view of traditional contract theory, the analysis of the (second best) optimal contract is straightforward if we make the standard assumption that the principal and the agent both want to maximize their material payoffs. If the agent is just offered a fixed wage, he has no incentive to provide any effort above $e = \underline{e}$. Thus, if the verification technology is not too expensive, the principal will offer an Incentive Contract which uses the maximum feasible fine $\overline{f}$ and requires the agent to choose the highest incentive compatible effort level $\overline{e}$. Furthermore, the principal will offer a wage $w = c(\overline{e})$ which compensates the agent for his effort cost and leaves him just indifferent whether or not to accept the contract.

However, this analysis rests on the important assumption that both players are interested only in their own material payoffs. To see the implications of this assumption note that effort is observable by both parties. Thus, as an alternative to the above incentive contract, the principal could simply "ask" the agent to put in $e^* > \underline{e}$ and "promise" him a reward in return. This could be done in two different ways:

(i) *Trust Contract (TC):* The principal offers the agent an unconditional payment $w > c(\underline{e})$. In return, she asks the agent to put in effort $e^* > \underline{e}$. However, if the agent accepts a trust contract, he cannot be forced to choose $e = e^*$. The monetary payoff from a trust contract (w,e^*) is given by $M^A = w - c(e)$, for the agent, and $M^P = v(e) - w$, for the principal, where e is the agent's actual effort level.

(ii) *Bonus Contract (BC):* In a BC (w,e^*,b^*) the principal offers a base wage $w \geq c(\underline{e})$ and asks the agent to spend effort $e^* > \underline{e}$. Furthermore, the principal announces to pay a bonus b^* if the agent chooses $e \geq e^*$. However, neither the agent's effort nor the principal's bonus payment are enforceable. If the agent accepts a bonus contract, he chooses any effort e at date 1. Then, at date 2, the principal is informed about e and chooses the actual bonus b. The principal is not obliged to pay $b = b^*$; she can choose any $b \geq 0$. A bonus contract implies monetary payoffs $M^A = w - c(e) + b$, for the agent, and $M^P = v(e) - w - b$, for the principal.[3]

Obviously, according to the self-interest model the trust contract and the bonus contract are doomed to fail. With a trust contract the agent knows that his wage is fixed and independent of his effort level. Therefore, he chooses $e = \underline{e}$. In a bonus contract the principal never pays the promised bonus at date 2. Anticipating this, the agent again chooses $e = \underline{e}$ at date 1.

[3] In all three types of contracts the principal was constrained to choose $w \geq c(e^*)$. We imposed this constraint in order to make sure that loss aversion does not affect the behavior of the agents.

If, however, principals and agents are not only self-interested but are also motivated by reciprocity and concerns for fairness, the outcome is less clear. By offering a generous trust contract the principal can appeal to the reciprocity of the agent, and the agent may indeed reciprocate by providing $e > \underline{e}$. If the agent is offered a bonus contract he may choose a high effort level in order to appeal to the reciprocity of the principal, and the principal may indeed reciprocate by paying a bonus voluntarily. Thus, both the TC and the BC may be more efficient than predicted by the self-interest model. Whether they will be more efficient than the IC is, however, an open question that cannot be answered on the basis of general, qualitative notions of fairness.

III. Experimental Design and Procedures

In the experimental design we chose the following parameters for this simple principal agent problem. The agents could choose effort $e \in \{1, \ldots, 10\}$ with effort costs given by Table 1.

Table 1: Effort cost function

e	1	2	3	4	5	6	7	8	9	10
c(e)	0	1	2	4	6	8	10	13	16	20

An effort of e yields a gross profit $v(e)=10 \cdot e$ to the principal. If the principal invests in the verification technology she can verify the agent's effort with probability $p=1/3$ and has to incur the fixed cost $k=10$. The fine that can be imposed on the agent is bounded above by $\overline{f} = 13$. Note that in a first best world the total surplus would be maximized if the principal did not invest in verification and the agent chose $e=10$ which would yield $v(e)-c(e) = 80$.[4]

Given the parameters of the experiment, a self-interested agent who maximizes his expected payoff can be induced to spend an effort level of at most 4 by imposing the maximum fine of 13. Thus, if both parties are self-interested, the optimal incentive contract to be offered by the principal stipulates $f=13$, $e^*=4$ and $w=4$ which holds the agent down to his reservation utility. In equilibrium monetary payoffs are $M^A = 0$ and $M^P = 26$. If the principal was restricted to offering a fixed wage, a

[4] All payoffs are in "tokens". For the conversion rate into DM see Section II.C below.

self-interested agent would always choose $e=1$, so the principal will offer w=0 which gives rise to monetary payoffs of $M^A = 0$ and $M^P = 10$.[5]

The experiments were conducted at the University of Munich in 1999 and 2001. The participants were undergraduate students from the University and the Technical University of Munich (students of law, political science, engineering, etc.). The subjects were recruited with the announcement that they could earn, depending on their decisions, a considerable amount of money during the experiment.

In 1999 we conducted four experimental sessions. In sessions S1 and S2 the principal could choose between offering a trust contract and an incentive contract (TI-treatment), in sessions S3 and S4 they could choose between offering a bonus contract and an incentive contract (BI treatment). We used the different treatments in order to generate enough data on each type of contract. In 2001 we conducted another three control sessions in which the principal could choose between a bonus contract and a contract that combined the bonus and the incentive contract by using both, a bonus and a fine.

In each session we had 20-24 subjects, half of them principals ("employers") the other half agents ("employees"). The two groups were located in separate rooms. Before the experiment started, all subjects had to read detailed instructions and to solve several exercises to make sure that all of them understand the rules of the experiment. In each session we had ten rounds. In each round an employee was matched with a different employer. Thus, for each subject in each experimental session we have ten contracts with ten different contracting partners.

After each round the subjects had to compute their own payoff and the payoff of their partner. To rule out the possibility of reputation building, the outcome of each round was strictly confidential, that is, each principal-agent pair only observed what happened in their relationship. They did not observe the contracts chosen by or offered to the other subjects in the room. Nor did they observe the past behavior of their current partner. Furthermore, the matching was random and anonymous. Finally, at the end of the session the subjects collected their total monetary payoffs privately and

[5] Note that the agent is indifferent whether to accept or to reject the contract. However, in equilibrium he must accept with probability one, otherwise the principal would have been better off by offering a slightly higher wage, $w'+\varepsilon$, that the agent would accept for sure. However, $w' + \varepsilon$ cannot be an equilibrium because $w' + \varepsilon/2$ would also be accepted for sure by the agent and yields a strictly higher payoff to the principal. Thus, the only equilibrium is the one described above.

anonymously. Each session lasted between two and two and a half hours. A complete set of the instructions for all our experiments can be found on our webpage.[6]

In each session all participants received an initial endowment of DM 20.-. The experimental (token) payoffs were exchanged into money at the rate of 1 token = 0.2 DM. Thus, an employer and an employee could jointly earn a maximum surplus of DM 16 (≈US $ 10 at the time of the experiment) in each of the ten rounds. The highest total income of one individual was DM 115.80 (US $72), an hourly wage of ca. DM 50 (US $31). However, the subjects could also make substantial losses. In order to avoid the possibility that somebody ends up with negative earnings, a subject had to drop out of the experiment if his accumulated earnings fell below DM 5.- (US $3.12), which happened three times.

IV. Experimental Results

A. The TI-Treatment

We first present the results of the TI-treatment where principals could choose between a trust contract *(w,e*)* and an incentive contract *(w,e*,f)*. We observed a total of 195 contractual choices in sessions S1 and S2. Ten incentive contracts and two trust contract have been rejected so that, in total, 183 effort choices have been made by the agents.

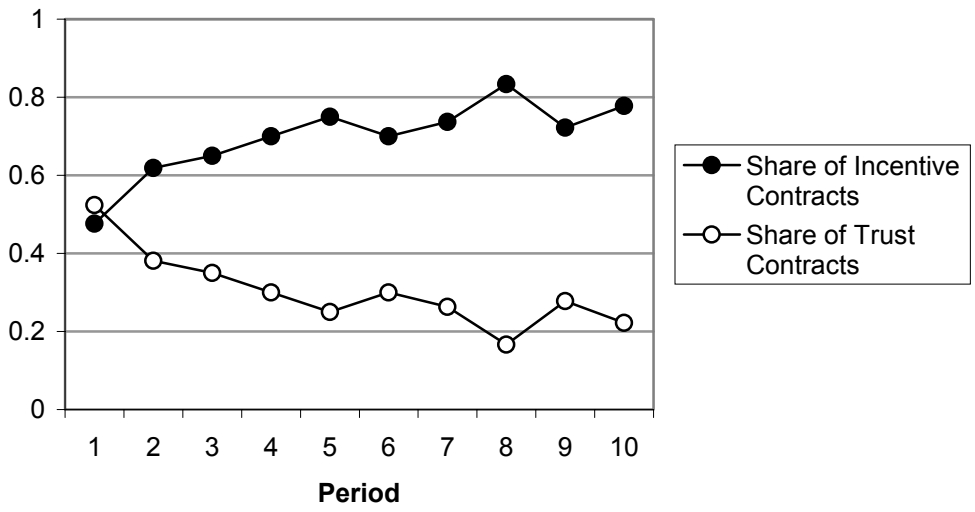
Result 1(a): *In the TI-treatment the clear majority of the contracts are incentive contracts and the share of incentive contracts increases substantially over time.*
(b) *The average effort of the agents and the average payoff of the principals are both higher in the incentive contracts.*

Overall, 135 of the 195 contracts are incentive contracts (69 percent) and only 60 (31 percent) are trust contracts. These numbers hide, however, a strong time trend in the share of incentive contracts. Figure 1, which depicts the evolution of contractual choices in the TI-treatment, shows this trend. At the beginning slightly less than 50 percent of the proposed contracts were incentive contracts. Yet, from period 4 onwards the fraction of incentive contracts is never below 70 percent and in the final three periods roughly 80 percent of all contracts are incentive contracts. Although 71 percent of the principals tried the trust contract at least once, only 33 percent did so in more than

[6] The full set of all our experimental instructions, in German and translated into English, are available at

three periods. This indicates that most principals experimented somewhat with the trust contract until they settled for the incentive contract.

Figure 1: Share of incentive and trust contracts (S1 and S2)



Result 1b is illustrated by Figure 2 that depicts the evolution of average effort levels (and average desired effort levels) over time for both contracts. The figure shows that in almost all periods the average effort is higher in the incentive contracts. Moreover, in those periods in which average effort is somewhat higher in the trust contract, the fraction of trust contracts is already small, so that this is driven by very few observations. On average, the principals earned a payoff of –0.87 when they proposed an incentive contract and –2.23 when they proposed a trust contract. These differences are not statistically significant.[7] To better understand why the principals did so poorly, let us first look at the incentive contracts.

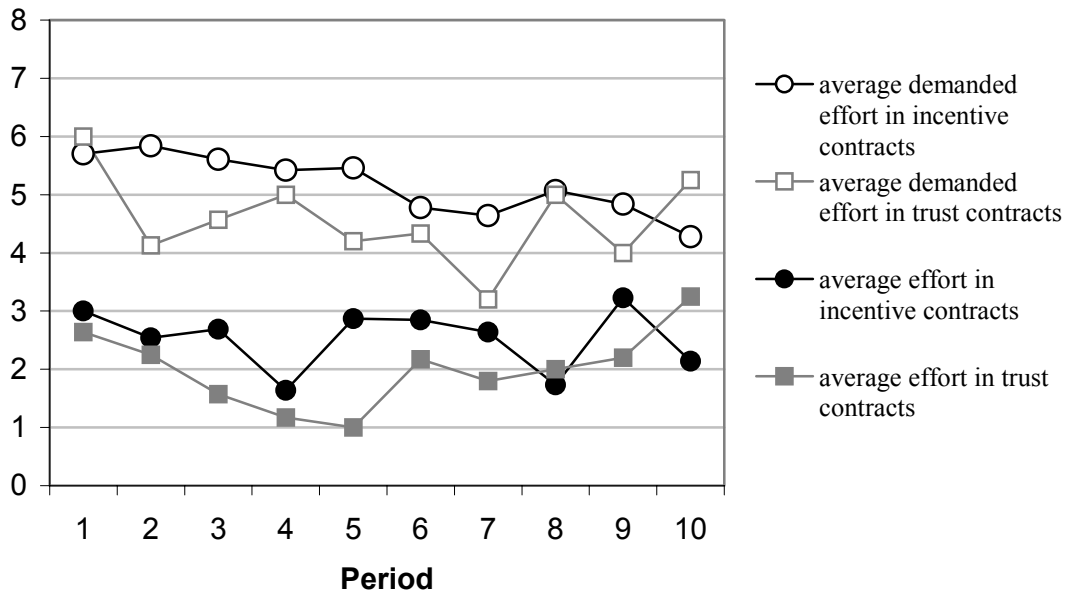
Result 2: *Although most incentive contracts stipulate the maximal fine, the majority of incentive contracts violates the no-shirking condition because the principals demand too high effort levels. There is, however, a strong increase in the share of incentive compatible contracts over time. In roughly 1/4 of all incentive compatible contracts agents shirk.*

http://www.vwl.uni-muenchen.de/ls_schmidt/experiments/incomplete_contracts/index.htm .

[7] We conducted a Wilcoxon signed-rank test. The null hypothesis that both distributions are the same cannot be rejected at any level below 79,89%.

The average fine is 12.3. Yet, in 79 (58.5 percent) of the 135 incentive contracts the no-shirking condition is violated, i.e. principals demanded too high effort levels. This is also illustrated in Figure 2 which shows that the average demanded effort level in the incentive contracts is persistently above the maximal enforceable effort of $e^* = 4$. However, demanded effort levels decline over time. In period 1 the demanded effort level is, on average, around $e^* = 6$, in the final period it is only slightly above the predicted level of $e^* = 4$. This suggests that, over time, the fraction of non-incentive compatible contracts decreases, which is indeed the case. In period one, 90 percent of all incentive contracts are not incentive compatible. This fraction decreases to 36 percent in period ten.

Figure 2: Average effort and average demanded effort in the trust and incentive contracts (S1 and S2)



A similar time pattern is exhibited by average wages. During the first few periods, when the share of incentive compatible contracts was still low, incentive contracts stipulated on average wages well above 20. The average wage decreases, however, strongly over time and reaches a level of 11.9 in period ten which is relatively close to the predicted level of 9.2. The strong time trend in the share of incentive compatible contracts and the average wage suggests that, initially, the principals tried to elicit non-incentive compatible effort levels by paying generous wages but, as these attempts failed, they converged slowly towards the predicted incentive contracts.

Table 2 presents more detailed information on the relative profitability of the different contract types and the agents' effort behavior. The final row of Table 2 shows that the average profit in the incentive compatible contracts is 8.6 while it is –7.6 in the non-incentive compatible contracts. Thus, while incentive compatible contracts are considerably more profitable than trust contracts, the non-incentive compatible contracts are less profitable than the trust contracts.[8] This explains both the strongly increasing time trend in the fraction of incentive contracts (relative to the trust contract) and the strong increase in the fraction of incentive compatible contracts (relative to the non-incentive compatible contracts).

Table 2: Wages, Effort and Principals' Payoff in the Trust-Incentive Treatment

Wage Offer	Incentive Compatible Incentive Contracts					Non-Incentive Compatible Incentive Contracts					Trust Contracts				
	#	rej.	e<e*	e≥e*	P's payoff	#	rej.	E =1	e>1	P's payoff	#	rej.	e =1	e>1	P's payoff
low w < 10	29	8	6	15	8.5	0	n.a.	n.a.	n.a.	n.a.	17	2	15	0	3.7
medium 10≤w≤ 20	26	1	6	19	9.8	33	1	20	12	-1.4	13	0	9	4	-1.0
high 20 < w	1	0	0	1	-20.0	46	0	28	18	-12.0	30	0	13	17	-6.4
All	56	9	12	35	8.6	79	1	48	30	-7.6	60	2	37	21	-2.4

Remark: In the column # the table shows the number of contract offers. The sum of the columns e<e* (e=1) and e≥e* (e ≥1) gives the number of accepted contract offers.

There are two reasons why the incentive compatible contracts are more profitable than the non-incentive compatible contracts. Principals pay much higher wages when they offer non-incentive compatible contracts and, despite their generosity, shirking is much more frequent in these contracts. Table 2 shows that in all 79 offered contracts that are *not* incentive compatible the wage is above w = 10 while in the majority of the incentive compatible contracts (in 29 of 56 cases) the wage is below w = 10. This suggests that the principals wanted to elicit reciprocal effort choices from the agents when they proposed non-incentive compatible contracts. Table 2 also shows, however, that these

[8] All these differences are statistically significant according to a non-parametric Mann-Whitney test ($p < 0.005$ when the trust contract is compared to the incentive compatible contract; $p < 0.036$ when the trust contract is compared to non-

attempts frequently failed. In 48 (62 percent) of the 78 accepted non-incentive compatible contracts the agents chose the minimal effort level.[9] This contrasts sharply with those contracts that meet the no-shirking condition. Only in 12 (26 percent) of the 47 accepted incentive compatible contracts do the agents shirk.[10]

Table 2 also shows that for trust contracts the principal's payoff is monotonically decreasing in the offered wage.[11] For wages below w = 10 the principals earns 3.7, for wages in the middle interval ($10 \leq w \leq 20$) the payoff declines to $M^P = -1.0$ and for high wages ($w > 20$) $M^P = -6.4$. A similar relation holds for the non-incentive compatible incentive contracts. In the middle interval $M^P = -1.4$ while for high wages $M^P = -12.0$. This is summarized in the following result.

Result 3: *Increasing the generosity of the wage offer to induce non-incentive compatible effort levels decreases the principals' average payoff.*

To summarize the results of the TI treatment, the clear majority of contracts uses the explicit incentive option, in most cases the fine is maximal, after an initial learning period the majority of incentive contracts demands incentive compatible effort levels, and shirking is much more prevalent in trust contracts and in incentive contracts that violate the no-shirking condition. Moreover, incentive compatible contracts are much more profitable, and it does not pay for the principals to elicit reciprocal effort choices from the fair agents by paying high wages.

B. The BI-Treatment

Next we turn to the results in the BI-treatment. In total we observed 230 contract offers in the BI-treatment. Four bonus contracts and two incentive contracts were rejected so that we have 224 accepted contracts in the BI-treatment. While in the TI treatment the incentive contract outperformed the trust contract that relied on reciprocity, the exact opposite happened when the principals could choose a bonus contract.

Result 4: (a) *In the BI-treatment the overwhelming majority of all contracts are bonus contracts.*

---

incentive compatible incentive contracts).

[9] Note that the number of accepted contracts is given by the sum of the the two effort columns. For example, for the non-incentive compatible contracts it is given by the 48 contracts with e = 1 plus the 30 contracts with e > 1.

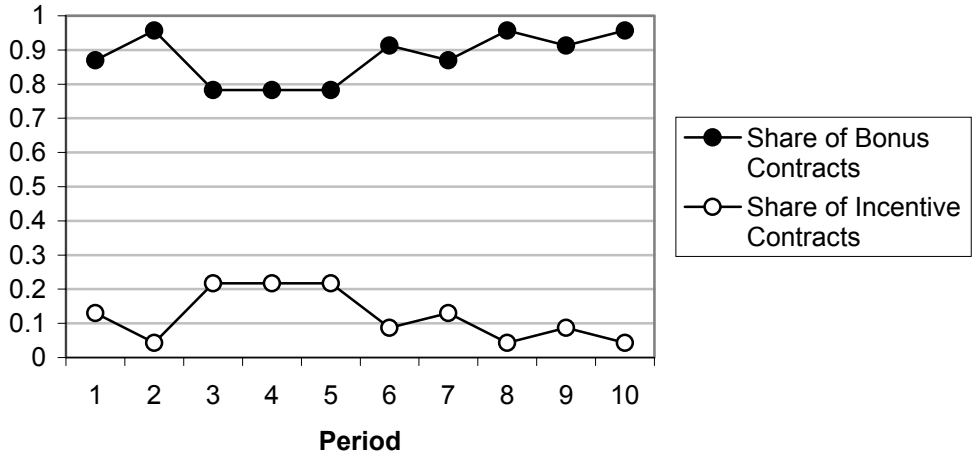
[10] In all 12 cases the agents shirked fully.

[11] This result is also confirmed by a simple OLS-regression of effort on wages that yields $e = 1.08 + 0.04\, w + \varepsilon$ where ε denotes the error term. The t-value for the constant is 3.11 while the t-value for the coefficient on *w* is 3.65. According to this regression, effort significantly increases with wages but the increase is associated with a loss: A wage increase by 10 units raises effort only by 0.4 and, hence, the expected revenue increases only by 4.

(b) *The average effort and the average payoff of the principals is much higher in the bonus contract as compared to the incentive contract.*

Figure 3 presents the evidence in favor of R4a. It shows the evolution of the share of bonus contracts over time. Already in period one 87 percent of all contracts are bonus contracts. The share of bonus contracts drops slightly below 80 percent in periods three to five because a few principals experimented with the incentive contract in these periods.[12] Yet, from period six onwards the share of bonus contracts is roughly 90 percent and approaches even 96 percent in the final period. There can thus be little doubt that principals strongly prefer the bonus contract.[13]

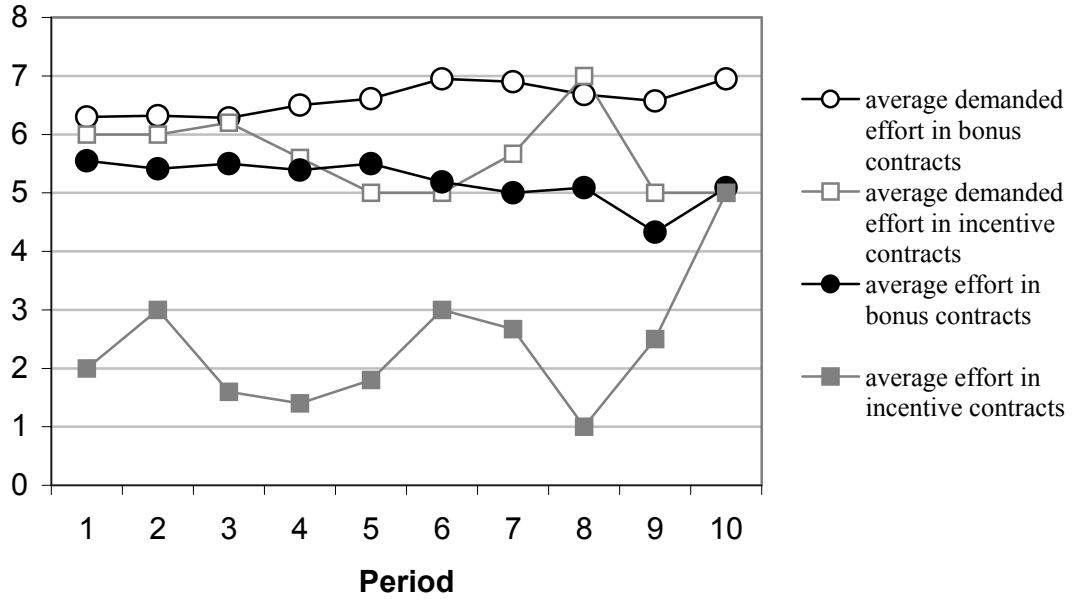
Figure 3: Share of bonus and incentive contracts (S3 and S4)



To examine the reasons for this preference we compare the average effort level in bonus and incentive contracts (see Figure 4). The figure shows that the average effort is considerably higher in the bonus contracts in nine out of ten periods.[14] While in the incentive contracts the average effort is, in general, between $e = 2$ and $e = 3$, effort in the bonus contracts is, in general, above $e = 5$. Figure 4 also indicates that agents' effort in the bonus contracts is somewhat below the desired effort level but the gap between actual and desired effort levels is much smaller than in the incentive contracts. In fact, as in the TI-treatment, many incentive contracts are not incentive compatible. This is indicated

by the fact that the desired average effort is always above $e^* = 4$. The large effort differences between the contracts are also translated into large profit differences. Principals' average *profit* from bonus contracts, taken over all ten periods, is 27 tokens while the incentive contract generates an average *loss* of 9 tokens. In each of the ten periods the average profit from bonus contracts is *always* above 20 tokens while in six of the ten periods the incentive contract causes losses and in the remaining periods the profits are, in general, rather low. In view of these large profit differences it is no longer surprising that principals exhibit a strong preference for bonus contracts.

Figure 4: Average effort and average demanded effort in the bonus-incentive treatment (S3 and S4)



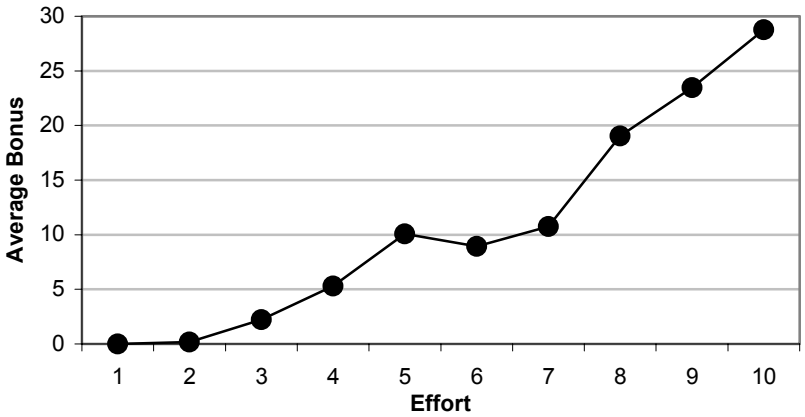
The higher effort level in the bonus contracts imply a higher surplus. To what extent did the agents receive part of this increase in the surplus relative to the incentive contracts? On average agents earned an income of 14.4 in the incentive contracts while in the bonus contracts their payoff was 17.8. Thus, agents received a small part of the surplus increase while the bulk of the increase is reaped by the principals. This shows that the option to pay a bonus yields a substantial efficiency increase *and* causes sizable changes in the distribution of the surplus.

[12] 57% of all principals tried an incentive contract at least once, but only 4% did so more than three times.
[13] In the BI-treatment trust contracts, i.e. contracts in which the principals promise and pay no bonus, are completely absent.
[14] The exemption is period ten where the effort difference is negligible. Furthermore, in the incentive contracts the average effort in period ten is based on a single data point.

Why is the bonus contract doing so much better than the incentive contract? Our next result shows that on average principals respond reciprocally to the agents' effort choice.

Result 5: (a) *The average bonus is strongly increasing in the effort level so that non-minimal effort choices are profitable for the agents.*
(b) *Principals offer on average wages of 15 and pay an average bonus of 10.4 while the agents average effort is equal to 5.2.*

Figure 5: Bonus-effort relation in the bonus-incentive treatment (S3 and S4)



Evidence for Result 5a is presented in Figure 5 which shows the average bonus as a function of the actual effort level. The figure exhibits a strongly increasing bonus-effort schedule. If the agents provide minimal or close to minimal effort levels the bonus is indeed zero while for high effort levels the average bonus approaches $b = 30$ tokens. The positive slope of the bonus-effort schedule is also confirmed by the following regressions:

$$b = \alpha_0 + \alpha_1 e + \alpha_2 e^* + \alpha_3 w + \alpha_4 b^* + \varepsilon$$

where ε represents the error term. Table 3 reports the outcome of this regression for each session and for S3 and S4 together.

Table 3: Determinants of bonus payments

Session	*N*	α_0	α_1	α_2	α_3	α_4	*Adj.* R^2

S3	85	-10.36 (-3.71)	3.06 (10.39)	1.16 (1.80)	-0.43 (-3.22)	0.15 (1.65)	0.68
S4	113	-1.51 (-0.57)	2.69 (9.84)	-0.01 (-0.01)	-0.24 (-1.59)	0.04 (0.54)	0.50
S3-S4	198	-5.58 (-2.97)	2.86 (14.23)	0.33 (0.87)	-0.30 (-2.98)	0.12 (2.15)	0.57

Note: *N* denotes the number of observations, t-values are in parentheses.

All regressions indicate a sizable and significantly positive effort (α_1) coefficient. On average, an increase in the effort level by one unit increases the expected bonus payment by 2.86 tokens. Note that this is higher than the marginal cost of effort for all effort levels $e \leq 7$, i.e., a rational and selfish agent chooses an effort level of $e = 7$ if he faces this bonus-effort relation. The impact of the desired effort level (α_2) is small and not significant, suggesting that e^* is considered to be cheap talk. The fixed wage enters the regression with a significantly negative sign, suggesting that if the actual wage increases by 1 token, the principal will reduce the bonus payment by 0.3 tokens on average. The announced bonus enters significantly with a positive, but very small coefficient. An increase in the announced bonus by 10 tokens increases the average actual bonus by only 1.2 tokens. Thus, it seems that principals feel somewhat, but not very much committed to their bonus announcements and that the effort level is the major determinant of the principals' bonus choice.

Even though principals do reciprocate quite strongly on average, it is important to notice that there are big differences in individual behavior. Considering only those (162) contracts where the agents chose a non-minimal effort level ($e > 1$), the principals did not pay any bonus in 34 cases (21%). Among those who did pay a bonus, many paid very little even if a high effort level was chosen. However, there were also some principals who reciprocated very generously to high effort levels.

Taken together the TI- and the BI-treatment show that the principals strongly prefer the bonus contract. If this contract is not available they prefer the incentive over the trust contract. In terms of the average effort and the average surplus associated with the three types of contracts the same ranking holds. Note, that this ranking implies a non-monotonic relationship between performance and contractual incompleteness. The bonus contract is the most incomplete contract. The agent is free which effort to choose and the principal is free what bonus to pay. The trust contract is somewhat less incomplete. Here the agent is still unconstrained in his effort choice, but the total compensation is fixed in advance and is not conditional on the effort level. The most complete contract is the incentive contract which restricts the principal and the agent. Our experiment shows,

that the most incomplete contract induces the highest effort and generates the highest profit for the principals, while the less incomplete trust contract does slightly worse than the complete incentive contract.[15]

C. The Control Treatment

A possible objection to the above interpretation is that we restricted the principals to use either a fine or to promise a bonus, but that they were not allowed to use both instruments simultaneously. Therefore, we conducted three sessions of a control treatment in which the principals could choose between a pure bonus contract (as in the BI treatment) and an incentive contract that was augmented by the option of a bonus payment. We will call the latter contracts "bonus+fine" contracts. In total we observed 339 contract offers in this treatment.

Result 6: *In the Control Treatment on average two thirds of all contractual offers are pure bonus contracts. This share is slowly increasing over time.*

Figure 6: Share of Pure Bonus and Bonus+Fine Contracts

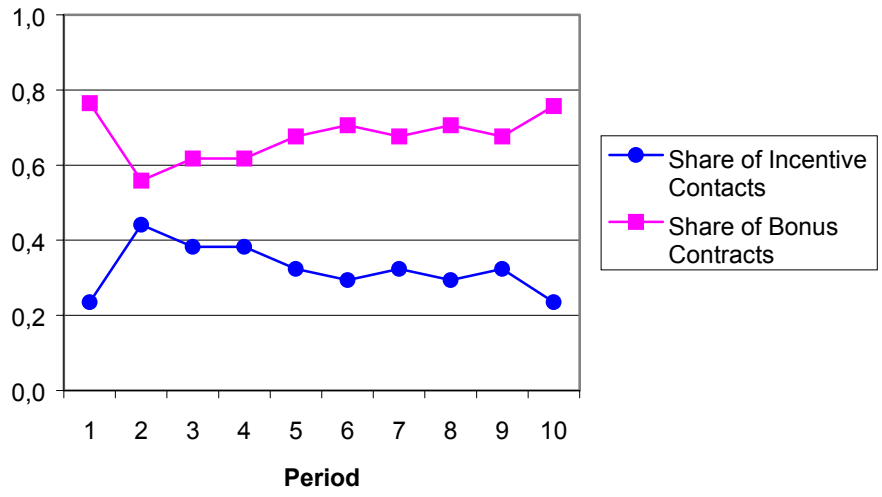


Figure 6 shows the evolution of the share of pure bonus and grand contracts. Clearly, the principals' preference for a pure bonus contract is less pronounced than in the BI treatment. A closer look at the data reveals that the performance of these two types of contracts is fairly similar.

[15] Recall that the principal had to pay a fixed verification cost of 10 if he chose the incentive contract. Without this cost, the incentive contract would do considerably better than the trust contract, but still much worse than the bonus contract.

- The average actual effort level with a pure bonus contract is 5.26, while the average actual effort level with a bonus+fine contract is 5.82. This difference is small and not statistically significant.[16] The average desired effort level is virtually identical (7.30 for pure bonus contracts vs. 7.22 for bonus+fine contracts). There is no time trend in these data.
- On the other hand, the offer of a bonus+fine contract gets rejected more often. 8 out of 110 bonus+fine contracts have been rejected (7.3%), while only 5 out of 229 pure bonus contracts (2.2%) are turned down by the agents. This difference is statistically significant at the 5% level.
- The principals' average payoff is 24.71 with a pure bonus contract, while they earned on average 24.02 with a bonus+fine contract. Thus, pure bonus contracts did marginally better, but this difference is not statistically significant.[17]
- However, the agents' average income was 19.22 with a pure bonus contract but only 12.47 with a bonus+fine contract. If we look at efficiency, i.e. at the sum of principal's and agent's payoff, then the average total surplus is 43.93 with pure bonus, but only 36.49 with bonus+fine contracts. This difference is significant at the 5% level.

Comparing the average effort level of the bonus+fine contracts (5.82) with the average effort spent under incentive contracts (2.23) in the BI-Treatment it is apparent that adding the option of a bonus payment does significantly increase the performance of the incentive contract. Note, however, that 104 out of 110 bonus+fine contracts are not incentive compatible. This fraction is even larger than with pure incentive contracts. Thus, the motivation of the agents to spend effort must be largely driven by the prospect of getting the bonus.

The bonus+fine contract requires that the principals invest in the verification technology at cost $k=10$. Without this fixed cost the principals would have done slightly better with the bonus+fine contracts.

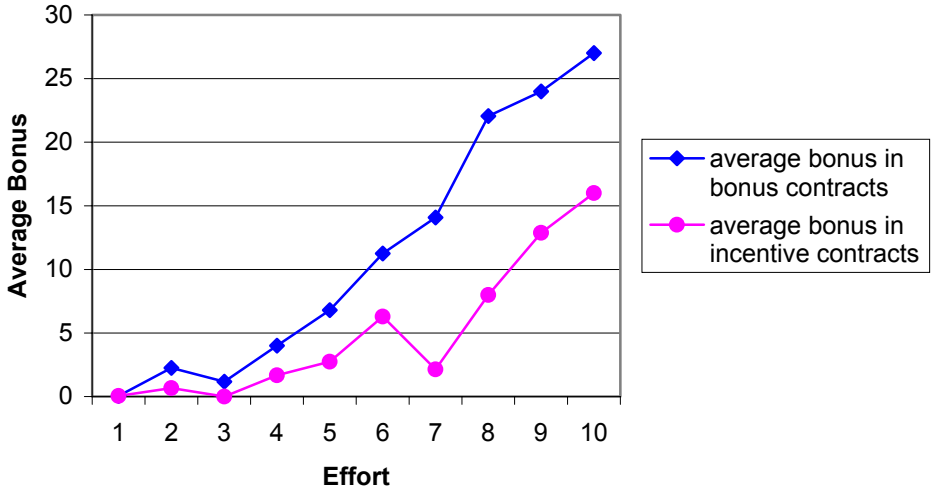
The difference in the payoffs of the agents becomes more transparent if we look at the actual bonus payments of the principals.

[16] We conducted the Wilcoxon signed-rank test. The null hypothesis that both distributions are the same cannot be rejected at any level below 20.26%.

[17] The Wilcoxon signed-rank test shows that the null hypothesis that both distributions are the same cannot be rejected at any level below 87.85%.

Result 7: *With both types of contracts the average bonus is increasing in the agent's effort level. However, bonuses are significantly more generous if a pure bonus contract had been offered. The average bonus is 10.92 in pure bonus contracts while it is only 6.16 in bonus+fine contracts.*

Figure 7: Average bonus payments conditional on effort



Evidence for Result 7 is presented in Figure 7 which shows the average bonus as a function of the actual effort level. For pure bonus contracts this function is strongly increasing for effort levels larger than 3. For bonus+fine contracts, the function is virtually flat for effort levels smaller than 7 and increases strongly only thereafter. The difference between the bonus functions is confirmed by the following regression that we carried out separately for pure bonus and bonus+fine contracts :

$$b = \alpha_0 + \alpha_1 e + \alpha_2 e^* + \alpha_3 w + \alpha_4 b^* + \varepsilon$$

Table 4 reports the outcome of this regression for sessions S5-S7.

Table 4: Determinants of bonus payments

Contracts	*N*	α_0	α_1	α_2	α_3	α_4	*Adj.* R^2
Pure Bonus	224	-3.21 (-1.56)	3.03 (18.99)	0.19 (0.53)	-0.18 (-2.42)	0.01 (-0.18)	0.64
Bonus + Fine	102	-6.02 (-1.45)	1.61 (6.85)	0.97 (1.48)	-0.29 (-2.08)	0.04 (0.45)	0.32

Note: *N* denotes the number of observations, t-values are in parentheses.

The regressions report a sizable and significantly positive effort (α_1) coefficient for both types of contracts. However, this coefficient is almost twice as large with pure bonus contracts than with bonus+fine contracts.

Again it is important to notice that there are strong differences in individual behavior. If we look at the pure bonus contracts and consider only those cases (169) where the agents chose a non-minimal effort level ($e > 1$), the principals did not pay any bonus in 38 cases (22%). This fraction is considerably larger with bonus+fine contracts (38%, 31 out of 81 contracts).

V. Interpretation of the Results

How can we explain the behavior of the subjects in these experiments? The self-interest model predicts that principals always offer the incentive contract with $e^* = 4$, $w = 4$ and $f = 13$, that this contract induces the agents to choose the incentive compatible effort level $e = 4$ and that all the surplus goes to the principals. This seems to be roughly consistent with the TI-Treatment, where principals predominantly choose the incentive contract and learn over time to make the contract incentive compatible.[18] However, the experimental results of the BI-Treatment and the Control Treatment are clearly inconsistent with the self-interest model. The overwhelming majority of principals chooses the bonus contract, many principals pay a generous bonus, and, anticipating this, agents choose considerably higher effort levels than with the Incentive Contract. Thus, in contrast to the prediction of the self-interest model, the bonus contract turns out to be the most efficient and most profitable contractual choice.

The results of the BI- and the Control Treatment suggest that fairness or reciprocity may play an important role in the principal-agent relationship. However, it is not clear why the trust contract, which also relies on reciprocity, fares so poorly, while the bonus contract does very well. If the subjects are strongly motivated by fairness or reciprocity one might have expected that the agents respond to generous wage offers by the principals in the same way as the principals respond to generous effort choices by the agents.

This puzzle can be explained by the Theory of Inequity Aversion of Fehr and Schmidt (1999). This theory captures some aspects of reciprocal fairness in a tractable way and is consistent

[18] There are three observations in the TI-Treatment that are not consistent with the self-interest model, however. First, low wage offers are frequently rejected, or, if they are accepted, agents often choose $e=1$ even if the contract is incentive compatible. Second, there are many generous wage offers with wages between 10 and 20. Finally, many agents choose effort levels larger than 1 in response to generous wage offers in trust contracts and in incentive contracts that are not incentive compatible. We will show below that these results are consistent with inequity aversion.

with the outcomes of many different classes of experimental games.[19] It has two main ingredients: First, the theory assumes that some people are not only concerned about their own material payoff but also care about inequity or, in our context, inequality.[20] Second, the theory acknowledges that people differ. Some people are very much concerned about inequality and have a high willingness to pay in order to reduce it, while others only care about their own material payoff. In the two-player case the utility function of inequity averse (fair) players is given by

$$U_i(x) = x_i - \alpha_i \max\{x_j - x_i, 0\} - \beta_i \max\{x_i - x_j, 0\},$$

$i \in \{1,2\}$, $i \neq j$, where $x=(x_1,x_2)$ denotes the vector of monetary payoffs and $\beta_i \leq \alpha_i$, $0 \leq \beta_i < 1$. In this utility function, the term weighted with α_i measures the utility loss that stems from inequality to i's disadvantage, while the term weighted with β_i measures the loss from advantageous inequality. We use a grossly simplified version of this theory. Following Fehr and Schmidt (1999) we assume that there are 60 percent self-interested types ($\alpha_i = \beta_i = 0$) and 40 percent "fair" types. Fair subjects exhibit α_i, $\beta_i > 0.5$, i.e., they have a willingness to pay in order achieve equality.[21] If the inequality is to their disadvantage, they are prepared to engage in costly "punishments" in order to reduce the payoff of their opponent. If the inequality is to their advantage, they are willing to spend resources in order to benefit the other player. Subjects with α_i, $\beta_i > 0.5$ are willing to share the surplus of a contract equally and reject offers that give them less than 25 percent of the surplus.

The assumption of just two types and the distribution of 60% selfish and 40% inequity averse types is, of course, somewhat arbitrary. However, in this particular game it turns out that the restriction to two types is very natural. In this game it only matters whether a player wants to equalize payoffs, which is the case if and only if α_i, $\beta_i > 0.5$. The assumption that 60% of all players are not prepared to equalize payoffs is derived from the distribution of types calibrated in Fehr and Schmidt (1999) with experimental data on the ultimatum game. Fehr and Schmidt used this distribution to explain the experimental results in many different classes of games, so we want to use it for this game as well. However, the qualitative results that follow are robust to changes in this

[19] Fehr and Schmidt apply the theory e.g. to ultimatum, market, public good, and gift exchange games, and show that the interaction of the distribution of types and the strategic environment can explain the major facts in these games.

distribution as long as there is a significant fraction of at least 1/3 of all players that belong to each type.

On the basis of these assumptions, our principal agent problem can be analyzed using standard game theoretic tools. The full analysis is not difficult but somewhat lengthy and is therefore relegated to an appendix that can be found on our webpage.[22] Here, we want to report the main predictions that follow from this analysis and give the intuition for them:

Proposition 1 [Trust contracts]: *Increasing the wage in a trust contract increases the effort of the fair agents, but, on average, the effort increase is too small to make a wage increase profitable for the principal.*

To see the intuition for this proposition consider a fair agent who accepted a generous trust contract. He will choose an effort level that equalizes the monetary payoff of the principal with his own monetary payoff:

$$M^P = 10 \cdot e - w = w - c(e) = M^A$$

Using the implicit function theorem, we get

$$\frac{de}{dw} = \frac{2}{10 + c'(e)}$$

Thus, for a fair agent e increases with w, but, if the fraction of fair agents in the population is $q = 0.4$, then an increase of w by 1 token increases average effort by at most $\Delta e = 0.4 \cdot 2/11 = 0.07$ which increases the principals gross profit by at most $10 \cdot 0.07 = 0.7$ tokens. Hence, a wage increase does not pay off for a selfish principal.[23] What about an inequity averse principal? A generous wage will not pay off in monetary terms, and it will generate inequality to the principal's disadvantage

[20] There is no generally accepted notion of *fairness*, but probably all fairness definitions imply that equals should be treated equally. In our experiments, the subjects enter the laboratory as equals. They have no information about their opponents and do not know with whom they trade. Thus, in these very simple environments, it seems natural to define *equality* as the reference point for a fair payoff distribution.

[21] See Fehr and Schmidt (1999) for a more extensive discussion of the experimental evidence on the distribution of inequity averse types. When Fehr and Schmidt calibrate their model to explain the quantitative evidence in the different games they use four different types, but aggregated they also have that 40 percent of subjects exhibit $\alpha_i \geq \beta_i > 0.5$ and that 60 percent exhibit $0.5 > \alpha_i \geq \beta_i$.

[22] Please visit: http://www.vwl.uni-muenchen.de/ls_schmidt/experiments/incomplete_contracts/index.htm .

[23] In other environments paying generous wages in a trust contract may be profitable for the principal. For example, it has been shown experimentally (e.g., Fehr and Falk 1999) that if the principal's payoff function is $M^P = (v-w)e$ with $v = 120$ and $e \in \{0.1, 0.2, \ldots, 1\}$, the payment of generous wages increases profits. The reason is that with this payoff function the principal's effective wage cost is $we < w$ for all $e < 1$. Thus, paying generous wages is considerably less costly if the agents choose low effort levels. In contrast, with our payoff function the wage cost is independent of the agent's effort which seems to be more natural.

whenever a selfish agent chooses *e=1*. Hence, an inequity averse principal will not pay a higher wage either. The reason why the trust contract does not work is that the fraction of fair agents is just too small. It would have to be at least *2/3* in order to make it profitable for the selfish principals to offer generous wages that induce the fair agents to choose $e > 4$.[24]

Proposition 2 [Incentive Contracts]: *The optimal incentive contract stipulates the maximal fine,* $\overline{f} = 13$*, and demands the maximal incentive compatible effort level, e*=4. Selfish principals offer w=4, while fair principals offer w=17, so the average wage offer is w=9.2. Selfish agents accept the contract and choose e=e* if and only if the contract is incentive compatible. Otherwise they shirk. If the wage offer is generous (w ≥ 17), fair agents accept it and choose e ≥ 4, where e is increasing with w. If w=4, fair agents reject the contract. If 4 < w < 17, fair agents will either reject or shirk, even if the contract is incentive compatible.*

We know already that a selfish principal would offer *(w=4, e*=4,* $\overline{f} = 13$*)* if all agents were selfish. With some fair agents, the selfish principal runs the risk that the agent is fair and rejects her offer. Thus, she may want to increase *w* in order to increase the probability that her offer will be accepted. However, it turns out that with *q=0.4* increasing the wage does not pay off. Furthermore, increasing *w* somewhat, but not up to the level that aims at equality, is dangerous, because the agent may accept the contract but consider it as unfair and shirk. Therefore, *(w=4, e*=4,* $\overline{f} = 13$*)* is optimal for a selfish principal. Note that with probability 0.4 the contracts of the selfish principals will be rejected. This explains the frequent rejection of low wage offers and why the optimal incentive contract becomes less efficient as predicted by the self-interest model due to the existence of fair agents. A fair principal offers (*w=17, e*=4,* $\overline{f} = 13$*)* which shares the surplus equally if the agent chooses *e=4*. The maximum fine is necessary to induce the selfish agents to choose *e=4*. A fair principal could pay a higher wage in order to induce the fair agents to choose $e > 4$ (which is no longer incentive compatible), but, for the same reasons as in the trust contract, this strategy will lose money on average.

Given Propositions 1 and 2 it is clear that trust contracts do not work while incentive contracts can at least be used to induce the selfish agents to choose *e=4*. Hence, the next hypothesis is no longer surprising.

[24] Note that for e > 4, c'(e) ≥ 2. An effort level $e \leq 4$ can be implemented with an incentive contract at a lower risk of suffering from inequality to the principal. Note also, that even if q > 2/3, the inequity averse principals need not offer generous wages because they may still be afraid to suffer from the inequality caused by the selfish agents.

Proposition 3 [TI-Treatment]: (a) *Both types of principals prefer incentive to trust contracts.*
(b) *Incentive contracts are more efficient and give a higher monetary payoff to the principal because they elicit, on average, a higher effort level than trust contracts.*

Thus, the main conclusion from the model of inequity aversion is the same as from the self-interest model: The incentive contract outperforms the trust contract. However, the model of inequity aversion is consistent with several observations in the TI-Treatment that are not consistent with the self-interest model. First, it explains why low wage offers are frequently rejected, or, if they are accepted, why agents often choose $e=1$ even if the contract is incentive compatible. Second, it correctly predicts that there will be many generous wage offers with wages between 10 and 20 (offered by fair principals). Finally, it offers an explanation for why many agents choose effort levels larger than 1 in response to generous wage offers in trust contracts and in incentive contracts that are not incentive compatible.

The analysis of the *bonus contract* is a little more complicated. The problem is, that the principal moves twice, first when he offers the contract and second when he chooses which bonus to pay. Thus, the agents may take the contract offer as a signal about the principal's type and update the probability that a bonus will be paid. However, it can be shown that no separating equilibrium exists in this signaling game and that both types of principals must offer the same bonus contract in equilibrium. This contract can be characterized as follows:

Proposition 4 [Bonus contract]: (a) *There does not exist a separating equilibrium in which the selfish principal offers a different contract from the contract offered by the fair principal.*

(b) If the a higher wage offer of the principal is not interpreted as a signal that the principal is selfish, then there exists a unique pooling equilibrium in which both types of principals offer w=15. The selfish agent chooses e=7 and is rewarded by the reciprocal principal with a bonus of 25, while the selfish principal does not pay a bonus. The fair agent chooses e=2 and neither type of principal pays a bonus. There does not exist a separating equilibrium.

To see that there cannot exist a separating equilibrium suppose that the fair and the selfish principal would offer different contracts. In this case the agents would know from the contract offer when they face a fair principal in which case they would choose e=10, while they would choose e=1

when facing a selfish principal. Hence, a selfish principal always wants to mimic the contractual offer of the fair principal.

At the last stage of the game it is obvious that a selfish principal will not pay a bonus while a fair principal pays a bonus that equalizes payoffs:

$$10 \cdot e - w - b = w + b - c(e)$$

Using the implicit function theorem, we get

$$\frac{db}{de} = \frac{10 + c'(e)}{2}$$

If agents believe that q percent of all principals are fair and choose b in this fashion while $(1-q)$ percent are self-interested and choose $b=0$, the expected monetary payoff of the agent as a function of e is given by

$$M^A(e) = q \cdot [w + b(e) - c(e)] + (1-q) \cdot [w - c(e)]$$

Differentiating with respect to e yields

$$\frac{dM^A}{de} = q \cdot \frac{db}{de} - c'(e) = q \cdot \frac{10 + c'(e)}{2} - c'(e)$$

This expression is positive if q is large enough as compared to $c'(e)$. Recall that, according to the cost schedule in Table 1, $1 \leq c'(e) \leq 4$. For $c'=1$, the critical value for q is 0.18, for $c' = 2$ it is 0.33, for $c' = 3$ it is 0.46 and for $c' = 4$ it is 0.57. Hence, in a pooling equilibrium, where the agents believe that they face a fair principal with probability $q = 0.4$, profit-maximizing agents will choose the maximal effort level for which the marginal effort cost does not exceed 2, that is, they choose $e = 7$. Note that this result holds for all $q \in (0.33, 0.57)$. If $q \geq 0.57$, $e=10$ maximizes the agent's expected payoff.

It is important to note that the theory implies that only self-interested agents choose $e = 7$, hoping that they will be rewarded with a generous bonus payment by the principal. Fair agents choose $e = 1$ or $e = 2$ depending on the wage offered (for $w \geq 10$ the fair agent chooses $e = 2$). The reason for this interesting implication is that a fair agent suffers more than a self-interested agent if he meets a selfish principal who does not pay the bonus. Hence, even if it is profitable from a monetary perspective to choose $e = 7$, a fair agent prefers $e = 1$ or $e = 2$ in order to ensure that equality prevails and to avoid the disutility from disadvantageous inequality when the bonus is not paid. Thus, the presence of fair principals induces selfish agents choose high effort levels while the presence of selfish principals induces the fair agents to provide low effort levels. This is an interesting example of the sometimes surprising effects that arise in a heterogeneous population with fair and selfish subjects.

There are many pooling equilibria in this game that differ in the unconditional base wage (and therefore also in the bonus to be paid ex post). If we impose the mild condition on out of equilibrium beliefs that higher wage offers are not taken as a signal that the principal is more likely to be selfish, then the set of pooling equilibria shrinks to a singleton in which all principals offer $w=15$.[25] If the selfish agent chooses $e=7$, then a bonus of 25 just equalizes payoffs. However, this bonus is paid only by the fair principals, so the expected bonus is $0.4 \cdot 25=10$. The fair agents choose $e=2$, so the expected effort level is $0.6 \cdot 7 + 0.4 \cdot 2=5$.

Proposition 5 [BI-Treatment]: *Both types of principals prefer bonus to incentive contracts. The bonus contract elicits, on average, a higher effort level than the incentive contract and both, principals and agents, achieve a higher material payoff.*

Propositions 4 and 5 are in sharp contrast to the self-interest model, but they explain the experimental results of the BI-Treatment surprisingly well. The model of inequity aversion not only predicts that bonus contracts outperform incentive contracts. It also offers an explanation for several observations that may have been puzzling at first glance and even the quantitative predictions fit the experimental data very well: (i) The model explains why some principals pay a generous bonus while others don't. (ii) The predicted average bonus payment is 10 and the predicted wage is 15, which matches the average observed behavior ($b=10.4$ and $w=15.0$) quite closely. (iii) The model explains why some agents do not choose a high effort level even though this would have been profitable in expected terms. (iv) The average predicted effort level is 5, again very close to the average actual effort level in the experiments ($e=5.2$).

Furthermore, the above analysis also shows why the bonus contract is so much better than the trust contract, even though both contracts appeal to fairness. There are two reasons for this results: First, with a trust contract the principal has to pay the generous wage upfront before observing the agent's effort, so he has to reward the agent even if the agent shirked. The bonus contract offers additional opportunity for reciprocation, because the principal can make the reward contingent on the effort spent by the agent. Second, with a bonus contract the agents have to make the first step and to trust the principals. Note that even if the agent engages in the maximum effort of 10, the cost of this effort $(c(10)=20)$ is small as compared to the upfront payment that the principal would have to make

[25] Note that in a pooling equilibrium with bonus contracts wages cannot exceed 15. For $w > 15$, $b(7) = 40 - w < 25$ and, hence, the expected bonus payment is $0.4 \cdot b(7)$ which is less than 10. Therefore, the selfish agent will no longer be willing to provide $e = 7$.

in a trust contract in order to induce a fair agent to choose $e=10$. The wage that equalizes payoffs at $e=10$ is $w=60$, so the principal risks to loose 50 if the agent is not trustworthy, while the agent looses only 20 if the principal does not pay the bonus. This suggests a general principle for gift exchange games: The party for whom it is less costly to engage in trustworthy behavior should always move first.

Let us finally briefly consider the Control Treatment. Unfortunately, the model of inequity aversion offers less clear-cut predictions in this case. If the principal can choose between a pure bonus and a bonus+fine contract, then there exists a continuum of pooling equilibria. For example, in the equilibrium which gives the highest payoff to the principal (both the selfish and the fair type) all principals offer a bonus+fine contract $(w=20, e^*=10, f=13, b^*=35)$ which is accepted by both types of agents. The selfish agent chooses $e=10$ while the reciprocal agent chooses $e=3$. The selfish principal does not pay a bonus, the fair principal pays $b=35$ if $e=10$. But there are many other equilibria as well. In particular, it is an equilibrium that both principals offer the pure bonus contract characterized by Proposition 4. The crucial question is how the agents interpret the fact that they are offered a bonus+fine contract. If they take this offer as a signal that the principal is likely to be selfish, then a pure bonus contract will be offered.

When the principals offered a bonus+fine contract in the experiments, they faced a trade-off. On the one hand, the additional instrument of the fine can be used to induce the agents to spend more effort. If they shirk, they do not only loose the expected bonus, they may also be fined. On the other hand, using the fine requires the investment in the verification technology which is not necessary for a pure bonus contract. Even though the cost of the verification technology was fairly small, it may have induced some principals to go for the pure bonus contract, in particular if they did not believe that the threat of the fine is going to be very effective. Finally, and perhaps most importantly, offering a bonus+fine contract may have been interpreted as a signal by the agents that they face a selfish principal. In fact, for any given effort level of the agents those principals who opted for the bonus+fine contract paid only half as much bonus as the principals who offered a pure bonus contract. Thus, at least ex post it seems to have been the case that principals who offered pure bonus contracts are more trustworthy.

VI. Concluding Remarks

Our experiments have shown that the existence of fair subjects may have important consequences for the optimal provision of incentives. Incentive contracts that are optimal when there are only selfish

actors perform less well when some agents are concerned about fairness. On the other hand, incomplete bonus contracts that cannot work when all actors are selfish provide powerful incentives and become superior when there are also fair-minded players. Principals seem to understand that fairness matters and predominantly choose the bonus contract that relies on fairness as an enforcement device. This contract gives more discretion to the principals to reciprocate to the agent's effort, which renders this contract more efficient than more complete contracts that rely on explicit incentives only.

There are several other points that deserve to be emphasized. First, it is important to keep in mind that only some subjects are "reciprocators" who are prepared to reward friendly behavior and to punish unfriendly behavior even if this is costly to them. There is also a considerable fraction of subjects who seem to be interested only in their own material payoff. Second, whether or not appealing to reciprocal fairness works as an enforcement device depends on the fraction of fair types in the population and on the strategic situation in which the subjects interact. We have seen, for example, that the trust contract is not a good alternative to the more complete incentive contract. Third, our theoretical results show that simple and tractable models of fairness can yield interesting and non-obvious insights into the problems of incentive provision. The Fehr-Schmidt (1999) model turns out to be surprisingly accurate in explaining the details of our experimental results and it organizes the qualitative data pattern very well.

Our experiments and the theoretical analysis show that the presence of fair types does not automatically provide a solution to every contracting problem and may sometimes even exacerbate incentive and contracting problems. Fair types are much more afraid of "exploitative" situations in which the other party may take advantage of them. The reason is that they do not only value their material payoffs but they also value the fairness of the behavior of their opponents and the equity of the final outcomes. For example, our theoretical analysis shows that self-interested agents respond much stronger to the implicit incentives provided by a bonus contract. The reason is that – in case that the principal does not reward the agent with a bonus – a fair agent experiences *additional* disutility from the unfairness of the behavior of the principal while the selfish agent "only" suffers from the reduced material payoff. This shows that the presence of fair types may complicate the task of incentive provision because – in addition to the conventional incentive compatibility constraints – the "fairness compatibility" of the contract also has to be taken into account.

It may be surprising that some of the results of the model of inequity aversion (Propositions 1-3) are quite similar to the predictions of the self-interest model, while Propositions 4 and 5 get very

different conclusions. This illustrates a point of general importance:[26] The presence of a substantial fraction of fair types does not guarantee that the aggregate outcome diverges from the predictions of the self-interest model. However, there are also circumstances where the reverse is true: The presence of a majority of self-interested types does not guarantee that the aggregate outcome is in line with the predictions of the self-interest model. The application of the Fehr-Schmidt model to the BI-treatment illustrates this nicely. Although we still maintain the assumption that 60 percent of the subjects are completely self-interested (and rational) the predictions of Propositions 4 and 5 are very different compared the predictions of the self-interest model.

To conclude, our experiments show that concerns for fairness have an important impact on the optimal provision of incentives. These effects have been neglected by traditional contract theory but have to be taken into account if we want to fully understand the functioning of real world incentives schemes. Our theoretical analysis shows that it is possible to model these effects explicitly and offers a first attempt how to develop richer models that may become part of a new paradigm, "behavioral contract theory".

[26] In Fehr and Schmidt (1999) several other examples of this kind are provided.

References

Aghion, Philippe, Dewatripont, Matthias and Rey, Philippe, 1994. "Renegotiation Design with Unverifiable Information." *Econometrica* 62, 257-282.

Allen, Franklin and Gale, Douglas, 1992. "Measurement Distortion and Missing Contingencies in Optimal Contracts." *Economic Theory* 2, 1-26.

Baker, George, Gibbons, Robert and Murphy, Kevin J., 1994. "Subjective Performance Measures in Optimal Incentive Contracts." *Quarterly Journal of Economics* 109, 1125-56.

Bernheim, B. Douglas and Whinston, Michael D., 1998. "Incomplete Contracts and Strategic Ambiguity." *American Economic Review* 88, 902-32.

Bolton, Gary E. and Ockenfels, Axel, 2000. A Theory of Equity, Reciprocity and Competition. *American Economic Review* 100, 166-193.

Camerer, Colin F., 2000. *Behavioral Game Theory*, forthcoming, Princeton, Princeton University Press.

Charness, Gary, and Rabin, Matthew, 2000. "Social Preferences: Some Simple Tests and a New Model." *mimeo*, University of California at Berkeley.

Che, Yeon-Koo and Hausch, Donald B., 1999. "Cooperative Investments and the Value of Contracting." *American Economic Review* 89(1), 125-47.

Dufwenberg, Martin and Kirchsteiger, Georg, 1998. "A Theory of Sequential Reciprocity." Discussion Paper. CentER, Tilburg University.

Edlin, Aaron S. and Reichelstein, Stefan, 1996. "Holdups, Standard Breach Remedies, and Optimal Investment." *American Economic Review* 86(3), 478-501.

Falk, Armin and Fischbacher, Urs, 1998. "A Theory of Reciprocity." Discussion Paper. University of Zürich.

Fehr, Ernst, Gächter, Simon and Georg Kirchsteiger, 1997. "Reciprocity as a Contract Enforcement Device." *Econometrica* 65(4), 833-860.

Fehr, Ernst and Falk, Armin, 1999. "Wage Rigidity in a Competitive Incomplete Contract Market", *Journal of Political Economy* 107(1), 106-134.

Fehr, Ernst and Schmidt, Klaus M., 1999. "A Theory of Fairness, Competition and Co-operation." *Quarterly Journal of Economics* 114, 817-868.

Fehr, Ernst and Schmidt, Klaus M., 2000. "Theories of Fairness and Reciprocity: Evidence and Economic Applications", forthcoming in: M. Dewatripont, L.P. Hansen, S. Turnovski, *Advances in Economic Theory, Eigth World Congress of the Econometric Society*, Cambridge: Cambridge University Press.

Hart, Oliver and Moore, John, 1999. "Foundations of Incomplete Contracts." *Review of Economic Studies* 66, 115-138.

Holmström, Bengt, 1982. "Moral Hazard in Teams." *Bell Journal of Economics* 13, 324-340.

Holmström, Bengt and Milgrom, Paul, 1991. "Multi-task Principal-Agent Analyses." *Journal of Law, Economics, and Organization* 7 (Sp.), 24-52.

Laffont, Jean-Jacques and Tirole, Jean, 1993. *A Theory of Regulation and Procurement*. Cambridge (Mass.): MIT-Press.

MacLeod, W. Bentley and Malcomson, James M., 1989, "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment." *Econometrica* 57, 447-480.

Nöldeke, G., Schmidt, K.M., 1995. Option Contracts and Renegotiation: A Solution to the Hold-Up Problem. *Rand Journal of Economics* 26, 163-179.

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

Rabin, Matthew, 2000. "Risk Aversion, Diminishing Marginal Utility, and Expected-Utility Theory: A Calibration Theorem", Econometrica 68(5), 1281-1292.

Schmidt, Klaus M. and Schnitzer, Monika, 1995. "The Interaction of Explicit and Implicit Contracts". *Economics Letters* 48, 193-99.

Segal, Ilya, 1999. "Complexity and Renegotiation: A Foundation for Incomplete Contracts." *Review of Economic Studies* 66(1), 57-82.

Spier, Kathryn, 1992. "Incomplete Contracts in a Model with Adverse Selection and Exogenous Costs of Enforcement." *RAND Journal of Economics* 23, 432-443.