Performance Incentives and the Dynamics of Voluntary Cooperation[†]

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

12 February 2009

PRELIMINARY VERSION – comments welcome!

Abstract

We investigate how explicit performance incentives in incomplete employment contracts interact with agents' voluntary cooperation in one-shot and repeated gift-exchange experiments. If contracts are incentive compatible, agents choose their best-reply effort and there is no voluntary cooperation. By contrast, there is substantial voluntary cooperation if the contract is not incentive compatible. Experiencing incentive contracts reduces voluntary cooperation even after incentives are abolished. Incentive contracts have no lasting negative effects on voluntary cooperation if agents experience voluntary cooperation before being exposed to incentive contracts or if there are implicit incentives in repeated interactions. Implicit incentives increase voluntary cooperation strongly and make explicit incentives unnecessary.

Key words: principal-agent games; gift exchange experiments; incomplete contracts, explicit incentives; implicit incentives; repeated games.

JEL-Codes: M5, C7, C9

[†] This paper is part of the MacArthur Foundation Network on Economic Environments and the Evolution of Individual Preferences and Social Norms. Support from the EU-TMR Research Network ENDEAR (FMRX-CT98-0238) and from the Grundlagenforschungsfonds at the University of St. Gallen is gratefully acknowledged. We benefited from excellent research assistance by Christian Thöni and Eva Poen and from helpful comments in numerous workshops and seminars.

[*] University of Nottingham, School of Economics, Sir Clive Granger Building, University Park, Nottingham NG7 2RD, United Kingdom. Simon Gächter is also affiliated with CESifo and IZA. E-mail: simon.gaechter@nottingham.ac.uk.

[**] University College London, Institute of Child Health, Behavioural & Brain Sciences Unit, 30 Guilford Street, London WC1N 1EH, United Kingdom. E-mail: esther.kessler@ich.ucl.ac.uk.

[***] Universität Erfurt and IZA. Professur für Angewandte Mikroökonomie, Nordhäuser Str. 63, D-99089 Erfurt. E-mail: manfred.koenigstein@uni-erfurt.de.

1. INTRODUCTION

The employment relation is an incomplete contract, which leaves many important aspects, in particular the content of work effort, unregulated and therefore non-enforceable by third parties. Contractual incompleteness gives opportunistic agents an incentive to shirk and therefore leads to an inefficiently low surplus. Thus, many scholars have argued that voluntary cooperation ('gift exchange', 'loyalty', 'good will', 'initiative', 'organizational citizenship') is necessary to ensure efficiency (Akerlof (1982); Williamson (1985); Organ (1988); Simon (1997); Bewley (1999); Bowles (2007)).

In this paper we investigate how explicit and implicit performance incentives affect voluntary cooperation in simple one-shot and repeated games that model an employment relationship with incomplete contracts. Arguably, most employment relationships contain some explicit or implicit performance incentives. Explicit incentives are incentives in which compensation depends on observed output in an ex ante announced way; implicit incentives arise through strategic incentives to behave cooperatively in a repeated relationship. If voluntary cooperation is crucial for ensuring the smooth functioning of employment relationships it is important to understand how explicit and implicit incentives interact with voluntary cooperation.

The reason why there might be an interaction effect between performance incentives and voluntary cooperation is that voluntary cooperation and incentives operate on different psychological mechanisms. The psychological sources of voluntary cooperation are non-selfish motives like fairness and equity concerns (Adams (1965); Akerlof (1982); Bolton and Ockenfels (2000); Fehr and Schmidt (1999)), reciprocity (Rabin (1993); Dufwenberg and Kirchsteiger (2004); Falk and Fischbacher (2006)); let-down aversion (Dufwenberg and Gneezy (2000); Charness and Dufwenberg (2006)), or loyalty and goodwill (Simon (1991), Bewley (1999)). By contrast, performance incentives give the agent a selfish motive to perform. These selfish motives may or may not point into the same direction as the non-selfish motives. For instance, explicit incentives may transform a good-will based relationship into a monetized relationship which is governed by selfish cost-benefit considerations rather than good will, reciprocity and other intrinsic motivations (Gneezy and Rustichini (2000); Frey and Jegen (2001)). Explicit incentive contracts may also signal distrust and introduce an element of control that might undermine voluntary cooperation (Falk and Kosfeld (2006)). By contrast, implicit incentives may support voluntary cooperation because the strategic incentives in a repeated relationship do not contradict other-regarding motivations.

We approach the question how incentives and voluntary cooperation interact by means of laboratory experiments in the framework of gift exchange games[1]. We chose laboratory experiments for two reasons: (i) only the laboratory allows for the comprehensive investigation of all interaction effects we are interested in and (ii) controlling for selfish incentives, which will be crucial for our approach, is hardly feasible in the field. We chose gift exchange experiments because they have proved useful tools for investigating voluntary cooperation in experimental employment relationships[2].

The gift exchange game is a version of a sequential prisoner's dilemma game. Specifically, it is a two-player game in which a principal ('an employer') offers a fixed wage to an agent ('an employee'). The agent can accept or reject the fixed wage offer. If the agent accepts, he or she chooses an output level. Output is costly for the agent and beneficial for the principal. In the standard version of the gift exchange game parameters are such that efficiency calls for the maximal output whereas selfish incentives induce the agent to provide the minimal output irrespective of the wage he or she has accepted (that is, there is no voluntary cooperation). The results of numerous experiments have refuted this prediction and have demonstrated the relevance of voluntary cooperation as wages and output are positively correlated even in one-shot games.[3] We will replicate this finding in a version of the gift-exchange game that we call 'Trust game'. The results from the Trust games will provide the necessary benchmark for the comparisons we are mainly interested in.

Our main interest is in understanding how explicit and implicit incentives influence voluntary cooperation. For this reason we run our games both as sequences of one-shot games (with randomly changing pairs of principal and agent) and as repeated games (with fixed pairs). In the one-shot games only explicit incentives are feasible, whereas in the repeated games explicit and implicit incentives might be operative.

The *explicit* incentives take a very simple form in our experiments: The principal offers a contract that specifies a desired output level and two compensation levels, one for the case in which the agent delivers at least the contractually pre-specified output and a second (lower) compensation level if the agent's output falls short of the pre-specified output. We design the set of feasible contracts such that the maximally enforceable output (by means of incentive compatible contracts if agents are selfishly rational) is substantially less than the efficient

[1] See Falk and Fehr (2003) for a general methodological discussion of experiments in labor economics and Fehr and Falk (2002) for a discussion on the psychology of incentive provision.
[2] See, for instance, Fehr, Kirchsteiger and Riedl (1993); Fehr, Gächter and Kirchsteiger (1997); Fehr, Kirchler, Weichbold and Gächter (1998); Gächter and Falk (2002); Charness (2004); Brandts and Charness (2004).
[3] For an overview of laboratory gift exchange experiments see Fehr and Gächter (2000). Evidence on gift exchange is not confined to the laboratory. See Gneezy and List (2006) and Falk (in print) for recent field experiments on gift exchange.

output level. Thus, there is room for efficiency-enhancing voluntary cooperation beyond the maximally enforceable output level. In general, we define voluntary cooperation as any output the agent puts forward that exceeds his or her selfishly-rational (best-reply) output level. Since given previous evidence we expect to see substantial voluntary cooperation in the Trust games, we can compare voluntary cooperation under incentive contracts and under Trust games.

In the one-shot games our design also has the feature that contracts that are not incentive compatible induce the same best-reply output levels as pure fixed wage contracts (Trust contracts). There is an interaction effect of incentives and voluntary cooperation if the output levels under non-incentive compatible Incentive contracts are different than under Trust contracts, holding the offered compensation constant. There are three logically possible interaction effects: output is higher, lower or equal than under comparable Trust contracts. We borrow terminology from psychological motivation crowding theory (Frey and Jegen (2001)) and speak of 'crowding out' ('crowding in') if, holding the offered compensation constant, output is reduced (increased) due to the experience of performance incentives. We expect that in the one-shot games explicit incentives will lead to a crowding out of voluntary cooperation because explicit incentives appeal to the agents' selfishly-rational motivations, whereas Trust contracts appeal to the agents' other-regarding motivations (see also Fehr and Gächter (2002)). In the repeated games we expect a crowding in of voluntary cooperation because the strategic incentives give all types of players, the self-regarding and the other-regarding ones, an incentive to cooperate (see also Gächter and Falk (2002)).

We study two forms of 'crowding effects': short-run crowding and long-run crowding. *Short-run crowding* occurs simultaneously to experiencing explicit performance incentives, whereas *long-run crowding* occurs *after* explicit incentives have been abolished. Understanding short-run crowding is important because it informs us about the extent of voluntary cooperation one can expect in the presence of explicit and/or implicit performance incentives. Long-run crowding effects are interesting because they tell us about the long-run impact of explicit incentives on voluntary cooperation even beyond the point in time in which explicit incentives have been operative (Gneezy and Rustichini (2000)).

We will present a comprehensive design to evaluate possible interaction effects of incentives with voluntary cooperation in one comparable design that also allows us to gauge which effects are of first-order importance and which ones are of second-order importance.

To our knowledge this has not been done before[4]. Specifically, after having established the extent of voluntary cooperation in pure Trust games we will (i) assess the power of limited explicit performance incentives to influence output and the role of framing of incentives (we will compare 'fines' versus 'bonuses' as incentives); (ii) estimate the short- and long-run crowding effects of explicit incentives in one-shot games; (iii) gauge the role of experience of Trust contracts before being exposed to Incentive contracts for short- and long-run crowding effects, and (iv) measure the power of implicit incentives and the interplay with explicit incentives in repeated interactions.

In our benchmark experiments we find substantial voluntary cooperation in one-shot Trust games. In particular, we find a positive wage-output relationship that is stable over the thirty rounds of our benchmark experiments. Our main new results are as follows:

1. Explicit performance incentives work almost exactly as standard economic theory predicts provided the contract is incentive compatible. The framing of incentives is unimportant. Higher incentives lead to higher output. There is virtually no voluntary cooperation if the contract is incentive compatible. By contrast, there is substantial voluntary cooperation if the contract is not incentive compatible.
2. We find no evidence for a short-run crowding out effect but a substantial long-run crowding out effect if people are inexperienced with Trust contracts before being exposed to Incentive contracts.
3. The experience of Trust contracts before being exposed to Incentive contracts eliminates also the long-run crowding out effect.
4. The implicit incentives are very strong in our setup and lead to a strong crowding *in* of voluntary cooperation. The explicit incentives become unimportant as the implicit incentives are strong enough to make the use of explicit incentives unnecessary.

In this paper we focus on the behavior of agents because we are interested in how incentives affect voluntary cooperation by agents. Optimal contract design by principals, given the behavioral effects we will observe in this paper, is left for future research. Given its focus on the behavior of agents, our paper contributes to an emerging economic literature on the psychology of incentives (Fehr and Falk (2002); Gneezy and Rustichini (2000); Frey and Jegen (2001); Falk and Kosfeld (2006); Fehr, Klein and Schmidt (2007)). It is also related to

[4] Aspects of our comprehensive design have been studied before. Gächter and Falk (2002) compare one-shot and repeated gift exchange games. Fehr and Gächter (2002) investigate a version of short-run crowding out of voluntary cooperation by incentive contracts. However, the behavioral effects of these papers are hardly comparable because the designs differ in too many ways. To our knowledge, no study has looked at long-run crowding effects and the role of experience and implicit incentives for crowding effects of voluntary cooperation.

a psychological literature on the "crowding out of intrinsic motivation" (e.g., Deci, Koestner and Ryan (1999)). Our paper differs from this literature in two important ways. First, the psychological literature focuses on the (detrimental) impact of intrinsic motivation for performing a particular task, whereas we are interested in voluntary cooperation. Second, our comprehensive design permits conclusions about efficiency implications, which is out of focus in the psychological literature and allows us to draw conclusions about first and second order effects.

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 gift exchange 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 to reward low or high output, respectively.

We first describe the rules and then derive a benchmark solution. The benchmark solution characterizes behavior of rational and money-maximizing players. Specifically, we determine the paths of subgame perfect equilibria of the games. Later on we will compare these benchmark predictions with alternative predictions and actual behavior.

2.1 *The games*

We summarize the games and parameters in Table 1. Each game consists of three stages. The principal first designs a contract which is offered to the agent. In the *Trust game* the contract comprises two components, a fixed wage w and a desired effort e^d (effort can also be interpreted as output in our games). The contract offer has to obey the following restrictions: $1 \leq e^d \leq 20$ and $-700 \leq w \leq 700$. 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 can accept or reject the contract. If he rejects the contract, the game ends and both the principal and the agent earn nothing. If the agent accepts the contract, he chooses effort (or output) e in integers (where $1 \leq e \leq 20$). That is, the third stage is entered only if the agent has accepted the offered contract. Again, the agent is not restricted by e^d in his choice of e. This feature of the design reflects contractual incompleteness because e^d is not enforceable. The stage game ends after effort has been chosen.

In the Trust game the principal's profit is $35e - w$ and the agent's profit is $w - c(e)$ in case of acceptance. Since the payment w can not be conditioned upon effort, we refer to this game as the 'Trust game'. 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.

TABLE 1
GAMES AND PARAMETERS

	Trust game	Fine game	Bonus game
Wage	$w \in [-700, 700]$	$w \in [-700, 700]$	$w \in [-700, 700]$
Desired effort (=output)	$e^d \in [1,20]$	$e^d \in [1,20]$	$e^d \in [1,20]$
Performance incentive		$f \in \{0, 24, 52,80\}$	$b \in \{0, 24, 52,80\}$
Agent's payoff	$w - c(e)$	$w - c(e)$ if $e \geq e^d$ $w - c(e)$-f if $e < e^d$	$w - c(e) + b$ if $e \geq e^d$ $w - c(e)$ if $e < e^d$
Principal's payoff	$35e - w$	$35e - w$ if $e \geq e^d$ $35e - w + f$ if $e < e^d$	$35e - w - b$ if $e \geq e^d$ $35e - w$ if $e < e^d$

Effort cost: $c(e) = 7e - 7$
Payoff if contract rejected: 0 for both

In the *Fine game* the principal may punish the agent if effort falls short of e^d. The offered contract in the Fine game consists of w, e^d and f, where f represents a fine (it can be interpreted as 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 principal's payoff. If $e \geq e^d$, the fine is not imposed.

In the *Bonus game* the principal has the opportunity to reward the agent if the chosen effort e is equal to or larger than the desired effort e^d. Thus, the offered contract features w, e^d and b, where b is a bonus (a wage increase) with $b \in \{0, 24, 52, 80\}$[5]. If $e \geq e^d$, the bonus is added to the agent's payoff and subtracted from the principal's payoff. If $e < e^d$, the bonus is not due. Under standard assumptions the Fine game and the Bonus game have exactly the same incentive effects. Thus the comparison of effort levels in the experiments will inform us about framing effects in incentive provision. We include the comparison of Fine and Bonus contracts into our design because there are theoretical and empirical arguments that frames might influence beliefs and beliefs influence motivations (see Dufwenberg, Gächter and Hennig-Schmidt (2006)). Thus, Fine and Bonus contracts might have different behavioral effects.

[5] In our setup the bonus (as well as the fine) is enforceable. This is in contrast to Fehr, Klein and Schmidt (2007), where paying the bonus is at the discretion of the principal after the agent has chosen the effort.

2.2 *Benchmark solutions*

Trust game: Since providing effort is costly and payment does not depend on effort, it is optimal for an agent to choose minimal effort $e = e^{min} = 1$. Anticipating the minimal effort choice, 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 $e = e^{max} = 20$.

Fine game and Bonus game: In choosing effort the agent has to consider two alternatives, $e = e^d$ or $e = 1$. All other effort choices are clearly suboptimal. A higher effort is suboptimal since it causes higher cost without increasing payment. A rational and selfish agent will choose the desired effort only if the fine or the bonus is higher than the cost of the desired effort. In case of $e < e^d$, the agent either has to pay the fine or loses the bonus, and this is independent of the exact value of e. So, conditional on $e < e^d$ minimal effort $e = 1$ is best. The optimal effort level is defined as:

$$e^* = \begin{cases} e^d \text{ if } w - c(e^d) \geq w - f - c(1) \;\Leftrightarrow\; f \geq c(e^d) \text{ or } w + b - c(e^d) \geq w - c(1) \Leftrightarrow b \geq c(e^d); \\ 1 \quad \text{otherwise.} \end{cases} \tag{1}$$

Thus, the agent will perform at the desired effort level if the fine or bonus is larger than the costs of contractual compliance. Notice that the best reply efforts are the same in the Fine game and the Bonus game; any behavioral difference for a given contract is therefore due to a framing effect.

The agent's best reply function (1) is the incentive compatibility constraint for the principal's contract design problem. For each level of fine or bonus there exists a maximal level of effort that satisfies $f, b \geq c(e^d)$. The maximal enforceable effort is 12. Before choosing effort the agent has to accept an offered contract in the first place. This requires that the following participation constraints are satisfied ((2) in the Fine game and (3) in the Bonus game).

$$\begin{aligned} &w - c(e^d) \geq 0 && \text{if } e^* = e^d \text{ or} \\ &w - f - c(1) \geq 0 && \text{if } e^* = 1, \end{aligned} \tag{2}$$

$$\begin{aligned} &w + b - c(e^d) \geq 0 && \text{if } e^* = e^d \text{ or} \\ &w - c(1) \geq 0 && \text{if } e^* = 1, \end{aligned} \tag{3}$$

The principal chooses w, f or b and e^d in order to maximize his profit subject to (1) and (2) or (1) and (3). One can see easily that it is optimal for the principal to set w such that (2) or (3) holds with equality, and furthermore that the solution to the principal's problem is f, b = 80, e^d = 12 and $w_f = c(12) = 77$ or $w_b = b - c(12) = -3$ (where w_f (w_b) denotes the wage in the Fine (Bonus) game).. Accordingly, the agent will accept the contract and will indeed choose e = 12 as desired by the principal. The solution is more efficient as the solution without incentives but it does not generate the maximal surplus. It induces a rather asymmetric distribution of payoffs: 343 money units for the principal and 0 money units for the agent.

3. RESEARCH QUESTIONS, EXPERIMENTAL DESIGN AND PROCEDURES

3.1 *Research Questions and Experimental Design*

In the experiment the participants played multiple rounds of the above games. Specifically, they played two, or respectively three, blocks ('phases') 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. Table 2 lists our research questions and all respective treatments. For several of our empirical analyses treatment TTT will be the baseline against which we compare other treatments. In all three sets of experiments we compare Fine and Bonus games to gauge the relevance of framing effects for incentive provision and crowding effects.

Table 2 also indicates the respective matching procedure. In most sessions subjects (principal and agent) were matched randomly across games. In some sessions subjects were matched with the same player during the whole experiment ('Partner' matching).

We conducted three sets of experiments that correspond to the three main purposes of our experiments. The first set of experiments aims at measuring (i) the impact of explicit incentives on performance; and (ii) the short- and long-run crowding effects of voluntary cooperation that are neither confounded with strategic incentives, nor with prior experience of Trust contracts before being exposed to incentive contracts. Therefore, subjects played one-shot experiments in two phases of ten periods each. In the first phase of principals could design either Fine or Bonus contracts, where in the second phase only Trust contracts were feasible. We refer to these treatments as FT or BT treatments.

The second set of experiments investigates the relevance of prior experience of Trust contracts before being exposed to Incentive contracts. All subjects play three phases of ten one-shot games with random matching. The first and the last phase consist of Trust contracts, whereas the second phase allows for either Fine or Bonus contracts. We refer to these experiments are TFT or TBT treatments. These experiments are interesting for two reasons. First, we can investigate the impact of experience of Trust contracts on the robustness of short- and long-run crowding out. Second, it is also of practical relevance to understand what happens to voluntary cooperation if Incentive contracts are introduced into an existing Trust environment.

The final set of experiments allows for both explicit and implicit incentives. Participants play either the TTT, the TFT or the TBT experiments in fixed pairs ('Partner' matching) for three phases of ten periods each. We refer to these experiments as TTT-R, TFT-R and TBT-R, where the suffix 'R' stands for repeated games. Although subjects are aware that they play finitely repeated games there are theoretical and empirical reasons why there are implicit (that is strategic) incentives to cooperate: First, if selfishness and rationality are not common knowledge cooperation can be sequentially rational (Kreps, Milgrom, Roberts and Wilson (1982)). Second, boundedly rational play can also lead to cooperation (Selten and Stoecker (1986)). Third, previous experimental evidence suggests that cooperation in the finitely repeated gift exchange game is higher than in one-shot games (Gächter and Falk (2002)).

TABLE 2

MAIN RESEARCH QUESTIONS, EXPERIMENTAL DESIGN, NUMBER OF SUBJECTS AND NUMBER OF MATCHING GROUPS

A. *Establishing voluntary cooperation and the effects performance incentives: Are there short-run and long-run crowding out effects? Does the framing of incentives matter?*					
Treatment label	Phase 1 (Period 1-10)	Phase 2 (Period 11-20)	Phase 3 (Period 21-30)	# Subjects	# independent matching Groups
TTT	Trust	Trust	Trust	78	6
FT	Fine	Trust	-	80	6
BT	Bonus	Trust	-	78	6
B. *Introducing incentives after experiencing trust contracts: What is the impact of performance incentives on short-run and long-run crowding out effects? Does the framing of incentives matter?*					
TFT	Trust	Fine	Trust	86	6
TBT	Trust	Bonus	Trust	84	6
C. *What is the impact of a repeated interaction on voluntary cooperation, performance incentives, and short-run and long-run crowding out effects? Does the framing of incentives matter?*					
TTT-R	Trust	Trust	Trust	24	12
TFT-R	Trust	Fine	Trust	36	18
TBT-R	Trust	Bonus	Trust	34	17

The main benchmark is treatment TTT. In these experiments subjects play three phases where each phase comprises ten one-shot Trust games played in randomly matched pairs. Based on previous experiments on the gift exchange game we predict that the fixed wage will be positively correlated with effort. If effort is higher than predicted according to the benchmark solution (i.e., $e > e^*$), we refer to this as 'voluntary cooperation'. Thus, contrary to the benchmark solution, we predict a positive amount of voluntary cooperation in the games of treatment TTT. We are now in a position to give a precise definition of crowding effects.

Definition 1: A *short-run crowding effect* occurs in phase 1 (phase 2) of treatment FT (TFT) if $(\overline{e}^{Trust} - \overline{e}^{Fine}_{NIC})|_{w,\hat{e},f} \neq 0$, and in BT (TBT) if $(\overline{e}^{Trust} - \overline{e}^{Bonus}_{NIC})|_{w,\hat{e},b} \neq 0$, where $\overline{e}$ is the average effort (over all ten periods) in phase 1 (phase 2) of the Trust game and the Fine or Bonus game, respectively; the subscript *NIC* indicates that we only look at non-incentive compatible contracts.

In words, a short-run crowding effect exists if the average phase 1 effort under non-incentive compatible Fine or Bonus contracts differs from the average effort in phase 1 (phase 2) of TTT, *holding the offered contract constant*. We will test for crowding effects by means of an econometric analysis. We will say there is short-run crowding out (crowding in) if the inequality is statistically significantly negative (positive). This definition focuses on non-incentive compatible contracts because the best-reply effort is $e^* = 1$ as under a Trust contract. We also only compare phase 1 (phase 2) of either FT or BT (TFT or TBT) to phase 1 (phase 2) of TTT. Holding the phase and offered contract constant ensures that we compare the exact same situations.

A long-run crowding effect is defined similarly:

Definition 2: *Long-run crowding* occurs if $(\overline{e}^{Trust\ \text{AFTER}\ Trust} - \overline{e}^{Trust\ \text{AFTER}\ Fine/Bonus})\Big|_{w} \neq 0$, where $\overline{e}$ indicates the average effort (across all periods of the respective phase).

In other words, a long-run crowding effect exists if the average effort in the Trust games of phase 2 (phase 3) of either FT or BT (TFT or TBT) differs from the average effort in phase 2 (phase 3) of TTT, *holding the offered wage and the phase of comparison constant*. Long-run crowding out (crowding in) occurs if the econometric analysis returns a significantly negative (positive) inequality.

In the one-shot games of the first set of experiments (treatments FT and BT) we expect a crowding out of voluntary cooperation both in the short-run and the long-run. First, previous related experiments suggest the existence of short-run crowding out (Fehr and Gächter (2002)). Second, incentive contracts appeal to an agent's selfishly rational motivations, whereas a Trust contract appeals to an agent's other-regarding motivations. We expect that the experience of Trust contracts (treatments TFT and TBT) will reduce both short- and long-run crowding out.

The implicit incentives in the repeated games of our third sets of experiments (treatments TTT-R, TFT-R and TBT-R) should eliminate crowding out and may even lead to a crowding in of voluntary cooperation (Gächter and Falk (2002)). The reason is that strategic incentives and other-regarding motivations are not in opposition to one another in repeated games. If the implicit incentives are strong enough they may overpower any crowding out effect that might exist in the one-shot games.

3.2 *Procedures*

We conducted 20 sessions at the University of St. Gallen with a total of 500 participants (first-year undergraduates of business, economics or law). We recruited our subjects from a large data base of people who volunteered to participate in experiments. In a typical session 28 participants 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, participants 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 ask questions privately. In the instructions the principal was called participant X and the agent was called participant Y. To facilitate understanding of the decision problem we indicated that one may think of X as an employer and of Y as an employee. Role assignment was random and fixed throughout the session. It was explained that all decisions would be anonymous during the whole experiment. A short summary of the instructions was read aloud by the experimenter.

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’ (Fischbacher (2007)). Participants 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. RESULTS

We organize our analysis in the order of our research questions. First, we will look at voluntary cooperation under Trust contracts (section 4.1). We will investigate the behavioral reactions to incentive contracts in Section 4.2. Section 4.3 will analyze to what extent there is crowding out of voluntary cooperation both in the short-run and the long-run. Section 4.4 will investigate how the experience of Trust contracts before being exposed to incentive contracts influences short- and long-run crowding out. Finally, in Section 4.5 we will look at the impact of repeated interactions on incentive provision and (crowding out of) voluntary cooperation.

4.1 *Voluntary Cooperation under Trust Contracts*

We set the stage for our analysis by looking at the results of our benchmark TTT treatment. We are in particular interested in how the actual effort depends on the offered wage. Based on previous evidence on closely related variants of our Trust game (e.g., Fehr, Kirchsteiger and Riedl 1993; Fehr, Gächter and Kirchsteiger 1997) we predict that wage and effort will be positively correlated. The results confirm this prediction. Figure 1 is a scatter plot of fixed wage and effort choices in the three phases of treatment TTT; each dot is a single observation and the solid line is a locally weighted regression of actual effort on wage.[6]

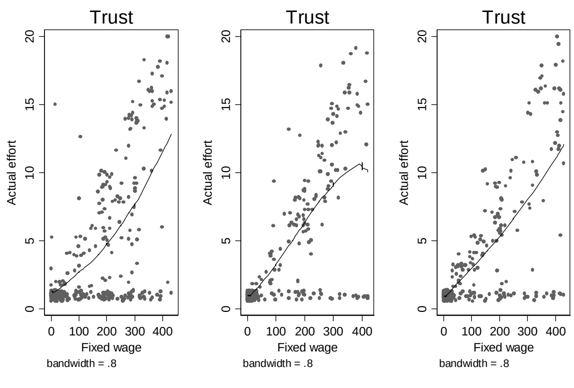


FIGURE 1: Relation between offered wage and actual effort in each of the three phases of TTT

[6] Figure 1 shows the relation between actual effort and fixed wage for wages between 0 and 450. Each dot represents one data point. We produced this scatter plot (and several similar ones below) by using the 'lowess' estimation command in the statistical software package Stata. We had a few wages above 450 (up to wages of 700), but they comprised less than 2 percent of cases. For expositional clarity we therefore restrict our attention in this (and subsequent figures) to wages below 450. In the econometric analyses we include all accepted contracts, not just those with wages up to 450.

Consistent with money-maximization some agents choose minimal effort independent of the fixed wage level. But Figure 1 also shows voluntary cooperation ($e > e^*$) and, in line with previous evidence, a clear positive correlation between fixed wage and actual effort. Thus, more trust (higher fixed wage payments) induces more effort.

A somewhat novel feature of our design is that we observe behavior in the Trust game for 30 periods, whereas most previous evidence on the wage-effort relationship was based on ten to twelve periods only. The fact that we observe behavior in three phases allows us to investigate the role of experience for the stability of the wage-effort relationship. Figure 1 (and the econometric analysis reported in Table B2 in Appendix B) shows that the wage-effort relationship is stable across the three phases of our experiment. This is an important observation for our purposes because it suggests that voluntary cooperation under Trust contracts is a stable phenomenon (at least for the time span of our experiment). Conditional on offered wages more experience does not lead to more selfish behavior. Thus, when we observe different patterns in other treatments we know that these differences are due to treatment effects and not due to a systematic temporal change in voluntary cooperation (and the wage-effort relationship).

4.2 *The Behavioral Effects of Incentive Contracts*

We start our main analysis by looking at behavioral effects of incentive contracts in phase 1 of FT and BT. Figure 2 is a scatter plot showing the relation between the money-maximizing best-reply effort and the actual effort in the first phase of FT (left panel) and BT (right panel). The size of dots is proportional to the number of underlying observations. The behavioral reactions are remarkably similar in both treatments. There is obviously an almost linear correlation between best-reply effort and actual effort for $e^* > 1$, proving that performance incentives are effective in this environment.[7] Provided the principal offers an incentive-compatible contract (which happens in 67.8 percent of all cases in FT, and in 67.4 percent of cases in BT), the agent chooses the best-reply effort in 69.2 (76.9) percent of the cases in FT and BT respectively. Figure 2 also shows that if the contract is incentive compatible and above the minimum effort ($e^* > 1$) there is almost no voluntary cooperation as only very few contracts exceed e^*. If agents deviate from the best reply effort (given $e^* > 1$), they tend to choose minimal effort.

[7] The variation in best-reply efforts visible in Figure 2 is mainly due to desired effort levels rather than fines and bonuses: in FT (BT) 82.0 percent (73.1 percent) of contracts are at the maximal fine or bonus level. See Figure B1 in Appendix B for a development of desired effort over time.

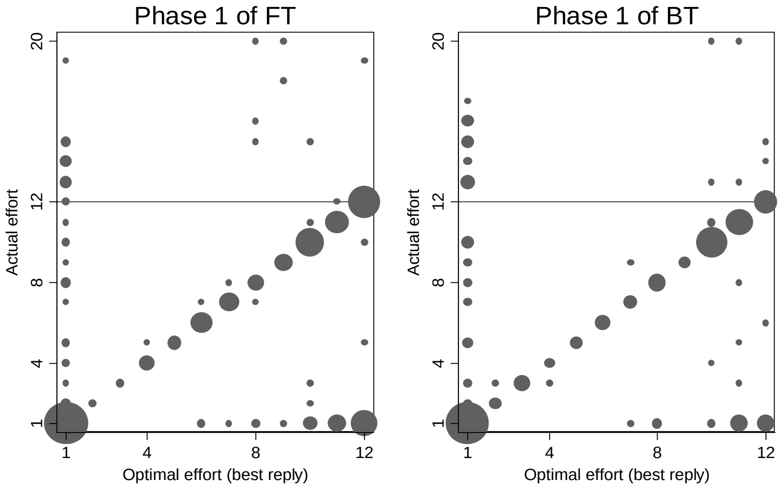


FIGURE 2: Relation between actual effort and best reply effort in phase 1 of FT (left panel) and phase 1 of BT (right panel). The size of dots is proportional to the number of underlying observations.

If the contract is not incentive compatible (which implies $e^* = 1$), many agents choose a substantially higher than minimal effort level. It is interesting to note that an effort above level 12 – which is the maximal level that can be enforced via incentives – is more frequent for $e^* = 1$ than for $e^* > 1$. Thus, non-incentive compatible contracts may allow for effort levels that are largely infeasible with incentive contracts. In summary, there is crowding-out of voluntary cooperation if the contract is incentive compatible, but substantial voluntary cooperation for non-incentive compatible contracts.

Tobit regression analyses (with robust standard errors, clustered on independent matching groups) confirm the observations from Figures 1 and 2 (see Table 3).[8] We estimate three models, one each for the data of TTT, BT and FT, respectively.

We regress actual effort on the "offered compensation" (in TTT) and in addition on the best-reply effort in FT and BT. The first model – using the phase 1 data of TTT – shows that actual effort and the offered fixed wages are highly significantly positively correlated. By contrast, the offered compensation has no significant impact on actual effort in phase 1 of treatments FT and BT.[9] Thus, in stark contrast to TTT fixed wages do not exert any significant influence on actual effort in the presence of incentive contracts in phase 1 of both FT and BT. We will return to this observation in more detail below.

[8] We chose a Tobit regression in all effort regressions because effort is censored at 1 and 20 which are binding in many cases.

[9] Note that in BT and FT the offered compensation measures the effect of payments beyond the incentive effect (which is captured by the variable 'Best-reply effort').

TABLE 3

TOBIT REGRESSION OF EFFORT ON OFFERED COMPENSATION AND BEST REPLY EFFORT FOR PHASE 1 OF TREATMENTS TTT, BT AND FT

	Dependent variable: effort		
	TTT	FT	BT
Offered compensation	0.040	0.011	0.007
	(0.005)***	(0.008)	(0.006)
Best-reply effort (e^*)		0.648	0.702
		(0.0821)***	(0.142)***
Constant	-6.671	-1.639	-0.520
	(1.594)***	(1.093)	(0.679)
Observations	316	343	332
Wald chi2	69.97***	159.62***	50.18***

Robust standard errors in parentheses; *** p < 1%

Actual effort and best-reply effort are also highly significantly positively correlated in both FT and BT. The estimation results are highly similar for FT and BT, which suggests that framing is unimportant. We summarize our findings in our first result.

Result 1: *Performance incentives increase the agent's work effort. If a contract is incentive compatible, voluntary cooperation is crowded out because agents choose the best-reply effort in most cases. By contrast, there is substantial voluntary cooperation for non-incentive compatible contracts. With regard to best-reply behavior framing does not matter as there is no behavioral difference between Fine and Bonus contracts.*

Result 1, in combination with the observations from Figure 1, shows that both voluntary cooperation and performance incentives are behaviorally relevant. Thus, the question arises how performance incentives and voluntary cooperation interact. Result 1 already suggests one form of a 'crowding out effect': if the contract is incentive compatible there is hardly any voluntary cooperation, whereas such voluntary cooperation exists – and can be quite substantial – if the contract is not incentive compatible.

4.3 *Short-Run and Long-run Crowding Effects*

As a first step we will provide a graphical analysis which indicates the impact of incentive contracts on the development of effort over time. This analysis only looks at average effort levels and does not control for the offered compensation, which varies between treatments (see Figure B2 in Appendix B). In a second step we will provide an econometric analysis of short- and long-run crowding effects by controlling for the offered compensation.

Figure 3 shows the development of average effort choices over time for treatments TTT versus FT and BT. In phase 1 of FT or BT incentives are present whereas in phase 2 only Trust games are played. In the phase 1 data of FT and BT we do not distinguish between incentive compatible and non-incentive compatible contracts. The phase 1 data therefore tells us what the average effort level is in the presence of incentive contracts (be they incentive compatible or not) relative to the effort observed in the Trust games. We see that on average effort levels in phase 1 are lower in TTT than in FT and BT (with no systematic differences between the latter). Thus incentive contracts improve performance on average, in particular in the second half of phase 1; the framing of incentives is unimportant.

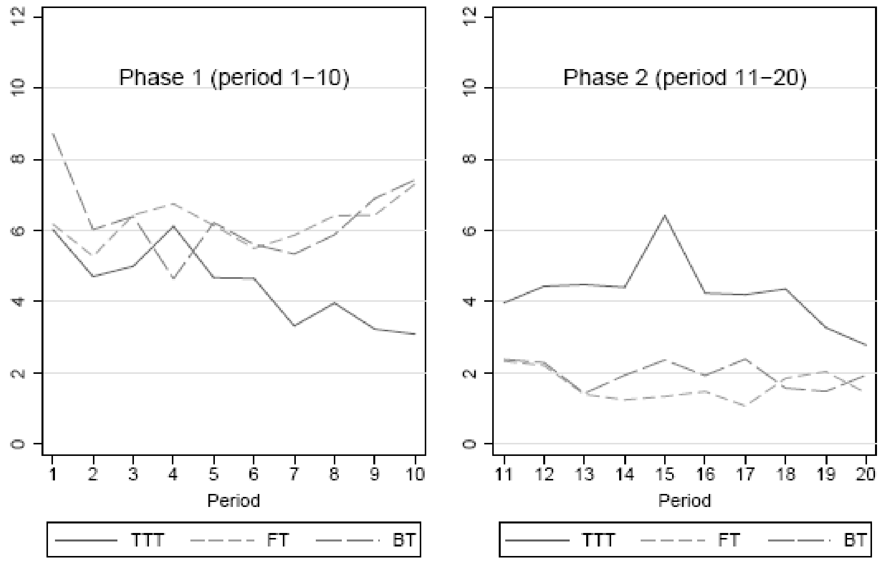


FIGURE 3: Mean effort levels over time in phase 1 and 2 of TTT, FT and BT

In phase 2 of FT and BT incentives are abolished and participants play Trust games for ten periods. We find that the average effort in phase 2 of both FT and BT is lower than in phase 2 of TTT. Thus, reverting to Trust contracts after a phase of incentives reduces effort compared to effort in TTT. This observation suggests the existence of a long-run crowding out effect, provided it also holds once we control for offered wages (recall Definition 2). The regression analysis below will estimate the average long-run crowding out effect.

We turn now to the analysis of short-run crowding effects. Figure 4 shows effort levels in phase 1 of TTT and phase 1 of FT and BT in all those cases in which the offered contracts are *not* incentive compatible. These cases are interesting for two reasons. First, Figure 2 shows that if $e^* > 1$ people supply their best-reply effort in the large majority of cases, so there is not much left to explain. There is only substantial voluntary cooperation if the best-reply effort is $e^* = 1$ and this voluntary cooperation needs an explanation. Second, like in TTT

$e^* = 1$ if the contract is not incentive compatible. Thus, given that people are prepared to supply non-minimal effort levels under Trust contracts and given there is no short-run crowding out, average effort levels with non-incentive compatible contracts in FT and BT should be the same as in TTT (provided the offered wages are the same).

Our results in Figure 4 suggest that if incentives are offered in a non-incentive compatible manner, effort slightly deteriorates compared to Trust contracts. However, an unambiguous identification of a short-run crowding out effect requires controlling for the offered contract (recall Definition 1).

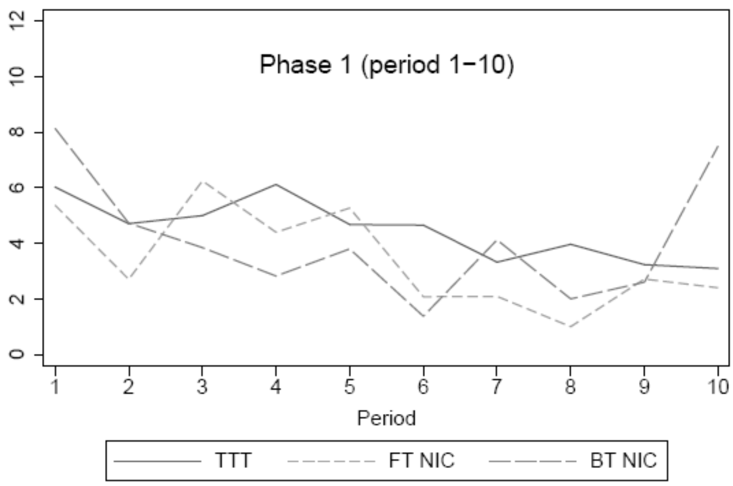


FIGURE 4: Mean effort levels over time for non-incentive compatible contracts.

We now turn to an econometric analysis of short-run and long-run crowding out by controlling for the offered compensation. We ran a Tobit regression analysis which we report in Table 4. We estimate separate models for phase 1 and for phase 2 and we use the data of the respective phase of TTT as the benchmark. For phase 1 data we only use non-incentive compatible contracts. The dummy variables 'Dummy BT Phase 1 NIC' and 'Dummy FT Phase 1 NIC' measure the average short-run crowding out effects and 'Dummy BT Phase 2' and 'Dummy FT Phase 2' measure the average long-run crowding out effects. Crowding out predicts negative coefficients for these dummies.

The results suggest the existence of a significant long-run crowding out effect but no short-run crowding out effects. There is no systematic short-run crowding out because both phase 1 dummies are individually (and jointly, $\chi^2(1)=0.45$, $p=0.501$) insignificantly different from zero. By contrast, the phase 2 dummies, which identify a long-run crowding out effect, are significantly negative. This holds individually and jointly ($\chi^2(1)=14.42$, $p<0.001$). The phase 2 dummy coefficients also differ weakly significantly from one another ($\chi^2(1)=3.58$, p=0.059). This suggests a framing effect as the long-run crowding out effect seems to be (weakly) stronger after Fine contracts than after Bonus contracts.

TABLE 4
MEASURING SHORT- AND LONG-RUN CROWDING OUT EFFECTS

	Dependent variable: effort	
	Phase 1	Phase 2
Offered compensation	0.033	0.044
	(0.005)***	(0.003)***
Dummy Fine Phase 1 NIC	-2.225	
	(1.939)	
Dummy Bonus Phase 1 NIC	-0.134	
	(2.193)	
Dummy Fine Phase 2		-3.859
		(0.694)***
Dummy Bonus Phase 2		-2.234
		(1.084)**
Constant	-6.36	-5.741
	(1.742)***	(1.003)***
Observations	545	876
Wald chi2(3)	49.69***	309.84***

Robust standard errors in parentheses; * p < 10%; ** p < 5%; *** p < 1%.

Figure 5 sheds light on the mechanisms behind the crowding effects by depicting the scatter plots of offered compensation and effort for phase 1 (Figure 5a) and phase 2 (Figure 5b).

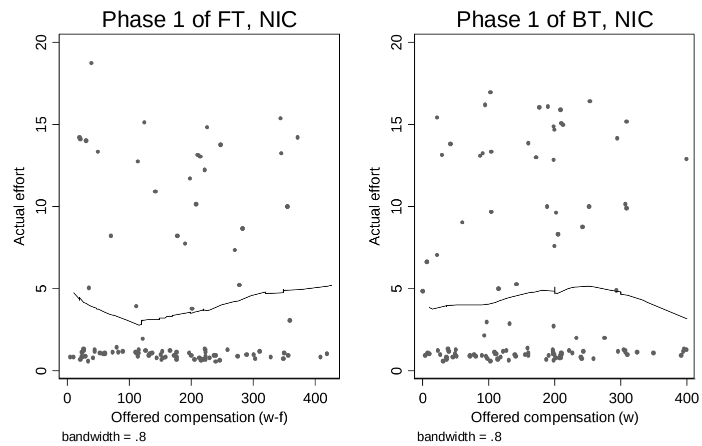


FIGURE 5A: Offered compensation and effort under non-incentive compatible contracts in phase 1 of FT and BT

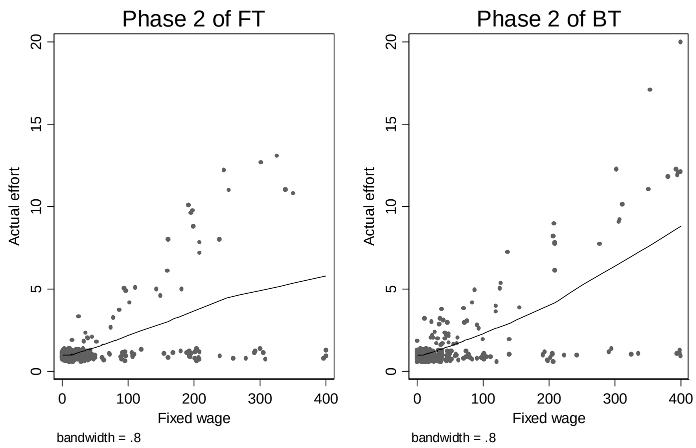


FIGURE 5B: Wages and effort in phase 2 of FT and BT

For phase 1 Figure 5a shows that the positive correlation between offered compensation and effort as displayed in TTT is destroyed under both non-incentive compatible Fine and Bonus contracts. An econometric analysis corroborates this finding (see Table B3 in Appendix B). Agents' effort is uninfluenced by the offered compensation but it is not equal to $e^{*} = 1$ in many cases: effort is all over the place. By contrast, in phase 2 effort choices follow a reciprocal pattern: effort is positively correlated with offered wages on average (see Figure 5b). An econometric analysis (see Table B2 in Appendix B) shows that this relationship is highly significantly positive in both FT and BT.

In summary, there is no short-run crowding out effect despite a destroyed relationship between offered compensation and effort, whereas there is a long-run crowding out effect despite a significantly positive wage-effort relationship. Thus, there are two explanations for the average effort level as observed in phase 2 (see Figure 3): a genuine long-run crowding out effect (for a *given* wage, effort is significantly lower after the experience of an incentive contract) and overall wages are lower in phase 2 of either FT or BT compared to phase 2 of TTT. A comparison of phase 2 data in TTT (see Figure 1) and FT and BT (Figure 5b) suggests this is the case: in both FT and BT there is a stronger concentration on wages below 100 than in phase 2 of TTT (see also Figure B1 in Appendix B which shows the temporal patterns of wage payments). We collect our findings in our second main result.

Result 2: *Holding the offered compensation constant, we find evidence for long-run crowding out but not for short-run crowding out. Furthermore the long-run crowding out is* (*weakly*) *significantly stronger under Fine contracts than under Bonus contracts. The long-run crowding out effect is due to a weakened wage-effort relationship compared to TTT.*

There is no short-run crowding out under non-incentive compatible contracts because agents on average choose similar effort levels $e > e^*$ as in TTT which in FT and BT are unrelated to the offered compensation.

Result 2 suggests that the positive wage-effort relationship is only destroyed in the presence of incentive contracts; if performance incentives are not feasible anymore the positive wage-effort relationship 'appears' albeit weaker than under Trust contracts. How can we interpret this result? One possibility is that the destroyed wage-effort relationship under non-incentive compatible contracts might be due to the ambiguity of the signal sent by a non-incentive compatible contract. To see this, notice that under a Trust contract the level of an offered fixed wage is an unambiguous signal of the principal's generosity. People can readily interpret this signal and respond accordingly (Figure 1). Likewise, an incentive-compatible contract also sends an unambiguous signal by appealing to people's self-interest. If a contract uses incentives but does so in a non-incentive compatible way the contract neither appeals to an agent's self-regard nor to his or her other-regarding motivations, so agents either choose the minimal effort or some non-minimal effort they see as appropriate. In the second phase of FT and BT the only feasible contracts are Trust contracts, which send an unambiguous signal. This is why a positive wage-effort relationship appears.

The fact that the relationship is weaker than in phase 2 of TTT suggests that the experience of appeals to self-regard under incentive contracts has left its traces by making people less other-regarding compared to no such appeal to self-regard in TTT. If this explanation is correct then the crowding out effects should be mitigated if people first experience Trust contracts before they are exposed to Fine or Bonus contracts (and Trust contracts again). We investigate this issue in the next section.

4.4 *Short-Run and Long-Run Crowding Out After the Experience of Trust Contracts*

To investigate the role of experience of Trust contracts for crowding out subjects in a new set of experiments first play a first phase of ten rounds of a Trust game before they are exposed to ten rounds of either a Fine game or a Bonus game in the second phase. In the third phase all play a Trust game again for ten rounds. The matching of principal and agent is again random in each period. We refer to these new experiments as TFT and TBT (see Table 2B for further details).

Figure 6a documents the relation between the offered compensation (the fixed wage w in phases 1 and 3 and $w - f$ in phase 2) and effort choices. Figure 6b shows the analogous results for TBT (where the offered compensation is the fixed wage w in all phases).

In contrast to the results from Figures 5a and 5b we find no strong change in the correlation of offered compensation and effort in phase 2 of both TFT and TBT. The econometric analysis of Table B1 in Appendix B supports this conclusion statistically. There is a positive correlation between offered compensation and effort in all phases, and there is a substantial amount of voluntary cooperation.

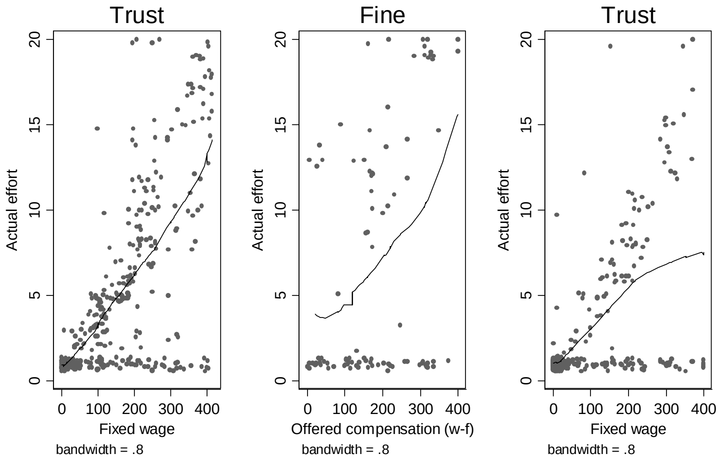


FIGURE 6A: Offered compensation and effort under non-incentive compatible contracts in TFT

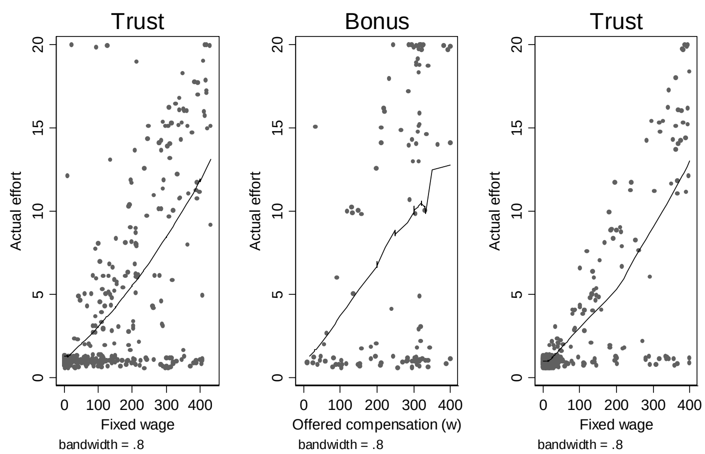


FIGURE 6B: Offered compensation and effort under non-incentive compatible contracts in TBT

To identify the crowding out effects econometrically we estimate, separately for each phase, exactly the same models as in Table 4 by using the phase 2 and 3 TFT and TBT data (with non-incentive compatible contracts) and the TTT data as the benchmark. Our econometric analysis (the details can be found in Table B3 in Appendix B) shows that there is neither a statistically significant short-run crowding out effect nor a long-run crowding out effect. Thus, experiencing Trust contracts in phase 1 eliminates both forms of crowding out. We collect our observations on the role of experience of Trust contracts for crowding out in our third main result.

Result 3. *The experience of Trust contracts in phase 1 of TFT and TBT eliminates both the short-run and the long-run crowding out effect. This holds for TFT and TBT.*

Result 3 shows that the experience of Trust contracts before being exposed to incentive contracts matters for the occurrence of crowding out effects. The experience of Trust contracts makes it easier for agents to interpret the generosity of a contract even if it sends a mixed signal as it is the case for non-incentive compatible contracts. Experience of Trust contracts also eliminates the long-run crowding out effect possibly because agents are reminded of their own other-regarding behavior before they had been exposed to incentive contracts. Apparently, the appeal to self-interest has no long-lasting effects if one has practiced some other-regarding behavior beforehand.

4.5 *The Behavioral Consequences of Implicit Incentives*

So far we analyzed the data of random matching sessions. The main reason for this design feature was to minimize strategic effects that might confound the crowding effects we are interested in. However, in reality most employment relations are long-term relations where strategic effects might influence the amount of voluntary cooperation we see.

In a repeated interaction there might be implicit incentives that are absent in a one-shot interaction. As our above analysis has shown, in one-shot interactions high effort levels can only be achieved by either high incentives or very generous non-incentive compatible contracts. If implicit incentives are possible (which is the case, at least in principle, in repeated games) they may also give selfish people who in one-shot games would exert at most 12 (provided $e^* = 12$), an incentive to provide more effort. We also see no reason why the voluntary cooperation of non-selfish people should be adversely affected by implicit incentives, quite the contrary. Strategic incentives might even reinforce reciprocal motivations. Thus, explicit incentives may not be necessary anymore as long as the implicit incentives are at least as strong as the explicit incentives.

To study the role of implicit incentives we adapt the design of the previous section by changing the interactions from random matching to so-called ‘Partner’-matching. Under Partner-matching a randomly formed pair of a principal and an agent remains paired for the whole experiment, and they know this. We can study repeated game effects by comparing treatments TFT-R, TBT-R and TTT-R with the respective random matching treatments TFT, TBT and TTT. See Table 2C for further details.

We proceed in four steps in presenting the results. We first compare the setting of incentives in one-shot and repeated games. Second, we look at the impact of implicit incentives on average effort levels, and thirdly we will investigate crowding effects. Finally, we will compare the relationship between offered wages and effort in the one-shot and the repeated games.

Incentive setting

We start by looking at how the presence of implicit incentives influences the use of explicit incentives. We find a clear difference with respect to the design of incentive contracts (Bonus contracts and Fine contracts). With random matching 74.6 percent of contracts are incentive compatible, whereas with repeated matching only 26.3 percent of the offered contracts are incentive compatible. The reason for this striking difference is not in the design of fines and bonuses (60.4 (71.2) percent of contracts stipulate maximal fines (bonuses)) but in the desired effort levels. In the one-shot experiments the average desired effort level is 11.1 (11.6) in phase 2 of TFT (TBT), whereas it is 16.7 (16.1) in phase 2 of TFT-R (TBT-R). See Figure B2 in Appendix B for further details.

Figure 7 shows the relation of best-reply effort and actually chosen effort. The left panel depicts the data from the one-shot games of TFT and TBT analyzed above and the right panel shows the data from the repeated games of TFT-R and TBT-R.[10] If the contract is incentive compatible, agents choose their best reply effort in most cases. This is in particular true in the repeated games. Besides similarities between the two graphs important differences are that in repeated games the bulk of observations are for non-incentive compatible contracts and above the maximally enforceable effort level (e = 12). This is not the case for one-shot games.

[10] We pool the data of TFT and TBT, as well as of TFT-R and TBT-R since there are no differences the incentive effects of Fine and Bonus contracts both in the one-shot and the repeated games.

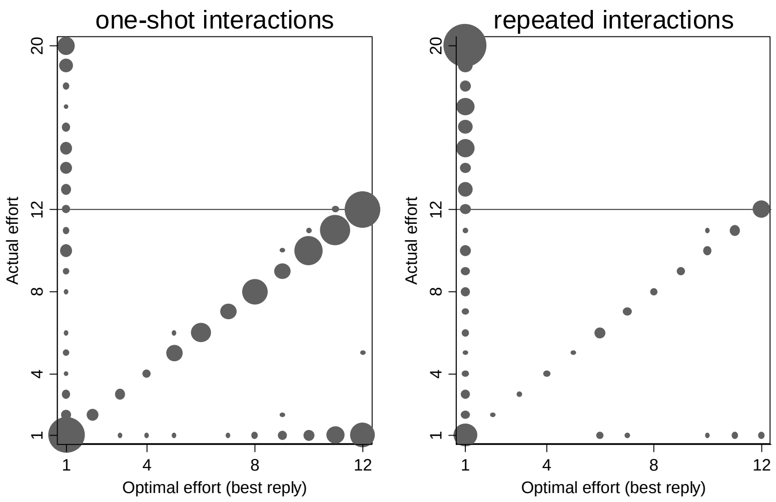


FIGURE 7: Relation between actual effort and best reply effort in phase 2 of one-shot interactions (TFT, and TBT) and repeated interactions (TFT-R and TBT-R).

Thus, as predicted Figure 7 suggests that explicit and implicit incentives are indeed substitutes. To corroborate this finding econometrically we run a regression similar to the one in Table 3. We run this analysis separately for TFT and TBT and TFT-R and TBT-R. We record the results in Table 6.

TABLE 6
TOBIT REGRESSION OF EFFORT ON OFFERED COMPENSATION AND BEST REPLY FOR PHASE 2 OF TREATMENTS TFT & TBT AND TFT-R & TBT-R

	Dependent variable: Actual effort	
	TFT & TBT Phase 2 (One-shot interactions)	TFT-R & TBT-R Phase 2 (Repeated interactions)
Offered compensation	0.030	0.079
	(0.006)***	(0.009)***
Best-reply effort (e*)	0.427	0.290
	(0.109)***	(0.152)*
Constant	-0.766	-5.711
	(0.751)	(2.223)**
Observations	697	296
Wald chi2(2)	77.47***	82.88***

Robust standard errors in parentheses; * p < 10%; ** p < 5%; *** p < 1%.

We find that the coefficients for ‘Offered compensation’ are highly significant for both the one-shot and the repeated interactions but the coefficient for the repeated game is more than twice the coefficient of the one-shot games. The 95-percent confidence bounds also do not overlap, which suggests that offered compensation is significantly higher in the repeated games than in the one-shot games (more on this below). The coefficient for best-reply effort is highly significant for the one-shot interactions and only weakly significant (and much smaller) for the repeated interactions.

The impact of implicit incentives on effort levels

Figures 8a-c illustrates the impact of implicit incentives on average effort levels by showing the time series of mean effort levels. We compare one-shot and repeated games separately for each treatment ((a) TTT & TTT-R; (b) TFT & TFT-R; and (c) TBT & TBT-R). Effort is much higher in the repeated games than for in one-shot games in all treatments.

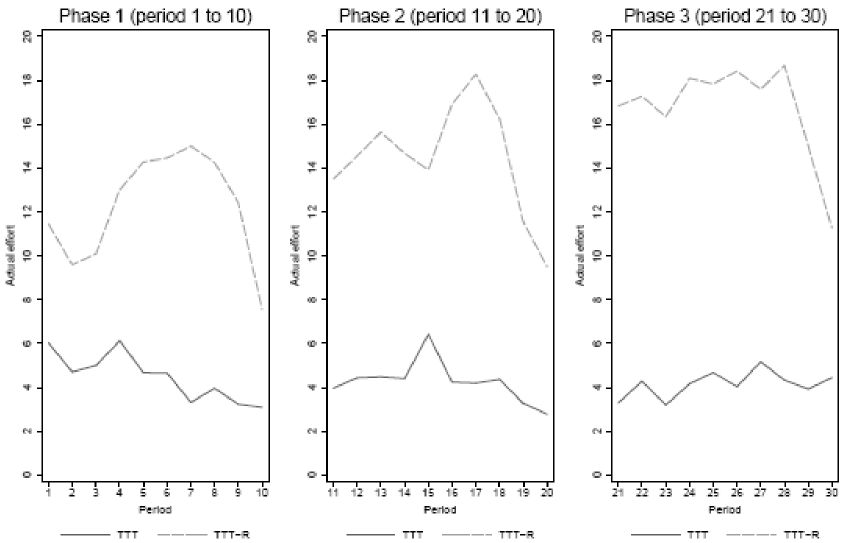


FIGURE 8A: Mean effort over time in the one-shot games TTT and in the repeated games TTT-R.

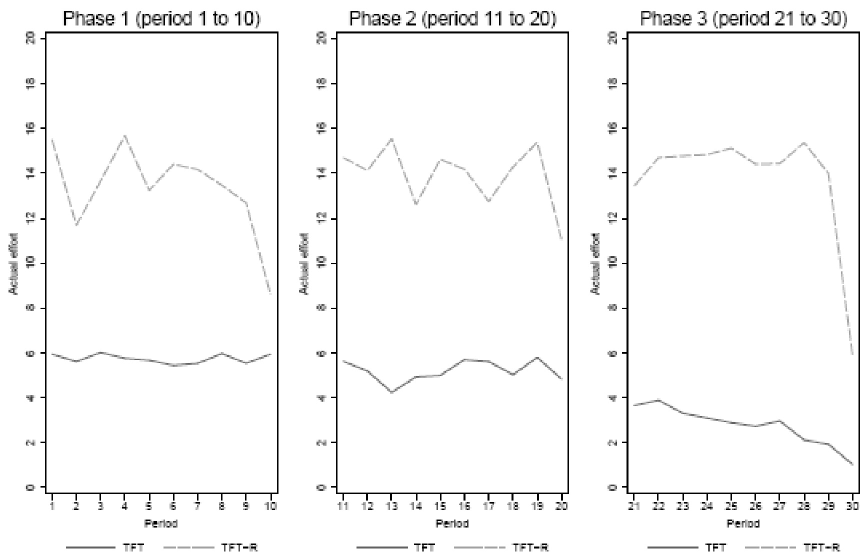


FIGURE 8B: Mean effort over time in the one-shot games TFT and in the repeated games TFT-R.

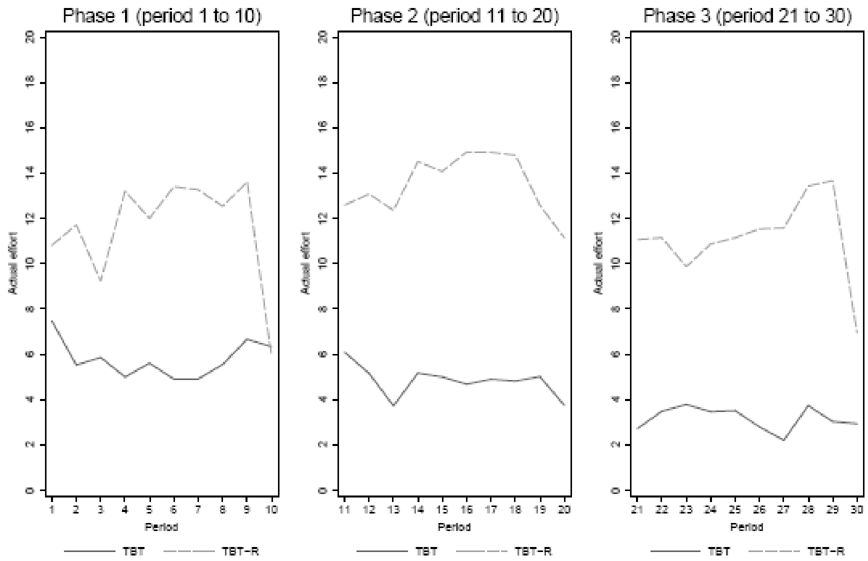


FIGURE 8C: Mean effort over time in the one-shot games TBT and in the repeated games TBT-R.

In treatment TTT-R mean effort is increasing over time within each phase and across phases, except for a pronounced end-game effect in each phase. In phase 3, mean 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 explicit incentives – voluntary cooperation almost reaches the maximal level.

As a next step we measure the impact of implicit incentives econometrically by holding the offered compensation fixed and by comparing with the respective one-shot treatment. Thus, we estimate how, *for a given offered compensation*, effort changes relative to the one-shot experiment. We model the repeated game effects as follows. We pool the data of the one-shot and the repeated games. We add a dummy ('Dummy Repeated Game') that is one if the treatment is TTT-R, or TFT-R or TBT-R, and zero if TTT, TFT or TBT. We also include a dummy that captures endgame effects in the repeated games (this dummy equals 1 in period 8-10, 18-20 and 28-30 of the repeated experiments, and 0 in all other periods of repeated games and all periods of the one-shot experiments). We also include dummies for phase 2 and phase 3 in the one-shot games ('Dummy Phase 2 (3) one-shot') as well as dummies for phases 2 and 3 in the repeated games ('Dummy Phase 2 (3) repeated'). We estimate our model separately for treatments TTT & TTT-R, TFT & TFT-R and TBT & TBT-R. Thus, given our construction, the phase dummies in the one-shot experiments measure the change in contribution relative to phase 1 of the respective one-shot treatment, whereas the phase dummies in the repeated games measure the change in effort levels relative to phase 1 of the respective repeated game. 'Dummy Repeated Game' measures the mean increase in effort levels across all phases compared to the respective one-shot treatment. We report the results in Table 7.

We start by comparing treatments TTT & TTT-R. We find that there is no strong change in average effort levels over the three phases of the one-shot games, although we do get a significantly higher effort in phase 2 compared to phase 1. The dummy for the repeated game is highly significant. In TTT-R we observe an increase in effort over the three phases; in particular we find that effort in the third phase is significantly higher than in phase 1 of TTT-R. In summary, holding offered wages constant, the implicit incentives inherent in the repeated Trust games elicit substantially higher and even increasing effort levels compared to the one-shot games.

TABLE 7
MEASURING THE IMPACT OF IMPLICIT INCENTIVES

	Dependent variable: Actual effort		
Data used:	TTT & TTT-R	TFT & TFT-R	TBT & TBT-R
Offered compensation	0.042	0.053	0.052
	(0.003)***	(0.004)***	(0.003)***
Best-reply effort (*e**)		0.430	0.421
		(0.109)***	(0.191)**
Dummy Phase 2 one-shot	0.908	3.563	2.663
	(0.460)**	(0.999)***	(0.778)***
Dummy Phase 3 one-shot	0.595	-2.504	-0.790
	(0.608)	(0.698)***	(1.157)
Dummy Repeated Game	7.563	4.652	5.192
	(1.500)***	(1.292)***	(1.529)***
Dummy Phase 2 repeated	0.429	-0.691	0.514
	(0.907)	(1.422)	(0.960)
Dummy Phase 3 repeated	2.494	3.733	1.379
	(1.063)**	(1.133)***	(1.362)
Endgame effect	-2.625	-1.357	-0.487
	(1.076)**	(0.114)	(0.848)
Constant	-6.147	-6.746	-7.727
	(0.982)***	(1.196)***	(1.645)***
Observations	1216	1541	1422
Wald chi2	488.47***	237.07***	484.97***

Robust standard errors in parentheses; * p < 10%; ** p < 5%; *** p < 1%.

The results of TFT & TFT-R and TBT & TBT-R are similar to one another. Incentives in phase 2 of the one-shot game increase effort highly significantly compared to phase 1. Average effort in phase 3 of the one-shot TFT is significantly lower than the average effort in phase 1 of TFT; in TBT the decrease in effort is not significant.[11] Although still substantial and highly significant, the average increase in the repeated games is weaker in TFT-R and TBT-R than in TTT-R. This is no surprise given that phase 2 effort is significantly higher in TFT and TBT than in TTT. There are some interesting differences between TFT-R and TBT-R. While in phase 2 of TFT-R the average effort is weakly significantly lower than in phase 1, *ceteris paribus*, this is not the case in TBT-R. In phase 3 of TFT-R effort is significantly higher than in phase 1; this is not the case in TBT-R.

In summary, the implicit incentives are very strong and induce agents to put forward much higher effort levels than in the one-shot games. The implicit incentives are so strong that they induce agents even in Trust games to provide effort above level 12, which is the highest effort level that can be induced by incentives. This power of the implicit incentives in our experiments explains why principals deliberately designed contracts that were not incentive compatible.

Crowding effects in repeated games

The above analysis has already shown that the implicit incentives inherent in repeated interactions lead to a strong increase in reciprocity. This also holds for phase 2 games. The purpose of the following analysis is to calculate the increase in effort levels that is due to a 'crowding in', that is, an increase in effort levels compared to the one-shot TFT and TBT games in phases 2 and 3, respectively. Our methodology is exactly the same as in the previous analyses of crowding effects (in Tables 4 and B3 in Appendix B). We measure the mean crowding effects by dummies of the respective phase (phase 2 for short-run crowding and phase 3 for long-run crowding). In phase 2 we only take non-incentive compatible contracts. The benchmark games are the one-shot experiments in TFT or TBT, respectively. Table 8 reports the results.

[11] One may view this decrease in effort as a crowding out effect *relative to voluntary effort in phase 1.* Our definition of crowding out compared *phase 3* efforts in TTT and TFT, respectively. According to this latter (and more conservative) definition, we do not find a significant crowding out effect.

TABLE 8
CROWDING EFFECTS IN REPEATED GAMES
(COMPARED TO THE BENCHMARK ONE-SHOT GAMES)

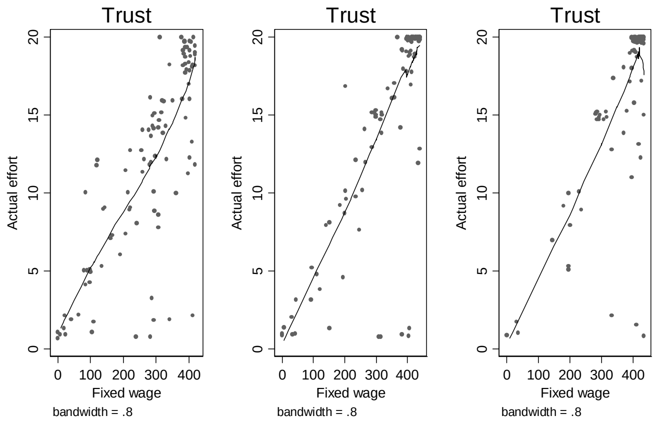
	Dependent variable: Actual effort			
Data used:	TFT & TFT-R		TBT & TBT-R	
	Phase 2	Phase 3	Phase 2	Phase 3
Offered compensation	0.086	0.064	0.078	0.056
	(0.016)***	(0.006)***	(0.017)***	(0.004)***
Dummy TFT-R Phase 2 NIC	7.216			
	(2.813)**			
Dummy TBT-R Phase 2 NIC			9.225	
			(3.722)**	
Dummy TFT-R Phase 3		8.397		
		(1.888)***		
Dummy TBT-R Phase 3				4.696
				(1.589)***
Endgame effect	2.279	-3.374	3.241	0.704
	(1.930)	(1.344)**	(2.848)	(0.830)
Constant	-13.892	-11.399	-13.800	-7.452
	(5.113)***	(2.072)***	(5.970)**	(1.976)***
Observations	227	493	224	467
Wald chi2(3)	31.84***	114.5***	22.68***	217.27***

Robust standard errors in parentheses; * p < 10%; ** p < 5%; *** p < 1%

We find highly significant and substantial crowding in of voluntary cooperation. This holds for both phase 2 (short-run) and phase 3 (long-run).

The relationship between offered compensation and effort in one-shot and repeated games

Figure 9 describes the relation between offered compensation and actual effort for the treatments TTT-R, TFT-R and TBT-R. Like in all previous similar analyses we restrict our attention to non-incentive compatible contracts in phase 2.



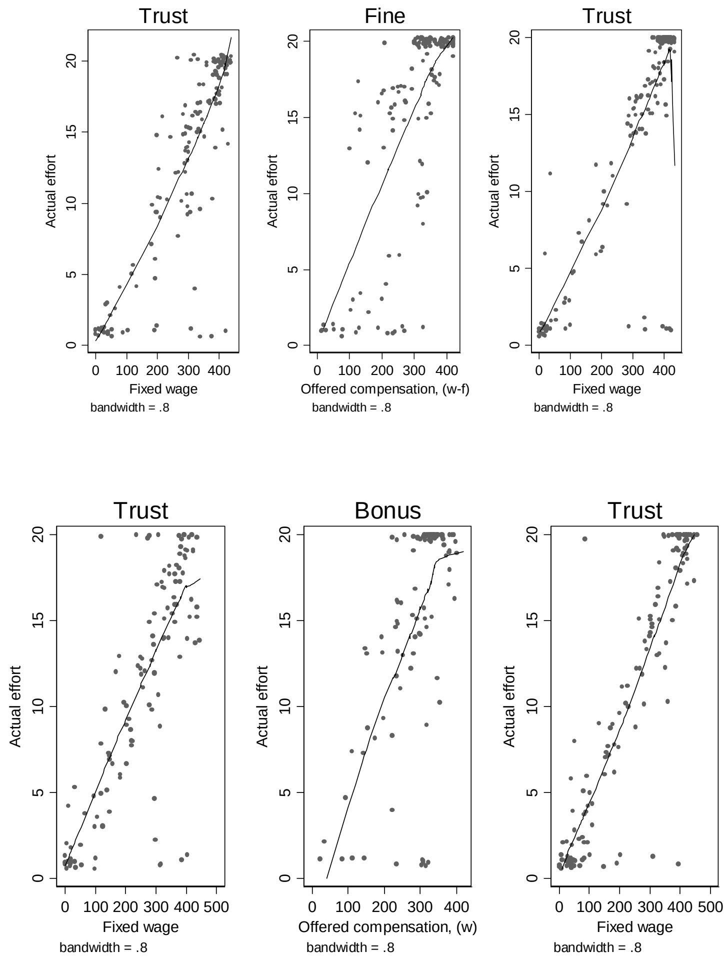


FIGURE 9: Actual effort and offered compensation in the three phases of TTT-R (upper panel) TFT-R (middle panel) and TBT-R (lower panel).

There is a very strong and stable wage-effort relationship in all repeated games, which is steeper in all cases than in the respective one-shot game. An econometric analysis corroborates this conclusion (see Section B6 in Appendix B). If we interpret this relationship as a reciprocity-relationship then we have to conclude that the implicit incentives have led to a crowding-in of reciprocity.

We collect the observations on the repeated games in our final main result.

Result 4: *The implicit incentives inherent in repeated interactions have a very strong behavioral impact on voluntary cooperation in our experiments. We find that the implicit incentives are strong enough to make explicit incentives unnecessary. A comparison of repeated games and one-shot games shows that explicit and implicit incentives are substitutes. Implicit incentives lead to a crowding in of voluntary cooperation that far exceeds* *the crowding out effects we observe in the one-shot games. The implicit incentives also strengthen the reciprocal relationship between offered compensation and effort.*

5. SUMMARY

In this study we analyzed the data of a comprehensive experimental study on a principal-agent game that allows for either incentive contracts or pure fixed wage contracts. We looked at the data from 13420 games played by 500 participants. We investigated the influence of trust, performance incentives, framing of incentives, experience, repeated interaction, and, more specifically, crowding effects within a unified framework. It is to our knowledge the most comprehensive laboratory study of this kind.

A synopsis of our results suggests that voluntary cooperation, the effects of incentive compatible explicit incentives, and the crowding in of voluntary cooperation by implicit incentives are first-order effects; short- and long-run crowding out effects and in particular framing effects are of second-order importance.

Appendix A: Instructions

Here we document the instructions of the Trust game and the Fine game 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:**

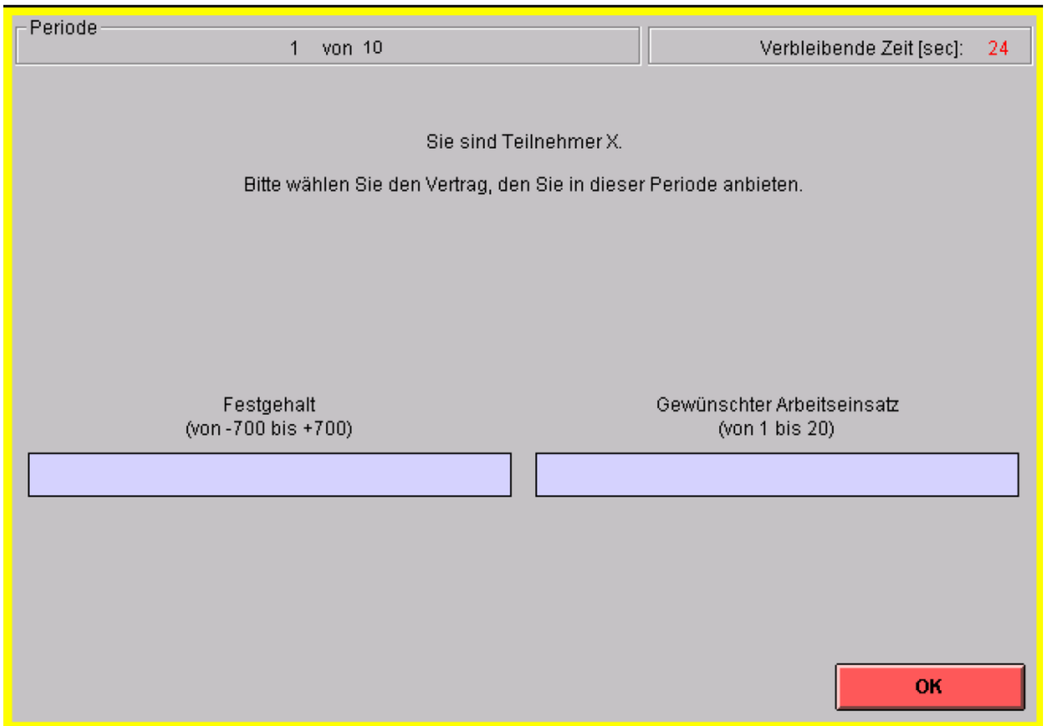
$$-700 \leq \text{fixed wage} \leq 700$$

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

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.

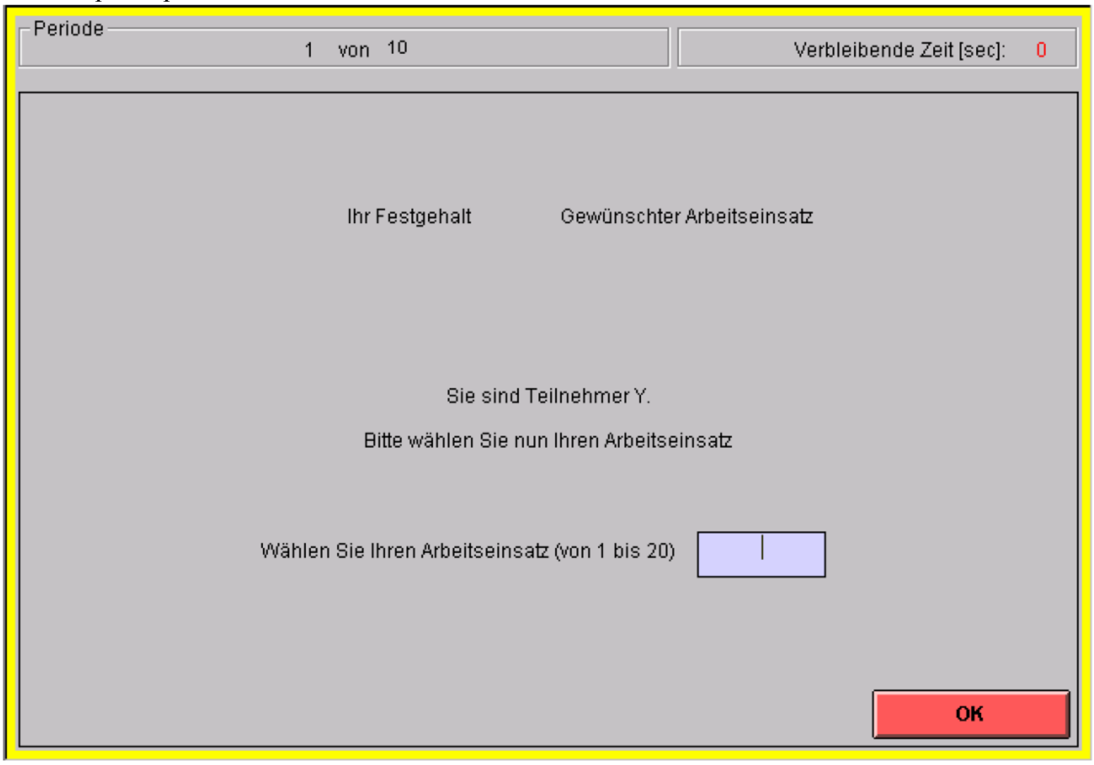
3.2 Employment contract: Acceptance of the contract

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 **<u>effective</u> 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:

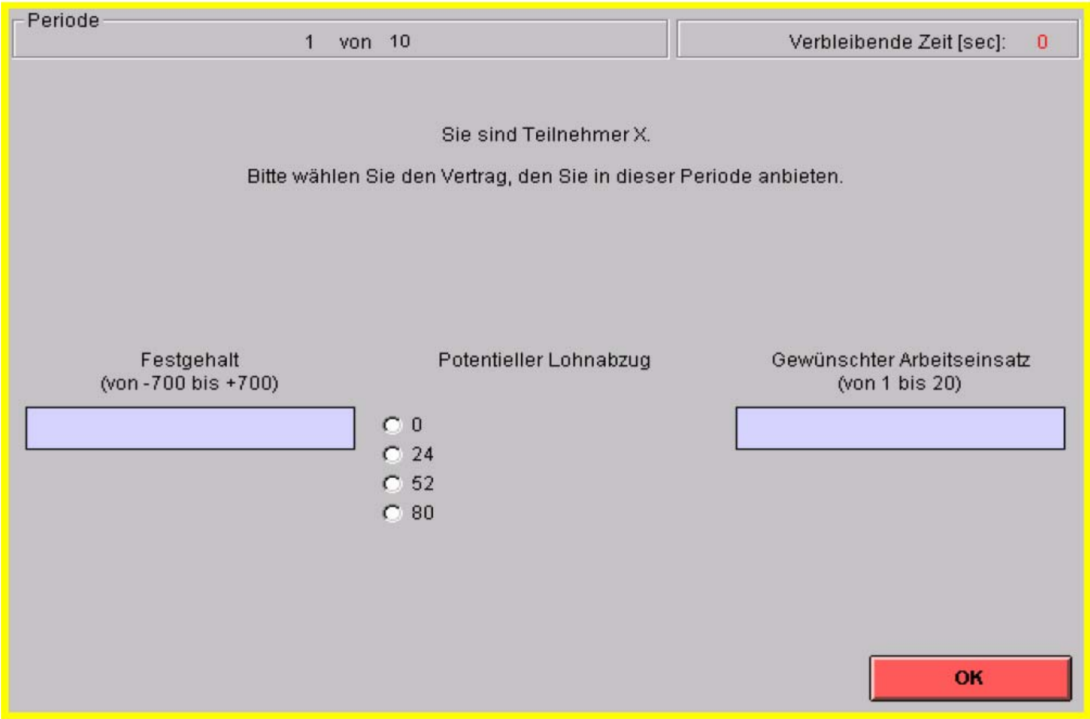
–700 ≤ fixed wage ≤ 700

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

1 ≤ desired work effort ≤ 20

All numbers must be integers. 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.

APPENDIX B: SUPPORTING ANALYSES

B1. *The Stability of the Wage-Effort Relationship in Treatments TTT, TFT and TBT*

We start with the TTT data. The scatter plots of Figure 1 show that the offered wages and actual effort in the TTT treatments are positively correlated on average. This holds in all three phases. Moreover, Figure 1 suggests that the wage-effort relationship is very similar in all three phases. We investigate the stability of the wage-effort relationship by running a Tobit estimation (with robust standard errors clustered on the independent matching groups). The explanatory variables are "Offered compensation"[12], dummies for the respective phase ("Dummy Phase 2" and "Dummy Phase 3") and interaction variables of offered compensation with the respective phase dummy ("Offered compensation × Dummy Phase 2 (3)"). Thus, by construction, "Offered compensation" measures the relationship between the offered compensation (i.e., the fixed wage) and actual effort in phase 1 of TTT. The interaction variables "Offered compensation × Dummy Phase 2 (3)" measure the slope differential of the wage-effort relationship in phase 2 (3) relative to phase 1. The phase dummies measure the fixed effect differences between phase 2 (or 3) relative to phase 1. We report the results in Table B1.

TABLE B1
TOBIT REGRESSION OF WAGE-EFFORT RELATIONSHIP IN TREATMENTS TTT, TFT, TBT.

	Dependent variable: actual effort		
	TTT	TFT	TBT
Offered compensation	0.038	0.049	0.046
	(0.003)***	(0.003)***	(0.003)***
Dummy Phase 2	-1.626	4.309	1.535
	(1.162)	(2.718)	(3.582)
Dummy Phase 3	-0.390	-2.417	-3.283
	(1.103)	(1.401)*	(1.745)*
Offered compensation × Dummy Phase 2	0.013	-0.005	0.008
	(0.005)***	(0.008)	(0.015)
Offered compensation × Dummy Phase 3	0.005	0.007	0.022
	(0.005)	(0.007)	(0.007)***
Constant	-5.662	-8.558	-8.427
	(0.848)***	(1.544)***	(1.925)***
Observation	881	1180	1202
	434.25***	494.68***	415.89***

Robust standard errors in parentheses; * p < 10%; ** p < 5%; *** p < 1%.

We find that wages and effort are highly significantly positively correlated in phase 1 of TTT. There are no fixed effect differences in average actual effort in phases 2 and 3 relative to phase 1. Similarly, the slopes of the wage-effort relationships in phases 2 and 3 are not significantly different from the slope of the wage-effort relationship in phase 1. We conclude, therefore, that the wage-effort relationship is stable over the three phases of our benchmark treatment TTT.

We estimated the same model also for treatments TFT and treatment TBT. These estimations complement the graphical analyses of Figures 6a (for TFT) and 6b (for TBT). In TFT we find no significant changes in the slopes of the offered compensation-effort relationships in phases 2 and 3 relative to phase 1. In TBT the phase 2 slope is statistically the same as the phase 1 slope; the phase 3 slope is highly significantly steeper than the phase

[12] Since we only look at non-incentive compatible contracts "Offered compensation" equals the fixed wage w in all phases of treatments TTT and TBT. In the treatments TFT "offered compensation" is equal to the fixed wage w in phases 1 and 3, and equals $w - f$ in phase 2.

1 slope. The phase 2 dummies are not significantly different from zero in both TFT and TBT. The phase 3 dummies, however, indicate a weakly significantly lower effort level in phase 3 compared to phase 1. Thus, we have a weakly significant reduction of effort *compared to phase 1.*[13]

B2. *Offered and accepted contracts in TTT (phases 1 and 2) and FT and BT*

When we calculate short- and long-run crowding out effects we hold the offered compensation fixed, because only in this way we can identify whether incentives crowd out voluntary cooperation, *ceteris paribus.* Of course, the offered compensation may be different as well between treatments and thereby exert an additional influence on the overall effort as observed in Figures 3 and 4. Figure B1 (left panel) documents the average offered compensation of accepted contracts (incentive compatible and non-incentive compatible ones). For the sake of completeness we also document the mean desired effort levels.

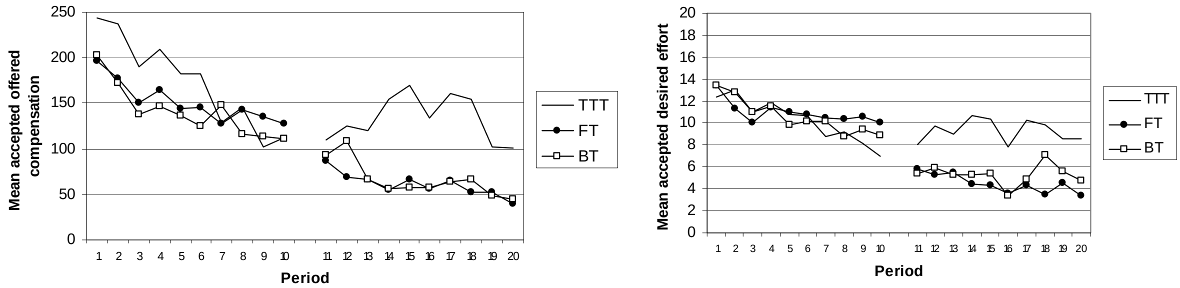


FIGURE B1: Offered and accepted contracts in the first two phases of TTT and in FT and BT. Left panel: Mean accepted offered compensation over time; Right panel: mean accepted desired effort levels over time.

B3. *The relationship of offered compensation and effort under non-incentive compatible contracts in FT and BT*

The analysis in Table B2 complements Figures 5a and 5b. Figure 5a shows that for non-incentive compatible contracts in phase 1 of both FT and BT there is no positive relation between offered compensation and actual effort, which is in stark contrast to phase 1 of TTT (see Figure 1). Figure 5b shows there is a positive relationship of wage and effort in phase 2 of both FT and BT.

TABLE B2

TOBIT REGRESSIONS OF WAGE-EFFORT RELATIONSHIPS FOR NON-INCENTIVE COMPATIBLE CONTRACTS

	Dependent variable: Actual effort			
	FT Phase 1	BT Phase 1	FT Phase 2	BT Phase 2
Offered compensation	0.018	0.006	0.045	0.037
	(0.016)	(0.009)	(0.005)***	(0.003)***
Constant	-8.379	-3.222	-10.873	-6.476
	(3.096)***	(1.621)**	(1.063)***	(2.190)***
Observations	109	120	281	311
Wald chi2(1)	1.18	0.46	93.13***	132.49***

Robust standard errors in parentheses; * p < 10%; ** p < 5%; *** p < 1%

[13] One may view this reduction in effort also as evidence for long-run crowding out. Recall, however, that we measured a long-run crowding out effect in the main text by comparing effort levels in *phase 3* of TTT with effort levels of phase 3 of either TFT or TBT (that is, by holding the phase and therefore the level of experience constant).

The Tobit regressions (with robust standard errors clustered on the independent matching groups) simply relate the actual effort to the offered compensation. We run four models: FT Phase 1 and BT Phase 1 (which complement Figure 5a) and FT Phase 2 and BT Phase 2 (to complement Figure 5b). We find that offered compensation has no significant impact on actual effort in phase 1 for both FT and BT but a highly significant one in phase 2 of both FT and BT.

B4. *Short-run and long-run crowding out after the experience of Trust contracts*

Table B3 reports the econometric analysis of the non-incentive compatible contracts in TTT, TFT and TBT. The goal is to identify the short-run and the long-run crowding out effects. The rationale for the estimation model is exactly the same as in Table 4. Thus the phase 2 variables “Dummy TFT (TBT) Phase 2 NIC” measure the average short-run crowding out effect relative to phase 2 in the benchmark TTT. The phase 3 variables “Dummy TFT (TBT) Phase 3” measure the average long-run crowding effect relative to phase 3 of TTT.

TABLE B3
SHORT-RUN AND LONG-RUN CROWDING OUT EFFECTS (TTT, TFT, TBT)

	Dependent variable: Actual effort	
	Phase 2	Phase 3
Offered compensation	0.060	0.049
	(0.008)***	(0.003)***
Dummy TFT Phase 2 NIC	2.707	
	(2.098)	
Dummy TBT Phase 2 NIC	2.073	
	(2.800)	
Dummy TFT Phase 3		-1.113
		(1.317)
Dummy TBT Phase 3		-0.058
		(1.564)
Constant	-10.864	-7.280
	(2.434)***	(1.084)***
Observations	492	929
Wald chi2(3)	67.43***	268.16***

Robust standard errors in parentheses; * p < 10%; ** p < 5%; *** p < 1%.

The results for phase 2 estimations show that the respective dummies are neither individually (TFT: $\chi^2(1)=1.66$, p= 0.1970), (TBT: $\chi^2(1)=0.55$, p= 0.4590) nor jointly significantly different from zero ($\chi^2(1)=1.57$, p= 0.2099). A similar conclusion holds for phase 3 ($\chi^2(1)=0.26$, p=0.609). Thus, we conclude that the experience of Trust games in phase 1 has eliminated crowding effects, in particular the long-run crowding out effect.

B5. *Accepted contracts in one-shot and repeated games*

Figure B2 is the analogue to Figure B1. It depicts the development of desired effort levels (right panel) and the offered compensation (left panel) for the three-phase one-shot and repeated game experiments.

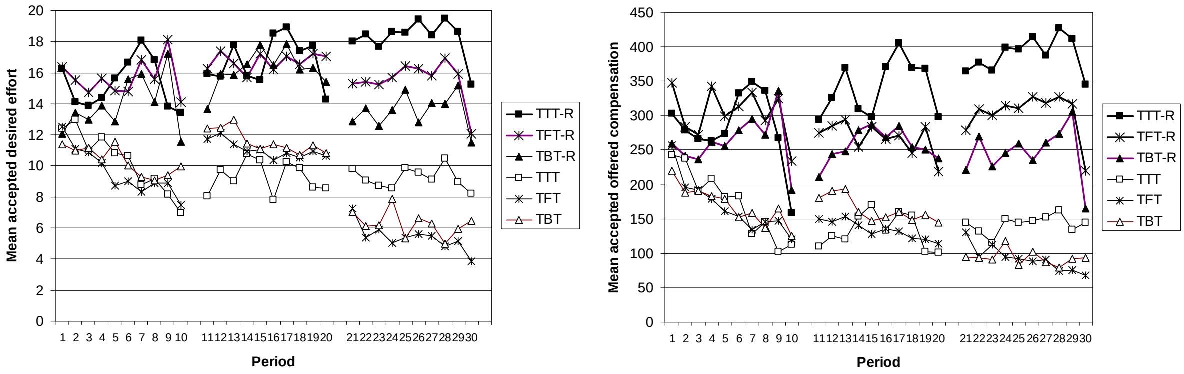


FIGURE B2: Accepted contracts in the one-shot and repeated games. Left panel: Mean accepted desired effort levels over time; Right panel: mean accepted offered compensation over time.

B6. *The relationship between offered compensation and effort in one-shot and repeated games*

Inspection of Figures 1, 6a and 6b (for one-shot games) and Figures 9a-c (for repeated games) suggests that the relationship between offered compensation is remarkably similar across treatments and phases. This holds for both the one-shot and the repeated games. Moreover, a comparison of the relationship between offered compensation and repeated games suggests that this relationship is steeper in all nine cells of the repeated game treatments than in the comparable cells of the one-shot treatments. In the following supplementary econometric analysis we pool all non-incentive compatible contracts of all one-shot treatments, and all repeated game treatments, and run simple Tobit regressions of actual effort on offered compensation. We then calculate predicted effort levels (by taking into account the censoring at effort levels 1 and 20) and plot them in Figure B3.

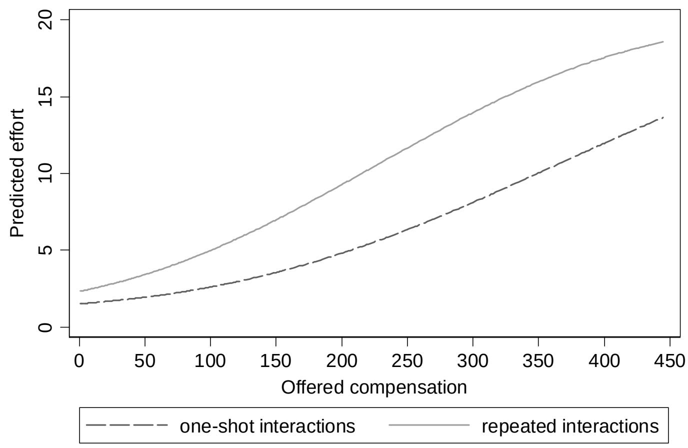


FIGURE B3: The predicted relationship between offered compensation and actual wages in one-shot and repeated games

We find that the predicted relationship between the offered compensation and actual effort across all phases of the one-shot game and the repeated game is indeed steeper in the repeated games. Thus, implicit incentives "crowd in" reciprocal behavior.

REFERENCES

ADAMS, J. S. (1965): "Inequity in Social Exchange," New York: Academic Press, 267-299.

AKERLOF, G. A. (1982): "Labor Contracts as Partial Gift Exchange," *Quarterly Journal of Economics*, 97, 543-569.

BEWLEY, T. (1999): *Why Wages Don't Fall in a Recession.* Cambridge: Harvard University Press.

BOLTON, G. E., and A. OCKENFELS (2000): "Erc: A Theory of Equity, Reciprocity, and Competition," *American Economic Review*, 90 1, 166-93.

BOWLES, S. (2007): "Social Preferences and Public Economics: Are Good Laws a Substitute for Good Citizens?," *Working paper, Santa Fe Institute and University of Siena*.

BRANDTS, J., and G. CHARNESS (2004): "Do Labour Market Conditions Affect Gift Exchange? Some Experimental Evidence," *Economic Journal*, 114, 684-708.

CHARNESS, G. (2004): "Attribution and Reciprocity in an Experimental Labor Market," *Journal of Labor Economics*, 22, 665-688.

CHARNESS, G., and M. DUFWENBERG (2006): "Promises and Partnership," *Econometrica*, 74, 1579-1601.

DECI, E. L., R. KOESTNER, and R. M. RYAN (1999): "A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation," *Psychological Bulletin*, 125, 627-668.

DUFWENBERG, M., S. GÄCHTER, and H. HENNIG-SCHMIDT (2006): "The Framing of Games and the Psychology of Strategic Choice," *CeDEx Discussion Paper 2006-20, University of Nottingham*.

DUFWENBERG, M., and U. GNEEZY (2000): "Measuring Beliefs in an Experimental Lost Wallet Game," *Games and Economic Behavior*, 30, 163-182.

DUFWENBERG, M., and G. KIRCHSTEIGER (2004): "A Theory of Sequential Reciprocity," *Games and Economic Behavior*, 47, 268-298.

FALK, A. (in print): "Gift Exchange in the Field," *Econometrica*.

FALK, A., and E. FEHR (2003): "Why Labour Market Experiments?," *Labour Economics*, 10, 399-406.

FALK, A., and U. FISCHBACHER (2006): "A Theory of Reciprocity," *Games and Economic Behavior*, 54, 293-315.

FALK, A., and M. KOSFELD (2006): "The Hidden Costs of Control," *American Economic Review*, 96, 1611-1630.

FEHR, E., and A. FALK (2002): "Psychological Foundations of Incentives," *European Economic Review*, 46, 687-724.

FEHR, E., and S. GÄCHTER (2000): "Fairness and Retaliation: The Economics of Reciprocity," *Journal of Economic Perspectives*, 14, 159-181.

— (2002): "Do Incentive Contracts Undermine Performance?," *Mimeo, Unversity of Zurich*.

FEHR, E., S. GÄCHTER, and G. KIRCHSTEIGER (1997): "Reciprocity as a Contract Enforcement Device: Experimental Evidence," *Econometrica*, 65, 833-860.

FEHR, E., E. KIRCHLER, A. WEICHBOLD, and S. GÄCHTER (1998): "When Social Norms Overpower Competition: Gift Exchange in Experimental Labor Markets," *Journal of Labor Economics*, 16, 324-351.

FEHR, E., G. KIRCHSTEIGER, and A. RIEDL (1993): "Does Fairness Prevent Market Clearing - an Experimental Investigation," *Quarterly Journal of Economics*, 108, 437-459.

FEHR, E., A. KLEIN, and K. M. SCHMIDT (2007): "Fairness and Contract Design," *Econometrica*, 75, 121-154.

FEHR, E., and K. M. SCHMIDT (1999): "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics*, 114 3, 817-68.

FISCHBACHER, U. (2007): "Z-Tree: Zurich Toolbox for Readymade Economic Experiments," *Experimental Economics*, 10, 171-178.

FREY, B. S., and R. JEGEN (2001): "Motivation Crowding Theory," *Journal of Economic Surveys*, 15, 589-611.
GÄCHTER, S., and A. FALK (2002): "Reputation and Reciprocity: Consequences for the Labour Relation," *Scandinavian Journal of Economics*, 104, 1-26.
GNEEZY, U., and J. A. LIST (2006): "Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments," *Econometrica*, 74, 1364-1985.
GNEEZY, U., and A. RUSTICHINI (2000): "A Fine Is a Price," *Journal of Legal Studies*, 29, 1-17.
KREPS, D., P. MILGROM, J. ROBERTS, and R. WILSON (1982): "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma," *Journal of Economic Theory*, 27, 245-252.
ORGAN, D. (1988): *Organizational Citizenship Behavior*. Lexington, Mass.: Lexington Books.
RABIN, M. (1993): "Incorporating Fairness into Game-Theory and Economics," *American Economic Review*, 83, 1281-1302.
SELTEN, R., and R. STOECKER (1986): "End Behavior in Sequences of Finite Prisoners-Dilemma Supergames - a Learning-Theory Approach," *Journal of Economic Behavior & Organization*, 7, 47-70.
SIMON, H. (1991): "Organizations and Markets," *Journal of Economic Perspectives*, 5, 25-44.
SIMON, H. A. (1997): *Administrative Behavior. A Study of Decision-Making Processes in Administrative Organizations.* New York: Free Press.
WILLIAMSON, O. (1985): *The Economic Institutions of Capitalism*. New York: Free Press.