Performance Incentives and the Dynamics of Voluntary Cooperation

Simon Gächter[*], Esther Kessler[*] and Manfred Königstein[**]

31 January 2006 – VERY PRELIMINARY VERSION

Abstract

How reversible are incentive systems? – Can firms revert to "trust and reciprocity" after practicing "pay for performance"? – These questions highlight the motivation underlying our work here. It is fashionable nowadays for firms to establish performance based pay systems. There seems to be a common understanding among business people that incentive contracts are essential in order to get employees to well-behave. This view is in line with the purely economic view of man, but it is challenged by the idea that monetary incentives may crowd out intrinsic motivation, an idea that has been put forward by social psychologists and, more recently, a few economists.

Intrinsic motivation of employees can not be directly observed. However it has been shown in laboratory experiments on principal-agent games that economic efforts of participants may be well above the levels that are economically rational. Thus, there is over-provision of effort by agents, which may be seen as indirect evidence of intrinsic motivation and which we refer to as "voluntary cooperation". We replicate and extend this line of research by investigating whether voluntary cooperation is affected by incentive contracting in the short run and in the long run. Our main question is whether voluntary cooperation can be re-established after experiencing incentive pay for some time. We find that voluntary cooperation is much smaller after experiencing incentive pay and refer to this effect as crowding out in a dynamic sense. Furthermore, the positive correlation between payment and voluntary cooperation, which can be interpreted as reciprocity, deteriorates. The observed effects are stronger for fine contracts than for bonus contracts.

Key words: principal-agent, contract, experiment, intrinsic motivation

JEL-Codes: M5, C7, C9

[* *] University of Nottingham, School of Economics, Sir Clive Granger Building, University Park, Nottingham NG7 2R, United Kingdom Simon Gächter is also affiliated with CESifo and IZA.

[**] Universität Erfurt and IZA. Professur für Angewandte Mikroökonomie, Nordhäuser Str. 63, D-99089 Erfurt.

1. Introduction

How reversible are behavioral consequences of incentive systems? – Can firms revert to "trust and reciprocity" after practicing "pay for performance"? – These questions highlight the motivation underlying our study.

It is fashionable nowadays for firms to establish performance based pay systems. Business people seem to share a common understanding that incentive contracts are essential in order to induce employees to behave well. This view is in line with the purely economic view of man, but it is challenged by the idea that monetary incentives may crowd out intrinsic motivation, an idea that has been put forward by social psychologists and, more recently, a few economists.

Intrinsic motivation of employees, provided it exists, can not be easily observed. In a narrow definition a person may be considered as intrinsically motivated, if her behavior is not driven by any kind of economic or social reward or punishment. For employees it is hardly conceivable that such situations exist; in all their activities they may consider prospective earnings, promotions, possible layoff or social approval by superiors or fellow workers. Similarly, it is hard to conceive any kind of social interaction which does not allow for considerations of reward or punishment.

Thus, instead of studying whether there is "crowding out of intrinsic motivation by economic or social reward or punishment", we investigate whether there is "crowding out of voluntary cooperation by explicit economic incentives". By voluntary cooperation we mean generally the difference between a person's actual work effort and the work effort that can be enforced in a contract. This definition of voluntary cooperation is applicable in employer-employee relationships as well as other social relationships. Specifically, the connection between voluntary cooperation and explicit economic incentives can be studied in laboratory experiments; e.g., it has been studied in Fehr and Gächter (2002) and Anderhub, Gächter and Königstein (2002). Laboratory experiments allow to control the terms of incentive contracts and give easy access to individual performance or effort measures. Both is difficult within natural work relationships. Therefore we think that experiments are suitable for investigating crowding out theory. If crowding out happens in employer-employee relationships, it should also happen in lab experiments on social relationships. Furthermore, the economic paradigm, which contrasts crowding out theory, is not restricted to natural work relationships; it is also applicable in lab experiments (see also Falk and Fehr 2002, and Fehr and Falk 2003).

Our study aims at two things: Reproducing several findings of previous studies within a unified framework, and adding another bit to the crowding out story that has not been

investigated before. We study a principal-agent relationship that allows for different kinds of contracts: incentive contracts, in which a bonus or fine may be conditioned on work effort (Bonus game, and respectively, Fine game) or a pure fixed wage contract (Trust game). Moreover, the participants experience different treatments in different phases of the experiment, e.g., Trust in phase 1, Bonus in phase 2 and Trust in phase 3 (treatment TBT). Among other things this design allows to investigate whether voluntary cooperation can be reestablished after an intermediate phase of incentive pay. We find that this is not the case. Voluntary cooperation does not return to the same level it had before application of the incentive treatment. Such reduction in voluntary cooperation does not occur in the baseline treatment TTT (Trust in each phase). It seems that experiencing incentive pay has a lasting deteriorating effect on voluntary cooperation and the underlying behavioral mechanism of trust-and-reciprocity. To our knowledge, this has not been studied in principal-agent experiments before. We also find crowding out when participants experience incentive contracts in the beginning of the experiment and trust contracts later on. Furthermore we reproduce findings of other studies on the effectiveness of bonuses and fines as well as the effectiveness of pure fixed wages.

In section 2 we describe the experimental design. In section 3 we derive the behavioral predictions and describe the experimental procedures. The results are presented in section 4. In section 5 we summarize our findings and provide concluding remarks.

2. Description of games and benchmark solutions

We investigate three versions of principal agent games based on the gift exchange game studied by Fehr, Kirchsteiger and Riedl (1993) and Fehr, Gächter and Kirchsteiger (1997). Specifically, we contrast a principal agent game in which only fixed wage contracts are feasible (called the “Trust” game) with games in which fixed wages and in addition fines (“Fine” game) or bonuses (“Bonus” game) are feasible. Fines and bonuses can be used to punish or reward low or high effort.

For each game we first describe the rules and then derive a benchmark solution. The benchmark solution characterizes behavior of principal and agent that should be observed if both are selfish and rational individuals. Specifically, we determine the paths of subgame perfect equilibria of the games. Later on we will compare these benchmark decisions – we refer to this as the “rational solution” – with alternative predictions and actual behavior. Finally, we describe the sequencing of games applied in the experiment.

2.1 Trust game

The game consists of three stages. First, the principal designs a contract to be offered to the agent. The contract comprises two components, a fixed wage w and a desired effort e^d, which is not binding for the agent. The contract offer has to obey the following restrictions.

$$1 \leq e^d \leq 20 \text{ and } -700 \leq w \leq 700. \quad (1)$$

Both e^d and w, as well as all other experimental choice variables described below are restricted to integer numbers.

Second, the agent is informed about the offer and may accept or reject the contract. If he rejects the contract, the game ends and both, principal and agent, earn nothing. If the agent accepts the contract, third, he chooses effort e. That is., the third stage is entered only if the agent has accepted the offered contract. Again, in his choice of e ($1 < e < 20$, in integers) the agent is not restricted by e^d. After effort has been chosen, the game ends. The principal's profit is given by:

$$\pi^P = \begin{cases} 35e - w & \text{if the contract is accepted.} \\ 0 & \text{if the contract is rejected.} \end{cases} \quad (2)$$

The agent's profit is:

$$\pi^A = \begin{cases} w - c(e) & \text{if the contract is accepted.} \\ 0 & \text{if the contract is rejected.} \end{cases} \quad (3)$$

Thus, effort induces a return that accrues to the principal and causes a cost (disutility of effort) to be borne by the agent. The cost function is increasing and linear in effort: $c(e) = 7e - 7$. It is a game of complete information. Each player knows the rules, including the payoff functions for both players, and is informed about all choices made in the game. Since the payment w can not be conditioned upon effort, we refer to this game as "Trust game".

Benchmark Solution: Since providing effort is costly and payment does not depend on effort, it is optimal for the agent to choose minimal effort $e = e^{min} = 1$. Anticipating this, the principal offers the wage that just ensures the agent's acceptance, namely $w = 1$ (or $w = 0$ if one assumes acceptance in case the agent is indifferent between acceptance and rejection). These decisions result in payoffs of 34 money units for the principal and 1 money unit for the

agent. This solution is highly inefficient, since the efficient surplus is 567 and requires maximal effort $e = e^{max} = 20$.

2.2 Fine game

In the Fine game the principal may punish the agent if effort falls short of e^d. A contract now stipulates w, e^d and f, where f represents a fine (a wage reduction). The principal may choose one of four fine levels: $f \in \{0, 24, 52, 80\}$. If $e < e^d$, then f is subtracted from the agent's payoff and added to the principals payoff. If $e \geq e^d$, the fine has no effect. The principal's profit is given by:

$$\pi^P = \begin{cases} 35e - w & \text{if the contract is accepted and } e \geq e^d \\ 35e - w + f & \text{if the contract is accepted and } e < e^d \\ 0 & \text{if the contract is rejected.} \end{cases} \qquad (4)$$

The agent earns:

$$\pi^A = \begin{cases} w - c(e) & \text{if the contract is accepted and } e \geq e^d \\ w - c(e) - f & \text{if the contract is accepted and } e < e^d \\ 0 & \text{if the contract is rejected.} \end{cases} \qquad (5)$$

In all other respects the Fine game is equivalent to the Trust game.

Benchmark Solution: In choosing effort the agent has to consider two alternatives, $e = e^d$ or $e = 1$. All other effort choices are clearly suboptimal. Namely, in case of $e = e^d$ there will be no fine. A higher effort is suboptimal since it causes higher cost without increasing payment. In case of $e < e^d$, the agent has to pay the fine, and this is independent of the exact value of e. So, conditional on $e < e^d$ minimal effort $e = 1$ is best. Taken together the agent will choose

$$e^{opt} = \begin{cases} e^d & \text{if } \; w - c(e^d) \geq w - f - c(1) \;\Leftrightarrow\; f \geq c(e^d) \;\text{ and} \\ 1 & \text{otherwise.} \end{cases} \qquad (6)$$

Thus, the agent will perform at the desired effort level if the fine is larger than the costs of contractual compliance. The agent's choice rule (6) is at the same time an incentive compatibility constraint for the principal's contract design problem. Namely, for each level of fine there exist a maximal level of effort that satisfies $f \geq c(e)$. This is the maximal effort level

which the principal may desire given the fine level. We refer to these as enforceable effort levels and summarize these in Table 1.[1]

Fine/Bonus	Enforceable Effort
0	1
24	4
52	8
80	12

Table 1: *Different levels of Fine/Bonus and corresponding enforceable effort levels*

Before choosing effort the agent has to accept an offered contract in the first place. This requires

$$\begin{aligned} w - c(e^d) &\geq 0 \qquad \text{if } e^{opt} = e^d \text{ or} \\ w - f - c(1) &\geq 0 \qquad \text{if } e^{opt} = 1, \end{aligned} \tag{7}$$

which are participation constraints for the principal's contract design problem. The principal chooses w, f and e^d in order to maximize profit (4) subject to (6) and (7). One can easily see that it is optimal for the principal to set w such that (7) holds with equality, and furthermore that the solution to the principal's problem is $f = 80$, $e^d = 12$ and $w = c(12) = 77$. Accordingly, the agent will accept the contract and will indeed choose $e = 12$ as desired by the principal. The solution is more efficient than the solution without fines, however, it does not generate maximal surplus. It induces a rather asymmetric distribution of payoffs: 343 for the principal and 0 for the agent.

2.3 Bonus game

In the Bonus game the principal has the opportunity to reward the agent if chosen effort e is equal or larger then desired effort e^d. Thus, the offered contract features w, e^d and b, where b is a bonus (a wage increase) with $b \in \{0, 24, 52, 80\}$. If $e \geq e^d$, then the bonus is added to the agents payoff and subtracted from the principal's payoff. If $e < e^d$, the bonus has no effect. The principal earns:

[1] We will show below that a similar reasoning holds for bonus contracts. Therefore the table also contains the results for bonus contracts.

$$\pi^P = \begin{cases} 35e - w - b & \text{if the contract is accepted and } e \geq e^d \\ 35e - w & \text{if the contract is accepted and } e < e^d \\ 0 & \text{if the contract is rejected.} \end{cases} \quad (8)$$

The agent receives:

$$\pi^A = \begin{cases} w + b - c(e) & \text{if the contract is accepted and } e \geq e^d \\ w - c(e) & \text{if the contract is accepted and } e < e^d \\ 0 & \text{if the contract is rejected.} \end{cases} \quad (9)$$

In all other respects the Bonus game is equivalent to the Trust game.

Benchmark Solution: Actually, the solution for the game with bonus contracts is quite similar to the solution with fine contracts. Again, in choosing effort the agent has to consider only $e = e^d$ or $e = 1$. In case of $e = e^d$ he receives the bonus. Higher effort is suboptimal. For all choices $e < e^d$, the bonus is lost, such that $e = 1$ is best. It follows:

$$e^{opt} = \begin{cases} e^d & \text{if } w + b - c(e^d) \geq w - c(1) \Leftrightarrow b \geq c(e^d) \text{ and} \\ 1 & \text{otherwise.} \end{cases} \quad (10)$$

Note, that (10) is equivalent to (6) except that the bonus b replaces the fine f. Consequently different bonus levels correspond to the enforceable effort levels as shown in Table 1. The agent will accept an offered contract only if

$$\begin{aligned} w + b - c(e^d) &\geq 0 \quad \text{if } e^{opt} = e^d \text{ or} \\ w - c(1) &\geq 0 \quad \text{if } e^{opt} = 1, \end{aligned} \quad (11)$$

The principal chooses w, b and e^d in order to maximize profit (8) subject to (10) and (11). Obviously, this implies to choose w such that (11) holds with equality. The solution to the principal's problem is $b = 80$, $e^d = 12$ and $w = c(12) - b = -3$. The agent will accept the contract and will choose $e = 12$ as desired by the principal. Thus, from a purely economic viewpoint there is no difference between bonus contracts and fine contracts. The assumption of selfish and rational players implies a maximal bonus, or respectively a maximal fine, an effort level of 12 and a profit of 343 for the principal and 0 for the agent.

3. Experimental design and behavioral predictions

3.1 Design and hypotheses

In the experiment the participants played multiple rounds of the above games. Specifically, they played two, or respectively three, blocks of ten games. For instance, one group of subjects played ten Fine games first (phase 1) and ten Trust games thereafter (phase 2). Another group played ten Trust games, then ten Fine games and finally ten Trust games. All treatments are listed in Table 2. For several of our empirical analyses treatment TTT will be the baseline against which other treatments are compared.

Table 2 also indicates the respective matching procedure. In most sessions subjects (principal and agent) were matched randomly across games ("Stranger" matching). In some sessions subjects were matched with the same player within each block and with different players across blocks ("Partner" matching).

Treatment label	Phase 1 (Period 1-10)	Phase 2 (Period 11-20)	Phase 3 (Period 21-30)	Subjects	No. Independent matching Groups
A. *Benchmark: Establishing reciprocity, pay for performance, undermining and crowding out effects*					
FT	Fine	Trust	-	80	6
BT	Bonus	Trust	-	78	6
TTT	Trust	Trust	Trust	78	6
B. *Introducing incentives into a Trust environment: What is the impact on pay for performance, undermining, and crowding out?*					
TFT	Trust	Fine	Trust	86	6
TBT	Trust	Bonus	Trust	84	6
C. *What is the impact of strategic effects on reciprocity, pay for performance, undermining and crowding out in long-term relations*					
TTT-Partner	Trust	Trust	Trust	24	12
TFT-Partner	Trust	Fine	Trust	36	18
TBT-Partner	Trust	Bonus	Trust	34	17

Table 2: *Experimental design and numbers of subjects and matching groups*

Compared to other principal agent experiments this is a rather comprehensive experimental design. It allows studying the effect of monetary incentives on behavior in a static sense as well as in a dynamic sense. It allows for framing effects as well as repeated game effects.

First, we will determine the effectiveness of fine contracts and bonus contracts in stimulating agent's effort. To do so we will analyze phase-1-data of treatments FT and BT as

well as phase-2-data of treatments TFT and TBT. Standard economic theory as worked out above in the benchmark solutions of the games predicts a positive correlation between effort and the level of fine (bonus).

Second, we assess the effectiveness of trust contracts in inducing effort by looking at the data for treatment TTT. The rational solution predicts that effort will be minimal in this case. But this prediction is challenged by social psychology and more recent economic concepts of social preferences,[2] which predict a positive influence of fixed wage payments on effort. Namely, a fixed wage may be interpreted as a trust signal by the principal to the agent. The agent in turn may reciprocate a high (low) fixed wage by choosing a high (low) effort level. This mechanism of "trust and reciprocity" – also referred to as "gift exchange" (see Fehr, Kirchsteiger and Riedl 1993, Fehr, Gächter and Kirchsteiger 1997) – leads to a positive correlation between fixed wage and effort, and furthermore to higher than minimal effort levels. If effort is higher than predicted according to the benchmark solution, we refer to this as "voluntary cooperation". Thus, contrary to the benchmark solution, we predict a positive amount of voluntary cooperation in treatment TTT.

The crowding out hypothesis generally proposes that experiencing monetary incentives reduces voluntary cooperation, with voluntary cooperation defined as $(e - e^{opt})$ and with e^{opt} representing individually rational effort. Therefore, third, we investigate several different issues related to crowding out. We look at the phase-1-data of FT and BT as well as the phase-2-data of TFT and TBT to determine the level of voluntary cooperation "while" subjects simultaneously experience incentive pay ("short run crowding out"). Then we use the phase-2-data of treatments FT, BT and TTT to investigate voluntary cooperation "after" subjects have previously experienced incentive pay ("long run crowding out"). And we also investigate the latter question using the data of treatments TFT, TBT and TTT. In this case subjects have the possibility to learn about the effectiveness of "trust and reciprocity" in phase 1 before experiencing incentive pay and before reverting to the trust game again. So we determine whether there is "crowding out despite learning voluntary cooperation".

Fourth, we investigate whether the above issues are influenced by the framing of incentives; i.e., we will compare fine contracts versus bonus contracts ("framing effect").

Fifth, all of the effects may be influenced by repeated interaction (Partner matching), since this allows for reputation formation. We will utilize the data of TFT-Partner, TBT-Partner and TTT-Partner together with the respective random matching data (TFT, TBT and

[2] see, e.g., Fehr and Schmidt 1999; Bolton and Ockenfels 2000; Falk and Fischbacher 1999; Dufwenberg and Kirchsteiger 2004.

TTT) to assess the importance of incentives, voluntary cooperation, crowding out and framing relative to the influence of repeated interaction.

3.2 Procedures and common features

We conducted 20 sessions at the University of St. Gallen with a total of 500 participants (first-year students of business, economics or law), which had been recruited from a large data base. In a typical session 28 subjects were present at the same time. In sessions with random matching (all treatments except TFT-Partner and TBT-Partner and TTT-Partner) we formed two independent matching groups of 14 subjects each. Subjects were not informed about the size of the matching groups but only that they will be randomly matched with another player present in the room.

After arrival at the lab, the subjects first had to read instructions (a translation can be found in the appendix). The instructions were the same for principal and agent. Subjects also had to answer a set of hypothetical questions to test their understanding of payoff calculations. The experiment did not start before all participants had answered all questions correctly. All participants could privately ask questions, which were answered individually. In the instruction the principal was called participant X and the agent was called participant Y. To facilitate understanding of the choice problem we indicated that one may think of X as a firm employer and of Y as an employee. Role assignment was random and fixed throught the session. It was explained that all decisions would be anonymous throught and after the experiment. A short summary of the instructions was read aloud by the experimenter to induce common knowledge of the game rules.

At the beginning of each session subjects were informed about the first block of ten games. After finishing the first block of games they were informed about the second block of ten games. And if there was a third block of games, they were informed about that block after completing the second block of games.

The experiments were computerized and conducted with the help of the experimental software “z-Tree” developed by Fischbacher (1999). The subjects were separated from each other by blinders and matched anonymously via a computer network. They never learned the identity of their opponent players. Each session lasted about two hours and the subject's average earnings were about 45 CHF (about 30 Euro).

4. Empirical Results

4.1 "Pay for performance" works

We first show that monetary incentives are effective in inducing effort. Figures 1A and 1B illustrate the relation between different levels of fine (bonus) and effort choices. Principal-agent theory predicts a positive correlation granted that contracts are incentive compatible. Our data are consistent with this prediction. Both figures depict the mean effort level (across all subjects and all periods of a single phase) for each value of fine and bonus, and for all accepted and incentive compatible contracts.[3] Figure 1A shows this for treatments BT and FT and Figure 1B for treatments TBT and TFT.

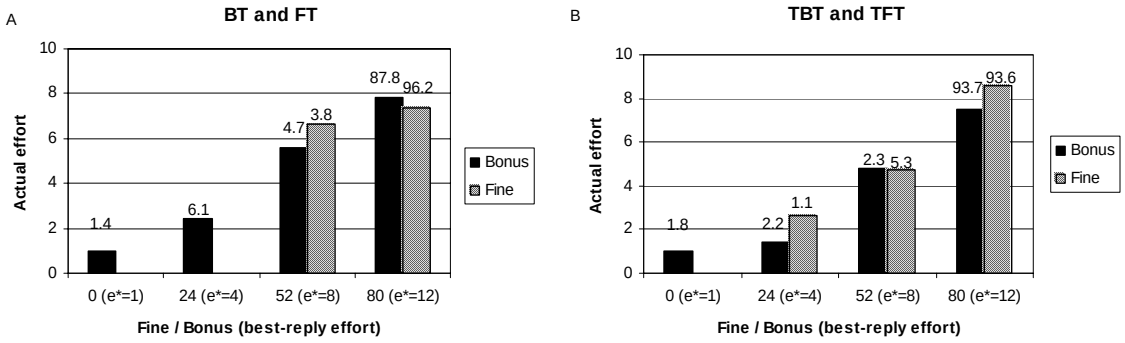


Figure 1: *Relation between incentives and effort in phase 1 of treatments BT and FT (panel A) and in phase 2 of the TFT and TBT experiments (panel B). Numbers above bars represent relative frequencies of observing the respective level of bonus and fine.*

The figures indicate a positive correlation. Spearman rank correlation coefficients reported in Table 3 are positive and significant for BT, TBT and TFT, while it is slightly negative but insignificant for FT. We conclude:

Result 1: *Performance incentives increase the agent's work effort.*

[3] All our analyses rely on accepted contracts unless explicitly noted otherwise. About 68,6 % of all incentive contracts (fine contracts and bonus contracts) are incentive compatible.

Treatment	Correlation Coefficient	# of obs	P-Value (two tailed)
BT	0.31	332	<0.001
FT	0.11	343	0.0467
TBT	0.16	338	0.0038
TFT	0.10	359	0.0543

Table 3: *Spearman rank correlations between bonus (fine) and effort for the different treatment.*

This result should be taken with some caution: First, the tests use individual decisions as units of analyses which is problematic since they may be correlated. Second, levels of bonus and fine smaller than 80 are scarce compared to levels of 80. The latter may be taken as indirect evidence for the effectiveness of monetary incentives. If the participants can design an incentive contract, they choose incentives maximally. This is reasonable if the participants anticipate a positive correlation between incentives and effort. We will investigate the influence of incentives in more detail below.

4.2 Trust induces voluntary cooperation

Figure 2 reports the relation between fixed wage and effort choices for treatment TTT. It is a bar chart showing mean effort conditional on different (classified) levels of fixed wage for each of the three phases of treatment TTT. The number above each bar is the relative frequency (in percent) of the respective fixed wage class. In all phases effort is positive, whereas it should be $e^* = 1$ according to the benchmark solution. Thus, there is a substantial amount of voluntary cooperation ($e - e^*$). Furthermore, effort and fixed wage are clearly positively correlated ($e'(w) > 0$). That is, more trust (higher fixed wage payments) induces larger effort. We refer to this as "reciprocity based voluntary cooperation" since the agent reciprocally responds with high or low effort depending on the granted wage.

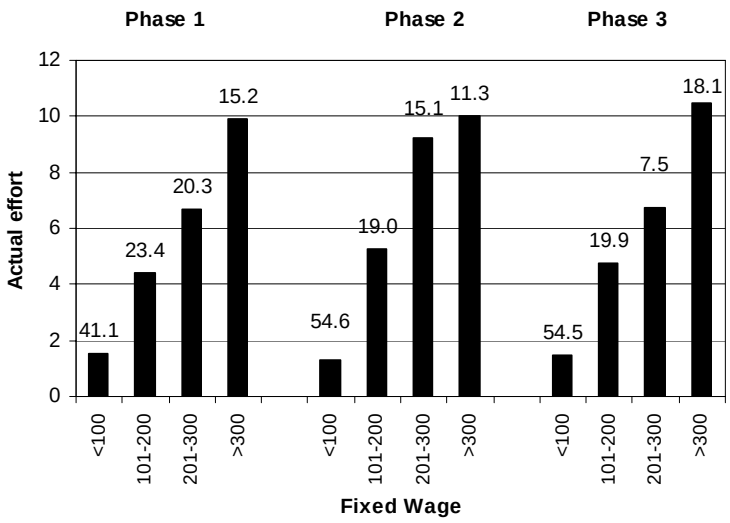


Figure 2: *Relation between fixed wage and effort in each phase of treatment TTT. Numbers above bars represent relative frequencies of observing the respective level of offered compensation.*

Table 4 reports the result of a Tobit regression analysis which explains the effort as a function of the offered fixed wages. We find a positive and highly significant relationship, and the correlation does not get weaker over time but is stronger in phases 2 and 3 than in phase 1.

	Dependent variable: effort			
	All data	Phase 1	Phase 2	Phase 3
Wage	0.043	0.04	0.05	0.041
	(0.003)**	(0.005)**	(0.006)**	(0.003)**
Constant	-6.32	-6.671	-6.709	-5.288
	(1.120)**	(1.594)**	(1.300)**	(0.878)**
Observations	881	316	284	281
Wald chi2 (2)	190.67**	69.97**	73.53**	149.33**

Robust standard errors in parentheses
* significant at 5%; ** significant at 1%

Table 4: *Tobit regressions of effort on fixed wages for all data and for the different phases in treatment TTT.*

Computing Spearman rank order correlations between fixed wage and effort separately for each matching group, results in 6 (out of 6) positive coefficients. Thus, our finding is also significant for a conservative statistical test (Binomial test, N = 6, p = 0.016, one-tailed). We conclude:

Result 2: *Voluntary cooperation exists. More trust induces more voluntary cooperation (reciprocity based voluntary cooperation). When participants gain experience this does not deteriorate but strengthen reciprocity-based voluntary cooperation.*

4.3 Short run crowding out

To determine the level of voluntary cooperation when subjects simultaneously experience incentive pay ("short run crowding out"), we look at the phase-1-data of FT and BT (Figures 3 A and B) as well as the phase-2-data of TFT and TBT (Figures 4 A and B).

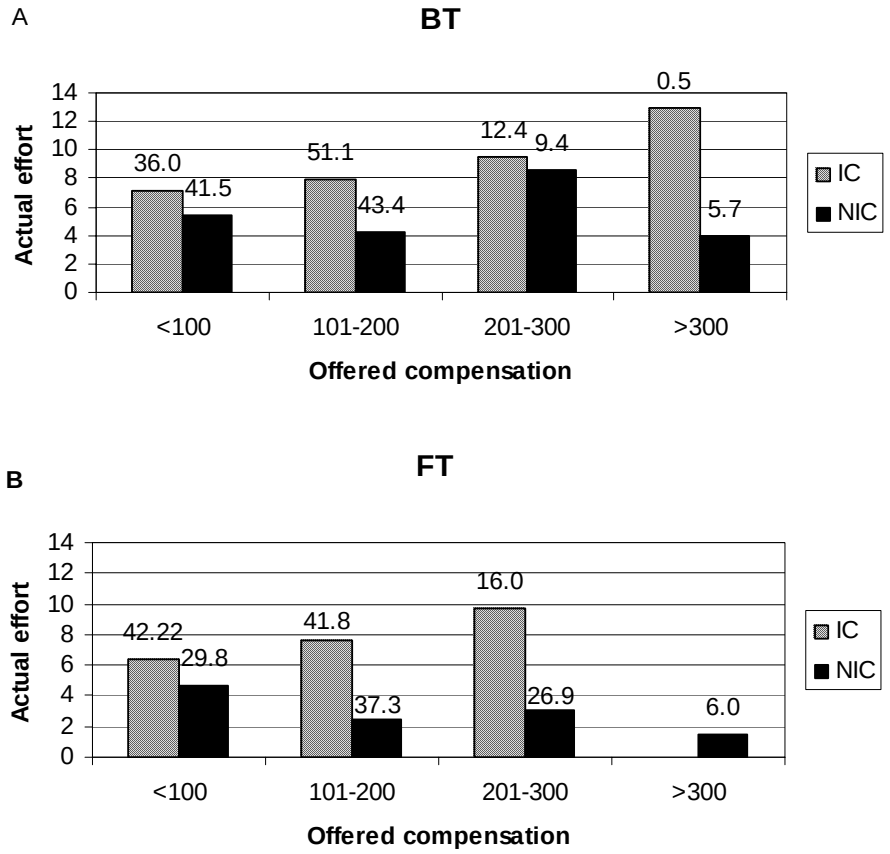


Figure 3: *Relation between offered compensation and effort in the first phase of treatment BT / FT. Numbers above bars represent relative frequencies of observing the respective level of offered compensation.*

The figures display bar charts of the effort chosen by agents for different levels of compensation offered by the principal and for different kinds of incentive contracts: bonus contracts versus fine contracts, as well as incentive compatible contracts versus non-incentive compatible contracts. For ease of comparison we use only those contracts that provide

maximal incentives (fine or bonus of 80).[4] Consequently, if the contract is incentive compatible, rational effort is 12 and voluntary cooperation calls for an even higher effort level. Contrary, Figures 3 and 4 show that effort is below 12 in most cases. Comparing effort under incentive contracts with effort under trust contracts (Figure 2), one finds that effort is about the same in both cases. That is, monetary incentives do not add to voluntary cooperation but rather crowd out voluntary cooperation. Incentives seem to substitute for trust.

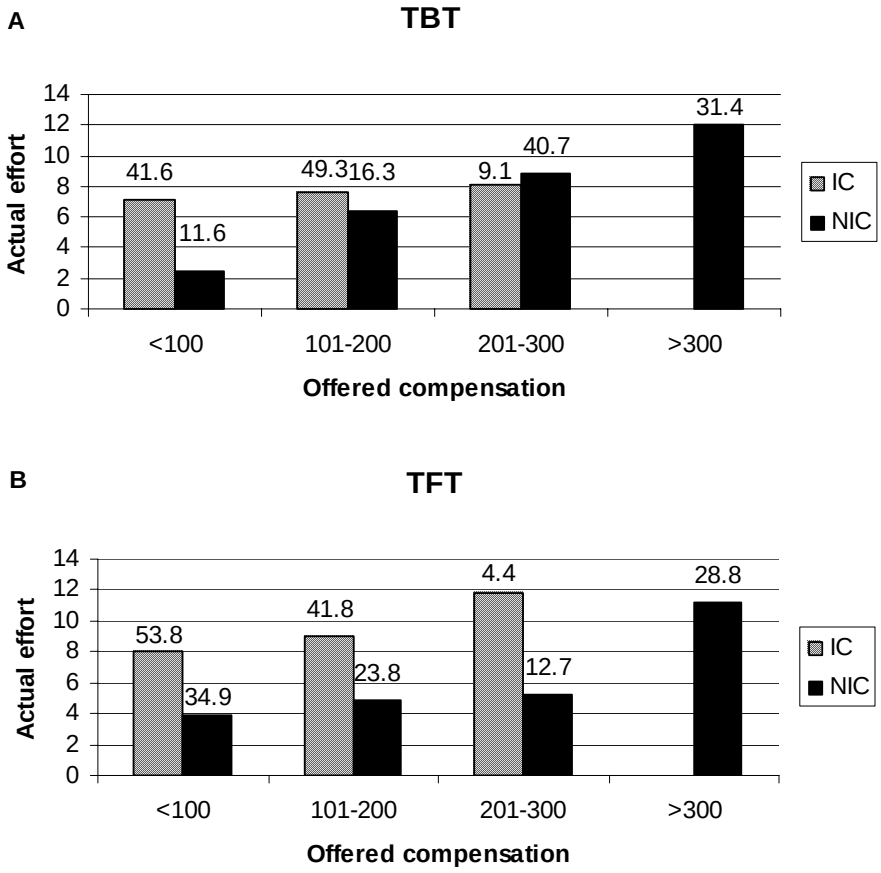


Figure 4: *Relation between offered compensation and effort in the second phase of treatment TBT / TFT. Numbers above bars represent relative frequencies of observing the respective level of offered compensation.*

If the contract is not incentive compatible, rational effort is 1. The situation is similar to a trust contract in the sense that effort above 1 can be attributed to voluntary cooperation. All Figures reveal voluntary cooperation when contracts are non-incentive compatible, but at a level that is smaller than in the TTT treatment (compare Figure 2). We conclude:

[4] 84,5 % of all incentive contracts specified a fine of 80, or respectively 80,7% a bonus of 80.

Result 3: *Experiencing incentive contracts has an immediate, short run crowding out effect on voluntary cooperation. Furthermore, if incentive contracts are not incentive compatible, effort is smaller than with trust contracts.*

4.4 Long run crowding out effect

Experiencing incentive contracts may have a lasting effect on subjects' willingness to voluntarily cooperate (long run crowding out). To investigate this we use the phase-2-data of treatments FT, BT and TTT; i.e., we look at behavior under conditions of trust contracts after previously experiencing incentives (FT, BT) or trust (TTT). If phase-2-effort in FT and BT is smaller than in TTT, this is evidence for long run crowding out.

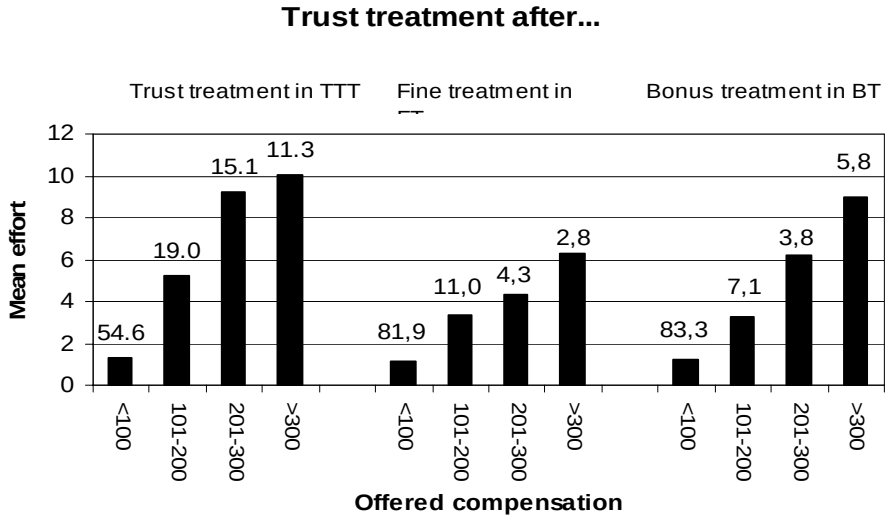


Figure 5: *Relation between fixed wage and effort in phase 2 of the BT, FT and the baseline treatment TTT. Numbers above bars represent relative frequencies of observing the respective level of wage*

It seems that effort for given levels of fixed wage is larger in TTT than in BT and FT. Furthermore effort seems to be larger in BT than in FT. A Kruskal-Wallis test supports our hypothesis that effort is lower after experiencing incentive pay (BT or FT) than in the TTT treatment (N = 14 groups, p = 0.067). Since this test does not control for possible differences in fixed wage across groups, we also run a Tobit regression analysis with effort as dependent variable and fixed wage as well as treatment dummies as independent variables. Estimation results are reported in Table 5.

	Effort
Fixed Wage	0.045
	(0.003)**
Dummy Bonus treatment	-1.949
	(1.258)
Dummy Fine treatment	-4.302
	(0.663)**
Constant	-5.916
	(1.078)**
Observations	683
Wald chi2 (3)	316.31**

Robust standard errors in parentheses
* significant at 5%; ** significant at 1%

Table 5: *Tobit regression of effort on wage and dummies for Bonus and Fine treatment in phase 2 of the BT and FT experiments.*

The treatment dummies for BT and FT are negative and confirm the result of the Kruskal-Wallis test. The difference between BT and FT indicates less effort (voluntary cooperation) for FT than for BT; i.e., framing of incentives matters. The difference is significant according to a Chi-Square test ($p = 0.021$). We conclude:

Result 4: *The data support the crowding out hypothesis: Voluntary cooperation is lower after experiencing incentive pay. Furthermore, crowding out is stronger in the Fine treatment than in the Bonus treatment (framing effect).*

4.5 Crowding out despite learning voluntary cooperation

A more subtle question regarding crowding out of voluntary cooperation is whether it occurs even when incentives are introduced after subjects have learned about the effectiveness of "trust and reciprocity". Specifically, we will compare effort choices in phase 1 and phase 3 for treatments TFT and TBT to investigate whether voluntary cooperation returns to the same level after an intermediate phase of incentive contracting. Treatment TTT will be used as a control treatment since there may be a general time trend in behavior.

We expect two effects here: First, similar to treatments FT and BT we expect a crowding out of voluntary cooperation due to monetary incentives. Second, we expect that subjects learn about the effectiveness of "trust and reciprocity" in phase 1. The latter effect works against the crowding out effect, such that crowding out may be less pronounced for treatments TFT and TBT than it was for FT and BT.

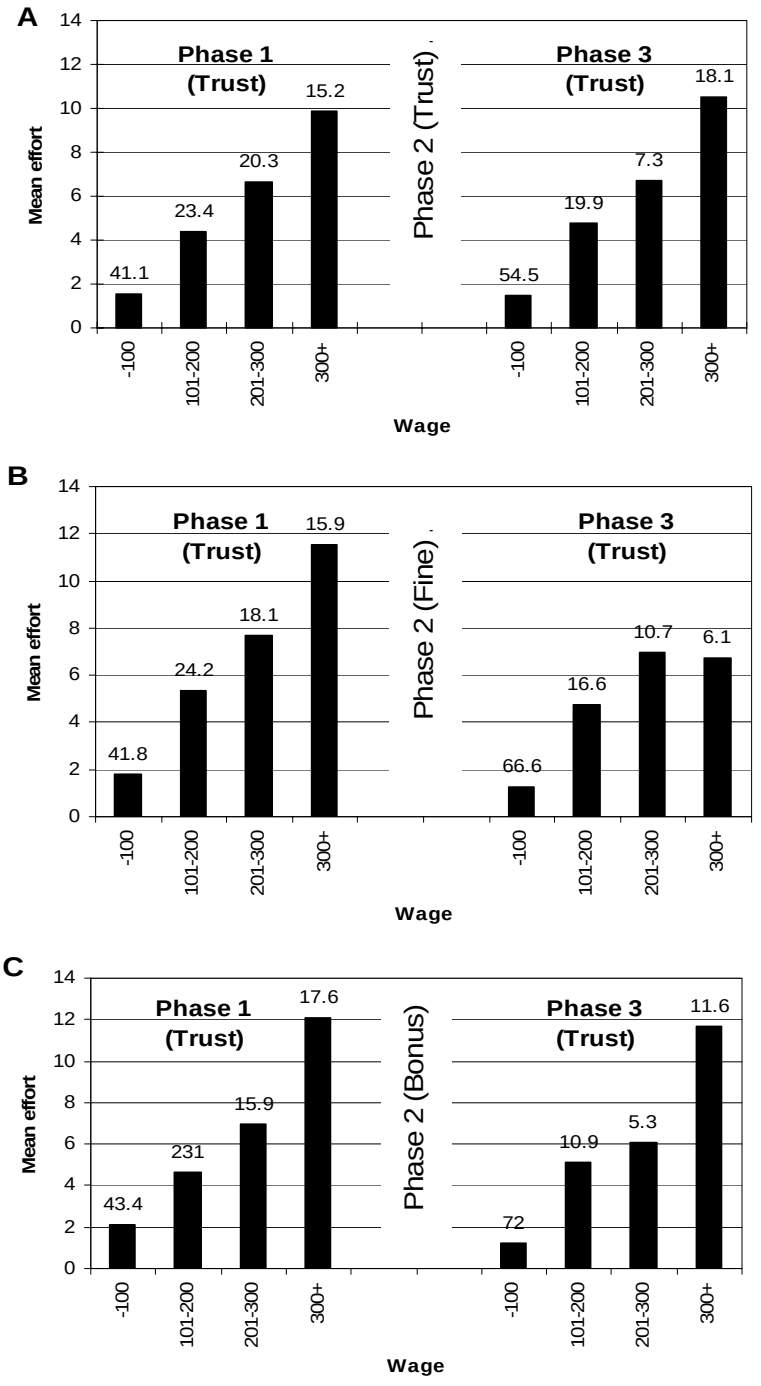


Figure 6: *The wage-effort relation in phase 1 compared to phase 3. Panel A: treatment TTT; Panel B: treatment TFT; Panel C: treatment TBT. Numbers above bars represent relative frequencies of observing the respective level of fixed wage.*

Figures 6A to 6C display the relation between effort and fixed wage in phase 1 versus phase 3 for treatments TTT (Fig. 6A), TFT (Fig. 6B) and TBT (Fig. 6C); i.e., in these phases all three treatments allow for fixed wage contracts only. While there is no reduction in effort in the control treatment TTT, there is a pronounced reduction in phase 3 compared to phase 1

for TBT and even more for TFT. Wilcoxon Matched-Pairs Signed Ranks tests based on matching groups as units of analyses indicate significant reductions for TBT ($p = 0.02$, $N = 6$) and TFT ($p = 0.02$, $N = 6$) but no significant difference for TTT ($p = 0.6$, $N = 6$). This confirms our prediction that crowding out of voluntary cooperation occurs even after subjects have learned about the effectiveness of "trust and reciprocity". Experiencing trust contracts beforehand does not preclude a lasting negative effect of monetary incentives. Voluntary cooperation does not immediately recover, when incentives are removed.

4.6 The impact of repeated interaction

So far we analyzed the data of random matching sessions. With random matching subjects are unable to build reputation or consider long-term strategies in general. It is known from other experiments that repeated interaction may increase cooperation. We can study repeated game effects by comparing treatments TFT-Partner, TBT-Partner and TTT-Partner with the respective random matching treatments TFT, TBT and TTT (Stranger treatments).

Figure 7 shows time-series of mean effort for different treatments. Effort is much higher for Partner treatments than for random matching treatments. The time-series are clearly separated according to matching conditions for all phases and all game types.

In treatment TTT-Partner effort is increasing over time within each phase and across phases, except for a pronounced end-game effect in each phase. In phase 3, effort is even close to the maximal and efficient effort level 20. Thus, under conditions of repeated interaction, long-term experience and trust – and without an intervening phase of incentive pay – voluntary cooperation almost reaches the collectively optimal level.

In treatments TBT-Partner and TFT-Partner effort does not reach similarly high levels as in TTT-Partner. Especially from phase 2 (incentives) to phase 3 (trust) there is no upward shift. This indicates crowding out of voluntary cooperation even under repeated game conditions. We have shown above that monetary incentives do work under Stranger conditions. Figure 7 displays this finding by clear upward shifts of the time-series' for TBT and TFT in phase 2. But no similar effect is shown for TBT-Partner and TFT-Partner. Note that in these treatments mean effort is larger than the maximal effort level that can be implemented by incentives ($e = 12$). So, incentives are largely ineffective when voluntary cooperation is already relatively high (due to repeated interaction).

Another impression one gets from Figure 7 is that in TBT and TFT learning drives toward minimal effort. Especially in phase 3, after experiencing incentive pay, effort levels decline.

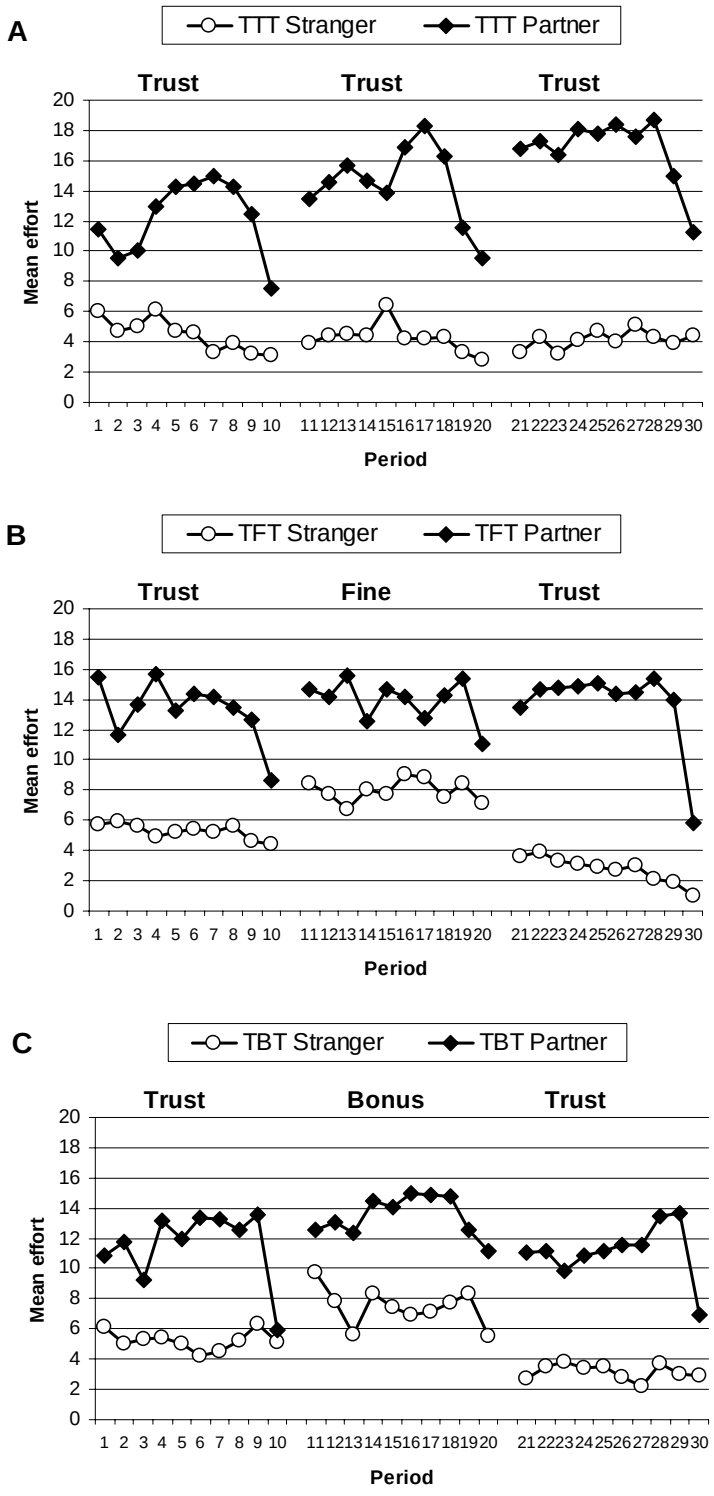


Figure 7: *Mean effort for periods 1-30 for different treatments. Panel A: TTT and TTT-Partner; Panel B: TFT and TFT-Partner; Panel C: TBT and TBT-Partner.*

4.7 Comparing effect sizes

While Figure 7 indicates the joint influence of monetary incentives, voluntary cooperation, framing and repeated interaction, we analyze this in more detail and compare relative effect sizes according to the regression reported in Table 6. It is a Tobit regression model with individual effort as dependent variable (with upper limit equal to 20 and lower limit equal to 1). We estimate the effects of fixed wages, incentives and crowding out separately for the Stranger condition and the Partner condition. The model and almost all partial effects are highly statistically significant.

We find a strong effect of repeated interaction (see the statistics on the variables "Repeated Game Dummy" and "Repeated Game End Effect Dummy"). Ceteris paribus effort is about 5.2 units higher with Partner matching than with Stranger matching. Paying a higher fixed wage does increase effort substantially both under Stranger conditions and under Partner conditions. We estimated this influence separately for the three phases and separately for TTT versus TBT, TFT in phase 2 and 3 to allow for interacting influences of time and treatment. Each of the 10 fixed wage coefficients is positive and highly significant, confirming our prediction of voluntary cooperation and the effectiveness of the "trust and reciprocity" mechanism. For instance, in phase 1 offering 100 money units more in fixed wage increases effort under Stranger (Partner) conditions by about 4.3 (4.7) units. Differences between pairs of fixed wage coefficients are insignificant in most cases. Nevertheless it is interesting to note that each coefficient in the Partner treatment is larger than the respective coefficient in the Stranger treatment; i.e., reciprocity is stronger under Partner conditions. The response to fixed wage payments is smallest in phase 2 of the Stranger treatment, when fixed wage is paid in addition to a bonus or a fine.

The influence of monetary incentives is captured by the variables "Maximal Bonus" and "Other Bonus" and, respectively, the variables for fines. "Maximal Bonus" is a dummy variable which is 1 if a bonus of 80 was offered to the agent, and it is 0 otherwise. "Other Bonus" is a dummy which is 1 if a bonus smaller than 80 was offered and which is 0 otherwise. Analogous definitions hold for the fine variables. Incentive effects are strongest for Stranger conditions and when maximal incentives are offered. In this treatment there is a minor framing effect indicated by the difference between "Maximal Bonus" and "Maximal Fine". But the difference is statistically insignificant. Under Partner conditions two of the four incentive effects are insignificant and the effect of "Maximal Bonus" is smaller than under Stranger conditions. Thus, in repeated games effort is less responsive to (short-run) incentives than under Stranger conditions. Maybe the most surprising incentive effect in the Partner

treatment is the negative effect of "Other Fine". Accordingly, a small fine is the worst incentive one may offer in repeated interaction.

	Coefficient	Robust Std. Err.	Z	P>|z|
Constant	-5.574154	0.6166901	-9.04	0.000
Repeated Game Dummy	5.174963	1.648113	3.14	0.002
Repeated Game Endeffect Dummy	-1.455428	0.516391	-2.82	0.005
Fixed Wage Effects Stranger				
Fixed Wage, Phase 1	0.0428338	0.0019543	21.92	0.000
Fixed Wage, Phase 2, TTT	0.0451899	0.0041816	10.81	0.000
Fixed Wage, Phase 2, TFT, TBT	0.0298034	0.0043056	6.92	0.000
Fixed Wage, Phase 3, TTT	0.0412252	0.0029006	14.21	0.000
Fixed Wage, Phase 3, TFT, TBT	0.0543764	0.003584	15.17	0.000
Incentive Effects Stranger				
Maximal Bonus	10.71716	1.143448	9.37	0.000
Other Bonus	5.073159	1.400207	3.62	0.000
Maximal Fine	10.06698	0.9791402	10.28	0.000
Other Fine	3.694406	1.30342	2.83	0.005
Long-Run Crowding Out Effects Stranger				
Crowding Out Bonus	-2.522937	1.195448	-2.11	0.035
Crowding Out Fine	-3.595018	1.144286	-3.14	0.002
Fixed Wage Effects Partner				
Fixed Wage, Phase 1	0.0470712	0.005532	8.51	0.000
Fixed Wage, Phase 2, TTT	0.0500328	0.0070754	7.07	0.000
Fixed Wage, Phase 2, TFT, TBT	0.0742324	0.0069151	10.73	0.000
Fixed Wage, Phase 3, TTT	0.0534859	0.0064762	8.26	0.000
Fixed Wage, Phase 3, TFT, TBT	0.0658101	0.0043333	15.19	0.000
Incentives Effects Partner				
Maximal Bonus	5.031759	2.036331	2.47	0.013
Other Bonus	-3.366104	2.345564	-1.44	0.151
Maximal Fine	-1.019057	2.872928	-0.35	0.723
Other Fine	-7.61217	2.925828	-2.60	0.009
Long-Run Crowding Out Effects Partner				
Crowding Out Bonus	-4.285159	1.821478	-2.35	0.019
Crowding Out Fine	-3.99699	1.900537	-2.10	0.035
Lnsigma	1.892575	0.0499354	37.90	0.000
Sigma	6.636438	0.3313931		

Table 6: *Tobit regression model*

Long-run crowding out effects are estimated separately for bonus and fine and for Stranger and Partner conditions. The dummy variable "Crowding Out Bonus" is 1 in phase 3 of the TBT (TBT-Partner) treatment and 0 otherwise. The other crowding out variables are defined accordingly. As predicted by the crowding out hypothesis each of the estimated coefficients is negative and statistically significant. For instance, crowding out reduces effort under Stranger conditions by about 2.5 (3.6) units in case of a bonus (fine). Qualitatively, crowding out is stronger in the Partner treatment than in the Stranger treatment, however, the differences are insignificant. Furthermore, framing has only a minor influence on crowding out. The differences between the coefficients for bonus and fine are insignificant.

Comparing effect sizes we find that under Stranger conditions incentives have a stronger effect than comparable increases in fixed wage. Nevertheless fixed wage payments induce substantial effort increases. Long-run crowding out effects are substantial as well. Thus, a policy of temporarily exposing agents to incentives has to trade of short-run increases and long-run decreases in effort.

Under repeated game conditions the effectiveness of bonuses and fines that are tied to period performance deteriorates, while the influence of fixed wages and crowding out become stronger. Furthermore, repeated interaction induces a strong upward shift in effort except for a counter acting but smaller end effect in the last few periods.

Compared to the effects of incentives, voluntary cooperation, crowding out and repeated interaction the influence of framing (bonus versus fine) is minor.

5. Summary and concluding remarks

In this study we analyzed the data of a comprehensive experimental study on a PA-game that allows for either incentive contracts or pure fixed wage contracts. It comprises 12560 games played by 500 participants. It investigates the influence of trust, monetary incentives, framing of incentives, learning, repeated interaction, and, more specifically, crowding out effects within a unified PA-framework. It is to our knowledge the largest laboratory study of this kind. We replicate several findings of previous studies, e.g., on the effectiveness of both, monetary incentives and trust, in inducing above minimal work effort. We find evidence for crowding out of voluntary cooperation through monetary incentives. Experiencing incentive pay reduces the effectives of trust contracts in future interactions. This effect does hold also when incentives are introduced after participants have experienced trust-and-reciprocity

before. Incentives seem to have a lasting negative effect on voluntary cooperation (irreversibility effect).

Our findings on crowding out and in particular the irreversibility effect suggest that playing around with pay systems is problematic. It may deteriorate future possibilities for fruitful cooperation.

Furthermore, an interesting result of our study is that it allows to compare the size and to evaluate the importance of several factors that influence behavior in PA-relationships. Repeated interaction seems to dominate other factors in increasing effort. Incentives and learning have positive effects, but are weaker than repeated play. Compared to these effects, the influence of framing (Bonus versus Fines) is minor.

References

Anderhub, Vital, Simon Gächter, and Manfred Königstein (2002): "Efficient Contracting and Fair Play in a Simple Principal-Agent Experiment." *Experimental Economics* 5(1), 5-25.

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

Dufwenberg, Martin and Georg Kirchsteiger (2004). A Theory of Sequential Reciprocity, *Games and Economic Behavior* 47, 268-298.

Falk, Armin, Simon Gächter, and Judit Kovács (1999): "Intrinsic Motivation and Extrinsic Incentives in a Repeated Game with Incomplete Contracts". *Journal of Economic Psychology* 20(3), 251-284.

Falk, Armin and Fehr, Ernst (2003): "Why labor market experiments?" *Labor Economics* 10, 399-406.

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

Fehr, Ernst, and Armin Falk (2002): "Psychological foundations of incentives." *European Economic Review* 46, 687-724.

Fehr, Ernst, and Simon Gächter (2000): "Fairness and Retaliation. The Economics of Reciprocity." *Journal of Economic Perspectives* 14(3), 159-181.

Fehr, Ernst, and Simon Gächter (2002): "Do incentives contracts undermine voluntary cooperation?" Working Paper No. 34, Institute for Empirical Research in Economics, University of Zurich.

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

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

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

Fischbacher, Urs (1999). "z-Tree, Toolbox for readymade economic experiments: Experimenter's manual." *Working paper* No. 21, Institute for Empirical Research in Economics, University of Zürich.

Frey, Bruno S., and Reto Jegen (2003): "Motivation Crowding Theory." Journal of Economic Surveys 15(5), 589-611.

Gächter, Simon and Ernst Fehr (2002): "Fairness in the Labour Market – A Survey of Experimental Results", Friedel Bolle, Marco Lehmann-Waffenschmidt (eds.): *Surveys in Experimental Economics. Bargaining, Cooperation and Election Markets*. Heidelberg: Physica-Verlag.

Gächter, Simon and Armin Falk (2002): "Reputation and Reciprocity – Consequences for the Labour Relation." *Scandinavian Journal of Economics* 104(1), 1-27.

Gneezy, Uri and Aldo Rustichini (2000a): "A Fine is a Price", *Journal of Legal Studies* XXIX, 1-17.

Appendix: Instructions

Here we document the instructions of the Trust treatment and the Fine treatment used in our TFT experiment. The instructions in the other treatments were adapted accordingly. The instructions were originally written in German.

The experiment in which you participate today is joint research with the Humboldt-University Berlin. It is financed by several science foundations.

During the experiment your income will be calculated in points. In the beginning you get a lump-sum of 1500 points. It is possible that some of your decisions lead to losses. You will have to finance them out of the gains from your other decisions, or, if necessary out of your lump-sum. **However, you can always make decisions that avoid any losses.**

The exchange rate of points into Swiss Francs is:

1 Point = 0.6 Rappen.

At the end of the experiment all points which you have earned will be summed up, exchanged into Swiss Francs and paid out to you in cash.

> **Please note that during the experiment communication is not allowed. If you have any questions, please raise your hand. We will answer your questions individually.**

Instructions

1. Introduction

In this experiment you will learn about a decision problem that involves two persons. The persons will be called participant X and participant Y. **All participants in this experiment are allocated into two groups: the group of the participants X and the group of the participants Y. After the experiment has started you can see on your computer's display whether you are participant X or participant Y.**

At the beginning you will be **randomly** matched with a participant of the other group. You will make your decisions at the computer. Your decisions will be transmitted via computer to the participant of the other group. This participant will only get informed about your decision. He will never learn about your name or your participant - number, i.e. your decisions remain anonymous.

2. An overview of the experiment

It may help you to understand the decision situation if you think about the following scenario. Participant X decides in the role of a "firm". The firm engages an employee (participant Y), who's work effort produces some period return. Y can choose his work effort freely in each period. Below we will explain what work effort means and how the period return comes about. A higher effort leads to a higher period return, but it also causes costs that Y has to bear.

Y's payment is determined in an **employment contract**. The employment contract consists of a **fixed wage** defined by X and a "desired effort". The fixed salary has to be paid by participant X to participant Y regardless of the period return.

Thus, each period consists of **three stages:**

1. Participant X proposes in accordance with the rules a employment contract including the fixed salary and the "**desired effort"**.
2. Participant Y decides to accept or reject the contract.
3. Y chooses his effective effort. The desired effort of X is not binding for Y.

Afterwards X and Y will be paid according to the rules. There are 10 periods. You will be matched in each period randomly with another person.

3. The experimental details

3.1 Employment contract: The proposal of participant X

At the beginning of each period a employment contract will be determined. For the employment contract the following rules hold:

The proposed employment contract consists of two components: a fixed wage and a desired effort. Participant X can design the contract according to the below-mentioned rules.

- The contract can contain a positive or a negative **fixed salary**. If the fixed salary is positive, this means that participant Y gets the wage form participant X, regardless of the period return. A negative fixed wage means that Y has to pay that amount to X, regardless of the period return.
- The proposed employment contract is only valid if participant Y accepts the labour contract. If Y accepts the contract, then Y decides about his effective work effort. X's desired work effort is not binding for Y. The participant Y can choose an effective work effort, which is higher, equal or lower than the desired effort.
- **For the contract design the following rules hold:**

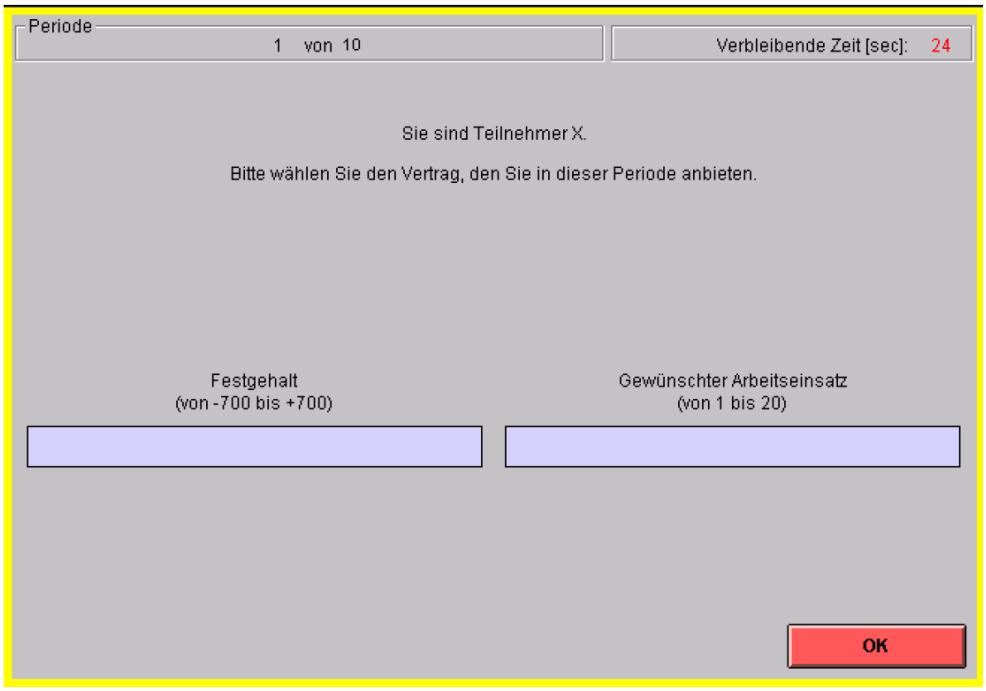
$$-700 \leq \text{fixed wage} \leq 700$$

$$1 \leq \text{desired work effort} \leq 20$$

All numbers must be integer.

In designing the contracts all combinations that are compatible with these rules are possible!

To make the rules clear to you, we depict the screen that will be shown to X at the beginning of period 1:



On this screen (as well as in all other screens in which you have to make a decision) you see the current period number on top left and the remaining time on top right. Participant X makes up his proposed employment contract on this screen.

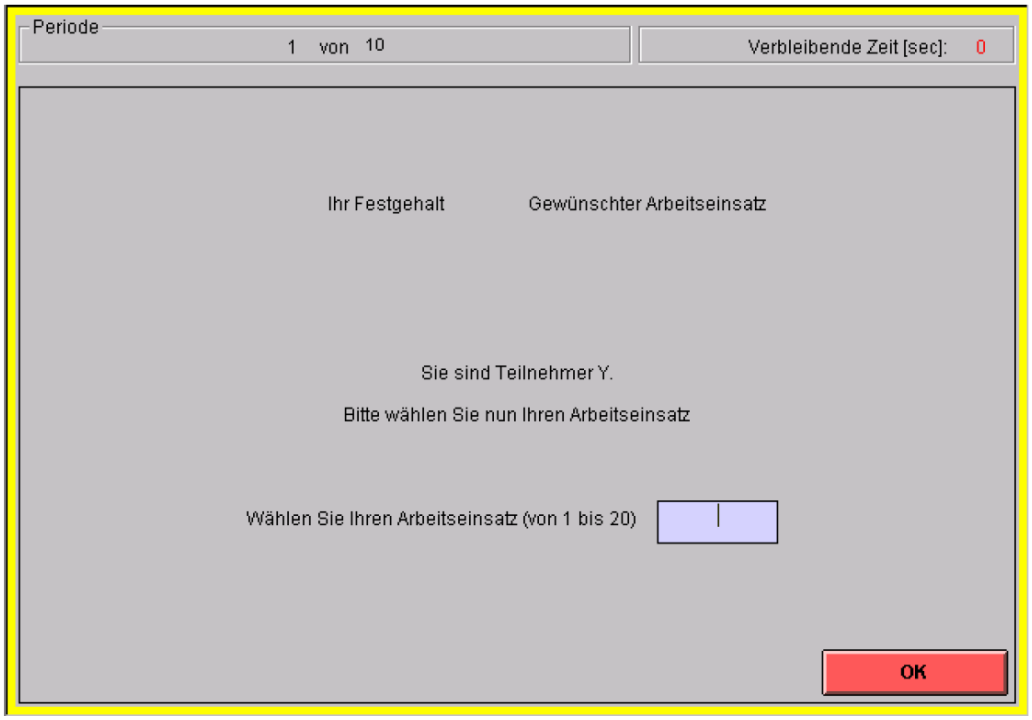
3.2 Employment contract: Acceptance of the contract

After participant Y has received the proposal of the contract, he decides whether he accepts or rejects the contact.

3.3 Work effort of participant Y

After the acceptance of the contract Y determines his **work effort**. The desired work effort stated by participant X in the contract is not binding for participant Y. The effort is expressed as a number. In the enclosed **table** all possible work effort (all integer numbers between 1 and 20) as well as the produced returns are given. The table also contains the **cost**s of the work that Y has to bear. The higher the work effort, the higher is the return, but also the costs of the work effort.

The screen of participant Y is shown below.



3.4 Period payoffs and end of period

After participant Y has entered his work effort into the computer, the period gains will be calculated and shown on the display. The following cases result for the calculation of the profits:

Profit of X:	**Profit of Y:**
Y rejects the contract:	
Zero	Zero
Y accepts the contract:	
Period return of the effective work effort – fixed wage	Fixed wage – cost of the effective work effort
Please note: For the profit only the **effective work effort** is relevant.	

After this-screen the period is finished and the next one starts. On the whole there are 10 periods.

Effort, period return of the work effort and costs of the work effort for Y:

Work effort :	**Period return** of the work effort	**costs** of the work effort for **Y**
1	35	0
2	70	7
3	105	14
4	140	21
5	175	28
6	210	35
7	245	42
8	280	49
9	315	56
10	350	63
11	385	70
12	420	77
13	455	84
14	490	91
15	525	98
16	560	105
17	595	112
18	630	119
19	665	126
20	700	133

Profit of Y: Fixed wage – costs of the effective work effort

Profit of X: Period return of the effective work effort – fixed wage

Profit of Y and X by rejection of the contract of Y: Zero

Only the effective work effort is relevant for the calculation of the profits!

Information about the new experiment

The new experiment also consists of 10 periods. In this experiment, too, you are matched randomly with another person in each period. Again you don't get to know the others person's identity. As before all decisions are anonymous. The only change compared to the first experiment is given by an additional parameter that X can offer in the proposed employment contract. In addition to the fixed salary and the desired effort participant X determines a potential deduction from wages. This is **the only difference** to the previous experiment. The potential deduction from wages is enforced if Y chooses a work effort that is lower than the desired effort of X. If Y choose a work effort which is higher or equal than the desired effort than the deduction from wages isn't enforced. There are four possible levels of potential deduction from wages. The deduction from wages can be either **0** or **24** or **52** or **80. The deduction from wages will only be enforced if the effective effort is lower than the desired effort!**

For the contract design the following rules hold:

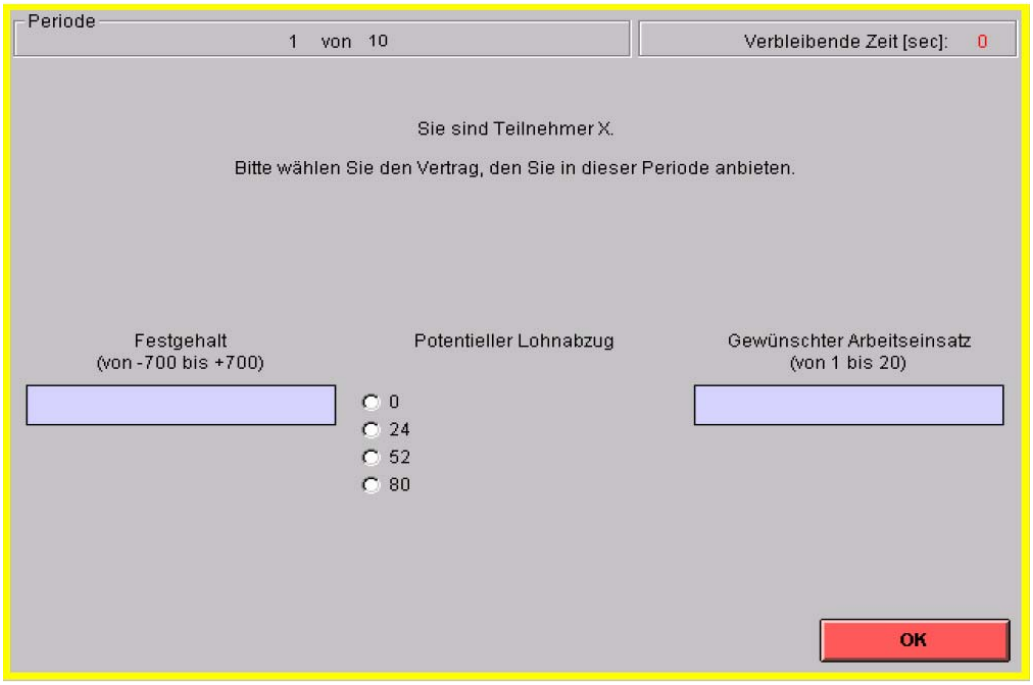
$-700 \leq$ fixed wage ≤ 700

Potential deduction from wages: *either* 0 *or* 24 *or* 52 *or* 80

$1 \leq$ desired work effort ≤ 20

All numbers must be integer. In designing the contract all combinations that are compatible with these rules are possible!

To make the rules clear to you, we depict the screen that will be shown to X at the beginning of the first period.



The profits are calculated as follows:

Profit of X:	**Profit of Y:**
Y rejects the contract:	
Zero	Zero
The effective work effort is higher or equal than the desired work effort.	
Period return of the effective work effort – fixed wage	Fixed wage – costs of the effective work effort
The effective work effort is lower than the desired work effort:	
Period return of the effective work effort – fixed wage + deduction from wages	Fixed wage –deduction from wage – costs of the effective work effort

The process of this experiment is identical with the previous experiment.