CONTRACTS, FAIRNESS, AND INCENTIVES

ERNST FEHR
ALEXANDER KLEIN
KLAUS M. SCHMIDT

CESIFO WORKING PAPER NO. 1215
CATEGORY 4: LABOUR MARKETS
JUNE 2004

CESifo *Working Paper No. 1215*

Contracts, Fairness, and Incentives

Abstract

We show experimentally that fairness concerns may have a decisive impact on both the actual and the optimal choice of contracts in a moral hazard context. Explicit incentive contracts that are optimal according to self-interest theory become inferior when some agents value fairness. Conversely, implicit bonus contracts that are doomed to fail among purely selfish actors provide powerful incentives and become superior when there are some fair-minded players. The principals understand this and predominantly choose the bonus contracts, even preferring a pure bonus contract over a contract that combines the enforcement power of explicit and implicit incentives. This contract preference is associated with the fact that explicit incentives weaken the enforcement power of implicit bonus incentives significantly. Our results are largely consistent with recently developed theories of fairness, which also offer interesting new insights into the interaction of contract choices, fairness and incentives.

JEL classification: C7, C9, J3.

Keywords: moral hazard, incentives, bonus contract, fairness, inequity aversion.

Ernst Fehr
Institute for Empirical Research in Economics
University of Zurich
Bluemlisalpstrasse 10
CH-8006 Zurich
Switzerland
efehr@iew.unizh.ch

Alexander Klein
Department of Economics
University of Munich
Ludwigstrasse 28
80539 Muenchen
Germany
alexander.klein@Lrz.uni-muenchen.de

Klaus M. Schmidt
Department of Economics
University of Munich
Ludwigstrasse 28
80539 Muenchen
Germany
klaus.schmidt@Lrz.uni-muenchen.de

A previous version of this paper was circulated under the title: “Fairness, Incentives and Contractual Incompleteness”. We would like to thank Sam Bowles, Colin Camerer, Mathias Erlei, Dan Friedman, Gilat Levy, Paul Milgrom, Massimo Motta, 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 Copenhagen, Stanford University, the University of Toulouse, the University of Washington, the University College London and the European University Institute 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 12-67751.02), the Ludwig Boltzmann Institute for Research in Economic Growth and the Network on the Evolution of Preferences and Social Norms of the MacArthur Foundation.

1. Introduction

This paper addresses the question of how concerns for fairness affect the actual and the optimal choice of contracts. We conducted a series of experiments where principals could choose which type of contract to offer to the agents. The optimal type of contract according to standard contract theory proves to be far less efficient than this theory predicts, while contracts forecast to be very inefficient if all agents are purely self-interested turn out to be superior. The experimental results suggest that this reversal in contract efficiency is due to the existence of fair subjects, as they exert a decisive impact on the incentive properties of different types of contracts. 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. Moreover, we also observe intriguing interactions between explicit and implicit incentives: the simultaneous use of both explicit and implicit incentives significantly weakens the enforcement power of the implicit incentive. Therefore, the large majority of the principals prefer contracts that rely solely on implicit incentives, i.e., they do not combine explicit with implicit incentives. In the final part of the paper, we show that these results are largely consistent with a simple model of fairness.

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 expend effort which is personally costly to her. Both parties can observe effort, but the courts can only verify it 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 explicit incentive contract is that the verification technology is imperfect and the fine that can be imposed on the agent is limited, so the highest effort level which can be implemented 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 promises a non-binding, voluntary bonus payment if the agent’s effort is satisfactory. This bonus contract is an implicit contract because third parties do not enforce the principal’s promise.

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, standard contract theory forecasts that the bonus contract is doomed to fail, while predicting the incentive contract to do much better. Yet, the overwhelming majority of principals offered a bonus contract in our experiments. Even though many principals did not pay the bonus, a substantial number of them made quite generous bonus payments, inducing the agents to spend much more effort than under an incentive contract. Thus, the bonus contract turns out to induce more efficient effort choices and, therefore, the principals predominantly prefer the bonus contract relative to the incentive contract.

These results contradict standard contract theory 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 wrong if applied to *all* people. In fact, there is considerable evidence indicating that a substantial percentage of people also care about fairness (see e.g. the surveys of Sobel 2002, Camerer 2003, and Fehr and Schmidt 2003). Our experiments indicate that the principals' contract choices differ from those predicted by the self-interest model because concerns for fairness strongly affect the incentive properties of the contracts.

However, we also conducted a second experiment in which the principal was restricted to choose between an incentive contract and a "trust contract". A trust contract offers a (generous) fixed wage to the agent and asks him to return this favor by choosing a high effort level. Standard contract theory again predicts that principals will choose the incentive contract. While the bonus contract appeals to the fairness of the principal to reward high effort, the trust contract appeals to the fairness of the agent to reciprocate a generous fixed wage. However, in contrast to the bonus contract the trust contract did rather poorly. Many principals experimented with the trust contract, but on average they incurred losses and eventually most of them shifted to the superior incentive contract.

The superiority of the bonus contract relative to the explicit incentive contract suggests that fairness concerns might be responsible for the fact that principals often do not use explicit incentives even though they are readily available.[1] To examine this question, we conducted a third experiment in which the principals could also propose a contract combining the explicit and the implicit incentives. In principle, this contract could implement very powerful incentives because a shirking agent incurs the cost of the fine *and* loses the bonus payment. Surprisingly, however, the use of the explicit incentive is typically associated with significantly lower bonus payments, i.e., the explicit incentive partially crowds out the bonus incentive. Therefore, pure bonus contracts are more efficient than those combining explicit and implicit incentives. Moreover, the large majority of the principals

[1] For example, a typical contract for a university professor does not make the salary directly contingent on easily measurable and verifiable measures of performance such as citations, teaching ratings or the placement of Ph.D. students.

indeed choose pure bonus contracts, suggesting the relevance of fairness concerns for explaining the absence of explicit incentives.

Our experimental evidence not only contrasts the viewpoint of standard contract theory, but also constitutes a challenge for theories of fairness. For example, why does the incentive contract outperform the trust contract – as standard contract theory predicts – while the bonus contract surpasses the incentive contract, contradicting standard theory. How can the remarkable performance difference between the trust and the bonus contract be explained as, after all, both contracts rely on fairness as an enforcement device? Why do principals voluntarily forgo the opportunity to combine the enforcement power of explicit incentives and implicit – fairness-based – incentives in favor of a purely implicit incentive? We provide a unified interpretation of our results in terms of a simple model of inequity aversion (Fehr and Schmidt, 1999) in the final part of the paper. We primarily chose this model because of its tractability. Inequity aversion is a simple extension of the standard self-interest model that takes the fact into account that some people are not only interested in their own material payoff but also dislike inequity. The model implies that the incentive contract, which is optimal when all actors are purely self-interested, is much less efficient when a share of people cares about fairness. Furthermore, bonus contracts that would be very inefficient if all actors were selfish achieve astonishingly high levels of efficiency when there are some fair-minded people. Thus, the major predictions of the model are consistent with the observed qualitative pattern of contract choices. In addition, the model makes surprisingly accurate quantitative predictions for the bonus contracts and, if we choose plausible out-of-equilibrium beliefs, it can also account for the fact that the principals prefer pure bonus contracts to those that combine explicit and implicit incentives. However, the model also shows that the presence of some fair-minded people alone does not suffice to implement efficient behavior. The incentive structure of the "gift exchange" is also very important. We show that, as a general principle, the person who loses less from trusting the other person should trust first. Thus, both our theoretical as well as our empirical results suggest that concerns for fairness can and should be taken into account in the design of optimal incentive schemes.

The model 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 percentage of fair principals is not too small, a purely self-interested agent will chose a high effort level as his expected return will be sufficiently large. This makes it profitable for a selfish principal to mimic the contract the fair principals offer, in order to benefit from the high effort levels of the selfish agents under a bonus contract without actually paying the bonus. However, 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. While the existence of fair principals might induce selfish agents to perform well, the presence of selfish principals encourages 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.

A large number of empirical papers have examined the effectiveness of different incentives schemes over the last 10 – 15 years. This literature was surveyed in Prendergast (1999) and in Chiappori and Salanié (2003). Both survey papers conclude that “incentives matter”, i.e., agents often seem to respond to changes in incentives in ways that are consistent with the predictions of prevailing principal-agent models. However, both papers also report that the evidence for the predictions of contract choices is much weaker. This is a main reason why we focused on the principals’ contract choices between different types of contracts in our experiments. In this way, our experiments may contribute to a better understanding of the forces determining which contracts prevail.

Chiappori and Salanié (2003) emphasize that problems of unobserved heterogeneity and endogenous selection often complicate clean inferences about the incentive effects of contracts. In fact, these problems result in an ambiguous interpretation of correlations between different contracts and different behaviors. Do the contracts induce the corresponding behaviors or are the behavioral differences across contracts the result of self-selection of heterogeneous individuals to different contracts? This problem is – in our view – particularly severe in the context of fairness preferences because there is little hope that non-experimental field data allow the control for such preferences. Therefore, an experimental approach to these questions can offer additional insights. In our experiments, for example, we had complete control of the selection of agents to contracts because principals and agents were randomly matched. Thus, we can unambiguously infer the incentive consequences of different types of contracts and how the principals responded to these incentive effects.

Previous work by Camerer and Weigelt (1988); Fehr, Kirchsteiger, and Riedl (1993); Berg, Dickhaut, and McCabe (1995); and Fehr, Gächter, and Kirchsteiger (1997) indicated that fairness concerns may play an important role in moral hazard contexts.[2] However, these papers neither studied the interaction between fairness concerns and explicit incentives nor how the principals choose between explicit and implicit incentives. There have been several experimental studies in the past few years which examined how the provision of explicit incentives affects the agents' behavior in a moral hazard context. DeJong, Forsythe, Lundholm and Uecker (1985) showed how different institutional remedies, such as liability rules, mitigate the moral hazard problem. Schotter, Bull and Weigelt (1987) study the effects of piece rates and tournament incentives, and Schotter and Nalbantian (1997) examine the performance of various group incentive schemes. Chaudhuri (1998) investigated the ratchet effect in a dynamic principal-agent experiment in which the principals chose output contingent wages. Likewise, Güth, Klose, Königstein, and Schwalbach (1998) examined a multi-period principal agent game in which the principals could offer linear profit-sharing contracts. Cooper, Kagel, Lo and Gu (1999) studied how Chinese students and managers respond to the incentives underlying the ratchet effect. Keser and Willinger (2000); Güth, Königstein, Kovacs, and Zala-Mezo (2001); and Anderhub, Gächter, and Königstein (2002) also studied the performance of output-contingent wages in a moral hazard context.[3]

Many of the studies mentioned above find indications that concerns for fairness and reciprocity affect the acceptance of explicit incentive contracts. However, the principals did not have the choice between *different types of contracts* in these papers – in particular, the choice between explicit incentive contracts and implicit bonus contracts. It was this setting which enabled us to identify the strength and the limits of the standard approach in contract theory by isolating conditions under which the model's contract choice predictions are met and conditions under which these predictions failed. Moreover, our fairness model explains why the predictions of the standard model are correct if incentive contracts compete with trust contracts but are completely at odds with the facts when bonus contracts become available. Our setting is also unique in the sense that it enabled us to study the interaction between explicit and implicit incentives. Since we implemented the possibility of choosing contracts which combined explicit and implicit elements we were able to find the puzzling crowding out of implicit incentives through explicit incentives. We are not aware of any other empirical work that addresses this question or that documented such a crowding out effect.

---

[2] More recently, Charness and Dufwenberg (2003) pointed out that there are important interactions between social preferences and communication opportunities in a moral hazard context. In their setting no explicit incentives are present.

[3] There are also a few experiments on the effects of incentives in environments with adverse selection (see Cabrales and Charness 2003 and the references therein).

The rest of this paper is organized as follows. Section 2 describes the principal-agent problem that we used in the experiments. Section 3 discusses the experimental design and procedures and Section 4 reports the results of the experiments. In Section 5 we offer a theoretical interpretation of the experimental results. We show that a simple fairness model is largely consistent with the data and can be quite useful in organizing and better understanding the data. Section 6 summarizes our main results and concludes.

2. A Simple Principal-Agent Problem

Consider a principal who hires an agent to carry out production. If the agent expends effort $e \geq \underline{e}$, he generates a gross profit $v(e)$ for the principal 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 by the courts. At date 0, before the agent chooses e, the principal can invest in a verification technology at a fixed cost k that permits partial verification of effort. 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 $p\overline{f} \geq c(\bar{e})$, 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 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 $\bar{e} < e^{FB}$.

The timing of events is as follows. At date 0, the principal decides whether to incur the verification cost and offers a take-it-or-leave-it contract 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.

If the principal does not invest in the verification technology, she can only offer a contract with a fixed wage w to the agent. If, however, she invests 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 shirking ($e<e^*$) is verified. Such a contract, relying on effort verification and explicit, enforceable incentives, will be called *Incentive Contract (IC)*. The agent's (expected) monetary payoff in an IC is given by $M^A = w - c(e)$ if $e \geq e^*$ and by $M^A = w - c(e) - pf$ if the agent shirked ($e<e^*$). The principal's expected monetary payoff is defined by $M^P = v(e) - w - k$ in case of $e \geq e^*$ and by $M^P = v(e) - w + pf - k$ if the agent shirked.

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 only 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 IC based on the maximum feasible fine $\overline{f}$ and requiring the agent to choose the highest incentive compatible effort level $\bar{e}$. Furthermore, the principal will offer a wage $w = c(\bar{e})$ which compensates the agent for his effort cost and leaves him indifferent whether to accept the contract or not.

However, this analysis rests on the important assumption that both players are only interested in their own material payoffs. To see the implications of this assumption, note that both parties can observe effort. 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) With a *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) With a *Bonus Contract (BC):* In a BC (w,e^*,b^*) the principal offers an unconditional base wage $w \geq c(\underline{e})$ and asks the agent to expend effort $e^*>\underline{e}$. Furthermore, the principal announces her intention to pay a bonus b^* if the agent chooses $e \geq e^*$. However, neither the agent's effort nor the principal's bonus payment is enforceable. If the agent accepts a bonus

contract, he chooses effort e at date 1. The agent is not obliged to choose $e = e^*$ but can choose any $e \geq \underline{e}$. Then, at date 2, the principal is informed about e and chooses the actual bonus b. The principal is not obliged to pay $b = b^*$ but 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.

Obviously, both the trust contract and the bonus contract are doomed to fail according to the self-interest model. The agent knows that his wage with a trust contract is fixed independently of his effort level. Therefore, he will choose $e = \underline{e}$. The principal will never pay the promised bonus at date 2 in a bonus contract. Anticipating this, the agent will again choose $e = \underline{e}$ at date 1. Thus, the self-interest model predicts that the IC will dominate both TCs and BCs, implying that a rational and self-interested principal will never propose a TC or a BC.

If, however, principals and agents are not only self-interested but also motivated by concerns for fairness and reciprocity, the outcome is less clear. By offering a generous trust contract, the principal can appeal to the fairness 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 fairness 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 the self-interest model predicts. Such a change in the relative efficiency of the different contracts may then induce the principals to prefer a TC or a BC over an IC. The question of whether TC and BC are more efficient than IC remains open, however, and cannot be answered on the basis of general, qualitative notions of fairness or in the absence of empirical evidence.

Therefore, we implemented a series of experiments in which the principals had the option of choosing between IC, TC, and BC. In a first step, we studied how the existence of fairness concerns affects the principals' choice between TC and IC by implementing the Trust-Incentive (TI) treatment, in which only a TC or an IC could be offered to the agents. In a second step, we examined how the availability of a nonbinding bonus affects the principals' relative preference for the incentive contract. We implemented the Bonus-Incentive (BI) treatment for this purpose, where all three contracts could be chosen. The BI treatment is thus well suited for examining whether the existence of fairness concerns affects the relative efficiency of the different types of contracts and whether the principals take such effects into account when they choose from the available contracts. The BI treatment does not enable us, however, to examine whether the principals voluntarily forgo the opportunity to use an available explicit incentive. This is due to the fact that they could only choose one of the three available contracts, i.e., they could not combine the IC with the BC. We can only argue that the principals voluntarily forgo the opportunity for using the incentive if the principals prefer a pure BC or a pure TC over the combination of a BC with the explicit incentive. For this reason, we implemented the extended Bonus-Incentive (EBI) treatment, in which TC, BC, IC, or a combined incentive-bonus contract could be chosen. The EBI treatment is thus well suited for examining the question of whether the principals voluntarily forgo the opportunity of using an explicit incentive. Moreover, this treatment enables us to study the interaction between explicit and implicit incentives if agents face both types of incentives simultaneously in a combined contract.

3. Experiment Design and Procedures

We chose the following parameters for our experiments. The agents could choose effort $e \in \{1, 2, \dots, 10\}$ 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 at cost k = 10, she can verify the agent's effort with probability p = 1/3. The maximum fine the agent can be charged 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 a total surplus of $v(e) - c(e) = 80$. The principal is constrained to choose $w \geq c(e^*)$ in all types of contracts. This rules out losses for the agents if they meet their contractual obligations. We imposed this constraint to ensure that loss aversion does not affect the agents' behavior.

Given the parameters of the experiment, a self-interested agent who maximizes his expected payoff can be induced to choose 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 the principal offers stipulates f

$= 13$, $e^* = 4$ and $w = 4$ which limits the agent to his reservation utility. In equilibrium, the monetary payoffs are $M^A = 0$ and $M^P = 26$. If the principal were restricted to offer a fixed (unconditional) wage, a self-interested agent would always choose $e=1$, so that the principal would offer $w = 0$; in this case the monetary payoffs are $M^A = 0$ and $M^P = 10$.[4]

The experimental subjects were students of the University of Munich and the Technical University of Munich (students of law, political science, engineering, etc.). We had 20-24 subjects in each session, half of them randomly assigned to the role of the principal and half to that of the agent. The two groups were located in separate rooms. All subjects had to read detailed instructions and to solve several exercises before the experiment started, to ensure that all of them understood the rules of the experiment. We had ten periods in each session. The agents were randomly matched with a different principal in each period. The randomization procedure ensured that no agent interacted more than once with the same principal. Thus, we had ten contracts with ten different contracting partners for each subject in each experimental session.

After each period, the subjects had to compute their own payoff and that of their partner. The outcome of each period remained strictly confidential in order to rule out the possibility of reputation building, that is, each principal-agent pair only observed what happened in their own relationship. They did not observe the contracts offered by the other subjects in the room. Nor did they observe their current partner's past behavior. Furthermore, the matching was random and anonymous, i.e., the subjects' identity was never revealed to the other players. Finally, the subjects collected their total monetary payoffs privately and anonymously at the end of the session. 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.[5]

We conducted seven experimental sessions. We implemented the TI treatment in Sessions 1 and 2, where the principals could choose between TC and IC. In Sessions 3 and 4, we applied the BI treatment in which the principal could choose TC, IC, or BC. Finally, we conducted three sessions (S5 – S7) in which the principals could also offer a contract combining the available explicit incentive with the nonbinding announcement of a bonus payment.

[4] Note that the agent is indifferent whether to accept or to reject this contract. Because wages are discrete, a second equilibrium exists in which the principal offers a wage that is one token higher. This increases the agent's payoff while decreasing the principal's payoff by 1.

[5] The full set of all our experimental instructions, in the original German and translated into English, are available at http://www.vwl.uni-muenchen.de/ls_schmidt/experiments/incomplete_contracts/index.htm .

All participants received an initial endowment of DM 20 (≈ US $12.5 at the time of the experiment) in each session. The experimental (token) payoffs were converted into real money at the publicized rate of 1 token = DM 0.2. Thus, a principal and an agent could jointly earn a maximum surplus of DM 16 (US $10) in each of the ten periods. The highest total income of one individual was DM 115.80 (US $72), corresponding to an hourly wage of roughly DM 50 (US $31). However, the subjects could also incur substantial losses. In order to avoid the possibility of having a subject finish with negative earnings, he was expelled from the experiment if his accumulated earnings fell below DM 5.- (US $3.12); this occurred, but only three times.

4. Experimental Results

4.1. The Trust-Incentive (TI) Treatment

We present the results of the TI-treatment in this section, 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 contracts were rejected meaning that, in total, the agents made 183 effort choices. Our first result concerns the principals' contract choices.

Result 1(a): *A clear majority of the contracts in the TI-treatment 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 higher under the incentive contracts.*

Figure 1 and the following numbers support Result 1a: 135 (69 percent) of the 195 offered contracts are incentive contracts while only 60 contracts (31 percent) are trust contracts. However, these numbers fail to demonstrate the strong time trend in the share of incentive contracts shown in Figure 1. While slightly less than 50 percent of the proposed contracts were incentive contracts in the first period of the experiment, this fraction never fell below 70 percent beginning in period 4 and exceeded 80 percent of all contracts in the final three periods. Although 71 percent of the principals tried the trust contract at least once, only 33 percent did so in more than three periods. This indicates that most principals experimented somewhat with the trust contract until settling for the incentive contract.

Figure 1: Share of incentive and trust contract (TI treatment)

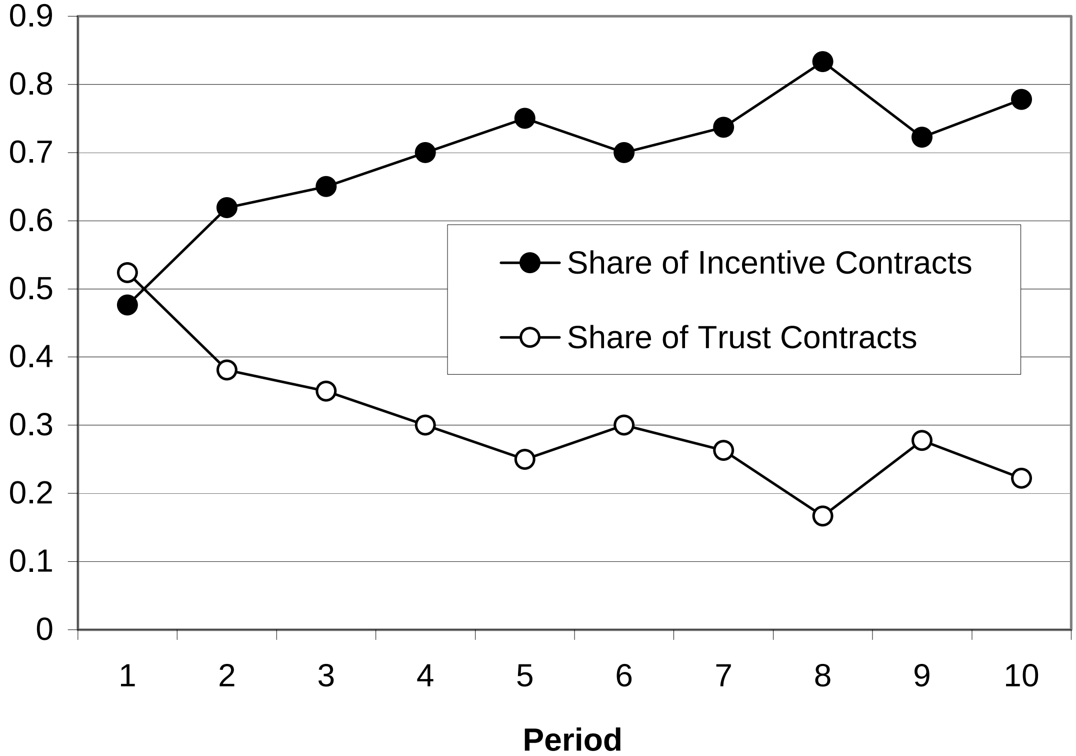


Figure 2: Average effort and average demanded effort in the Trust-Incentive treatment

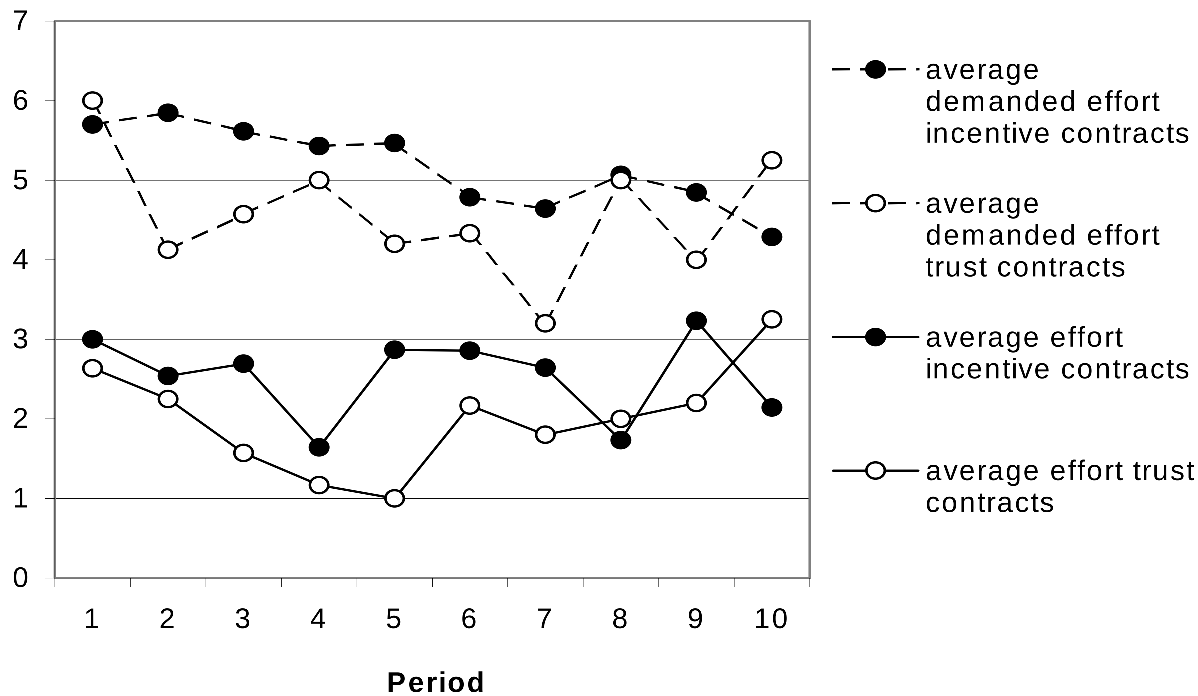


Figure 2, depicting the evolution of average effort levels (and average demanded effort levels) over time for both contract types, illustrates Result 1b. The figure shows that the average effort is higher in almost all periods in the incentive contracts. Moreover, the fraction of trust contracts is already small in those periods in which average effort is somewhat higher in the trust contract, meaning that this is driven by very few observations. The effort difference between ICs and TCs is significant (p = 0.028, Mann Whitney test). This difference in effort levels is also associated with differences in the principals' payoffs. On average, the principals earned a payoff of –0.87 when they proposed an incentive contract and –2.4 when they proposed a trust contract. These payoff differences are, however, not statistically significant ($p > 0.59$, Mann Whitney test).

Viewed from the perspective of the self-interest model, the rather low profits resulting from the incentive contracts are surprising because – recall from Section 3 – the predicted profit is $M^P = 26$. Moreover, it is also surprising that there is such a strong trend towards the incentive contracts in view of the small payoff differences between the incentive and the trust contracts. Why did the principals have such a strong preference for incentive contracts if these contracts performed so poorly? The next result shows that the distinction between incentive compatible and non-incentive compatible ICs is crucial in this context.

Result 2(a): *Although most incentive contracts stipulate the maximal fine, the majority of incentive contracts violate the no-shirking condition because the principals demand too high effort levels. In the majority of the cases, non-incentive compatible ICs induce the agents to shirk fully, implying negative payoffs for the principals.*
(b) *Incentive compatible ICs are, however, associated with significantly positive payoffs for the principals because the agents shirk much less in these contracts. The large payoff difference between incentive compatible and non-incentive compatible ICs is associated with a strong increase in the share of incentive compatible ICs over time.*

Our data supports R2a as follows: the average fine is 12.3, closely approximating the maximal fine of 13. However, the no-shirking condition, $pf \geq c(e^*)$, is violated in 79 (58.5 percent) of the 135 incentive contracts, i.e., principals demanded too high effort levels. Figure 2 also illustrates this fact, showing that the average demanded effort level in the incentive contracts persistently exceeds the maximal enforceable effort of $e^* = 4$. We present the agents' effort behavior and the principals' payoffs for incentive compatible ICs, non-incentive compatible ICs and TCs in Table 2, which shows that non-incentive compatible ICs are associated with a high rate of shirking and rather low payoffs for the principals. The last row of Table 2 indicates that there are 79 non-incentive compatible contracts, that only 1 of these contracts is rejected, and that the agents shirked fully by choosing the minimal effort level of $e = 1$ in 48 (62 percent) of the accepted contracts. This high rate of shirking has the consequence that the non-incentive compatible ICs cause on average a loss of –7.6 for the principals (see shaded area in last row of Table 2).

Figure 2 not only indicates that the average demanded effort in the ICs is too high relative to the enforceable effort level but also shows that the demanded effort level declines over time. The average demanded effort level in period 1 is close to $e^* = 6$ while it is only slightly above the incentive compatible level of $e^* = 4$ in the final period. This suggests that the share of incentive compatible ICs increases over time. In fact, only 10 percent of all ICs are incentive compatible in period one while this amount already exceeds 64 percent of all ICs in period ten. The profit differences between incentive compatible and non-incentive compatible ICs provide a natural explanation for this strong time trend. The shaded areas in the last row of Table 2 show that the average profit in the incentive compatible ICs is 8.6, which is much larger than the loss of –7.6 in the non-incentive compatible contracts. Thus, while incentive compatible ICs are considerably more profitable than trust contracts, the non-incentive compatible ICs are less profitable than the trust contracts.[6] The strong profit differences between the incentive compatible ICs and the TCs also explains why the share of trust contracts strongly declines over time.

There are two reasons why the incentive compatible ICs are more profitable than the non-incentive compatible ICs. First, the principals pay far lower wages when they offer incentive compatible contracts. Second, although the principals pay less when they offer incentive compatible ICs, shirking is much less frequent in these contracts. Table 2 shows that the wage is above $w = 10$ in all 79 offered ICs that are *not* incentive compatible while in the majority of the incentive compatible ICs (in 29 of 56 cases) the wage is strictly below $w = 10$. This suggests that the principals attempted to elicit reciprocal effort choices from the agents when they proposed non-incentive compatible contracts. Recall, however, that these attempts frequently failed. This contrasts

[6] These differences are statistically significant according to a Mann-Whitney test ($p = 0.005$ when the TC is compared to the incentive compatible IC; $p = 0.036$ when the TC is compared to non-incentive compatible IC).

sharply with those contracts that meet the no-shirking condition (see last row of Table 2). The agents shirk in only 12 (26 percent) of the 47 accepted incentive contracts.[7]

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				
	No. of Of-fers	reje ct	e<e *	e≥e *	P's payoff	No. of Of-fers	reje ct	e = 1	e > 1	P's payoff	No. of Of-fers	reje ct	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: n. a. means that no entries are available for the respective cells. The sum of the column indicating the number of contracts in which the agent shirked (e < e*) and in which the agent met the demanded effort (e≥e*) for the incentive compatible contracts yields the total number of accepted contract offers. Likewise, the sum of the column indicating how often the agents chose e = 1 and how often they chose e ≥1 for the non-incentive compatible ICs and the trust contracts gives us the total number of accepted contracts.

Table 2 also indicates that when trust contracts were offered, the principals paid relatively high wages – in 30 of the 60 trust contracts the principals offered a wage above 20. Thus, the strong decrease in trust contracts and non-incentive compatible ICs over time caused a decreasing trend in wages over time. The principals offered average wages well above 20 during the first few periods, when the share of incentive compatible ICs was still low. The average wage decreased, however, strongly over time and reached a level of 11.9 in period ten. The strong time trend in the share of

[7] Note that the number of accepted contracts is given by the sum of the two effort columns. For example, for the incentive compatible contracts this sum is given by the 12 contracts with $e < e^*$ plus the 35 contracts with $e \geq e^*$. In all

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 incentive compatible ICs.

There is a further noteworthy feature in Table 2. For the trust contracts the principals' payoff is decreasing in the offered wage. The principals earn 3.7 for wages below $w = 10$, the payoff declines to $M^P = -1.0$ for wages in the middle interval ($10 \leq w \leq 20$), and further diminishes to $M^P = -6.4$ for high wages ($w > 20$).[8] A similar relation holds for the non-incentive compatible ICs, where earnings amount to $M^P = -1.4$ in the middle interval while corresponding to $M^P = -12.0$ for high wages. We summarize this payoff pattern in

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

Results 1 – 3 in the TI treatment are interesting because many observed qualitative data patterns are consistent with the predictions of the self-interest model. This model predicts that the principals offer incentive compatible ICs by imposing the maximum fine and demanding the maximum enforceable effort level. This contract preference is based on the prediction that both the trust contract and non-incentive compatible ICs will induce shirking of the agents, implying that high wages will be associated with losses. In fact, the principals' contract choices do converge towards the prediction of the self-interest model and – as Result 3 shows – the payment of high wages is indeed not profitable. However, there are also some aspects of the data that violate the predictions of the self-interest model. If the principals offer incentive compatible ICs, they earn substantially less than predicted. Recall that – according to the prediction – they should earn $M^P = 26$; in fact they earn only 8.6. This misprediction has three reasons: (i) although the principals reap a larger share of the surplus, they rarely extract the whole rent from the agents, i.e., they still pay substantial wages. The principals offer on average 35% of the surplus that occurred if the worker met $e = e^*$ in the incentive contracts. Even if they propose incentive compatible ICs, the offered surplus is on average 31 %. (ii) The agents reject wages below $w = 10$ in 25% of the cases. (iii) The agents shirk by choosing the minimal effort in roughly 1/3 of the incentive compatible contracts. All three reasons suggest a role

---

12 cases with $e < e^*$ the agents chose $e = 1$.

[8] A simple OLS-regression of effort on wages also confirms this result, yielding $e = 1.08 + 0.04\, w + \varepsilon$ where ε denotes the error term. The t-value for the constant is 3.11 and 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 marginal loss: A wage increase

for fairness concerns. Likewise, the fact that the principals initially expressed a strong preference for trust contracts or non-incentive compatible ICs suggests that they attempted to elicit generous effort choices from the workers. However, the prevailing fairness motives were apparently not strong enough to render these contracts more efficient than the incentive compatible ICs. It remains to be seen whether fairness concerns can overturn the contract predictions of the self-interest model if the principals can announce a bonus payment.

4.2. The Bonus-Incentive (BI) Treatment

The only difference between the TI and the BI treatment is that principals have an additional contract option in the BI treatment: they can offer a nonbinding bonus to the agent in addition to the IC and TC. Note that the trust contract is a special type of bonus contract because the principal can always forgo the opportunity of announcing a bonus and offer a fixed wage to the agent. Thus, if a principal wants to propose a trust contract she just sets $b^* = 0$.

We observed a total of 230 contract offers in the BI treatment. Four bonus contracts and two incentive contracts were rejected, leaving 224 accepted contracts. While the incentive contract outperformed the trust contract in the TI treatment, the incentive contract performed very poorly when the principals could choose a bonus contract.

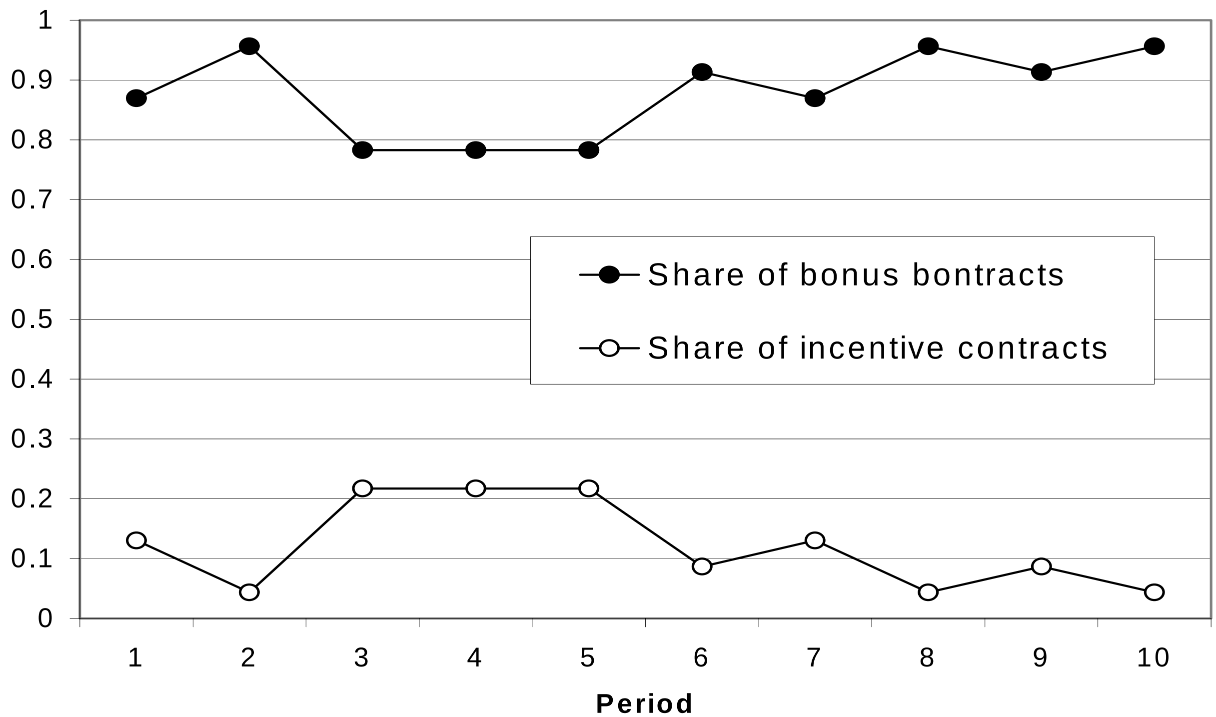
Result 4: (a) *the overwhelming majority of all contracts in the BI treatment are bonus contracts, while the incentive contract is rarely chosen and the trust contract is never chosen.*
(b) *The average effort and the average payoff of the principals are much higher in the bonus contract as compared to the incentive contract.*

Figure 3 presents the evidence for R4a. Trust contracts do not appear in this figure because they were never chosen. The figure shows the evolution of the share of bonus and incentive contracts over time. 87 percent of all contracts are already bonus contracts in period one. 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. However, the share of bonus contracts is roughly 90 percent from period six onwards and even approaches 96 percent in the final period. There can thus be little doubt that principals strongly prefer the bonus contract.

---

by 10 units raises effort only by 0.4 and, hence, the expected revenue increases only by 4 units.

Figure 3: Share of bonus and incentive contracts (BI treatment)

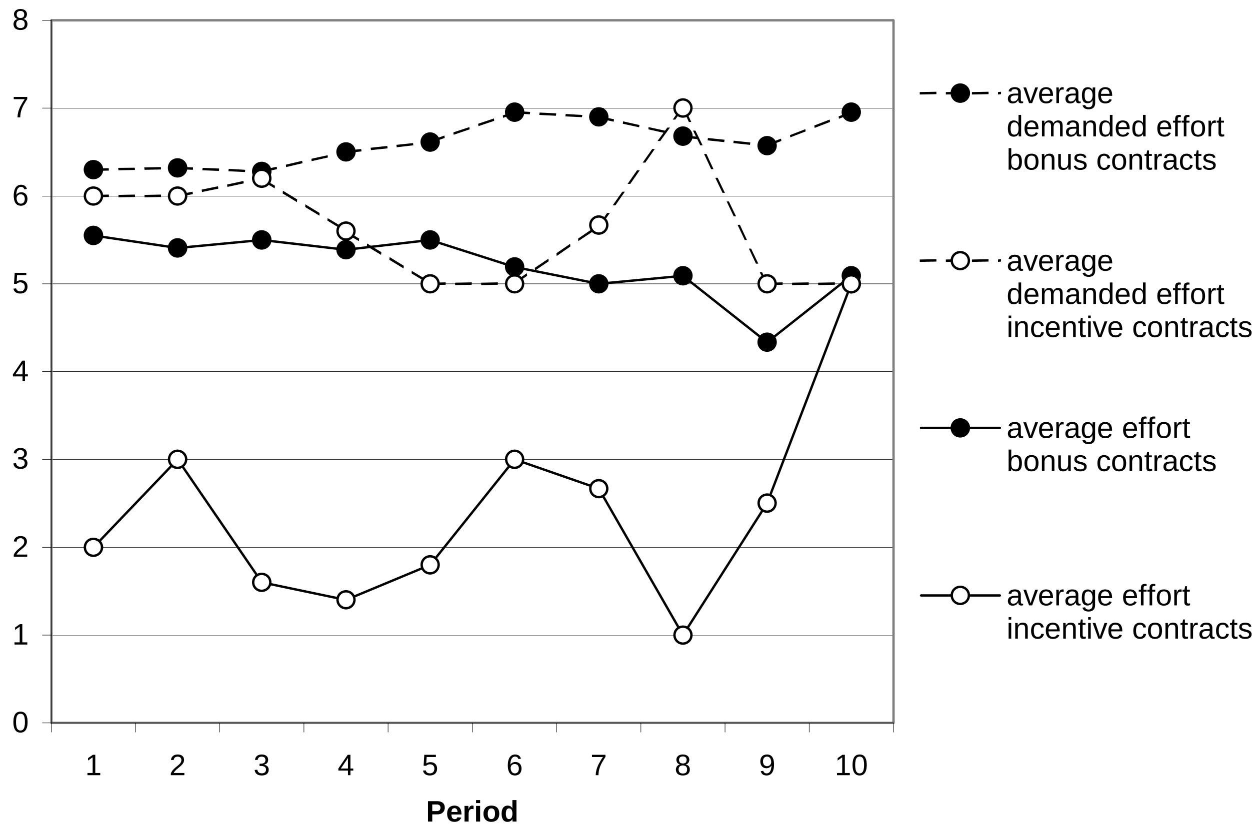


To examine the reasons for this preference, we compare the average effort level in bonus and incentive contracts (see Figure 4). The figure shows that the average effort is considerably higher in the bonus contracts in nine out of ten periods.[9] While the average effort in the incentive contracts is generally between $e = 2$ and $e = 3$, effort in the bonus contracts is, in general, above $e =5$. This difference across contract types is highly significant ($p < 0.001$, Mann Whitney Test). Figure 4 also indicates that agents' efforts in the bonus contracts are somewhat below the demanded effort level but the gap between actual and demanded 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 demanded average effort always exceeds $e^* = 4$. 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. The average profit from bonus contracts is *always* above 20 tokens in each of the ten periods while the incentive contract causes losses in six of the ten periods.

[9] The exception is period ten, where the effort difference is negligible. However, there was only one incentive contract in period ten, so that this data point has little relevance for the overall comparison.

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



There is an interesting difference in the performance of the incentive contracts across the TI and the BI treatment. The principals who offer incentive contracts in the BI treatment demand non-incentive compatible effort levels in all periods while they learned to make incentive compatible contracts over time in the TI treatment. Thus, it seems that the strong superiority of the bonus contract inhibited a learning process in favor of *incentive compatible* ICs. Note, however, that the average profit in the bonus contracts is more than three times higher than that in the incentive compatible ICs in the TI treatment. This suggests that – in view of the superiority of the bonus contracts – it simply did not pay to learn to make incentive compatible ICs in the BI treatment.

The higher effort level in the bonus contracts implies 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 principals reaped the bulk of the increase. This shows that the option to pay a bonus yields a substantial efficiency increase *and* causes sizable changes in the distribution of the surplus.

Why does the bonus contract prove to be vastly superior to the incentive contract? Our next result shows that the key for understanding this result lies in the principals' bonus payments.

Result 5: *The principals devote a substantial part of the agents' compensation to bonus payments. Moreover, the average bonus increases strongly with respect to the effort level so that non-minimal effort choices are profitable for the agents.*

Figure 5a and 5b support R5. Figure 5a shows the average wage offered in both the incentive and the bonus contracts; in addition the figure presents the average bonus payments in the BCs. The average wage in the BCs remains in the vicinity of w = 15 throughout the whole experiment and the bonus payment amounts to b = 10.4. Thus, the principals pay roughly 40% of the agents' compensation in the form of a bonus. However, this bonus payment strongly depends on the agent's effort (see Figure 5b). If the agent provides low effort at e = 1 or e = 2 the average bonus is zero, but the bonus approaches b = 30 for high effort levels. The positive slope of the bonus-effort schedule is also confirmed by the following regressions that related the bonus payment to the agent's effort e, the demanded effort e^*, the base wage w, and the announced bonus b* (see Table 3).

Figure 5a: Average wages and average bonus over time in the BI treatment

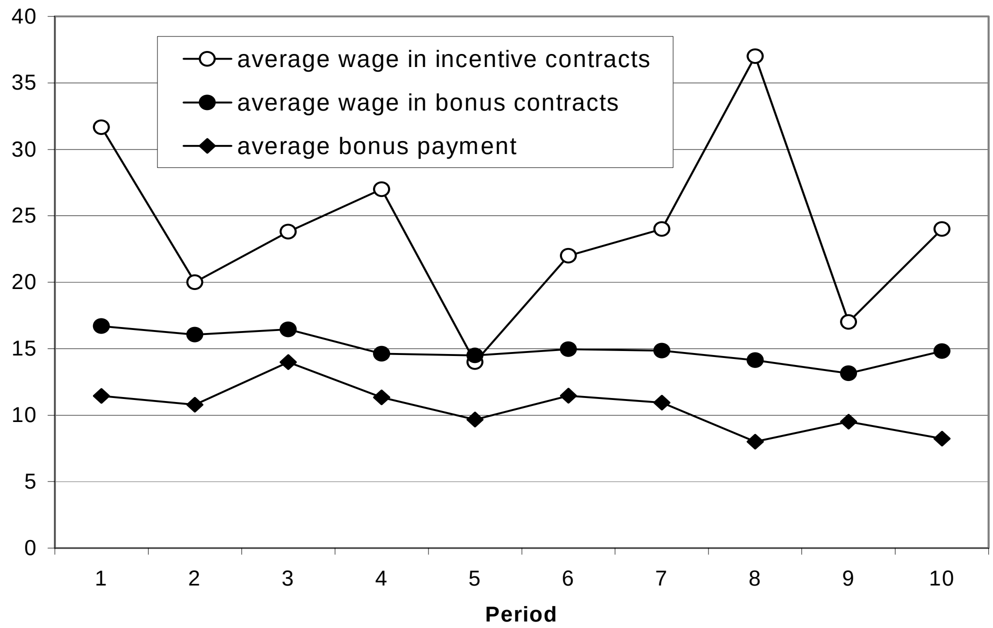


Figure 5b: Average bonus as a function of effort in the Bonus-Incentive treatment

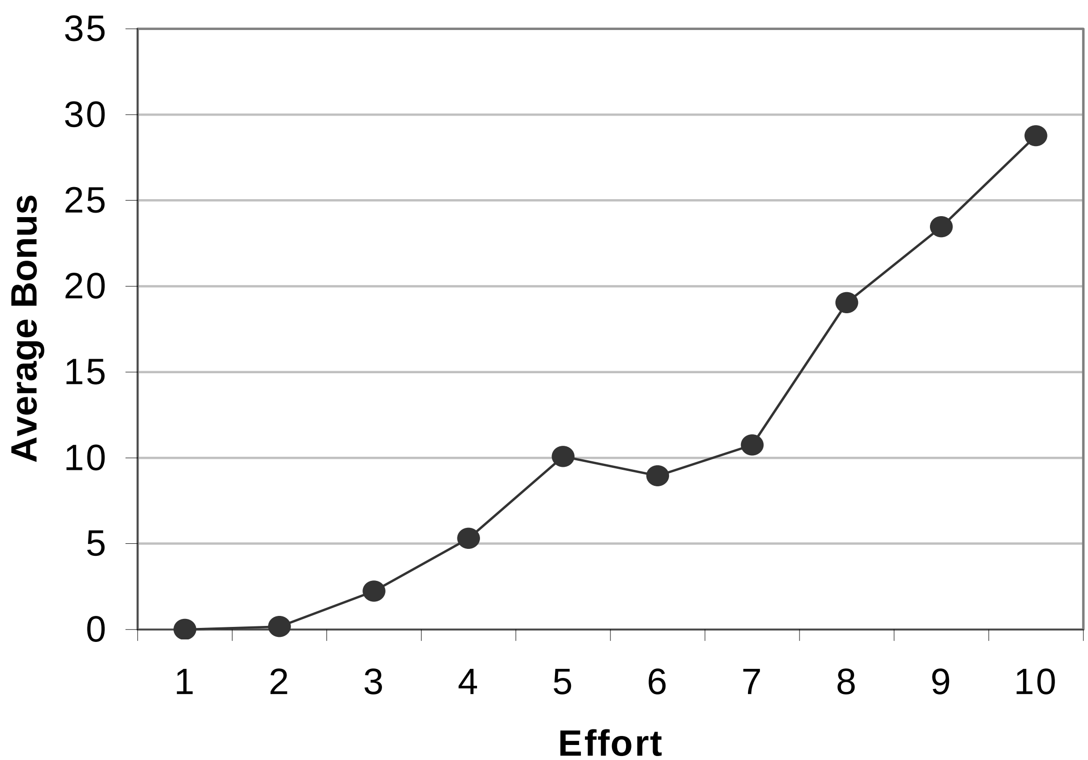


Table 3: Determinants of bonus payments in the Bonus-Incentive Treatment

Dependent variable: Bonus payments	(1) (robust standard errors)	(2) (robust standard errors & clusters)
Constant	-5.58*** (1.94)	-5.58** (2.59)
Effort	2.86*** (0.17)	2.86*** (0.33)
Demanded effort	0.33 (0.27)	0.33 (0.46)
wage	-0.30*** (0.09)	-0.30* (0.16)
Announced bonus	0.12* (0.06)	0.12 (0.08)
No. of observations	198	198
Adjusted R^2	0.57	0.57

Table reports the coefficients of OLS regressions. Robust standard errors are in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Table 3 reports the results of two regressions with the associated robust standard errors. The first is a simple OLS regression. In the second regression we treated the observations of individual principals as separate clusters because they may not be independent of each other. Thus, in the second regression the standard errors are based on the assumption that the bonus payments are independent across different principals but we allow for dependent observations within each cluster. The assumption that the bonus payments are independent across principals is reasonable because a principal could never observe what the other principals did. Moreover, since the bonus payment is the final choice in each period, the principals can respond to all previous actions that occurred in the match of that period.[10]

Both regressions in Table 3 show that an increase in the effort level by one unit increases the expected bonus payment significantly 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 demanded effort level is small and not significant, indicating 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 excessively committed to their bonus announcements and that the effort level is the major determinant of the principals' bonus choice.

Although the principals respond, on average, quite strongly to increases in the effort level, it is important to notice that there are big differences in individual behavior. In those 162 contracts where the agents chose a non-minimal effort level ($e > 1$), the principals did not pay any bonus at all in 34 cases (21%). Among those principals who did pay a bonus, many paid very little even if the agent selected a high effort level. However, there were also many principals who reciprocated high effort levels very generously.

Taken together, the TI and the BI treatments show that the principals strongly prefer the bonus contract. If this contract is not available, they prefer the incentive over the trust contract. The same ranking holds in terms of the average effort and the average surplus associated with the three types of contracts. These facts are puzzling from the viewpoint of the self-interest model. Recall that this model captures important qualitative aspects of the TI treatment quite accurately. Although there are several hints in the data suggesting that fairness concerns play a role in the TI treatment, these concerns are apparently too weak to overturn the basic prediction that the principals prefer the incentive contract relative to the trust contract. However, the mere addition of the possibility of announcing and paying a nonbinding bonus – which represents a completely innocuous change from the viewpoint of the self-interest model – suddenly transforms fairness concerns into a powerful

[10] To check the robustness of our results we also conducted Tobit regressions. All variables that are significant in the OLS regressions are also significant in the Tobit regressions and all variables that are insignificant in the OLS regressions remain insignificant in the Tobit regressions.

determinant of principals' contract choices: many principals reward generous effort levels with generous bonus payments and thus create powerful incentives for effort provision. We will provide a unified explanation for this puzzle with a fairness model that is based on the assumption of heterogeneous fairness preferences in Section 5 below.

Before doing so, we want to emphasize a further important aspect. A general prediction of traditional contract theory is that contracts should depend on all verifiable signals that contain statistical information about the agent's action or type (see Holmström 1982, Laffont and Tirole 1993). However, actual contracts frequently specify important obligations of the contracting parties in vague terms, and they do not explicitly tie the parties' monetary payoffs to measures of performance that would be available at a relatively small cost. Thus it seems that many contracts are left deliberately incomplete. In our BI treatment the principals expressed a strong preference for a fairly incomplete bonus contract in which bonus payments were only vaguely and implicitly tied to the agents' performance. This indicates that principals did not want to use the explicit incentive contract if they had to give up a vaguely defined, but nonetheless powerful, implicit incentive. In reality, however, implicit and explicit incentives can be combined. Thus, the question arises whether the principals would also forgo the opportunity to implement an explicit incentive if they could combine it with an implicit one. If we can show that the principals prefer a purely implicit incentive over a combination of explicit and implicit incentives, we indeed have direct evidence that they voluntarily do not utilize the explicit incentive. Moreover, since fairness concerns are a decisive force behind the bonus contracts in our context, we also have evidence that these concerns are important for the absence of explicit incentives.

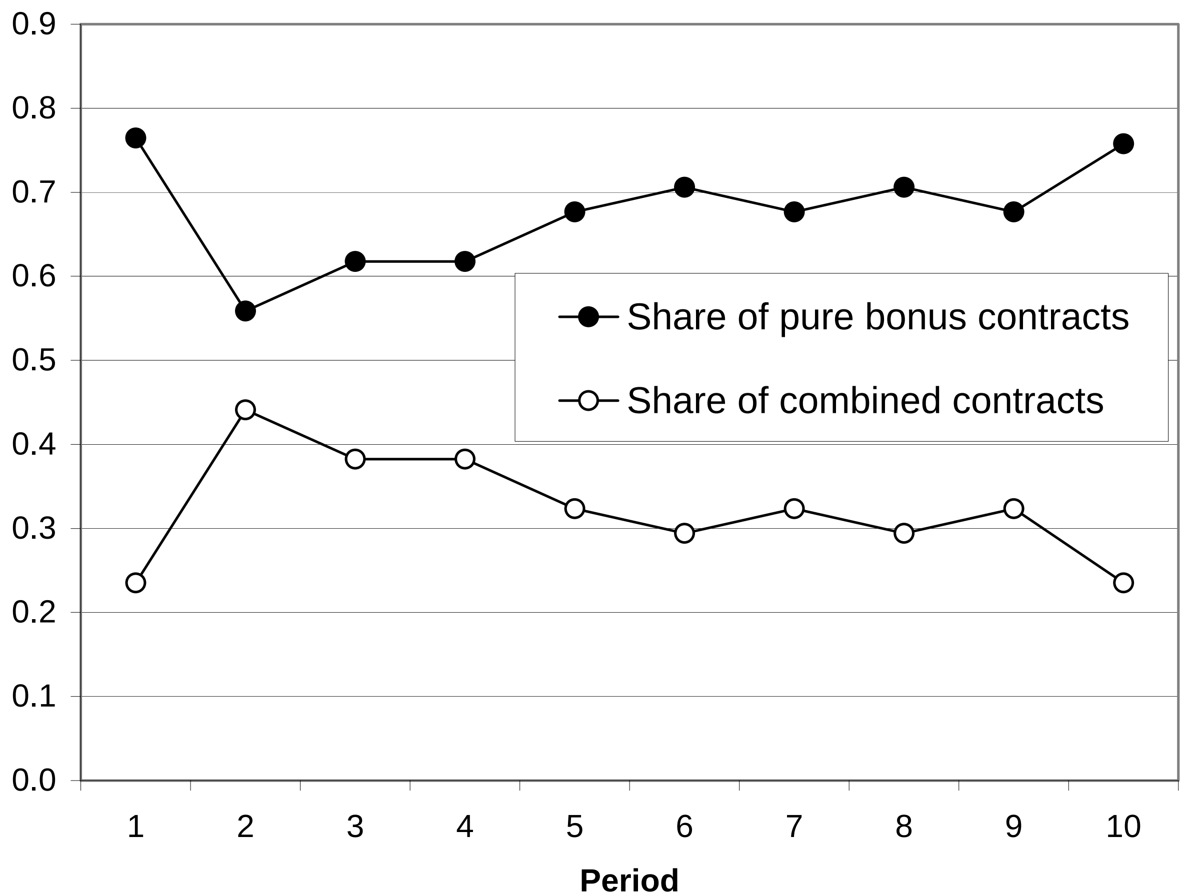
4.3. Do principals voluntarily forgo the use of explicit incentives?

To answer this question, we implemented an extended Bonus-Incentive treatment (henceforth called EBI treatment) in which the principals could also propose a combined contract. The combined contract (w, e^*, b^*, f) announces a bonus b* and imposes a fine $f \leq \overline{f}$ with probability 1/3 on the agent if $e < e^*$. The potential advantage of the combined contract is that it uses both the enforcement power of the explicit incentive and the implicit bonus incentive. Thus, the question is whether the combined contract elicits higher effort levels than a pure bonus contract and is, hence, more efficient. To be more efficient, the enforcement advantage of the combined contract has to be large enough to outweigh the verification costs of $k = 10$ that are associated with the implementation of the explicit incentive.

All the other contracts were still available in the EBI treatment, i.e., the principals could also offer a (pure) trust contract (i.e., $f = b^* = 0$), a (pure) bonus contract or a (pure) incentive contract (i.e., $f > 0$, $b^* = 0$). We conducted three sessions (S5 – S7) in the EBI treatment. In total we observed 339 contract offers in this treatment. 5 pure bonus contracts and 8 combined contracts were rejected, leaving 326 accepted contracts. The following result informs us about the principals' contract preferences in this setting.

Result 6: *We observe neither trust contracts nor incentive contracts in the EBI treatment. Roughly 2/3 of all contract offers are pure bonus contracts and 1/3 are combined contracts. In the final periods the share of pure bonus contracts even surpasses 70%.*

Figure 6: Share of pure bonus and combined contracts in the Extended Bonus-Incentive treatment



Support for R6 is provided by Figure 6 and the following numbers. 229 of the 339 contracts (67.6%) are pure bonus contracts, 110 contracts (32.4%) are combined contracts. Figure 6 also indicates that the principals' prefer the pure bonus contract in each period. From period 3 onwards, there is a slowly increasing trend in favor of the pure bonus contract which peaks in period 10 at a share of roughly 75 percent. Thus, the clear majority of principals voluntarily forgo the opportunity to implement an explicit incentive.

The strong preference for the pure bonus contract is puzzling because the enforcement power of implicit *and* explicit incentives can be used in the combined contract. Our next result shows, however, that this potential advantage of the combined contract does not generate significantly higher effort levels:

Result 7: *Effort is not significantly higher in the combined contracts than in the pure bonus contracts. Hence, pure bonus contracts are more efficient. The agents largely reap these efficiency gains.*

R7 is supported by Figure 7a and 7b and the following facts. The average effort is 5.8 in combined contracts and 5.3 in pure bonus contracts. This small difference is not significant ($p > 0.10$, Mann Whitney test) and does not suffice to outweigh the implementation costs for the explicit incentive.[11] This fact is also indicated by Figure 7a and 7b. In Figure 7b we plotted the principals' profit from both types of contracts. The figure shows that the principals earned more in most periods when they offered pure bonus contracts. In addition, their payoff from combined contracts exhibits considerably higher volatility, suggesting that the variance of profits was higher when combined contracts were offered. In fact, the standard deviation of the principals' payoff in the combined contract is 26.75 while in the pure bonus contract it is only 23.68. If we average over all periods, profits are 24.7 in the pure bonus contracts and 24.0 in the combined contracts. This difference is not significant ($p > 0.99$ Mann Whitney test). Figure 7b shows, however, that the agents earned considerably higher incomes in pure bonus contracts. Except for the final period, the agents' average income was roughly 20 units in the pure bonus contracts while they earned 15 units or less most of the time in the combined contracts. If we average over all periods, the agents earn 19.2 in the pure bonus contracts and 12.5 in the combined contracts. This difference is significant ($p < 0.001$) according to a Mann Whitney test.

[11] The average demanded effort level is virtually identical across contracts – 7.30 for the pure bonus contracts and 7.22 for the combined contracts.

Figure 7a: The principals' profits over time in the Extended Bonus-Incentive treatment

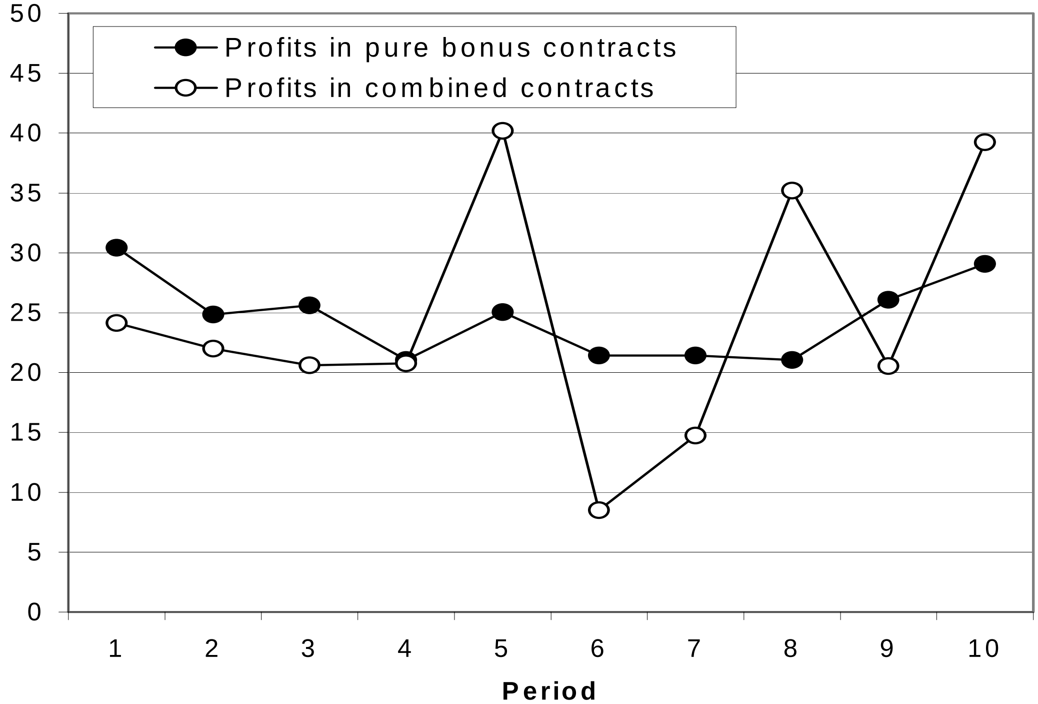
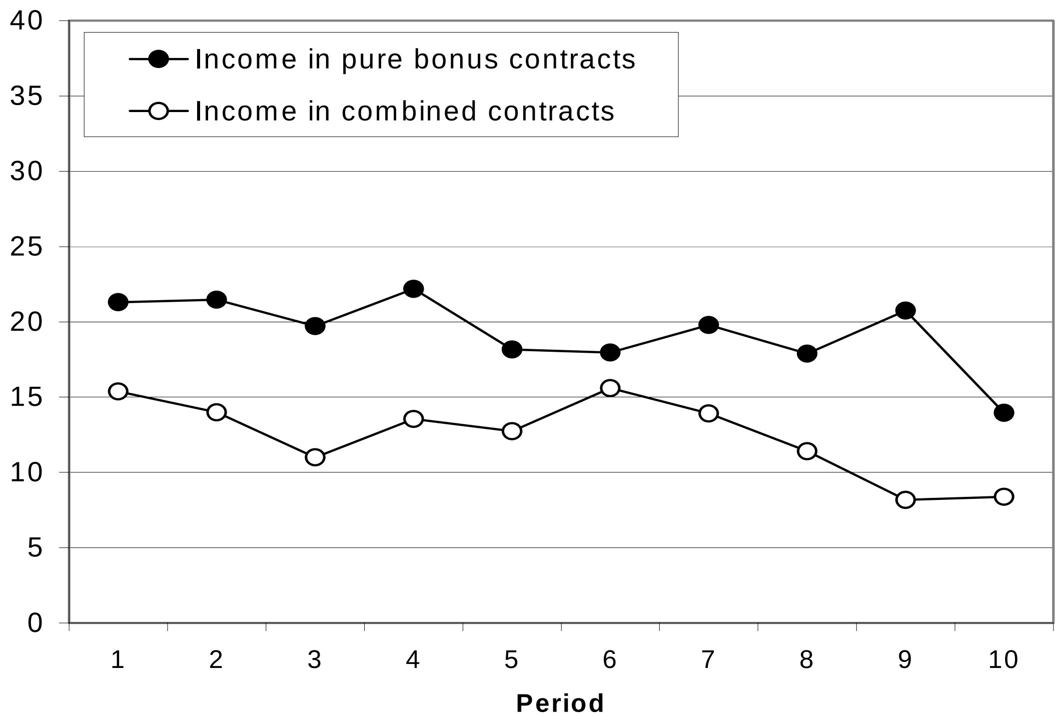


Figure 7b: The agents' income over time in the Extended Bonus-Incentive treatment



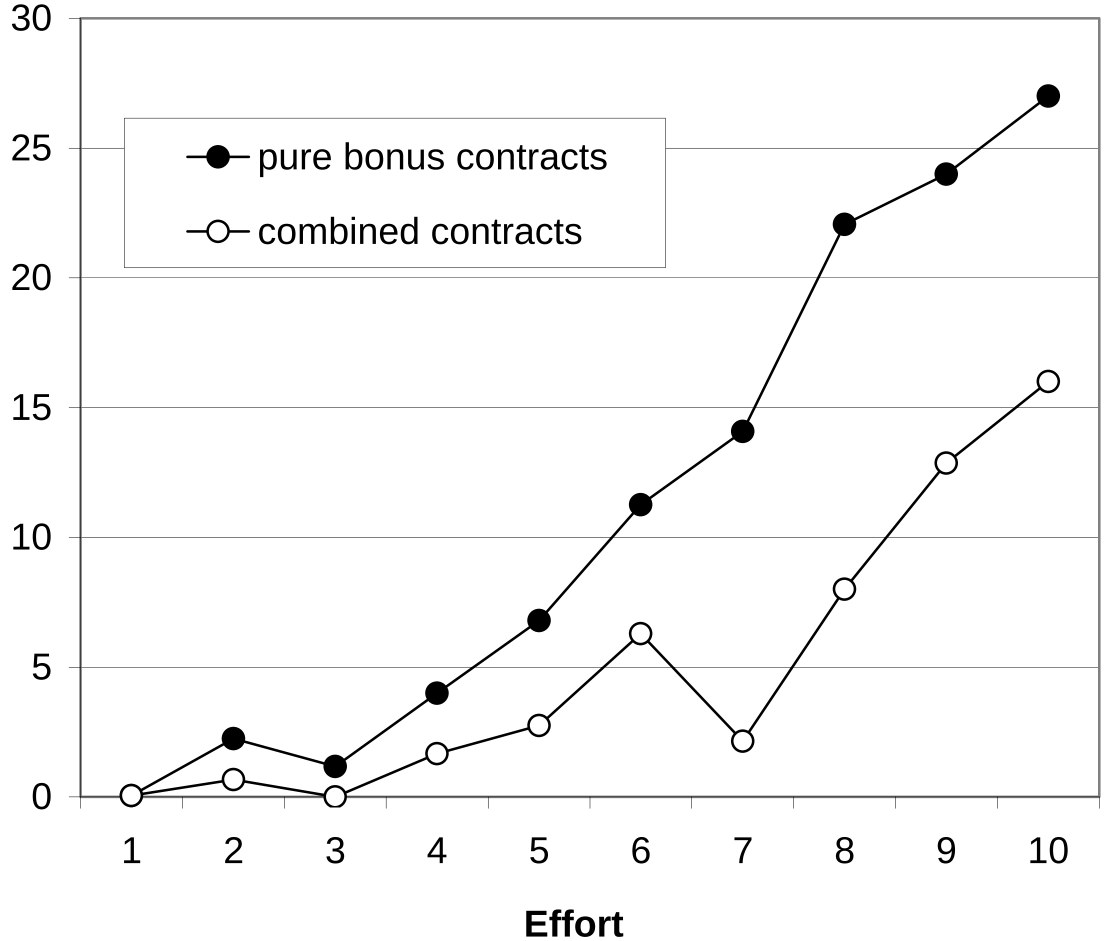
Due to the considerably higher earnings of the agents, total earnings are higher when pure bonus contracts are offered. Under the pure bonus contracts total earnings are 43.9 while they are only 36.5 under the combined contracts. This difference is again significant (p = 0.0012, Mann Whitney test).

These results suggest that the principals prefer pure bonus contracts because they generate slightly higher average profits and are less risky than combined contracts. Combined contracts require an initial investment of k = 10 which does not pay off because effort increases only insignificantly if workers face the threat of being fined. Thus, the key behind the preference for the bonus contract is that the combined contract does not raise effort sufficiently. Our next result indicates a plausible reason for this fact.

Result 8: *In the combined contracts principals reward high effort levels less generously than in pure bonus contracts. Thus, combined contracts provide lower implicit incentives.*

Support for R8 is provided by Figure 8 below and the associated regressions. Figure 8 unambiguously indicates that the principals paid a higher average bonus in the pure bonus contracts for any non-minimal effort level. Moreover, the difference is quite large for effort levels above e = 6 – on average the principals paid more than 10 tokens less in the combined contracts. This difference in the bonus payments is also a major source of the difference in the agents' income across contracts because the average bonus in the pure bonus contract is 10.92 while in the combined contract it is only 6.16. Note that this difference in the average bonus payments cannot be due to different effort levels because the actual distribution of effort is roughly identical across contract types.

Figure 8: Average bonus payments conditional on effort in the Extended Bonus-Incentive treatment



We also conducted multivariate regressions to examine whether the relationship between effort and bonus payments still holds if we control for other variables. In addition to effort we used the demanded effort e^*, the wage w, the announced bonus b* and interactions between these variables and a dummy indicating the pure bonus contract as regressors. The dummy is denoted by D and takes on a value of 1 if the observation is from a pure bonus contract. The results of the regression analyses are presented in Table 4.

Table 4: Determinants of bonus payments in the Extended Bonus-Incentive treatment

Dependent variable: Bonus payments	(1) pure bonus contracts	(2) combined contracts	(3) all contracts	(4) all contracts	(5) all contracts (clusters)
Constant	-3.21 (2.07)	- 6.01 (4.14)	-3.76** (1.64)	-3.83** (1.63)	-3.83* (2.03)
Effort	3.03*** (0.160)	1.61*** (0.24)	1.79*** (0.18)	1.58*** (0.22)	1.58*** (0.34)
Demanded effort	0.19 (0.36)	0.97 (0.66)	0.42 (0.28)	0.75* (0.44)	0.75 (0.56)
wage	-0.18*** (0.08)	0.29** (0.14)	-0.21*** (0.06)	-0.29** (0.13)	-0.29* (0.15)
Announced bonus	-0.01 (0.04)	0.04 (0.09)	0.01 (0.03)	0.03 (0.08)	0.03 (0.10)
D × effort			1.13*** (0.18)	1.46*** (0.26)	1.46*** (0.40)
D × demanded effort				-0.50 (0.48)	-0.50 (0.49)
D × wage				0.11 (0.15)	0.11 (0.17)
D × announced bonus				-0.03 (0.08)	-0.03 (0.09)
No. of observations	224	102	326	326	326
Adjusted R^2	0.64	0.32	0.593	0.596	0.596

Table reports the coefficients from OLS regressions. Robust standard errors are in parentheses. D denotes a dummy variable for the pure bonus contracts. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

The first important result is that the effort coefficient is sizeable and highly significant in all regressions. However, a comparison between regression (1) and (2) shows that the effort coefficient is much larger under a pure bonus contract. An increase in the effort level by one unit in a pure bonus contract increases the bonus by roughly 3 units which is consistent with the results we observed in the BI treatment. The bonus increase is only 1.6 units on average in the combined contracts. Regression (3) shows that this difference is highly significant as indicated by the coefficient on D × effort. This result remains robust when we introduce further interaction variables in regression (4). All the other variables show no significant interaction with the pure bonus dummy. Another noteworthy effect in Table 4 is that a higher base wage always reduces the bonus payments significantly. In regression (5) of Table 4 we checked the robustness of our results with regard to the clustering of individual principals. As already explained in the context of Table 3 the regression with clusters controls for potential dependencies in the bonus payments of individual principals. Regression (5) shows that this additional control does not affect the significance of the effort and the D × effort coefficient.[12]

This evidence leaves little doubt that principals who choose combined contracts reward high effort less generously. It seems that the use of the explicit incentive crowds out implicit incentives to some degree and thus weakens the enforcement power of the combined contracts. To examine how the crowding out of implicit incentives affects the agents' incentives, the fact that the principals chose the maximum fine in almost all combined contracts but also demanded effort levels that were not incentive compatible must also be taken into account. The average fine is given by 11.99 and, among the 110 combined contracts that were offered, only 6 contracts stipulated an incentive compatible effort level of $e^* = 4$. Thus, the combined contracts clearly also relied on the enforcement power of the implicit bonus incentive. This interpretation receives further support if we compare the actual average effort that was enforced in the incentive contracts in the TI or the BI treatment with the effort actually enforced in the combined contracts of the EBI treatment. Recall that the principals in the TI and BI treatment could enforce on average $e = 2.51$ and $e = 2.1$, respectively, with the explicit incentive alone, while they enforced $e = 5.8$ in the combined contracts. Therefore, a large part of the enforcement power of combined contracts must be due to the implicit incentive. However, the fact that the combined contracts rely on the implicit incentive also suggests that variations in the steepness of the bonus-effort relationship are associated with variations in the enforcement power of

[12] We also conducted Tobit regressions. Relative to the OLS regressions the same variables remain, and no other

these contracts. Therefore, the less generous reward policy of the principals in the combined contracts provides a natural explanation for the fact that the combined contracts did not elicit significantly higher effort levels than the pure bonus contracts.

5. A Unified Interpretation

The major puzzle posed by the evidence in Section 4 is that the incentive contract outperforms the trust contract while the bonus contract outperforms the incentive contract. Why are fairness concerns too weak to render the trust contract an effective tool but strong enough to render the bonus contract very effective? In addition, we also have to explain the remarkable result that principals predominantly choose the pure bonus contract even if they could combine the enforcement power of explicit and implicit incentives.

In view of the importance of fairness concerns in our experiments, it is natural to seek an explanation of these puzzles in the context of recently developed fairness models. We will show that the fairness approach is indeed capable of explaining the pattern of contract choices observed in Section 4; sometimes the approach even provides surprisingly precise quantitative characterizations of the observed behavior. In the following, we apply the theory of inequity aversion of Fehr and Schmidt (1999) to our experiments because it captures important aspects of fairness driven behavior in a tractable way and is consistent with the outcomes of many different classes of experimental games.[13] The choice of this theory does not mean that other theories of social preferences (e.g. Bolton and Ockenfels 2000, Levine 1998, Falk and Fischbacher 1999, Charness and Rabin 2002, Cox and Friedman 2002, Dufwenberg and Kirchsteiger, in press) may not also be able to rationalize the data.[14] A main reason for applying the Fehr-Schmidt model is tractability. In addition, we want to emphasize that it is not our aim to explain the dynamic pattern of the data over time. Instead, we

---

variables become, significant in the Tobit regressions.

[13] Fehr and Schmidt apply the theory to ultimatum, market, public good, and gift exchange games, for example, and show that the interaction of the distribution of types and the strategic environment can explain the major facts in these games. For the limits of this theory we refer to Charness and Rabin (2002) and Fehr and Schmidt (2003).

[14] For example, it is easy to see that the theories by Bolton and Ockenfels (2000) and Charness and Rabin (2002) make roughly the same predictions in our context as the Fehr-Schmidt model. All three theories have equality as the standard of fairness in two-player interactions. Like Fehr and Schmidt, Bolton and Ockenfels rely on inequality aversion. With just two players the two models yield qualitatively very similar results. Charness and Rabin have an additional term in the utility function assuming that people care about total surplus maximization. However, in our context the bonus payment of the principal is a simple transfer payment that leaves total surplus unaffected. Therefore, the surplus maximization motive in Charness and Rabin (2002) cannot explain the power of the bonus contract. Instead, it is the fairness motive in their theory that yields similar predictions as the Fehr-Schmidt model.

focus on the robust behavioral regularities that emerge in all treatment conditions in the final few periods.

The theory of Fehr and Schmidt (1999) has two main tenets: First, it assumes that some people are not only concerned about their own material payoff but also care about inequity or, in our context, inequality.[15] 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. For sufficiently high values of α_i and β_i players with this utility function want to achieve equality. If the inequality is to their disadvantage, they are prepared to incur costs 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.

We use a 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 "inequity-averse" types. In addition, we assume $\alpha_i, \beta_i > 0.5$, i.e., the inequity averse subjects are willing to share the surplus of a contract equally and reject offers that give them less than 25 percent of the surplus.[16] The restriction to two types for our games is quite natural because what matters is whether a player wants to equalize payoffs. The assumption that 40 percent of all players are 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

---

[15] 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, it seems natural to define *equality* in these very simple environments as the reference point for a fair payoff distribution.

[16] 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 in the aggregate these assumptions imply that 40 percent of subjects exhibit $\alpha_i \geq \beta_i > 0.5$ and that 60 percent exhibit $0.5 > \alpha_i \geq \beta_i$.

results that follow are robust to changes in this distribution, as long as the share of inequity averse types is at least 33 percent but not larger than 60 percent.

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 therefore relegated to an appendix that can be found on our webpage.[17] Here, we want to report the main predictions that follow from this analysis and expound on them. First, we focus on the trust contract.

Proposition 1 [Trust contracts]: *Increasing the wage in a trust contract increases the effort of the inequity-averse 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 an inequity-averse (i.e., 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, e increases with w for an inequity-averse agent, but, if the fraction of inequity-averse 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.[18] What about the fair principal? A generous wage will not pay off in monetary terms, and it will generate inequality to the principal's disadvantage whenever a selfish agent chooses $e=1$. Hence, a fair principal will not pay a higher wage either. The reason why the trust contract does not work is that the percentage of fair agents is simply too small. This percentage would have to be at least *60* percent in order to make it profitable

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

[18] Paying generous wages in a trust contract may be profitable for the principal in other environments. For example, it has been shown experimentally (e.g., Fehr, Kirchsteiger and Riedl 1993) that if the principal's payoff function is $M^P = (v\text{-}w)e$ with $v = 120$ and $e \in \{0.1, 0.2, \ldots, 1\}$, the payment of generous wages increases profits. The reason is that the principal's effective wage cost is $we < w$ for all $e < 1$ with this payoff function. Thus, paying generous wages is considerably less costly if the agents choose low effort levels. In contrast, the wage cost is independent of the agent's effort with our payoff function which seems to be more natural.

for the selfish principals to offer generous wages that induce the fair agents to choose $e > 4$.[19] Recall from Table 2 in Section 4.1. that – in a trust contract – higher wages are indeed associated with lower payoffs for the principals. Proposition 1 neatly rationalizes this fact. Next we turn to the analyses of the incentive contracts.

Proposition 2 [Incentive Contracts]: *Both the selfish and the fair principals stipulate the maximal fine, $f = 13$, and demand the maximal incentive compatible effort level, $e^* = 4$ in the optimal incentive contract. However, the selfish principals offer a wage $w = 4$ that gives the agents none of the surplus while fair principals offer $w = 17$ which distributes the surplus equally. Selfish agents accept and obey these contracts. The fair agents accept and obey generous contracts with $w \geq 17$ and choose $e \geq 4$, where e is increasing with w. Fair agents reject, however, $w = 4$. If $4 < w < 17$, they will either reject or shirk even if the contract is incentive compatible.*

We already know from Section 3 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 will reject her unfair 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 as long as $q < 0.6$, increasing the wage does not pay off. Furthermore, increasing w somewhat, but not up to the level of full equality, is dangerous, because the agent may accept the contract but consider it unfair and shirk. Therefore, $(w=4, e^*=4, \overline{f}=13)$ is optimal for a selfish principal. Note that the contracts of the selfish principals will be rejected with probability 0.4 due to the existence of fair agents. This explains the frequent rejection of low wage offers and why the optimal incentive contract becomes less efficient than predicted by the self-interest model. An inequity-averse 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 also 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 in expectation.

[19] Note that for $e > 4$, $c'(e) \geq 2$. Hence, the marginal revenue of a unit increase in wages at $e = 4$ equals $10 \cdot q \cdot (2/12)$ which exceeds 1 for $q > 12/20 = 6/10$. 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 $q > 0.6$, 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.

According to Proposition 1, trust contracts are very ineffective because it is too costly for the principals to offer generous wages to induce non-minimal effort levels. In contrast, both the fair and the selfish principal can enforce $e = 4$ in the optimal (accepted) incentive contracts. Hence, for the TI treatment, in which the principals can choose between trust and incentive contracts, the following proposition holds:

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: incentive contracts outperform trust contracts. However, the inequity aversion model 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 predicts correctly that incentive contracts are frequently associated with generous wages between 10 and 20 (offered by fair principals). Finally, it offers an explanation as to why many agents choose effort levels larger than 1 in response to generous wage offers in trust contracts and in incentive compatible incentive contracts.

We now turn to the analysis of the BI treatment. For this purpose, we first examine the equilibrium in a setting in which the principals can only choose a bonus contract. This gives us a characterization of the equilibrium bonus contract. Then we compare this contract with the optimal incentive contract from Proposition 2, which gives us the prediction for the case where the principals can choose between a bonus and an incentive contract.

The analysis of the bonus contract is a little more complicated than the analysis of the TC or the IC. 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. To see this suppose that the inequity-averse and the selfish principal offer different contracts. In this case, the agents know from the contract offer whether they face a fair principal or a selfish principal. If they face a fair principal they choose $e =$

10 because this principal pays a bonus that distributes the surplus equally. If they face a selfish principal they choose $e = 1$. 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 there are q percent fair principals who choose b in this way while $(1-q)$ percent of the principals 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 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 $2/11 \approx 0.18$, 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 $q = 0.4$, selfish agents will choose the maximal effort level for which the marginal effort cost does not exceed 2, that is, they choose $e = 7$.

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 (the fair agent chooses $e = 2$ iff $w \geq 10$). 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 the selection of $e = 7$ is profitable from a monetary perspective, 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 to 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 surplus-sharing 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 equilibrium outcomes shrinks to a singleton in which all principals offer $w=15$.[20] 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 $e=2$; therefore, the average effort level is $0.6{\cdot}7 + 0.4{\cdot}2=5$. We summarize our results regarding the bonus contract in

Proposition 4 [Bonus contract]: (a) *No separating equilibrium exists in which the selfish principal offers a different contract from that which the fair principal offers.*

(b) *If a higher wage offer of the principal is not interpreted as a signal that the principal is more likely to be selfish, then a unique pooling equilibrium outcome exists in which both types of principals offer w=15. The selfish agent chooses e=7 and is rewarded by the fair 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.*

Proposition 4 shows that the average effort level is higher in a bonus contract than in an optimal incentive contract. Moreover, the selfish principals reap the benefits of this high effort level without actually paying the bonus. Thus, the selfish principals can exploit the fact that there are fair principals. It is, therefore, obvious that they earn more by offering bonus contracts. However, the fair principals also earn more in the bonus contract because they share the surplus in the incentive and the bonus contract equally but in the latter the surplus is higher. Thus we get:

Proposition 5 [BI-Treatment]: *Both types of principals prefer the bonus relative to the incentive contract.*

---

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

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. Our fairness model not only predicts that bonus contracts outperform incentive contracts; it also offers remarkably precise quantitative predictions of the data. Recall from Figure 5a that the average wage is roughly 15 in the bonus contracts while the average bonus is roughly 10. This is exactly the prediction provided in Proposition 4. Moreover, the actual effort level in the bonus contracts is, on average, 5.2; the predicted average effort level according to Proposition 4 is e = 5. In addition, the model predicts that some principals pay a bonus while others don't – a fact that we emphasized already in Section 4.2.

The above analysis also shows why the bonus contract is so much better than the trust contract, even though both contracts rely on fairness as an enforcement device. Under a bonus contract, the agent has to trust first by expending more than the minimum level of effort. Note that even if the agent engages in the maximum effort of 10, the cost of this effort is relatively small *(c(10)=20)*. Thus, even if the principal does not pay the bonus, the agent does not lose too much. On the other hand, under a trust contract the principal has to pay a generous wage upfront. In order to induce a fair agent to expend the efficient level of effort (*e=10*) she has to offer *w=60*, i.e., the principal risks a loss of 50 if the agent is not trustworthy. Therefore, it is much more risky for the principal to appeal to the agent's fairness than the other way round. This suggests that, as a general principle, the player who loses less from trusting the other person should trust first.

Finally, we consider the EBI treatment briefly. The major fact that needs to be explained in this treatment is the strong preference in favor of the pure bonus contract. Moreover, this preference is probably associated with the fact that the effort is not significantly larger in the combined contract, so that it is not worthwhile to pay the verification cost for the explicit incentive. The surprisingly low enforcement power of the combined contract – relative to what seems, in principle, enforceable – is also related to the much lower bonus incentives in this contract. Recall that the principals paid much lower bonuses in the combined contracts.

Our fairness model is capable of rationalizing the principals' preference for the pure bonus contract. If the agents believe that those principals who propose a combined contract are more likely to be selfish and, hence, pay no bonus, the bonus incentive in the combined contract is weakened decisively. In fact, it is easy to show that if the agents take a combined contract offer as a signal that the principal is likely to be selfish and if a higher wage is taken as a signal that the principal is more likely to be fair, then the only equilibrium outcome in the EBI treatment is that the pure bonus contract of Proposition 4 will be offered. However, this is not the only equilibrium. Other pooling equilibria also exist and, in particular, there is an equilibrium in which all principals choose the combined contract. In such an equilibrium the principals are capable of enforcing almost the maximal effort level because they combine the power of the explicit and the implicit incentive in an optimal way: both the selfish and the fair principals offer a combined contract $(w=18, e^{*}=9, f=13, b^{*}=30)$ which is accepted by both types of agents. The selfish agent chooses $e = 9$ while the fair agent chooses $e = 3$. The selfish principal does not pay a bonus while the fair principal pays $b = 30$ if $e = 9$ and $b = 0$ *if* $e = 3$.

To assess the plausibility of the different equilibria, one has to make a judgment about the plausibility of the out-of-equilibrium beliefs that support the different equilibria. Here the fact that those principals who offered the combined contract indeed paid lower bonuses becomes important. This means that the agents had a strong reason to believe that these principals are more selfish. Once this belief was established among the agents, the principals have an incentive for choosing the pure bonus contracts. Therefore, we believe that the out-of-equilibrium beliefs, which support the pure bonus equilibrium are the more reasonable ones.

6. Conclusions

Our experiments have shown that fairness concerns 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, implicit bonus contracts that cannot work when all actors are selfish provide powerful incentives and become superior when there are also fair-minded players. Our results indicate that the principals understand that fairness matters and predominantly choose the superior bonus contract that relies on fairness as an enforcement device. Moreover, the principals even prefer the pure bonus contract over a contract that combines the explicit incentive with the implicit bonus incentive. This result suggests that fairness concerns may be one important reason why principals often do not use explicit incentives although they are readily available. In this context we also observe intriguing interactions between the explicit and the implicit incentive – the use of the explicit incentive substantially weakens the enforcement power of the bonus incentive. We conjecture that this is an important reason for the principals' preference in favor of the pure bonus contract.

There are several other points that deserve to be emphasized. First, the principals converge towards the most efficient contract in the set of available contracts. This observation is important because the "efficiency principle" provides the basis for much of modern contract theory. However, it remains to be seen whether this observation extends to other – more complicated – environments. Second, it is important to remember that only some subjects are concerned about fairness. A considerable percentage of subjects also seems to be only interested in their own material payoff. Whether fairness motives provide a good enforcement device depends on the percentage of fair persons in the population *and* on the strategic situation in which the subjects interact. We have shown that fairness concerns are too weak for contract enforcement in a setting where the trust contract competes with the incentive contract while they are strong enough if the bonus contract becomes available. This asymmetry in the impact of fairness concerns is due to the fact that it is less costly for the agent to trust in the bonus contract than for the principal to trust in the trust contract. Third, our theoretical results show that simple and tractable models of fairness can yield interesting and non-obvious insights into the problems of contract choice and incentive provision. Our fairness model is consistent with the major qualitative patterns in the data. In addition, it provides surprisingly accurate quantitative predictions of the details of the bonus contract.

Finally, 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 opponent's behavior and the equity of the final outcome. For example, our theoretical analysis shows that self-interested agents respond more strongly to the implicit incentives provided by a bonus contract than fair agents. 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 players 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.

To conclude, our experiments show that concerns for fairness have an important impact on the actual and the optimal choice of contracts. Traditional contract theory has neglected these effects, but they have to be taken into account if we want to fully understand the functioning of real world contracts and the associated incentives schemes. Our theoretical analysis shows that it is possible to model these effects explicitly and to develop richer models that may become part of a new paradigm of “behavioral contract theory”.

References

Anderhub, Vital, Gächter, Simon and Königstein, Manfred, 2002. „Efficient Contracting and Fair Play in a Simple Principal-Agent Experiment.“ *Experimental Economics*, 5, 5-27.

Berg, Joyce, Dickhaut, John W., and McCabe, Kevin A., 1995. Trust, Reciprocity, and Social History. *Games and Economic Behavior* 10, 122-142.

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

Cabrales, Antonio, and Charness, Gary, 2003. “Optimal Contracts, Adverse Selection, and Social Preferences: An experiment.” *Mimeo*, University of California at Santa Barbara.

Camerer, Colin and Weigelt, Keith (1988). Experimental Tests of a Sequential Equilibrium Reputation Model, *Econometrica*, 56, 1-36.

Camerer, Colin F., 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*, Princeton, Princeton University Press.

Charness, Gary and Dufwenberg, Martin, 2003. “Promises and Partnerhsips”, *Mimeo*, University of California at Santa Barbara.

Charness, Gary, and Rabin, Matthew, 2002. “Understanding Social Preferences with Simple Tests.” *Quarterly Journal of Economics*, 117, 817-869.

Chaudhuri, Ananish, 1998. “The Ratchet Principle in a Principal-Agent Game with Unknown Costs: An Experimental Analysis”, *Journal of Economic Behavior and Organization*, 37, 291-304.

Chiappori, Pierre-Andre and Salanié, Bernard, 2003. “Testing Contract Theory : A Survey of Some Recent Work.” In: M. Dewatripont, L.P. Hansen, S. Turnovski, *Advances in Economic Theory, Eigth World Congress of the Econometric Society*, Vol. 1, Cambridge: Cambridge University Press, 115-149.

Cooper, David J., Kagel John H., Lo W. and Qing Liang Gu (1999). “Gaming against Managers in Incentive Systems: Experiments with Chinese Students and Chinese Managers”, *American Economic Review*, 89, 781-804.

Cox, Jim and Friedman Daniel, 2002, “A Tractable Model of Reciprocity and Fairness”, Mimeo, University of Arizona at Tucson.

DeJong, Douglas V., Forsythe Robert, Lundholm Russel J., and Uecker Wilfred C., 1985. “A Laboratory Investigation of the Moral Hazard Problem in an Agency Relationship”, *Journal of Accounting Research* 23, 81 – 120.

Dufwenberg, Martin and Kirchsteiger, Georg, in press. “A Theory of Sequential Reciprocity.” *Games and Economic Behavior*.

Falk, Armin and Fischbacher, Urs, 1999. “A Theory of Reciprocity.” Working Paper No 6, Institute for Empirical Research in Economics, University of Zürich.

Fehr, Ernst, Kirchsteiger, George and Riedl, Arno, (1993). “Does Fairness Prevent Market Clearing? An Experimental Investigation”, *Quarterly Journal of Economics* 58, 437-460.

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

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., 2003. “Theories of Fairness and Reciprocity: Evidence and Economic Applications”, in: M. Dewatripont, L.P. Hansen, S. Turnovski, *Advances in Economic Theory, Eigth World Congress of the Econometric Society*, Vol. 1, Cambridge: Cambridge University Press, 208-257.

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

Güth, Werner, Königstein Manfred, Kovacs Judith, and Zala-Mezo Eniko, 2001. “Fairness within Firms: the Case of One Principal and Multiple Agents“, *Schmalenbach Business Review* 53, 82 – 101.

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

Keser, Claudia and Willinger, Marc, 2000. “Principals’ Principles when Agents’ Actions are Hidden”, *International Journal of Industrial Organization* 18, 163-185.

Laffont, Jean J. and Tirole, J., 1993. *A Theory of Incentives in Procurement and Regulation*, Cambridge MA: The MIT Press.

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

Prendergast, Canice, 1999. "The Provision of Incentives in Firms." *Journal of Economic Literature*, 37, 7-63.

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

Schotter, Andrew, Bull, Clive and Weigelt, Keith, 1987. "Tournaments and Piece Rates: An Experimental Study", *Journal of Political Economy*, 95, 1-33.

Schotter Andrew and Nalbantian Haig, 1997. "Productivity under Group Incentives: An Experimental Study", *American Economic Review* 87, 314 – 340.

Sobel, Joel (2002). Social Preferences and Reciprocity, mimeo, University of California, San Diego.

CESifo Working Paper Series

(for full list see www.cesifo.de)

1150 Giorgio Brunello and Francesca Gambarotto, Agglomeration Effects on Employer-Provided Training: Evidence from the UK, March 2004

1151 S. Brock Blomberg, Gregory D. Hess, and Athanasios Orphanides, The Macroeconomic Consequences of Terrorism, March 2004

1152 Bodo Sturm and Joachim Weimann, Unilateral Emissions Abatement: An Experiment, March 2004

1153 Wolfgang Ochel, Welfare-to-Work Experiences with Specific Work-First Programmes in Selected Countries, March 2004

1154 Jan K. Brueckner and Eric Pels, European Airline Mergers, Alliance Consolidation, and Consumer Welfare, March 2004

1155 Aaron Tornell, Frank Westermann, and Lorenza Martínez, NAFTA and Mexico's Economic Performance, March 2004

1156 George Economides, Sarantis Kalyvitis, and Apostolis Philippopoulos, Do Foreign Aid Transfers Distort Incentives and Hurt Growth? Theory and Evidence from 75 Aid-recipient Countries, March 2004

1157 Robert Fenge and Volker Meier, Are Family Allowances and Fertility-related pensions Siamese Twins?, March 2004

1158 Bruno S. Frey, Simon Luechinger, and Alois Stutzer, Valuing Public Goods: The Life Satisfation Approach, March 2004

1159 Jerome L. Stein and Guay C. Lim, Asian Crises: Theory, Evidence, Warning-Signals, March 2004

1160 Romain Ranciere, Aaron Tornell, and Frank Westermann, Crises and Growth: A Re-Evaluation, March 2004

1161 Assaf Razin and Efraim Sadka, Transparency, Specialization and FDI, March 2004

1162 Ludger Woessmann, How Equal Are Educational Opportunities? Family Background and Student Achievement in Europe and the United States, March 2004

1163 B.M.S. van Praag and Barbara E. Baarsma, Using Happiness Surveys to Value Intangibles: The Case of Airport Noise, March 2004

1164 Aaron Tornell, Frank Westermann, and Lorenza Martínez, The Positive Link Between Financial Liberalization, Growth, and Crises, March 2004

1165 Helge Berger and Carsten Hefeker, One Country, One Vote? Labor Market Structure and Voting Rights in the ECB, March 2004

1166 Clemens Fuest and Martin Kolmar, A Theory of User-Fee Competition, March 2004

1167 Friedrich Schneider and Robert Klinglmair, Shadow Economies around the World: What Do We Know?, April 2004

1168 Horst Raff and Nicolas Schmitt, Exclusive Dealing and Common Agency in International Markets, April 2004

1169 M. Hashem Pesaran and Allan Timmermann, Real Time Econometrics, April 2004

1170 Sean D. Barrett, Privatisation in Ireland, April 2004

1171 V. Anton Muscatelli, Patrizio Tirelli and Carmine Trecroci, Can Fiscal Policy Help Macroeconomic Stabilisation? Evidence from a New Keynesian Model with Liquidity Constraints, April 2004

1172 Bernd Huber and Marco Runkel, Tax Competition, Excludable Public Goods and User Charges, April 2004

1173 John McMillan and Pablo Zoido, How to Subvert Democracy: Montesinos in Peru, April 2004

1174 Theo Eicher and Jong Woo Kang, Trade, Foreign Direct Investment or Acquisition: Optimal Entry Modes for Multinationals, April 2004

1175 Chang Woon Nam and Doina Maria Radulescu, Types of Tax Concessions for Attracting Foreign Direct Investment in Free Economic Zones, April 2004

1176 M. Hashem Pesaran and Andreas Pick, Econometric Issues in the Analysis of Contagion, April 2004

1177 Steinar Holden and Fredrik Wulfsberg, Downward Nominal Wage Rigidity in Europe, April 2004

1178 Stefan Lachenmaier and Ludger Woessmann, Does Innovation Cause Exports? Evidence from Exogenous Innovation Impulses and Obstacles, April 2004

1179 Thiess Buettner and Johannes Rincke, Labor Market Effects of Economic Integration – The Impact of Re-Unification in German Border Regions, April 2004

1180 Marko Koethenbuerger, Leviathans, Federal Transfers, and the Cartelization Hypothesis, April 2004

1181 Michael Hoel, Tor Iversen, Tore Nilssen, and Jon Vislie, Genetic Testing and Repulsion from Chance, April 2004

1182 Paul De Grauwe and Gunther Schnabl, Exchange Rate Regimes and Macroeconomic Stability in Central and Eastern Europe, April 2004

1183 Arjan M. Lejour and Ruud A. de Mooij, Turkish Delight – Does Turkey's accession to the EU bring economic benefits?, May 2004

1184 Anzelika Zaiceva, Implications of EU Accession for International Migration: An Assessment of Potential Migration Pressure, May 2004

1185 Udo Kreickemeier, Fair Wages and Human Capital Accumulation in a Global Economy, May 2004

1186 Jean-Pierre Ponssard, Rent Dissipation in Repeated Entry Games: Some New Results, May 2004

1187 Pablo Arocena, Privatisation Policy in Spain: Stuck Between Liberalisation and the Protection of Nationals' Interests, May 2004

1188 Günter Knieps, Privatisation of Network Industries in Germany: A Disaggregated Approach, May 2004

1189 Robert J. Gary-Bobo and Alain Trannoy, Efficient Tuition Fees, Examinations, and Subsidies, May 2004

1190 Saku Aura and Gregory D. Hess, What's in a Name?, May 2004

1191 Sjur Didrik Flåm and Yuri Ermoliev, Investment Uncertainty, and Production Games, May 2004

1192 Yin-Wong Cheung and Jude Yuen, The Suitability of a Greater China Currency Union, May 2004

1193 Inés Macho-Stadler and David Pérez-Castrillo, Optimal Enforcement Policy and Firms' Emissions and Compliance with Environmental Taxes, May 2004

1194 Paul De Grauwe and Marianna Grimaldi, Bubbles and Crashes in a Behavioural Finance Model, May 2004

1195 Michel Berne and Gérard Pogorel, Privatization Experiences in France, May 2004

1196 Andrea Galeotti and José Luis Moraga-González, A Model of Strategic Targeted Advertising, May 2004

1197 Hans Gersbach and Hans Haller, When Inefficiency Begets Efficiency, May 2004

1198 Saku Aura, Estate and Capital Gains Taxation: Efficiency and Political Economy Consideration, May 2004

1199 Sandra Waller and Jakob de Haan, Credibility and Transparency of Central Banks: New Results Based on Ifo's *World Economic Survey*, May 2004

1200 Henk C. Kranendonk, Jan Bonenkamp, and Johan P. Verbruggen, A Leading Indicator for the Dutch Economy – Methodological and Empirical Revision of the CPB System, May 2004

1201 Michael Ehrmann, Firm Size and Monetary Policy Transmission – Evidence from German Business Survey Data, May 2004

1202 Thomas A. Knetsch, Evaluating the German Inventory Cycle – Using Data from the Ifo Business Survey, May 2004

1203 Stefan Mittnik and Peter Zadrozny, Forecasting Quarterly German GDP at Monthly Intervals Using Monthly IFO Business Conditions Data, May 2004

1204 Elmer Sterken, The Role of the IFO Business Climate Indicator and Asset Prices in German Monetary Policy, May 2004

1205 Jan Jacobs and Jan-Egbert Sturm, Do Ifo Indicators Help Explain Revisions in German Industrial Production?, May 2004

1206 Ulrich Woitek, Real Wages and Business Cycle Asymmetries, May 2004

1207 Burkhard Heer and Alfred Maußner, Computation of Business Cycle Models: A Comparison of Numerical Methods, June 2004

1208 Costas Hadjiyiannis, Panos Hatzipanayotou, and Michael S. Michael, Pollution and Capital Tax Competition within a Regional Block, June 2004

1209 Stephan Klasen and Thorsten Nestmann, Population, Population Density, and Technological Change, June 2004

1210 Wolfgang Ochel, Welfare Time Limits in the United States – Experiences with a New Welfare-to-Work Approach, June 2004

1211 Luis H. R. Alvarez and Erkki Koskela, Taxation and Rotation Age under Stochastic Forest Stand Value, June 2004

1212 Bernard M. S. van Praag, The Connexion Between Old and New Approaches to Financial Satisfaction, June 2004

1213 Hendrik Hakenes and Martin Peitz, Selling Reputation When Going out of Business, June 2004

1214 Heikki Oksanen, Public Pensions in the National Accounts and Public Finance Targets, June 2004

1215 Ernst Fehr, Alexander Klein, and Klaus M. Schmidt, Contracts, Fairness, and Incentives, June 2004