DISCUSSION PAPER SERIES

IZA DP No. 5774

The Roles of Incentives and Voluntary Cooperation for Contractual Compliance

Simon Gächter
Esther Kessler
Manfred Königstein

June 2011

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

The Roles of Incentives and Voluntary Cooperation for Contractual Compliance

Simon Gächter
University of Nottingham,
CESifo and IZA

Esther Kessler
University College London

Manfred Königstein
University of Erfurt
and IZA

Discussion Paper No. 5774
June 2011

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 5774
June 2011

ABSTRACT

The Roles of Incentives and Voluntary Cooperation for Contractual Compliance*

Efficiency under contractual incompleteness often requires voluntary cooperation in situations where self-regarding incentives for contractual compliance are present as well. Here we provide a comprehensive experimental analysis based on the gift-exchange game of how explicit and implicit incentives affect cooperation. We first show that there is substantial cooperation under non-incentive compatible contracts. Incentive-compatible contracts induce best-reply effort and crowd out any voluntary cooperation. Further experiments show that this result is robust to two important variables: experiencing Trust contracts without any incentives and implicit incentives coming from repeated interaction. Implicit incentives have a strong positive effect on effort only under non-incentive compatible contracts.

JEL Classification: C70, C90

Keywords: principal-agent games, gift-exchange experiments, incomplete contracts, explicit incentives, implicit incentives, repeated games, separability, experiments

Corresponding author:

Simon Gächter
University of Nottingham
School of Economics
Sir Clive Granger Building
University Park
Nottingham NG7 2RD
United Kingdom
E-mail: simon.gaechter@nottingham.ac.uk

* 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 by Nick Bardsley, Sam Bowles, Uri Gneezy and his students, Bernd Irlenbusch, Martin Sefton, and participants in numerous workshops and seminars. Simon Gächter gratefully acknowledges the hospitality of the Institute for Advanced Studies at Hebrew University in Jerusalem while working on this paper.

1. INTRODUCTION

Understanding the behavioral consequences of explicit and implicit incentives in contractual relations is a fundamental task in economics. Arguably, explicit incentives ('pay for performance') and implicit incentives (coming from repeated interactions) appeal to an agent's self-interest to perform in a particular way. Empirical and (field) experimental evidence (e.g., Lazear (2000); Anderhub, et al. (2002); Shearer (2004); Bandiera, et al. (2005)) shows that behavior is often consistent with predictions of self-interest based incentive theory. However, a large body of evidence also shows that many people are willing to act against their self-interest to benefit others.[1] Apparently, both motivations, following material incentives as set out in contracts and institutions, and social preferences, are behaviorally relevant, which raises the important question how they influence each other.

In this paper we provide a comprehensive answer to this question. We investigate experimentally how explicit and implicit contractual incentives affect voluntary cooperation in an environment where efficiency cannot be achieved by explicit incentives alone and hence the agents' voluntary cooperation is required. In reality many (potential) acts of desirable cooperation are not isolated from situations that give people self-interested motivations for contractual compliance. So how do incentives affect voluntary cooperation?

There are at least two (related) reasons why an answer to our question is important. First, many contracts in reality are incomplete, which leaves important aspects unregulated and therefore non-enforceable by third parties. Contractual incompleteness gives agents an incentive to act opportunistically which may impair efficiency. Thus, many scholars argue that voluntary cooperation is necessary to ensure efficiency.[2] Experimental evidence suggests that social preferences can induce such voluntary cooperation. For example, an early paper in this literature argues that the social preference of reciprocity can be a contract enforcement device that mitigates opportunism (Fehr, et al. (1997)). Could it be that performance incentives undermine reciprocity and thereby voluntary cooperation? Or do (some) incentives leave reciprocity unaffected or even reinforce it?

Second, on a more fundamental level, the question is whether 'material interests' and the 'moral sentiments' as expressed in voluntary cooperation are separable, that is, whether incentives and voluntary cooperation are independent of the levels of the other: can we add voluntary cooperation on top of what incentives induce the agent to do, or do incentives *per se* influence the extent of cooperation agents are willing to exert? On an abstract theoretical level there is no reason to assume that separability holds; whether it holds is an empirical

[1] See Camerer (2003); Gintis, et al. (2005) and Bewley (2007) for surveys.
[2] See, e.g., Akerlof (1982); Williamson (1985); Simon (1997); Bewley (1999); Bowles (2003); MacLeod (2007).

question. However, as Bowles and Hwang (2008) argue, separability is a frequently invoked assumption. If separability fails (as some existing evidence, surveyed in Bowles (2008) and Bowles and Polanía Reyes (2011), suggests), incentives may be overused or underused, which has important implications for mechanism design (Bowles and Hwang (2008)).

In this paper we provide a comprehensive experimental analysis of potential failures of separability. Our analyses are based on laboratory gift-exchange games.[3] The gift-exchange game is a two-player game in which a principal offers a fixed wage to an agent. The agent can accept or reject the offered wage. If the agent accepts, he or she chooses an effort level. Effort is costly for the agent and beneficial for the principal. Efficiency calls for the maximal effort whereas self-interest induces the agent to provide the minimal effort irrespective of the accepted wage (no voluntary cooperation). Numerous experiments refute this prediction and demonstrate the relevance of voluntary cooperation – wages and effort are positively correlated even in one-shot games.[4] We replicate this finding in a version of the gift-exchange game we call the 'Trust game'. This will provide the necessary benchmark for the comparisons we are mainly interested in.

The *explicit* incentives take the form of either a 'Fine contract', that is, a contractually agreed wage reduction in case actual effort falls short of the desired effort, or (in different experiments) of a 'Bonus contract' where the agent receives a contractually agreed additional wage payment if the actual effort is at least as high as the desired effort. Both contracts induce the same material incentives and hence any behavioral difference is a framing effect.

We design the set of feasible contracts such that the maximally enforceable effort (by means of incentive compatible contracts) is substantially less than the efficient level. Thus, there is room for efficiency-enhancing voluntary cooperation beyond the maximally enforceable level. Our design also allows for an easy distinction of incentive- and non-incentive compatible contracts; the latter are directly comparable to Trust contracts which are non-incentive compatible by design.

One fundamental reason why separability might fail is the following: Voluntary cooperation and incentives arguably operate on different psychological mechanisms and contracts send messages that appeal to these mechanisms in different ways. The psychological sources of cooperation are motives like fairness and equity, reciprocity, guilt

[3] We chose laboratory experiments for two reasons: (i) only the laboratory allows for the comprehensive investigation of all interaction effects we are interested in (Falk and Heckman (2009); Croson and Gächter (2010)) and (ii) controlling for self interest, which will be crucial for our approach, is hardly feasible in the field.
[4] A seminal paper is Fehr, et al. (1993). Evidence on gift exchange is not confined to the laboratory. See Gneezy and List (2006) and Falk (2007) for field experiments on gift exchange. See also Fehr, et al. (2009) and Charness and Kuhn (2011) for comprehensive surveys.

aversion, loyalty and goodwill, or social norms and social esteem (all now formalized in various theories).[5] By contrast, explicit incentives give agents a self-interested motive to perform. Incentives might also convey mistrust (Falk and Kosfeld (2006)). The general point is that Trust contracts, incentive-compatible contracts, and non-incentive compatible contracts, are psychologically different situations which agents may evaluate differently. Moreover, different frames may cue different responses and failures of separability may therefore also be frame-specific (e.g., Fehr and Gächter (2002); Dufwenberg, et al. (2011)). Failures of separability can take the form of 'crowding out' of cooperation but 'crowding in' is also possible, for instance when principals deliberately design non-incentive compatible contracts (see, e.g., Fehr and Rockenbach (2003)).

A second fundamental reason why separability might fail is that social preferences are 'endogenous' (Bowles (2008)): incentive contracts are appeals to self-interest, and hence people might become more selfish, even in the absence of incentives.[6] We test this as follows: in a first phase agents act in an environment where Fine or Bonus contracts are possible and in a second phase the possibility of explicit incentives is removed: only Trust contracts are feasible. To test separability we compare cooperation in the second phase after experiencing Fine or Bonus contracts to cooperation in the second phase after experiencing Trust contracts.

Our research strategy is based on eight experiments organized in three sets. In a first set of three experiments we establish some basic facts about the two possible fundamental failures of separability. We investigate how voluntary cooperation is affected (i) *while* agents are exposed to Bonus or Fine contracts and (ii) *after* agents experienced explicit incentives.

In a second set of two experiments we investigate how experience with Trust contracts *before* being exposed to Bonus or Fine contracts affects behavior under incentives. Experience with Trust contracts is an interesting contextual variable because the psychology of Trust contracts might set an important reference point before being exposed to incentives.

The third set of three experiments investigates how *implicit incentives* coming from repeated interaction affect the findings on separability observed in the first two sets of experiments, where we randomly change pairings across iterations to avoid confounds of separability issues with strategic incentives. Implicit incentives are arguably a very important

[5] For *fairness and equity* see Akerlof (1982), Fehr and Schmidt (1999), Bolton and Ockenfels (2000), Cox, et al. (2008); for *reciprocity* see Rabin (1993), Levine (1998), Dufwenberg and Kirchsteiger (2004), Sobel (2005), Falk and Fischbacher (2006); for *guilt aversion* see Dufwenberg and Gneezy (2000), Charness and Dufwenberg (2006); for *loyalty and good will* see Simon (1991), Bewley (1999); and for *social norms and social esteem* see Bénabou and Tirole (2006), Sliwka (2007), Ellingsen and Johannesson (2008), Andreoni and Bernheim (2009).

[6] Incentives might also change the nature of the relationship, from good-will based to a market-exchange based relationship. See, e.g., Gneezy and Rustichini (2000); Frey and Jegen (2001); Heyman and Ariely (2004).

feature of many ongoing contractual relationships and therefore it is important to understand how they, together with the explicit incentives, affect separability.

This comprehensive set of experiments allows us to draw robust conclusions on when separability holds and when it fails. Our most important results are as follows.

Our first set of results, in Section 4, demonstrates the existence of substantial voluntary cooperation but also shows that explicit performance incentives work as predicted by standard economic theory based on selfishness. We find both failures of separability discussed above: In the presence of incentive-compatible contracts agents choose their best-reply effort and there is no voluntary cooperation. Under non-incentive compatible contracts, we observe voluntary cooperation but less than under comparable Trust contracts. Separability also fails *after* experiencing incentive contracts: agents who experienced Fine or Bonus contracts are less cooperative than agents who experienced Trust contracts.

Our second set of results, documented in Section 5, shows that behavior under incentive-compatible contracts is robust to the prior experience of Trust contracts: like in the first set of experiments, incentive-compatible contracts lead to best-reply effort choices in the majority of cases and virtually no voluntary cooperation. Separability also fails under non-incentive compatible contracts, but the prior experience of Trust contracts strengthens voluntary cooperation in this case. Interestingly, however, with the prior experience of Trust contracts, separability after being exposed to incentive contracts now holds.

Our third set of experiments, reported in Section 6, shows a very strong positive effect of implicit incentives on voluntary cooperation. In the repeated games, most contracts are designed in a non-incentive compatible way. However, like before, we find crowding out of voluntary cooperation when contracts are incentive compatible: agents play their stage-game best replies in almost all instances. Thus, implicit incentives cannot be added on top of explicit incentives. Like in the second set of results, separability fails while experiencing non-incentive compatible contracts, but holds after being exposed to incentive contracts.

Taken together, the robust findings of three sets of experiments are twofold: (1) incentive-compatible contracts work as standard economic theory predicts, and (2) contrary to standard predictions, there is substantial voluntary cooperation in non-incentive compatible contracts and in Trust contracts. With regard to the separability issue at the heart of this paper, we observe that separability while being exposed to incentive contracts robustly fails under incentive-compatible and under non-incentive compatible contracts. Separability only holds *after* the exposure to incentive contracts but only for agents with prior experience of Trust contracts and/or those operating in an environment with implicit incentives.

2. DESCRIPTION OF STAGE GAMES AND BENCHMARK SOLUTIONS

2.1 *The Games*

Our tools are adapted gift-exchange games (Fehr, Gächter and Kirchsteiger (1997); Fehr and Gächter (2002)), summarized in Table 1. Each game consists of three stages. The principal first offers the agent a contract. In the *Trust game* the contract comprises a fixed wage w and a desired effort e^d (effort can also be interpreted as output). The contract has to obey the restrictions $1 \le e^d \le 20$ and $-700 \le w \le 700$, in integers.[7] In the *Fine* and *Bonus games*, the contract, in addition to w and e^d, also specifies a fine or bonus (details below).

Second, the agent can accept or reject the contract. If he or she rejects, the game ends and both earn nothing. If the agent accepts, he or she enters the third stage and chooses effort e in integers (where $1 \le e \le 20$). The agent is not restricted by e^d. This reflects contractual incompleteness because e^d is not enforceable. The stage game ends after the effort choice.

In all games the principal's return from effort is $35e$ and the agent's cost function is increasing and, for simplicity, linear in effort: $c(e) = 7e - 7$. Each player knows the rules, including all payoff functions, and is informed about all choices made in the game.

TABLE 1
GAMES AND PARAMETERS

Offered contract:	Trust game	Fine game	Bonus game
Fixed 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]
Fine/Bonus	-	$f \in \{0, 24, 52, 80\}$	$b \in \{0, 24, 52, 80\}$
Agent's payoff	$w - c(e)$	$w - c(e)$ if $e \ge e^d$ $w - c(e) - f$ if $e < e^d$	$w - c(e) + b$ if $e \ge e^d$ $w - c(e)$ if $e < e^d$
Principal's payoff	$35e - w$	$35e - w$ if $e \ge e^d$ $35e - w + f$ if $e < e^d$	$35e - w - b$ if $e \ge e^d$ $35e - w$ if $e < e^d$
Effort cost: $c(e) = 7e - 7$			
Payoff if contract rejected: 0 for both			

In the *Trust game* the offered contract only consists of w, e^d. Because w cannot be conditioned on effort, we refer to this game as the 'Trust game'. The principal earns $35e - w$ and the agent earns $w - c(e)$.

The offered contract in the *Fine game* consists of w, e^d, f, where f represents a fine (it can be interpreted as an announced wage reduction if $e < e^d$). The principal can announce one of four lump-sum fine levels: $f \in \{0, 24, 52, 80\}$. If $e < e^d$, f is subtracted from the agent's wage and the principal's wage bill is reduced accordingly. If $e \ge e^d$, the fine is not imposed.

[7] We allow for negative wages because in a benchmark solution (see next section) wages can become negative. To make the wage range fully symmetric we allow for wages between –700 and +700.

In the *Bonus game* the offered contract contains w, e^d, b, where b is a bonus (an announced wage increase if $e \geq e^d$) with $b \in \{0, 24, 52, 80\}$. 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.

We use lump-sum fine and bonus as incentives because they are simple and easy to understand.[8] Moreover, they have attractive properties for our purposes as we show next.

2.2 *Stage Game Benchmark Solutions for Self-Interested Agents*

Trust game: A self-interested agent will choose $e = e^{min} = 1$ irrespective of w and therefore the principal will offer the wage that just ensures the agent's acceptance: $w = 1$ (or $w = 0$). The resulting payoffs are 34 money units for the principal and 1 money unit for the agent. This solution is inefficient, since the efficient surplus is 567 at $e = e^{max} = 20$.

Fine game and Bonus game: In choosing effort the agent has to consider two alternatives, $e = e^d$ or $e = 1$. Effort $e > e^d$ is suboptimal since it causes higher cost without increasing payment. Conditional on $e < e^d$, minimal effort $e = 1$ is best because fine and bonus payments are independent of e. Hence, the optimal effort level is:

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

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 f or b there exists a maximal level of desired effort that satisfies $f, b \geq c(e^d)$. Given our parameters the maximally enforceable effort is 12. Before choosing effort the agent has to accept an offered contract. With the parameters from Table 1 it is optimal for the principal to set w such that the agent is just compensated for his or her effort cost $c(e^*)$; furthermore, 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 choose $e = 12$. This solution is more efficient than the solution without incentives but it does not generate the maximal surplus (the surplus is 343 money units which goes entirely to the principal).

We deliberately set the maximally enforceable effort under incentive-compatible contracts at 12, because this leaves room for voluntary cooperation beyond what incentives

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

can achieve. This design feature reflects contractual incompleteness that characterizes many contracts in reality, even if some aspects can be contractually regulated. By allowing for different fine and bonus levels (including zero) we give the principal the possibility to set the strength of the incentives he or she wants to apply to the agent.[9] Moreover, different combinations of *f*, *b*, and e^d that satisfy the incentive-compatibility constraint can induce different best-reply efforts and this variation allows for a sharper test of whether agents choose best-reply efforts than a more restricted (e.g., binary) set would have allowed for.

Notice also that in case the offered contract violates incentive compatibility, $e^* = 1$, like in the Trust game. This property will be important in our analysis because it makes Trust contracts and non-incentive compatible contracts directly comparable.

3. RESEARCH QUESTIONS, EXPERIMENTAL DESIGN, AND PROCEDURES

3.1 *Research Questions and Experimental Design*

Table 2 lists our research questions and our eight between-subjects experiments. The first experiment is TTT, our benchmark. In TTT subjects play three phases where each phase comprises ten one-shot Trust games played in randomly matched pairs. If effort is higher than predicted according to the benchmark solution (i.e., $e > e^*$), we refer to this as 'voluntary cooperation'. Based on previous gift-exchange experiments we predict that the wage will be positively correlated with effort and there will be substantial voluntary cooperation.

TABLE 2
MAIN RESEARCH QUESTIONS AND EXPERIMENTAL DESIGN

Experiment label	Phase 1 (Period 1-10)	Phase 2 (Period 11-20)	Phase 3 (Period 21-30)	# Subjects	# independent matching groups
0. *Establishing a benchmark of voluntary cooperation.*					
TTT	Trust	Trust	Trust	78	6
A. *Establishing the effects of explicit incentives: Are there crowding effects? Does framing matter?*					
FT	Fine	Trust	-	80	6
BT	Bonus	Trust	-	78	6
B. *Introducing incentives after experiencing Trust contracts: What is the impact of explicit incentives on crowding effects? Does framing matter?*					
TFT	Trust	Fine	Trust	86	6
TBT	Trust	Bonus	Trust	84	6
C. *How do implicit incentives available in repeated interactions influence voluntary cooperation, the use of explicit incentives and crowding effects? Does framing matter?*					
TTT-R	Trust	Trust	Trust	24	12
TFT-R	Trust	Fine	Trust	36	18
TBT-R	Trust	Bonus	Trust	34	17

[9] We included zero because there is evidence that deliberately abstaining from using incentives when incentives could have been used induces more cooperation (Fehr and Rockenbach (2003)).

Against this benchmark we conduct three sets of experiments. The first set of experiments (panel A in Table 2) aims at (i) measuring the impact of explicit incentives on effort choices; (ii) investigating how incentives affect voluntary cooperation in a setting that is neither confounded with strategic incentives, nor with prior experience of Trust contracts before being exposed to incentive contracts, and (iii) measuring the role of framing (Fine contracts vs. Bonus contracts). Therefore, subjects play one-shot experiments in two phases of ten periods each. In Phase 1 principals can design either Fine or Bonus contracts (in between-subjects treatments), whereas in Phase 2 (within-subjects) only Trust contracts are feasible. We label these experiments FT and BT, respectively.

The second set of experiments (panel B, labeled TFT and TBT) investigates the role of experience of Trust contracts before being exposed to incentive contracts. All subjects play three phases of ten one-shot games with random matching. Phase 1 and Phase 3 consist of Trust contracts, whereas Phase 2 allows for either Fine or Bonus contracts.

The final set of experiments (panel C, labeled TTT-R, TFT-R or TBT-R) allows for both explicit and implicit incentives in finitely repeated games ('R') with the same partner. Although subjects are aware of this there are theoretical and empirical reasons why there are implicit (i.e., strategic) incentives to cooperate: If selfishness and rationality are not common knowledge cooperation can be sequentially rational (Kreps, et al. (1982)). Bounded rationality can also lead to cooperation (Selten and Stoecker (1986)). Previous experimental evidence also suggests that cooperation in repeated gift-exchange games is higher than in one-shot games (e.g., Falk, et al. (1999); Brown, et al. (2004)).

We can now give a precise definition how we identify failures of separability (also called 'crowding-out' or 'crowding-in effects'). Because contracts can obey or violate incentive compatibility, we also have to distinguish these cases. Note that our definitions are only about measurement and not about the psychological sources of these crowding effects.

Definition 1 (failure of separability *while* experiencing incentive contracts):

(a) *Contracts are incentive compatible* ("IC"): Separability fails in Phase 1 of FT, BT if $(\overline{e}-1)^{Trust}\big|_{w,e^d} \neq (\overline{e}-e^*_{IC})^{Fine,Bonus}\big|_{w,e^d}$ where $\overline{e}$ is the average effort (over all ten periods) in Phase 1 of the Trust game and the Fine or Bonus game, respectively.

(b) *Contracts are not incentive compatible* ("NIC"): Separability fails in Phase 1 of treatment FT if $(\overline{e}^{Trust}-\overline{e}^{Fine}_{NIC})\,|_{w,e^d,f} \neq 0$, and in BT if $(\overline{e}^{Trust}-\overline{e}^{Bonus}_{NIC})\,|_{w,e^d,b} \neq 0$.

Thus, for assessing separability we compare average voluntary cooperation in Phase 1 of FT, BT with the average voluntary cooperation in Phase 1 of TTT, holding the offered contracts constant. Analogously, assessing separability in the presence of incentive contracts in TFT and TBT (or TFT-R and TBT-R) requires comparing Phase 2 of TFT and TBT (or TFT-R and TBT-R) with Phase 2 of TTT (or TTT-R), holding contracts constant.

Definition 2 (failure of separability *after* experiencing incentive contracts): Separability fails if $(\overline{e}^{Trust \text{ AFTER } Trust} - \overline{e}^{Trust \text{ AFTER } Fine/Bonus})\Big|_{w,e^d} \neq 0$.

Assessing separability after the experience of explicit incentives thus requires comparing the average effort in Phase 2 of TTT with Phase 2 of FT, BT, and Phase 3 of TTT with Phase 3 of TFT, TBT, and TFT-R and TBT-R, holding contracts constant.

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 subjects by drawing a random selection from a data base of volunteer subjects and invited them by email. In a typical session 28 participants were present at the same time.

After arrival at the lab, participants read the instructions (see Appendix A; the same for all) and then had to answer control questions on payoff calculations. The experiment did not start before all participants had answered all questions. Roles were assigned at random and fixed throughout the session. We explained that all decisions would be anonymous during the whole experiment. At the beginning of each session we told participants that there would be different parts and that they would learn about them one after the other.

The experiments were computerized and conducted with 'z-Tree' (Fischbacher (2007)). Participants were separated by partitions and matched anonymously. In sessions with random matching we formed two independent matching groups of 14 subjects each. Participants were not informed about the matching groups but only that they would be randomly matched with another person in the room. Participants also never learned the identity of their opponent. Each session lasted two hours. The average earnings were about CHF 45 (€30).

4. RESULTS I: HOW EXPLICIT INCENTIVES AFFECT VOLUNTARY COOPERATION

4.1 *The TTT Benchmark: Voluntary Cooperation under Trust Contracts*

In our benchmark we are interested in how actual effort depends on the offered wage and how this relationship changes over the up to thirty periods we have planned for our main experiments. Based on previous evidence on closely related variants of our Trust game (e.g., Fehr, Kirchsteiger and Riedl (1993); Fehr, Gächter and Kirchsteiger (1997); as well as the theories of social preferences referenced in footnote 5) we predict that wage and effort will be positively related. We expect these effects at least for the first ten periods (Phase 1), since this is about the duration of most previous gift-exchange experiments. The fact that we observe three phases allows us to investigate the role of experience for the stability of voluntary cooperation.

Figure 1 is a scatter plot of fixed wage and effort choices in each of the three phases of treatment TTT; each dot is a single observation.[10] In line with previous evidence Figure 1 shows extensive voluntary cooperation ($e > e^*$) and a clearly positive correlation between fixed wage and actual effort in a large number of choices. The wage-effort relationship appears to be stable across all three phases.

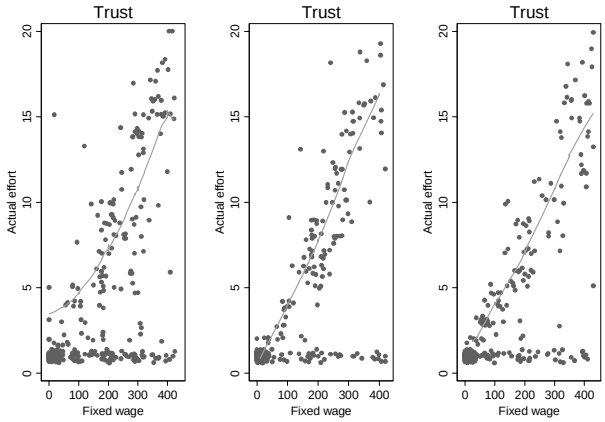


FIGURE 1. Relation between offered wage and actual effort in each of the three Phases of TTT.

While effort increases in wages for a large fraction of observations, there is a sizeable number of observations with $e = 1$ independent of the wage. It seems that decisions can be partitioned into two subgroups: a subgroup that is responsive to different levels of fixed wage and a subgroup that is unresponsive to wages. This pattern appears in our other experiments

[10] Figure 1 shows the relation between actual effort and fixed wage for wages between 0 and 450. We had a few wages above 450 (up to 700), but they comprised less than 2 percent of cases. For expositional clarity we therefore restrict our attention in this (and subsequent scatter plots) to wages up to 450. In the econometric analyses we include all accepted contracts, however. The lines in Figure 1 (as well as in all subsequent scatter plots) are locally weighted regressions of actual effort e on wage for $e > 1$.

as well (see below). It motivates our use of hurdle models (see Wooldridge (2002), p. 536). The hurdle model allows for the possibility that the decision for $e > 1$ versus $e = 1$ (the decision to pass the hurdle) follows different rules than the choice of e conditional on $e > 1$. Thus, in a first step we estimate a probit regression for the choice of $e > 1$ (value = 1) or $e = 1$ (value = 0), and in a second step we estimate a tobit regression with an upper bound of 20 for $e|e > 1$.[11] To account for conditioning on $e > 1$ choices we include the inverse Mills ratio into the tobit regressions. We will use a hurdle approach for all estimations reported below.

Table 3 shows the regression results for the three phases of TTT. The two explanatory variables in Table 3 are the two elements of a contract offer in a Trust contract – w and e^d.

TABLE 3
THE WAGE-EFFORT RELATIONSHIP IN THE BENCHMARK TTT EXPERIMENTS

	Phase 1		Phase 2		Phase 3	
	probit ($e>1$)	tobit ($e\|e>1$)	probit ($e>1$)	tobit ($e\|e>1$)	probit ($e>1$)	tobit ($e\|e>1$)
Fixed wage	0.003***	0.086***	0.006***	0.054***	0.005***	0.050***
	(0.000)	(0.027)	(0.002)	(0.009)	(0.001)	(0.005)
Desired effort	0.038***	1.236***	-0.004	0.140	0.006	0.181
	(0.013)	(0.401)	(0.020)	(0.089)	(0.027)	(0.116)
Inverse Mills ratio		39.156**		5.457***		8.509**
		(17.006)		(1.721)		(3.546)
Constant	-1.015***	-56.877**	-1.105***	-8.717***	-0.888***	-11.399**
	(0.208)	(24.151)	(0.175)	(2.445)	(0.147)	(4.571)
No. of obs.	316	142	284	116	281	123
LR χ^2	57.42***	111.83***	31.42***	313.33***	88.29***	1544.94***
Pseudo R^2	0.135	0.132	0.263	0.251	0.192	0.238

probit indicates a probit model whether $e>1$ or $e=1$. *tobit* indicates a tobit regression on $e>1$ choices only (censored at 20). Robust standard errors in parentheses; * p< 10%; ** p< 5%; *** p< 1%.

For Phase 1, Table 3 reports estimated coefficients for the influence of the fixed wage of 0.003 in the probit regression and 0.086 in the tobit regression; both are highly significant. This means that the probability of choosing above-minimal effort increases in the fixed wage and that effort conditional on $e > 1$ increases in fixed wages as well. Evaluating the two estimated models (probit and tobit) at mean values for all explanatory variables predicts that minimal effort ($e = 1$) is chosen with a probability of 0.419. With residual probability 0.581 players choose $e > 1$, and $e|e > 1$ is predicted at 8.9. The two parts of the regression analysis can be put together to provide a predicted value of effort $E(e)$ with

$$E(e) = prob(e = 1) \cdot 1 + (1 - prob(e = 1)) \cdot E(e \mid e > 1) = 5.6$$

with $prob(e = 1)$ calculated according to the probit regression and with $E(e \mid e > 1)$ calculated according to the tobit regression. We refer to this as the total effect. Similarly, in all

[11] This is similar to the Heckman two-step estimation procedure, just that in step two we apply a tobit regression rather than OLS to account for censoring at $e = 20$.

regression analyses below we will consider the partial effects of an explanatory variable on *prob*(e=1) and $e|e>1$ separately or combined (total effect).

Table 3 reports that fixed wages have a significantly positive effect on effort (partially and combined) for phases 2 and 3 as well. The estimations support the impression from Figure 1 that there is stable voluntary cooperation which depends on the offered wage. These behavioral patterns will serve as basis for comparison in our main experiments.[12]

4.2 *Separability in the Presence of Explicit Incentive Contracts*

We now look at the Phase 1 data of FT and BT to investigate separability under Fine contracts and Bonus contracts (see panel A in Table 2). Figure 2 is a scatter plot of actual effort choices against effort predicted by the best-reply effort e^* for FT (left panel) and BT (right panel). The size of dots is proportional to the number of underlying observations.

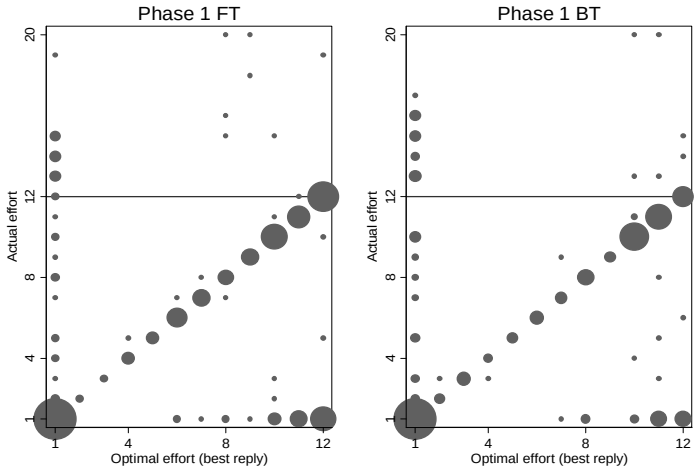


FIGURE 2. Relation between actual effort and best-reply effort in Phase 1 of FT (left panel) and Phase 1 of BT (right panel). The size of dots is proportional to the number of underlying observations. The horizontal line at 12 indicates the maximally enforceable effort level under incentive-compatible contracts.

Figure 2 shows a highly structured pattern that is very similar in both experiments. Many observations cluster exactly on the 45°-line (where $e = e^*$). Provided the principal offers an incentive-compatible contract (henceforth IC-contract; offered in 69.4 percent of all cases in FT, and in 67.6 percent of cases in BT), the agent chooses best-reply effort in 69.1 and 76.9 percent of the cases in FT and BT, respectively.[13] Given an IC-contract ($e^* > 1$), if

[12] Since desired effort is part of an offered contract our regression model also controls for desired effort. Table 3 reports a significant influence of desired effort in Phase 1, but not in later phases. Although we find that fixed wage and desired effort are positively correlated in all phases (all correlations exceed 0.70) we conclude that it is mainly the fixed wage that drives behavior under Trust contracts. Desired effort plays a minor role.

[13] The variation in best-reply efforts visible in Figure 2 is mainly due to desired effort levels rather than fines and bonuses. In FT the percentages of cases of fines of 80, 52, 24, 0 are 85.5, 8.9, 4.0, and 1.6 percent,

agents deviate from best-reply effort, they tend to choose minimal effort ($e = 1$). Effort levels at $e = e^*$ or $e = 1$ explain 92.1 percent of all effort choices.

There is a crowding out of voluntary cooperation, because only very few effort choices exceed e^*. If contracts are not incentive compatible (henceforth NIC) – which implies $e^* = 1$, like under Trust contracts – many agents choose a substantially higher than minimal effort level (33.5 percent of the choices; displayed at the left borders at $e^* = 1$). Moreover, for every e^*-level under IC-contracts we observe effort choices above that level under NIC-contracts. In particular, effort above 12 – which according to IC-constraint (1) is the maximal level that can be enforced by incentives – is relatively more frequent for $e^* = 1$ than for $e^* > 1$. Thus, NIC-contracts can induce effort levels that are largely infeasible with IC-contracts.

Figure 3 looks closer at NIC-contracts by showing a scatter plot of actual effort against offered compensation, with offered compensation defined as $w - f$ in Fine games and w in Bonus games.[14]

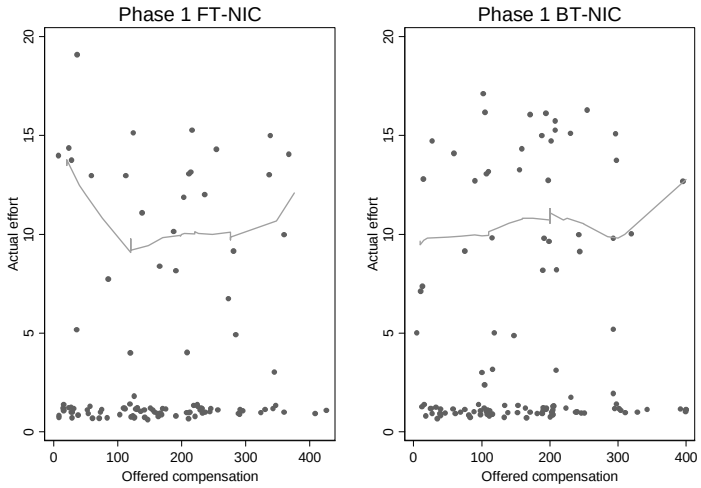


FIGURE 3. Offered compensation and effort under NIC-contracts in Phase 1 of FT and BT.

Unlike in Figure 1 we do not observe a clear positive correlation between effort and offered compensation anymore. It appears that the presence of a NIC-contract (Bonus or Fine) destroys reciprocity, but it does not entirely crowd out voluntary cooperation because a substantial number of effort choices are non-minimal.

Table 4 complements Figures 2 and 3 by an econometric analysis. We use a hurdle model that takes the Phase 1 data of FT, BT and TTT and estimates the influences of fixed wage, incentives and other variables for different contract types (Trust contracts, IC-contracts

respectively. In BT percentages for the respective bonus levels are 73.4, 9.0, 12.8, and 4.8, respectively. These distributions are significantly different from each other **(**$\chi^2(3)=26.25$, p=0.000**)**.

[14] We look at offered compensations rather than the fixed wage only because NIC-contracts give agents a selfishly rational incentive to shirk in which case they have to pay the fine or they forego the bonus.

and NIC-contracts) within a single regression model.[15] In a first step we estimate a probit regression for the choice of $e > 1$ (value = 1) or $e = 1$ (value = 0). The left column of Table 4 shows the estimation results. In a second step we estimate a tobit regression (with an upper bound of 20) for the choice of e conditional on $e > 1$.[16] Estimation results are shown in the right column. In step 2 we include the inverse Mills ratio as regressor.

TABLE 4
EXPLAINING PHASE 1 EFFORT CHOICES IN TTT, FT AND BT

	logit (e>1)	tobit (e|e>1)
Fixed wage (TTT)	0.003***	0.041***
	(0.001)	(0.008)
Desired effort (TTT)	0.038*	0.534***
	(0.020)	(0.113)
Fixed wage (BT, IC)	0.003	0.031**
	(0.002)	(0.015)
Dummy (BT, IC)	1.490***	17.509***
	(0.329)	(4.447)
Desired effort (BT, IC)	-0.007	
	(0.039)	
Fixed wage (FT, IC)	0.006***	0.030
	(0.002)	(0.022)
Dummy (FT, IC)	3.751***	18.983***
	(0.584)	(4.066)
Desired effort (FT, IC)	-0.278***	
	(0.054)	
Fixed wage (BT, NIC)	0.001	0.011**
	(0.001)	(0.005)
Dummy (BT, NIC)	1.103**	14.628***
	(0.444)	(4.764)
Desired effort (BT, NIC)	-0.040	-0.072
	(0.031)	(0.345)
Fixed wage (FT, NIC)	0.002	0.016
	(0.001)	(0.011)
Dummy (FT, NIC)	0.973*	17.858***
	(0.560)	(4.906)
Desired effort (FT, NIC)	-0.063	-0.518
	(0.039)	(0.403)
Inverse Mills ratio		11.266***
		(3.880)
Constant	-1.015***	-16.152***
	(0.148)	(5.465)
No. of obs.	988	253
	χ^2=219.99***	F=42.40***
Pseudo R^2	0.198	0.069

The upper part of Table 4 shows the estimated effects of the variables 'Fixed wage (TTT)' and 'Desired effort (TTT)' on behavior in Trust games. This reproduces the findings

[15] Most variables in this table are defined as treatment-specific variables. For example, 'Fixed wage (TTT)' has value = fixed wage when TTT and value = 0 when BT or FT. Similarly for 'Desired effort (TTT)'. 'Fixed wage (BT, IC)' has value = fixed wage when BT and if the contract is IC and value = 0 otherwise. And so on.

[16] While for step 2 we use only those observations for which $e > 1$ we further reduce the sample of IC-contracts by excluding observations in which effort is exactly equal to best-reply effort ($e = e^*$). The reason is that under these two conditions (an IC-contracts is offered and $e > 1$), $e = e^*$ in 90 percent of the cases.

from Table 3. The middle panel shows the estimated influences of IC-Bonus and IC-Fine contracts. For example, the significant coefficients 1.490 (Dummy variable (BT, IC)) and 3.751 (Dummy variable (FT, IC)) reported in the left column imply that *prob*($e > 1$) is larger, *ceteris paribus*, for both types of IC-contracts compared to the reference category (Trust contract). The corresponding effects in case of NIC-contracts reported in the lower part of Table 4 are also positive but substantially smaller. The right column shows that also conditional effort ($e|e>1$) is positively influenced by IC- and NIC-contracts. Similarly, all fixed-wage influences are positive in both steps of the estimation, although not all estimated influences of fixed wage and incentives are significant in both regression steps.

While our two-step analysis provides a detailed look at the decision to either provide $e = 1$ or $e > 1$, we are interested in putting these partial effects together in order to see how Trust contracts compare to payoff-equivalent IC- and NIC-contracts. We calculate predicted effort values based on our regression model and illustrate them in Figure 4.

For Trust contracts and NIC-contracts the predicted values $E(e)$ are calculated as

$$E(e) = p \cdot 1 + (1 - p) \cdot E(e \mid e > 1) \tag{2}$$

with $p = prob(e = 1)$ as estimated in step 1 and with $E(e|e > 1)$ being the expected value from step 2. For IC-contracts we have to consider that conditional on $e > 1$ there is a large fraction q of effort choices $e = e^*$. Therefore we calculate predicted values $E(e)$ as

$$E(e) = p \cdot 1 + (1 - p) \cdot \left[q \cdot e^* + (1 - q) \cdot E(e \mid e > 1)\right] \tag{3}$$

with p and $E(e|e > 1)$ as above. Furthermore we calculate all predictions for b, $f = 80$ (the most frequent choice) and for $e^d = 12$. Figure 4 shows predicted values for different contract types and for varying compensation levels holding other influences constant.

The left panel of Figure 4 shows that expected effort under accepted IC-contracts depends positively on the offered compensation, in particular for low compensation. This observation contradicts the IC-condition $f,\ b \geq c(e^d)$ (see (1)), which is independent of the offered compensation. Offering an IC-contract is therefore not enough to induce $e = e^*$; compensation also needs to be high enough to actually induce best-reply effort.

A second remarkable finding is that effort levels up to 12 can be implemented at much lower compensation levels with IC-Bonus or IC-Fine contracts than with Trust contracts. Bonus contracts perform somewhat better than Fine contracts, but the difference (framing) is relatively minor. However, effort above 12 seems infeasible with IC-contracts. Increasing compensation beyond 300 does not lead to substantially higher effort. This is different with Trust contracts were the regression model predicts further increases in effort even beyond 12.

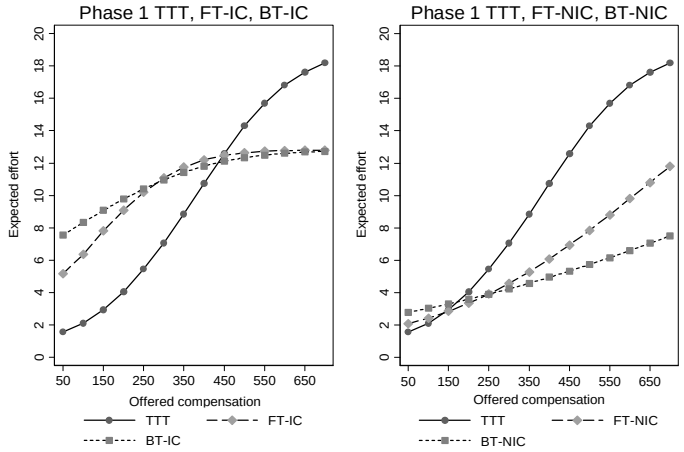


FIGURE 4. Predicted effort as a function of offered compensation (based on estimations in Table 4, for $e^d = 12$, and b, $f = 80$).

The right panel of Figure 4 compares predicted effort for Trust and NIC-contracts (recall that NIC-contracts share with Trust contracts that $e^* = 1$). We find that NIC-contracts induce substantially smaller effort levels than Trust contracts for all but very low compensations (according to Definition 1(a) this is crowding out). Effort is increasing in compensation, but its influence is not as strong as with Trust contracts.[17] Comparing the left and right panel we also see that NIC-contracts perform worse than IC-contracts. However, there is substantial voluntary cooperation even under NIC-contracts because predicted effort is larger than 1.

We summarize our findings in our first main result.

Result I.A. (separability *while* experiencing explicit incentive contracts) *Separability fails unambiguously – incentive contracts affect voluntary cooperation*: **(a)** *Given an IC-contract there is no voluntary cooperation. IC-contracts induce exact best-reply effort in 73 percent of cases.* **(b)** *By contrast, under NIC-contracts* (*and under Trust contracts*) *there is substantial voluntary cooperation.* **(c)** *NIC-contracts induce less expected effort than Trust contracts* (*that is, there is crowding out of voluntary cooperation*). **(d)** *Framing matters only for low compensation levels under IC-contracts and high compensation under NIC-contracts.*

Our next step is to investigate whether experiencing incentive contracts influences voluntary cooperation even when incentives are removed.

[17] This result might appear counter-intuitive, given the impression from Figure 3. However, Figure 3 does not control for other influencing variables (as is done in Table 4 and Figure 4). And it hides the facts that minimal effort is more frequent at low compensation levels and that $e|e > 1$ is increasing in compensation.

4.3 *Separability after Experiencing Explicit Incentive Contracts*

Do explicit incentive contracts also affect voluntary cooperation once incentives are abolished? Specifically we investigate behavior under the condition of a Trust contract in Phase 2 after experiencing Bonus or Fine contracts in Phase 1. Figure 5a shows the development of efforts over time and Figure 5b provides a scatter plot of effort and wages.

The left panel of Figure 5a depicts the average effort observed in Phase 1, separately for TTT, BT and FT, and without distinguishing whether the contract is IC or not. On average Bonus and Fine contracts lead to higher effort levels than Trust contracts. The right panel of Figure 5a shows the development of average effort in Phase 2, when there is no Bonus (Fine) contract any longer. For BT and FT average effort drops from more than 6 to about 2. By contrast, for TTT the average effort in Phase 2 is around 4.

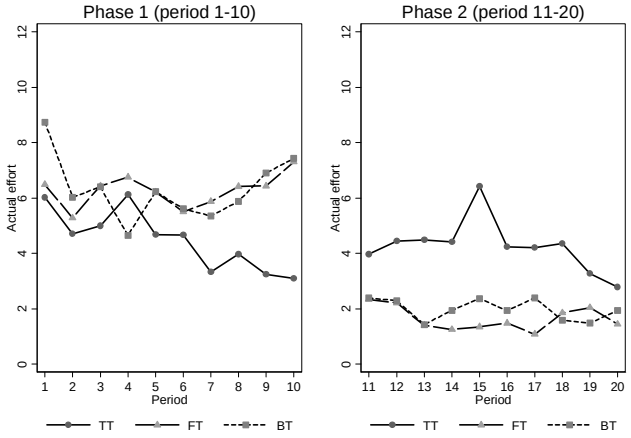


FIGURE 5A. Mean effort levels over time in Phase 1 and 2 of TTT, FT and BT.

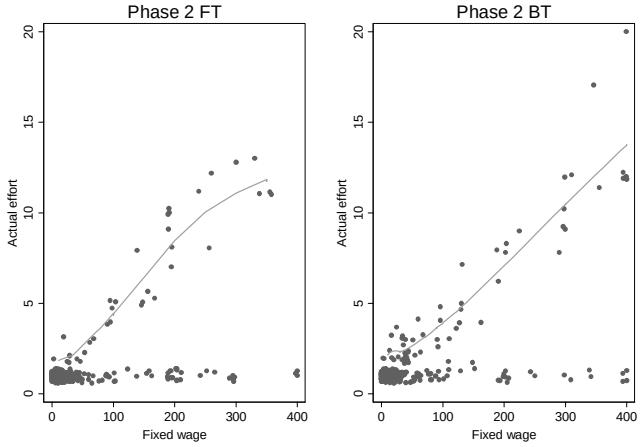


FIGURE 5B. Wages and effort in Phase 2 of FT and BT.

Figure 5b shows scatter plots of effort (in Phase 2) against fixed wage for FT (left panel) and BT (right panel). A comparison with Phase 2 of Figure 1 suggests that voluntary

cooperation is lower after the experience of Bonus (Fine) contracts. The regression in Table 5 and the predicted values graphs in Figure 6 (based on $e^d = 12$) confirm this impression.

TABLE 5
EXPLAINING PHASE 2 EFFORT CHOICES IN TTT, FT, AND BT

	probit (e>1)	tobit (e|e>1)
Fixed wage	0.006***	0.038***
	(0.001)	(0.012)
Desired effort	-0.008	0.097**
	(0.012)	(0.047)
Dummy BT	-0.295	-1.195*
	(0.210)	(0.696)
Dummy FT	-0.616***	-1.150
	(0.135)	(1.589)
Inverse Mills ratio		2.346
		(3.156)
Constant	-0.980***	-2.478
	(0.135)	(4.830)
No. of obs.	876	215
LR χ^2	123.57***	159.89***
Pseudo R^2	0.246	0.287

probit indicates a probit model whether e>1 or e=1. *tobit* indicates a tobit regression on e>1 choices only (censored at 20). Robust standard errors in parentheses; * p< 10%; ** p< 5%; *** p< 1%.

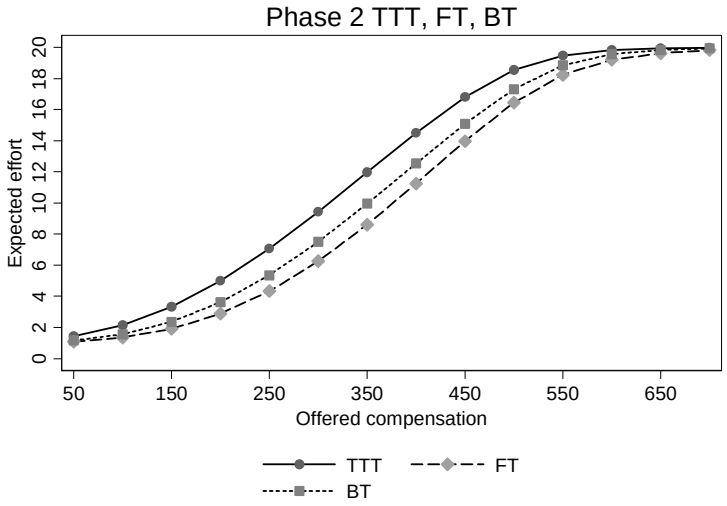


FIGURE 6. Predicted effort in Phase 2 (based on estimates of Table 5 and $e^d = 12$) as a function of wage after the experience of trust or incentive contracts.

The regression model uses the Phase 2 data of BT, FT and TTT. It is similar to the model shown in Table 3 except for introducing dummies for BT and FT. Accordingly, having experienced Fine contracts in Phase 1 has a significantly negative effect on *prob*(e >1) in Phase 2. Having experienced Bonus contracts also has a significantly negative effect on e|e>1. Thus, voluntary cooperation in Phase 2 is lower after experiencing Bonus (Fine)

contracts rather than Trust contracts in Phase 1. The differences between FT and BT are insignificant (comparing 'Dummy BT' with 'Dummy FT'; $F(1,210)=0.00$, $p=0.964$).

Result I.B summarizes our findings on separability after experiencing explicit incentives.

Result I.B. (separability *after* experiencing explicit incentive contracts). *Separability fails: The experience of Bonus or Fine contracts reduces voluntary cooperation under subsequent Trust contracts. The framing of incentives is unimportant.*

Result I.A. and I.B. document some basic effects of how explicit incentive contracts affect voluntary cooperation. Separability clearly fails. IC-contracts crowd out voluntary cooperation and the experience of explicit incentive contracts also has long-lasting negative effects on voluntary cooperation after incentives are abolished. Separability also fails under NIC-contracts where the incentives are the same than in Trust contracts.

We observe failures of separability in situations that are neither confounded with prior experiences of Trust contracts, nor with strategic incentives in repeated interactions. Yet, these are precisely two interesting and practically relevant situations, which we investigate in the remainder. The goal is to understand how the failures of separability in Results I.A. and I.B change if people are experienced with Trust contracts before they are exposed to explicit performance incentives (Section 5), and if contractual relations are embedded in an ongoing employment relationship which permits implicit incentives (Section 6).

5. RESULTS II: THE ROLE OF EXPERIENCE WITH TRUST CONTRACTS FOR FAILURES OF SEPARABILITY

In this second set of experiments we investigate how the experience of Trust contracts before being exposed to incentive contracts influences separability. For this purpose we analyze the TBT and TFT experiments in comparison to TTT (Table 2, panel B).

Experience with voluntary cooperation under Trust contracts might set a reference point that helps voluntary cooperation even under IC-contracts, and also under NIC-contracts. NIC-contracts are psychologically interesting because the principal neither sends a clear trust signal (as with a Trust contract), nor a clear signal that he relies on pay for performance (as with an IC-contract). Thus a NIC-contract is neither a clear appeal to self-interest, nor an unambiguous signal of the principal's generosity and thereby an appeal to the agent's generosity. This ambiguity may be the reason for the failure of separability observed by inexperienced subjects under NIC-contracts. However, in principle a NIC-contract that offers

a generous compensation could be understood as an appeal to cooperation as well. Agents might learn to interpret NIC-contracts this way, especially after experiencing Trust contracts.

Our analysis parallels the one in the previous section. Figure 7 illustrates behavior in Phase 2 of TBT and TFT, that is, when explicit incentives are present (analogous to Figure 2).[18] The results appear very similar to Figure 2. Under IC-contracts, agents either choose e^* (in 85.2 and 75.9 percent of cases in TFT and TBT, respectively) or they deviate to $e = 1$. Compared to Figure 2 this pattern is even more pronounced and clustering at $e = e^*$ for high levels of e^* (levels ≥ 10) is particularly strong. Together, $e = e^*$ choices and $e = 1$ choices comprise 98.6 percent of the decisions in IC-contracts. Thus, there is no voluntary cooperation at all under IC-contracts.

By contrast, there is substantial voluntary cooperation under a NIC-contract (where $e^* = 1$). Compared to Figure 2 we see more clustering at levels above 12 indicating that NIC-contracts perform better when subjects are experienced with Trust contracts.

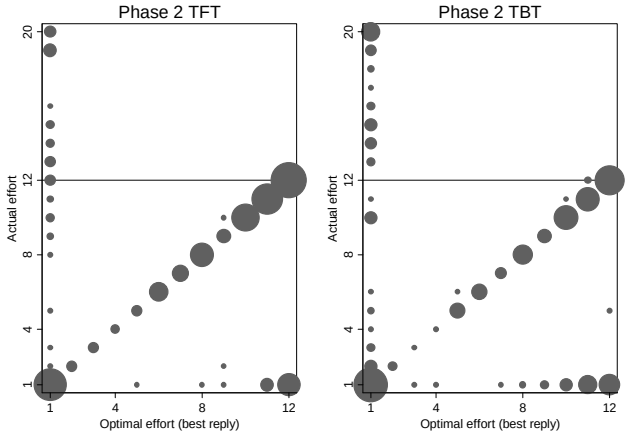


FIGURE 7. Relation between actual effort and best-reply effort in Phase 2 of TFT (left panel) and Phase 2 of TBT (right panel). The size of dots is proportional to the number of underlying observations.

Table B1 (Appendix B) reports an econometric model that is analogous to the Table 4 except that it uses the Phase 2 data of TTT, TFT and TBT. Figure 8 shows predicted based on the regression model (Table B1) and for varying compensation levels.

Looking at the left panel of Figure 8, we find similar effects as in the left panel of Figure 4: Expected efforts under IC-contracts depend positively on the offered compensation and IC-contracts are effective in implementing high effort for low compensation levels. Effort above 12 is infeasible. Trust contracts can implement higher effort than IC-contracts at high compensation levels.

[18] In Phase 2 of TFT the percentages of cases of fines of 80, 52, 24, 0 are 78.8, 11.9, 5.8, and 3.5 percent, respectively. In Phase 2 of TBT percentages for the respective bonus levels are 83.1, 9.0, 5.6, and 2.3, respectively. These distributions are significantly different from each other ($\chi^2(3)=12.39$, $p=0.006$).

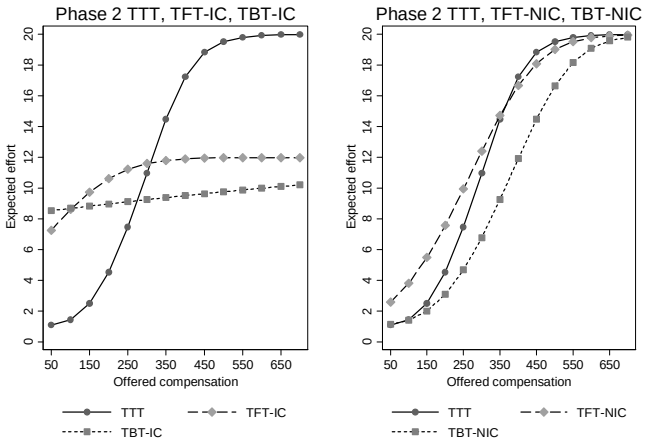


FIGURE 8. Predicted effort as a function of offered compensation (based on estimations in Table B1, for $e^d = 12$, and $b, f = 80$).

Yet, comparing the left panels of Figures 8 and 4, we also see an important difference: With experience, Trust contracts are superior to IC-contracts for a much wider range of compensation levels. Figure 8 (left panel) suggests that Trust contracts perform better than IC-contracts for all compensation levels above 270 (compared to 450 in Figure 4).

A surprising result is the framing effect under IC-contracts: under IC-Bonus contracts expected efforts are around 9 and increase only weakly in offered compensation. Under IC-Fine contracts, expected effort increases strongly at low compensation levels and approaches effort level 12 as compensation increases. This suggests that IC-Fine contracts perform better than IC-Bonus contracts when subjects are experienced with Trust contracts. A reason is that *prob*($e > 1$) depends differently on terms of the contract in TFT than TBT (Table B1).

The right panel of Figure 8 shows that NIC-contracts and Trust contracts perform similarly for experienced subjects in the sense that the correlation between offered compensation and effort is positive and similarly strong. Yet, the framing of incentives seems to matter: relative to Trust contracts there is some crowding in of effort under NIC-Fine contracts, and some crowding out under NIC-Bonus contracts.

To investigate how Trust contracts perform when applied *after* incentive contracts (Phase 2) and when subjects have experienced Trust contracts in Phase 1 we use the Phase 3 data of TTT, TFT and TBT to estimate a regression model reported in Table B2 (Appendix B) and illustrated in Figure 9. The model is analogous to Table 5.

Figure 9 reveals that compared to Figure 6 the differences between treatments have become very small. Separability holds: with the experience of Trust contracts in Phase 1, experiencing incentive contracts in Phase 2 does not crowd out voluntary cooperation in Phase 3 any longer.

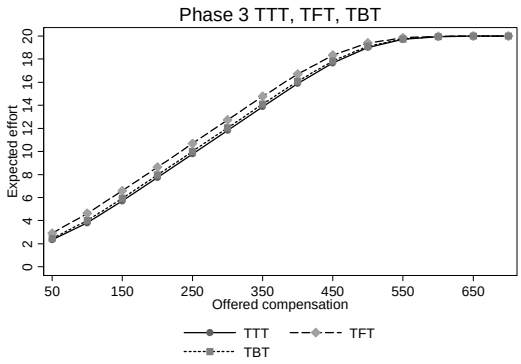


FIGURE 9. Predicted effort in Phase 3 (based on estimates of Table B2 and $e^d = 12$) as a function of wage after the experience of Trust or incentive contracts.

We summarize the findings of our second set of experiments in Result II:

Result II (The role of experience of Trust contracts for failures of separability): (a) *Under IC-contracts in Phase 2 separability fails also when agents experience Trust contracts in Phase 1. Agents choose their best-reply effort in 80.8 percent of cases and there is no voluntary cooperation.* **(b)** *Under NIC-contracts voluntary cooperation is stronger than without prior experience of Trust contracts, but separability fails and framing matters because there is crowding in* (*under Fine contracts*) *and crowding out* (*under Bonus contracts*) *relative to the Trust treatment.* **(c)** *The performance of Trust contracts in Phase 3 is independent of the prior experiences, that is, separability holds.*

Results I and II are derived from sessions with random matching to minimize strategic effects that might confound the separability issues we are interested in. However, in reality many contractual relations are long-term and the inherent implicit incentives might influence the extent of voluntary cooperation we see. How do implicit and explicit incentives interact and influence separability? We address these questions in the next section.

6. RESULTS III: THE BEHAVIORAL CONSEQUENCES OF IMPLICIT INCENTIVES

6.1 *Separability in the Presence of Implicit Incentives*

In this third set of experiments we study the role of implicit incentives by comparing the repeated game experiments TTT-R, TFT-R and TBT-R with the respective random matching experiments TTT, TFT and TBT (see Table 2, Panel C). Figures 10a-c display average effort across periods for all experiments and all phases.

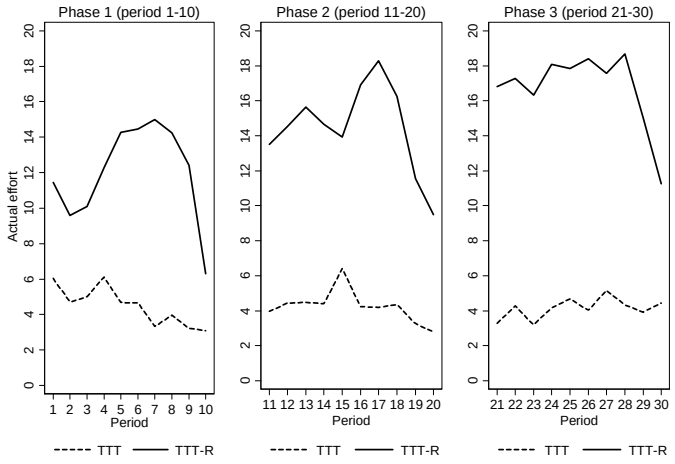


FIGURE 10A. Mean effort over time in the one-shot games TTT and in the repeated games TTT-R.

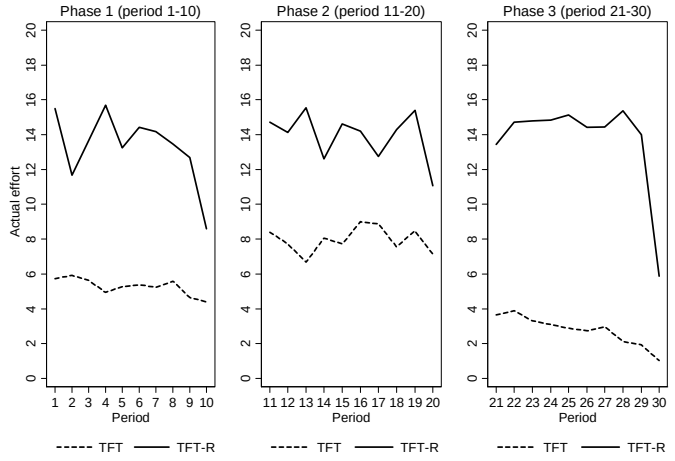


FIGURE 10B. Mean effort over time in the one-shot games TFT and in the repeated games TFT-R.

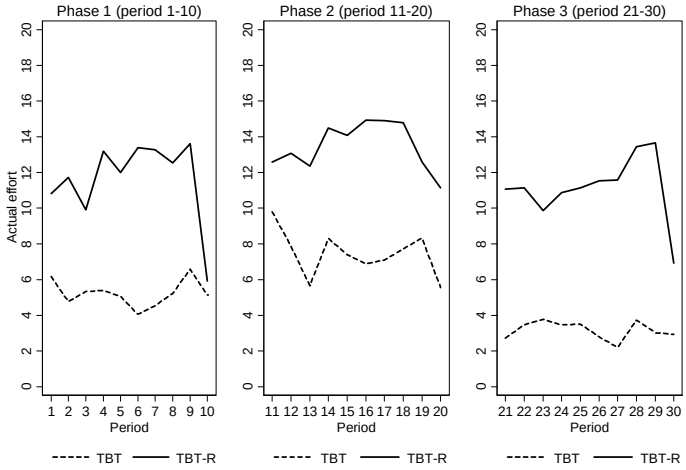


FIGURE 10C. Mean effort over time in the one-shot games TBT and in the repeated games TBT-R.

Figure 11 shows a bubble plot of actual effort against best-reply effort for the Phase 2 data of TFT-R and TBT-R. The data pattern looks the same as in Figures 2 and 7, but the

relative sizes of bubbles have changed. Contrary to Figures 2 and 7 the most frequent choice is maximal effort under NIC-contracts ($e^* = 1$). Furthermore, in the repeated games only 25.2 percent of contracts are IC whereas in TBT and TFT 74.5 percent of contracts are IC. The main reason for this striking difference is the desired effort levels. In TFT and TBT the average desired effort level is 10.8 (11.3) in Phase 2 of TFT (TBT), whereas it is 16.3 (14.9) in Phase 2 of TFT-R (TBT-R).[19]

These changes can be understood by noting that in repeated games implicit incentives can substitute for explicit incentives. Consistent with this conjecture is the observation that the relationship between offered compensation and effort is substantially stronger in all three phases of TTT-R, TFT-R and TBT-R, compared to TTT, TFT and TBT (see also Figure B1 in Appendix B). Yet, while the implicit incentives lead to very high effort levels under NIC-contracts, they do not increase effort under IC-contracts. As Figure 11 reveals, under IC-contracts, agents choose their stage game best-reply effort in the large majority of cases (75 and 89.7 percent in TFT-R and TBT-R, respectively). Agents choose $e = e^*$ despite the fact that, as the NIC-contracts show, people in the large majority of cases are willing to provide effort even beyond the maximally enforceable effort level $e = 12$.

Based on the repeated game data, Tables B4 and B5 in Appendix B report regression models analogous to those reported in Tables B1 and B2. Figures 12a and 12b provide predicted value graphs for varying compensation levels (and assuming b, $f = 80$, $e^d = 12$).

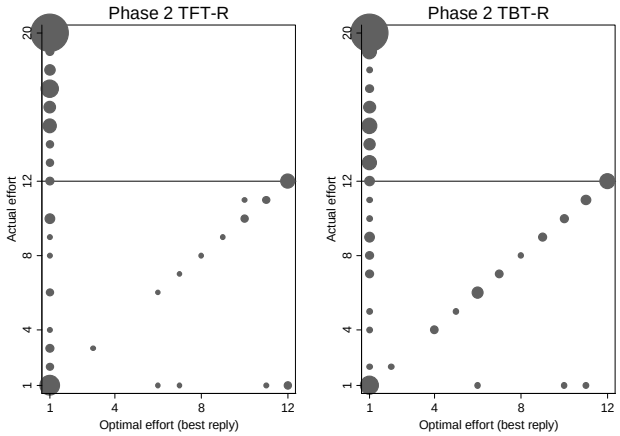


FIGURE 11: Relation between actual effort and best-reply effort in Phase 2 of TFT-R and TBT-R.

The left panel of Figure 12a highlights that in repeated games Trust contracts perform better than IC-contracts at most compensation levels. A further interesting observation is that

[19] There are some differences in the design of fine and bonuses, however. The distribution of fines and bonuses is as follows: In TFT-R the percentages of cases of fines of 80, 52, 24, 0 are 58.9, 15.4, 13.7, and 12.0 percent, respectively. In TBT-R percentages for the respective bonus levels are 70.5, 15.1, 11.4, and 3.0, respectively. These distributions are significantly different from each other ($\chi^2(3)=11.17$, p=0.011).

under IC-contracts expected effort is now independent of the offered compensation as it is predicted by the IC constraint (1). Recall that this prediction is refuted under IC-contracts in the random matching experiments of FT, BT, and TFT and TBT, where expected effort under IC-contracts depends positively on the offered compensation. Interestingly, the prediction holds however in an environment with implicit incentives.

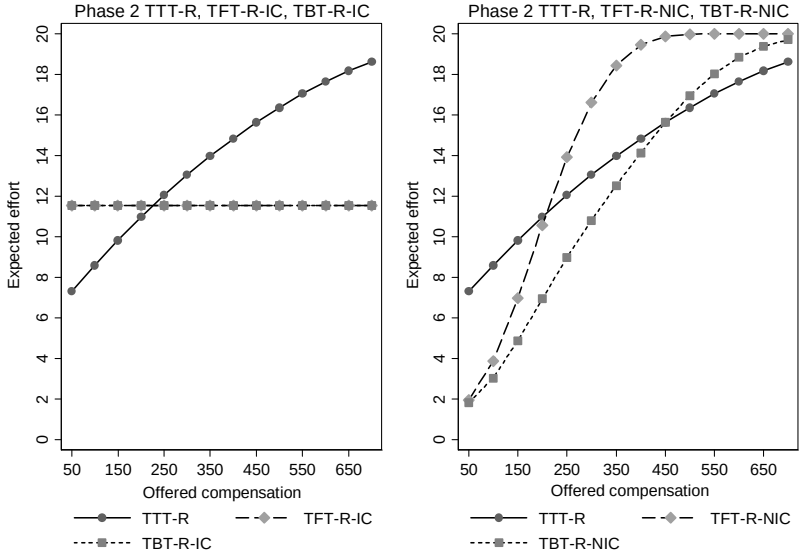


FIGURE 12A: Predicted effort as a function of offered compensation (based on estimations in Table B4, for $e^d = 12$, and b, $f = 80$.

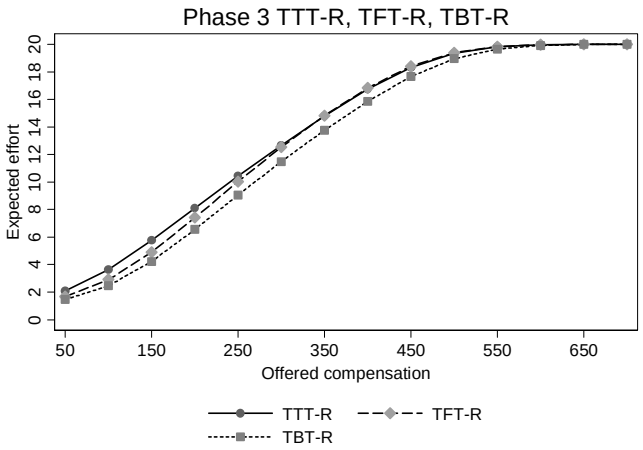


FIGURE 12B: Predicted effort in Phase 3 (based on estimates of Table B5 and $e^d = 12$) as a function of wage after the experience of trust or incentive contracts.

The right panel of Figure 12a shows an interesting failure of separability that occurs with NIC-contracts. At low compensation levels (up to about 230) Trust contracts perform substantially better than NIC-contracts. However, at larger compensation levels NIC-contracts outperform Trust contracts (and IC-contracts). Using NIC-contracts is a good

strategy if compensation is sufficiently high; if it is too low, a loss in expected effort results. We discuss the surprising effectiveness of NIC-contracts in the next subsection.

Figure 12b illustrates expected effort in Phase 3. Separability holds because there are no differences between treatments. This result makes perfect sense. Recall Result II(c), which shows that the experience of Trust contracts before being exposed to incentive contracts largely eliminates crowding in Phase 3. Moreover, in the repeated games IC-contracts are quite rare and therefore agents experience fewer consistent appeals to their self-interest which could induce them to become more selfish in Phase 3.

Our final result summarizes the behavioral consequences of implicit incentives.

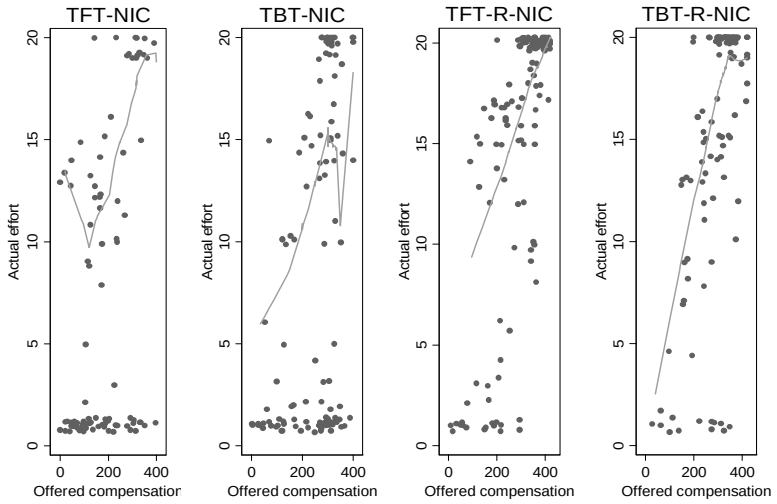
Result III (The role of implicit incentives for separability): (a) *Implicit incentives have a strongly positive impact on voluntary cooperation under Trust contracts and under NIC-contracts. 74.8 percent of contracts are NIC.* **(b)** *IC-contracts induce stage game best-reply effort in 83 percent of cases, and crowd out any voluntary cooperation: separability fails.* **(c)** *Separability also fails under NIC-contracts, which, compared to Trust contracts, induce lower efforts at low compensation but higher effort levels at high compensation. Expected effort also depends on framing.* **(d)** *Separability holds in Phase 3 because there is no crowding out effect.*

NIC-contracts perform worse than IC-contracts and Trust contracts under conditions of inexperienced subjects and random matching. But NIC-contracts perform best with experience and implicit incentives, provided compensation is sufficiently high. The question is why these factors have relatively more impact on behavior under NIC-contracts than under IC-contracts and Trust contracts. We address this question in the following subsection.

6.2 *The Surprising Effectiveness of NIC-Contracts*

If agents experience Trust contracts in Phase 1, the wage-effort relationship under NIC-Bonus and NIC-Fine contracts is strengthened, compared to a situation where agents are inexperienced with Trust contracts before being exposed to incentive contracts (compare Figures 3 and 4, with Figure 8). The presence of implicit incentives strengthens the wage-effort relationship even further. Figure 13 supports this claim (see also Figure B1 in Appendix B). Figure 13 displays scatter plots of effort against offered compensation for NIC-contracts in Phase 2 of TFT, TBT, TFT-R and TBT-R.

In the repeated games trust and its reward, strengthened by implicit incentives, are more important than explicit incentives. From this viewpoint high offered compensation should induce high effort for NIC-contracts as well as for IC-contracts. But if one notes in addition that voluntary cooperation usually does not mean that effort is chosen above desired effort but is typically chosen equal to desired effort or below (see Table B6 in Appendix B for the details), one understands why NIC-contracts outperform IC-contracts at high compensation levels. *Ceteris paribus*, NIC-contracts ask for higher effort than IC-contracts.



FIGURES 13: Offered compensation and effort under non-incentive-compatible contracts in Phase 2 of TFT, TBT, TFT-R and TBT-R.

A remaining question is why NIC-contracts outperform Trust contracts (for high compensation levels in repeated games) even though Trust contracts seem an even stronger, unambiguous appeal to rewarding trust than NIC-contracts. One reason may be the higher frequency of choices e with $1 < e < e^d$ under Trust contracts than under NIC-contracts (52.3 vs. 34.5 percent; see Table B6). By choosing effort between minimal and desired effort the agent may fine-tune the distribution of earnings between principal and agent. This is more difficult in NIC-contracts than in Trust contracts, because in NIC-contracts $e < e^d$ implies that the entire fine is to be paid or that the entire bonus is lost. Consequently, under NIC-contracts, if the agent has decided to provide higher than minimal effort at all, it makes less sense to marginally undercut desired effort. In turn, choosing $e \geq e^d$ is a stronger cooperative signal than choosing $1 < e < e^d$. The higher relative frequency of choices $e \geq e^d$ under NIC-contracts might therefore facilitate and improve cooperation relative to Trust contracts.

This interpretation is also supported by Figure 14 which displays *prob*($e > 1$) in the left panel and $E(e|e>1)$ in the right panel according to the regression model (see Table B4). For high compensation levels we see that *prob*($e > 1$) is about equal to 1 for Trust contracts as

well as NIC-contracts. But $E(e|e>1)$ is higher for NIC-contracts (especially NIC fine contracts) than for Trust contracts.

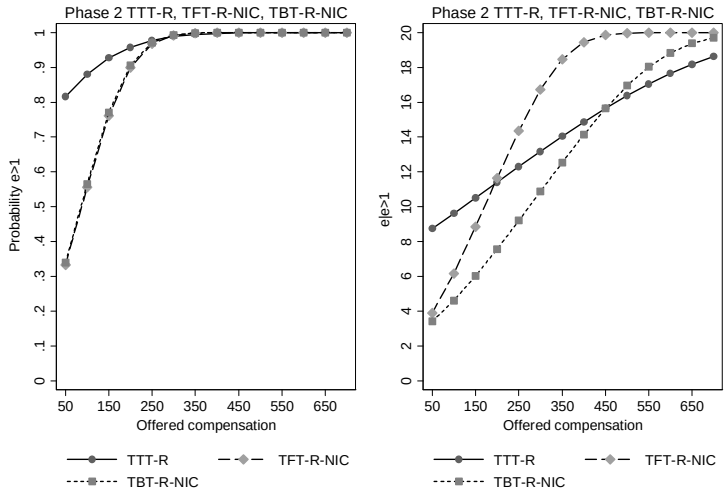


FIGURE 14: Behavior in NIC-contracts in the presence of implicit incentives. LEFT PANEL: Probability to choose non-minimal effort; RIGHT PANEL: predicted effort conditional on non-minimal effort.

Overall we conclude that, when backed up by implicit incentives, NIC-contracts perform better than IC-contracts in repeated games and for high compensation levels because NIC-contracts give more leverage to trust and its reward, whereas IC-contracts focus agents on their stage-game best reply. NIC-contracts perform better than Trust contracts because Bonus and Fine contracts reduce the frequency of $e < e^d$. While NIC-contracts are quite effective for high compensation levels they are risky when compensation is low because effort is rather low and even lower than under IC-contracts and Trust contracts.

7. SUMMARY AND CONCLUSIONS

In this paper we investigated the roles of explicit and implicit incentives and voluntary cooperation for contractual compliance. We focused in particular on the question whether incentives and voluntary cooperation are separable, that is, whether incentive contracts leave the extent of voluntary cooperation unchanged. This is an important question because arguably non-selfish voluntary cooperation is necessary and desirable in many real world contractual relations where self-regarding incentives are present as well. On an abstract theoretical level there is no reason to assume that separability between social preferences (which motivate voluntary cooperation) and incentives should hold. But separability is frequently assumed in theoretical models (Bowles and Hwang (2008)) and whether it holds is an empirical question, which we addressed in this paper.

We are certainly not the first to investigate issues of separability.[20] However, to our knowledge no study has investigated several potential failures of separability of explicit and implicit incentives in one comparable design with the goal to uncover robust conditions when separability holds and when it fails. Our main findings are as follows:

First, *separability fails robustly while being exposed to incentive contracts when they are designed in an incentive-compatible way*: As predicted by standard incentive theory, IC-contracts almost invariably induce agents to choose their best-reply effort and thereby unambiguously crowd out voluntary cooperation. Such best-reply/crowding out occurs in all six experiments with the possibility of incentive compatible contracts: we observe it in our first set of experiments (FT & BT); it is robust to experience with Trust contracts (TFT & TBT); and it also occurs in repeated games with otherwise strong implicit incentives (TFT-R & TBT-R). Except for TFT and TBT, the framing of incentives does not matter much.

Second, *separability while being exposed to incentive contracts also fails in all six experiments with incentive contracts, when the principal offers NIC-contracts*. However, experience with Trust contracts before being exposed to NIC-contracts strongly shapes the exact form the failures of separability take. The framing of incentives also matters.

Third, *separability after being exposed to incentive contracts only holds for agents who are experienced with Trust contracts prior to their exposure to incentive contracts*. Separability fails for agents inexperienced with prior Trust contracts (FT & BT), but it holds with experience (TFT & TBT) and/or in repeated games (TFT-R & TBT-R).

So, are incentives harmful for voluntary cooperation? On the basis of our eight experiments, we can give a differentiated answer. Implicit incentives coming from repeated interaction are strongly beneficial for voluntary cooperation. By contrast, explicit incentives are always harmful for voluntary cooperation under IC-contracts. But when contracts are not incentive compatible, even crowding in of voluntary cooperation, relative to Trust contracts *can* occur, provided agents are at least experienced with Trust contracts or implicit incentives are present and the principal offers high enough a wage.

Our overall conclusion is that assuming separability of the 'material interests' from the 'moral sentiments' in theoretical models of contractual compliance is not warranted, in particular not if contracts are assumed to be incentive compatible. Our results suggest that people will follow the incentives they face, at the detriment of voluntary cooperation, despite the fact that most people are not selfish. Utilizing the benefits of voluntary cooperation requires non-incentive compatible contracts and strong implicit incentives.

[20] For overviews of important findings see Frey and Jegen (2001); Fehr and Falk (2002); Bowles (2008); Bowles and Polanía Reyes (2011).

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.

General information

The experiment in which you participate today is conducted jointly with 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 an endowment of 1500 points. It is possible that some decisions lead to losses. You will have to finance them out of the gains from your other decisions, or, if necessary out of your endowment. **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 through your decisions will be summed up, exchanged into Swiss Francs and paid out in cash.

Please note that during the experiment communication is not allowed. If you have questions, please raise your hand. We will answer your questions in private.

Instructions

1. Introduction

In this experiment you will learn about a decision problem that involves two people. We will call them participant X and participant Y. **All participants in this experiment are allocated into two groups: the group of participants X and the group of participants Y. After the experiment has started you can see on your computer screen 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 on the computer. Your decisions will be transmitted via the 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 your understanding if you think about the following scenario. Participant X decides in the role of a "firm". The “firm” engages an “employee” (participant Y), whose 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 is determined. 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. In accordance with the rules participant X proposes an employment contract including the fixed salary and the "**desired effort"**.
2. Participant Y decides to accept or reject the contract.
3. Y chooses his actual 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 randomly matched with another person in each period.

3. The experimental details

3.1 Employment contract: The proposal of participant X

At the beginning of **each** period an **employment contract** will be determined. For the design of the contract the following holds:

The **proposed contract** consists of **two** components: a **fixed wage** and a **desired effort**. Participant X is free – in accordance with the rules mentioned below – to design any contract.

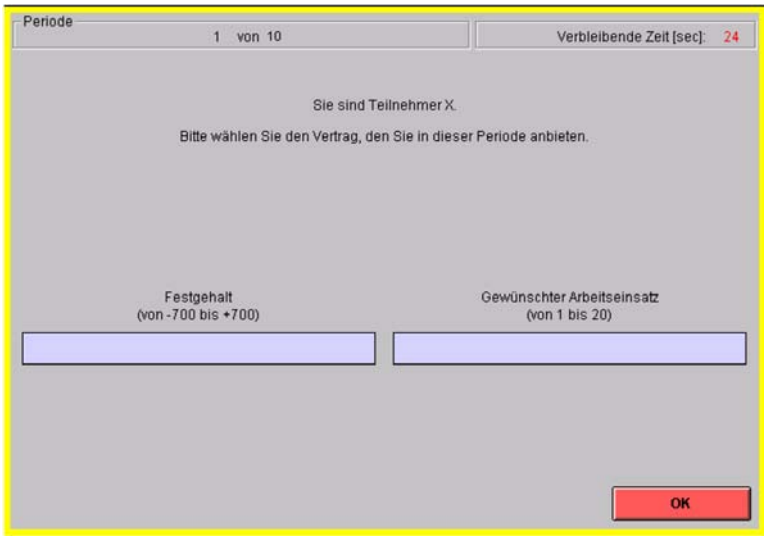
- The contract can contain a *positive* or a *negative* **fixed salary**. If the fixed salary is positive, this means that participant Y receives the wage from participant X, regardless of the period return. A negative fixed wage means that Y has to pay that amount to X, regardless of the period return.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

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

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

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

To clarify the rules, 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 the top right. Participant X makes his proposed employment contract on this screen.

3.2 Employment contract: Acceptance of the contract by participant Y

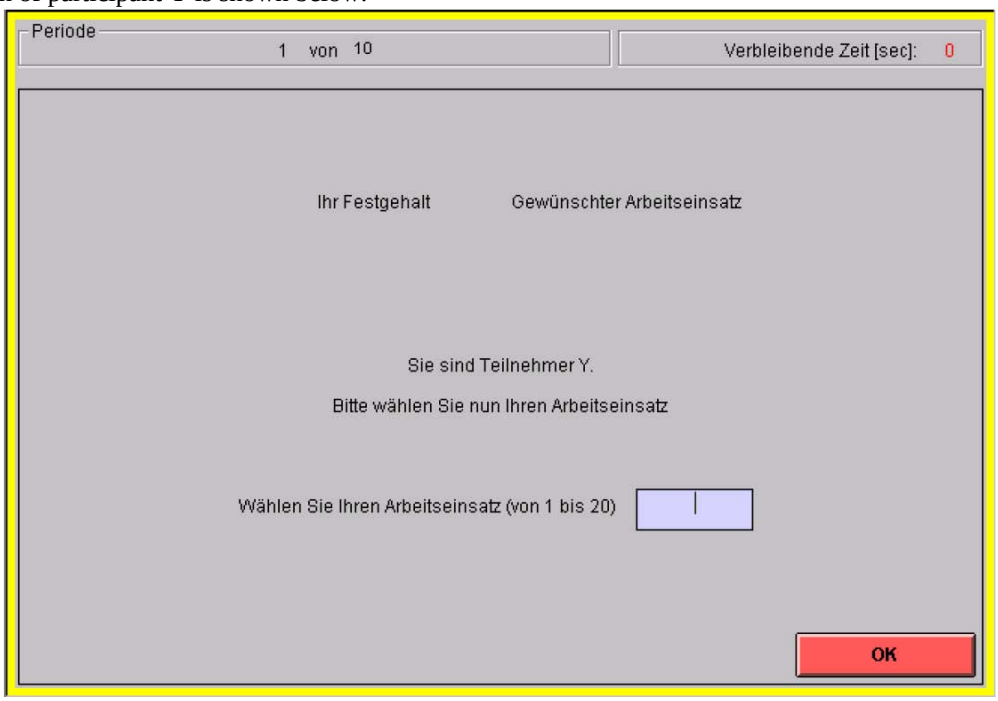
After participant Y has received the proposed contract, he has to decide whether to accept or reject the contact.

3.3 Work effort of participant Y

After Y has accepted the contract, Y determines his **work effort**. The desired work effort stated by participant X in the contract is not binding for participant Y. Work effort is symbolized by a number. In the enclosed **table** all possible work efforts (all integer numbers between 1 and 20) as well as the produced returns are given. The table

also contains the **cost**s of work effort 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 chosen his work effort, the period gains will be calculated and displayed on the screen. The following cases result for the calculation of the profits:

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

After this-screen the period is finished and the next one starts. There are 10 periods in total.

Work effort, period return from work effort and costs of work effort for Y:

Work effort :	**Period return** from 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

Period profit of Y: Fixed salary – costs of the effective work effort

Period profit of X: Period return of the effective work effort – fixed salary

Period profit of Y and X by rejection of the contract of Y: Zero

Only the <u>actual</u> work effort is relevant for the calculation of the profits!

Information on 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 do not get to know the other person's identity. As before all decisions are anonymous.

The **only change** compared to the previous experiment consists of the contract possibilities that X can offer. In addition to the fixed salary and the desired effort participant X determines a **potential wage reduction**, which is due if Y chooses a work effort that is *below* X's desired effort. If Y choose an actual work effort which is higher or equal than the desired effort than the wage reduction is not due. There are four possible levels of potential wage reductions: The potential wage reduction can be *either* **0** *or* **24** *or* **52** *or* **80. The wage reduction is <u>only</u> due if the actual effort is lower than the desired effort!**

For the contract design the following rules hold:

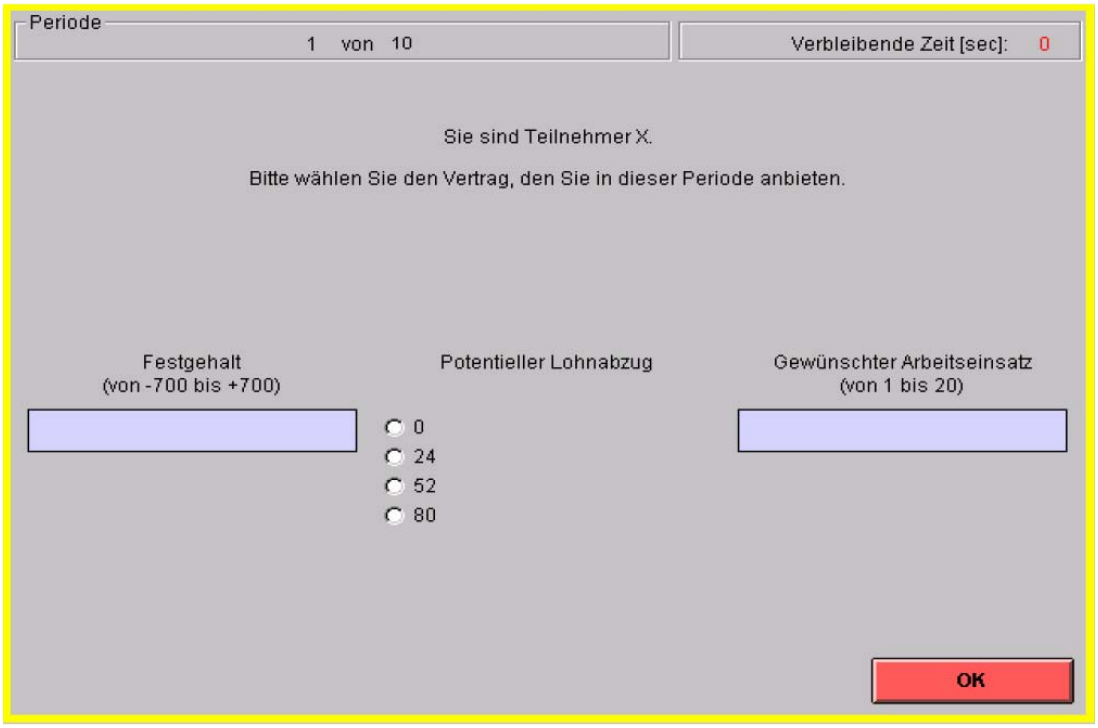
$-700 \leq$ fixed wage ≤ 700

Potential wage reduction: *either* 0 *or* 24 *or* 52 *or* 80

$1 \leq$ desired work effort ≤ 20

In designing the contract ALL integer combinations that are compatible with these rules are possible!

The rules are clarified by the following input screen of X:



The profits are calculated as follows:

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

Otherwise this experiment is entirely **identical** to the previous experiment!

APPENDIX B: SUPPORTING ANALYSES

APPENDIX B1: SUPPORTING ANALYSIS FOR RESULTS II

Measuring incentive effects and separability in Phase 2 of TFT and TBT

The regression model reported in Table B1 is similar to the one reported in Table 4 in the paper. One difference is that we include a variable "Individual cooperation level" to measure the average effort chosen by the individual agent in Phase 1. We included this to control for an individual (fixed effect) inclination to provide high effort. Unlike Table 4 the Tobit regression in the second column of Table B1 contains only Dummy (IC) as single regressor to capture the effect of IC-contracts. We refrained from adding more explanatory variables because there were only 7 observations $e \neq e^*$ (given $e>1$).

TABLE B1
EXPLAINING PHASE 2 EFFORT CHOICES IN TREATMENTS TTT, TFT AND TBT

	probit (e>1)	tobit (e|e>1)
Fixed wage (Trust games)	0.006***	0.066***
	(0.001)	(0.011)
Desired effort (Trust games)	-0.012	0.082
	(0.014)	(0.127)
Individual cooperation level in Phase 1	0.084***	0.481***
	(0.015)	(0.162)
Fixed wage (TBT, IC)	0.001	
	(0.002)	
Dummy variable (TBT, IC)	1.865***	
	(0.391)	
Desired effort (TBT, IC)	-0.029	
	(0.039)	
Fixed wage (TFT, IC)	0.007***	
	(0.002)	
Dummy variable (TFT, IC)	5.084***	
	(1.222)	
Desired effort (TFT, IC)	-0.352***	
	(0.114)	
Dummy variable IC		17.631***
		(3.114)
Fixed wage (TBT, NIC)	0.005***	0.045***
	(0.002)	(0.014)
Dummy variable (TBT, NIC)	0.728	3.097
	(0.458)	(2.773)
Desired effort (TBT, NIC)	-0.054	0.379
	(0.039)	(0.445)
Fixed wage (TFT, NIC)	0.005***	0.039***
	(0.002)	(0.011)
Dummy variable (TFT, NIC)	1.687***	8.874***
	(0.563)	(3.345)
Desired effort (TFT, NIC)	-0.126***	0.200
	(0.042)	(0.296)
Inverse Mills ratio		9.463***
		(2.880)
Constant	-1.408***	-15.963***
	(0.135)	(4.532)
No. of obs.	981	231
	$\chi2$=272.76***	F=230.27***
Pseudo R2	0.276	0.150

Measuring separability in Phase 3 of TFT and TBT

The regression model reported in Table B2 is the same as the one reported in Table 5 in the paper except for adding the variable "Individual cooperation level in Phase 1" (see Table B1).

TABLE B2
CROWDING EFFECTS IN PHASE 3 (COMPARING PHASE 3 DATA IN TFT AND TBT WITH TTT)

	probit (e>1)	tobit (e|e>1)
Fixed wage	0.006***	0.041***
	(0.001)	(0.009)
Desired effort	-0.014	0.117
	(0.017)	(0.073)
Dummy TBT treatment	-0.363	0.224
	(0.246)	(0.890)
Dummy TFT treatment	-0.598***	0.905
	(0.213)	(0.822)
Individual cooperation level in Phase 1	0.166***	0.351
	(0.018)	(0.260)
Inverse Mills ratio		2.781
		(2.451)
Constant	-1.689***	-6.292
	(0.247)	(4.747)
No. of obs.	929	300
LR chi2	279.34***	125.54***
Pseudo R^2	0.364	0.234

Probit (e>1) indicates a probit model whether e>1 or e=1. *tobit* (e|e>1) indicates a tobit regression on e>1 choices only (censored at 20).
Robust standard errors in parentheses; * p< 10%; ** p< 5%;
*** p< 1%.

APPENDIX B2: SUPPORTING ANALYSIS FOR RESULTS III

Measuring the impact of implicit incentives on effort levels (*comparing one-shot games with repeated games*)

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') which 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'). In phases 2 of treatments TFT, TBT, TFT-R and TBT-R we also included the best-reply effort e^*.

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 for simple tobit estimates because we are only interested in the means. We report the results in Table B3.

The most important result for your purposes is that the variable 'Dummy Repeated Game' (see bold face in Table B3) is highly significantly positive, holding everything else constant. Thus, implicit incentives increase cooperation in our experiment.

Further interesting findings are as follows. 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.[21] Although still substantial and highly significant, the average increase in the repeated games as measured by 'Dummy Repeated Game' 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.

TABLE B3
MEASURING THE IMPACT OF IMPLICIT INCENTIVES (COMPARING ONE-SHOT GAMES WITH REPEATED GAMES)

	Dependent variable: Actual effort		
Data used:	TTT & TTT-R	TFT & TFT-R	TBT & TBT-R
Offered compensation	0.044***	0.051***	0.052***
	(0.003)	(0.005)	(0.003)
Best-reply effort (*e**)		0.730***	0.815***
		(0.089)	(0.178)
Dummy Phase 2 one-shot	1.042**	0.729	3.318***
	(0.429)	(0.989)	(0.852)
Dummy Phase 3 one-shot	0.716	-2.627***	-0.531
	(0.593)	(0.717)	(1.105)
Dummy Repeated Game	**5.946*** **	**4.472*** **	**5.257*** **
	(1.381)	**(1.285)**	**(1.514)**
Dummy Phase 2 repeated	1.848***	0.015	3.328***
	(0.609)	(1.135)	(0.839)
Dummy Phase 3 repeated	3.444***	1.235	0.574
	(0.798)	(0.917)	(0.709)
Endgame effect	-1.686***	-1.023**	0.321
	(0.420)	(0.448)	(0.538)
Constant	-6.308***	-6.555***	-8.383***
	(1.028)	(1.231)	(1.683)
Observations	1214	1541	1418
Wald chi2	68.91***	33.22***	67.95***

Tobit regression (censored at 0 and 20). Robust standard errors in parentheses; * p < 10%; ** p < 5%; *** p < 1%.

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

Figure B1 presents predicted effort plots separately for each of the phases of treatments TTT and TTT-R; TFT and TFT-R, and TBT and TBT-R. We find that the predicted relationship between the offered compensation and actual effort across all phases of the one-shot games and the repeated games is indeed steeper in the repeated games than in the one-shot games. Thus, implicit incentives "crowd in" reciprocal behavior

[21] One may view this decrease in effort as a crowding out effect *relative to voluntary effort in Phase 1*. Our definition of crowding effects 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.

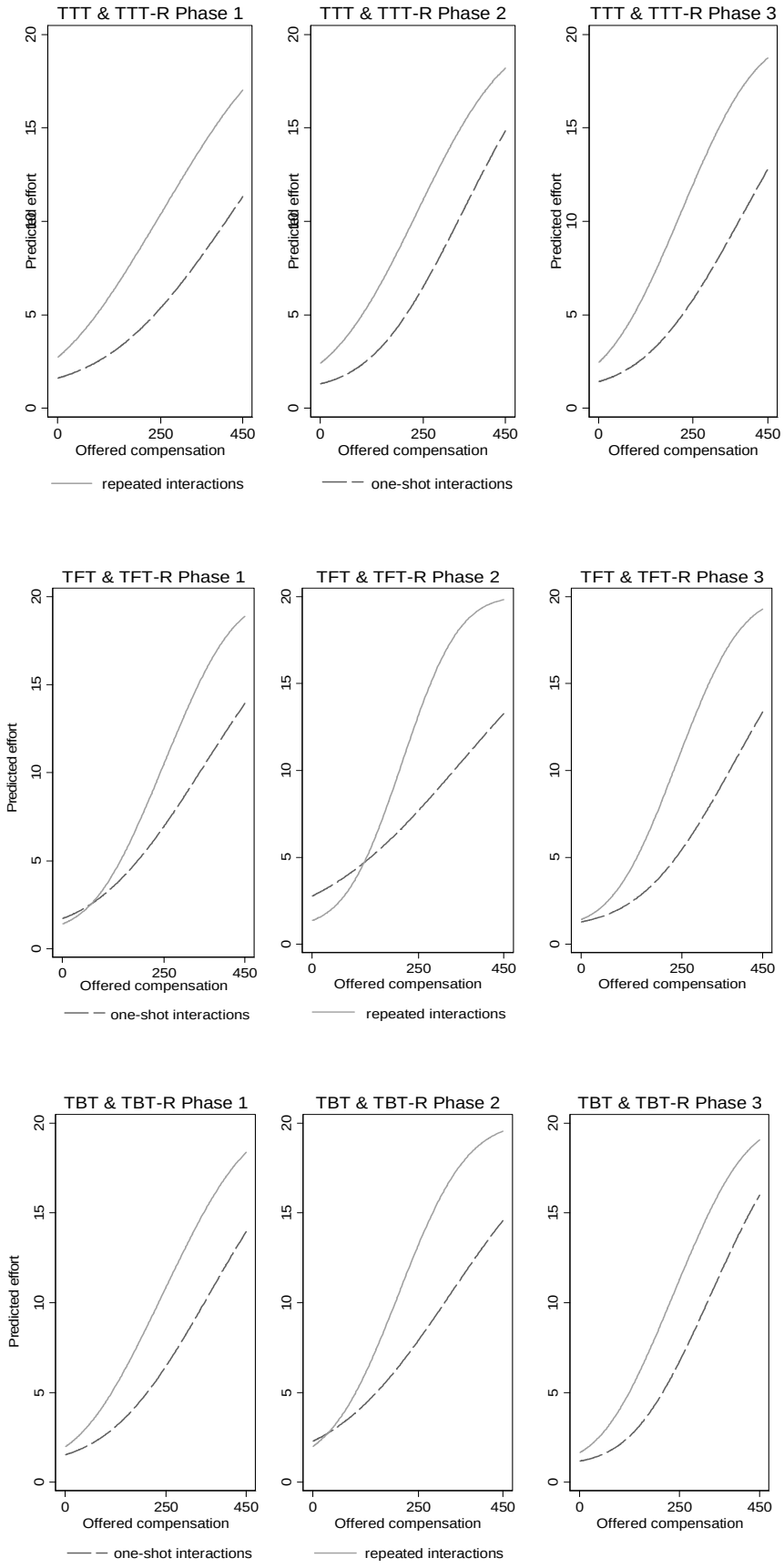


FIGURE B1: The predicted relationship between offered compensation and actual efforts in one-shot and repeated games

Measuring incentive effects and separability in Phase 2 of TFT-R and TBT-R

The regression model reported in Table B4 is analogous to the one reported in Table 4 in the paper and in Table B1. In these regressions we exclude the data of the final three periods of each phase to downplay influences of endgame effects. In contrast to Tables 4 and B1 we did not run the probit regression to estimate $prob(e = e^*|e > 1)$. Since there were almost no deviations (see Figure 11 in the paper) this probability is virtually equal to 1.

TABLE B4
EXPLAINING PHASE 2 EFFORT CHOICES IN TREATMENTS TTT-R, TFT-R AND TBT-R

	probit (e>1)	tobit (e|e>1)
Fixed wage (Trust games)	0.006	0.018
	(0.005)	(0.017)
Desired effort (Trust games)	-0.016	0.717*
	(0.077)	(0.366)
Individual cooperation level in Phase 1	0.079**	0.258
	(0.035)	(0.199)
Dummy variable IC	0.952	
	(0.684)	
Fixed wage (BT, NIC)	0.012**	0.034**
	(0.005)	(0.015)
Dummy variable (BT, NIC)	2.336	-3.476
	(1.857)	(4.465)
Desired effort (BT, NIC)	-0.268*	0.639*
	(0.155)	(0.367)
Fixed wage (FT, NIC)	0.011***	0.058**
	(0.002)	(0.025)
Dummy variable (FT, NIC)	-0.817	1.016
	(1.486)	(4.475)
Desired effort (FT, NIC)	-0.083	-0.000
	(0.089)	(0.466)
Inverse Mills ratio		1.857
		(5.669)
Constant	-0.176	-4.630
	(0.621)	(3.302)
No. of obs.	289	234
χ^2	94.05***	24.98***
Pseudo R^2	0.344	0.173

Phase 2 data of TTT-R, TFT-R and TBT-R. *Probit* (e>1) indicates a probit model whether e>1 or e=1. *Tobit* (e|e>1) indicates a Tobit regression on e>1 choices only (censored at 20). Robust standard errors in parentheses; * p< 10%; ** p< 5%; *** p< 1%.

Measuring separability in Phase 3 of TFT-R and TBT-R

The regression model reported in Table B5 is analogous to the ones reported in Tables 5 and B2.

TABLE B5
CROWDING EFFECTS IN PHASE 3 (COMPARING PHASE 3 DATA IN TFT-R AND TBT-R WITH TTT-R)

	probit (e>1)	tobit (e|e>1)
Fixed wage	0.009***	0.042***
	(0.002)	(0.010)
Desired effort	0.002	0.570***
	(0.034)	(0.219)
Dummy TBT-R treatment	-0.572	-0.964
	(0.541)	(0.89)
Dummy TFT-R treatment	-0.615	0.153
	(0.527)	(0.971)
Individual cooperation level in Phase 1	-0.046	0.103
	(0.03)	(0.09)
Inverse Mills ratio		10.965***
		(3.995)
Constant	0.313	-9.265***
	(0.614)	(2.922)
No. of obs.	304	270
LR chi2	57.24***	77.39***
Pseudo R^2	0.510	0.287

probit indicates a probit model whether e>1 or e=1. *tobit* indicates a tobit regression on e>1 choices only (censored at 20). Robust standard errors in parentheses; * p< 10%; ** p< 5%; *** p< 1%.

Explaining the surprising effectiveness of NIC-contracts in repeated games – the role of desired effort

TABLE B6
ACTUAL AND DESIRED EFFORT IN TTT-R, TFT-R AND TBT-R

	IC-contracts (TFT-R, TBT-R)		NIC-contracts (TFT-R, TBT-R)		Trust contracts (TTT-R)	
cases	#	percent	#	percent	#	percent
$e > e^d$	1	1.89%	87	10.13%	8	2.40%
$e = e^d$	44	83.02%	355	41.33%	133	39.94%
$1 < e < e^d$	-	-	296	34.46%	174	52.25%
$e = 1$ ($e^d > 1$)	8	15.09%	121	14.09%	18	5.41%
Sum	53	100%	859	100%	333	100%

REFERENCES

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

ANDERHUB, V., S. GÄCHTER, and M. KÖNIGSTEIN (2002): "Efficient Contracting and Fair Play in a Simple Principal-Agent Experiment," *Experimental Economics*, 5, 5-27.

ANDREONI, J., and D. B. BERNHEIM (2009): "Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects," *Econometrica*, 77, 1607-1636.

BANDIERA, O., I. BARANKAY, and I. RASUL (2005): "Social Preferences and the Response to Incentives: Evidence from Personnel Data," *Quarterly Journal of Economics*, 120, 917-962.

BÉNABOU, R., and J. TIROLE (2006): "Incentives and Prosocial Behavior," *American Economic Review*, 96, 1652-1678.

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

BEWLEY, T. F. (2007): "Fairness, Reciprocity, and Wage Rigidity," in *Behavioral Economics and Its Applications*, ed. by P. Diamond, and H. Vartiainen. Princeton: Princeton University Press, 157-188.

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

BOWLES, S. (2003): *Microeconomics: Behavior, Institutions, and Evolution*. Princeton: Princeton University Press.

— (2008): "Policies Designed for Self-Interested Citizens May Undermine "the Moral Sentiments": Evidence from Economic Experiments," *Science*, 320, 1605-1609.

BOWLES, S., and S.-H. HWANG (2008): "Social Preferences and Public Economics: Mechanism Design When Social Preferences Depend on Incentives," *Journal of Public Economics*, 92, 1811-1820.

BOWLES, S., and S. POLANÍA REYES (2011): "Social Preferences and Self-Interest: Why Do Economic Incentives Sometimes under-Perform?," Mimeo, Santa Fe Institute.

BROWN, M., A. FALK, and E. FEHR (2004): "Relational Contracts and the Nature of Market Interactions," *Econometrica*, 72 3, 747-80.

CAMERER, C. F. (2003): *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton: Princeton University Press.

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

CHARNESS, G., and P. KUHN (2011): "Lab Labor: What Can Labor Economists Learn from the Lab?," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card: Elsevier, 229-330.

COX, J. C., D. FRIEDMAN, and V. SADIRAJ (2008): "Revealed Altruism," *Econometrica*, 76, 31-69.

CROSON, R., and S. GÄCHTER (2010): "The Science of Experimental Economics," *Journal of Economic Behavior & Organization*, 73, 122-131.

DUFWENBERG, M., S. GÄCHTER, and H. HENNIG-SCHMIDT (2011): "The Framing of Games and the Psychology of Play," *Games and Economic Behavior*, in press, doi: 10.1016/j.geb.2011.02.003.

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.

ELLINGSEN, T., and M. JOHANNESSON (2008): "Pride and Prejudice: The Human Side of Incentive Theory," *American Economic Review*, 98, 990-1008.

FALK, A. (2007): "Gift Exchange in the Field," *Econometrica*, 75, 1501-1511.

FALK, A., and U. FISCHBACHER (2006): "A Theory of Reciprocity," *Games and Economic Behavior*, 54, 293-315.
FALK, A., S. GÄCHTER, and J. KOVACS (1999): "Intrinsic Motivation and Extrinsic Incentives in a Repeated Game with Incomplete Contracts," *Journal of Economic Psychology*, 20, 251-284.
FALK, A., and J. J. HECKMAN (2009): "Lab Experiments Are a Major Source of Knowledge in the Social Sciences," *Science*, 326, 535-538.
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 (2002): "Do Incentive Contracts Undermine Voluntary Cooperation?," IEW Working Paper No. 34, 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., L. GOETTE, and C. ZEHNDER (2009): "A Behavioral Account of the Labor Market: The Role of Fairness Concerns," *Annual Review of Economics*, 1, 355-384.
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 B. ROCKENBACH (2003): "Detrimental Effects of Sanctions on Human Altruism," *Nature*, 422, 137-140.
FEHR, E., and K. M. SCHMIDT (1999): "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics*, 114, 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.
GINTIS, H., S. BOWLES, R. BOYD, and E. FEHR eds. (2005): *Moral Sentiments and Material Interests. The Foundations of Cooperation in Economic Life*. Cambridge: MIT Press.
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.
HEYMAN, J., and D. ARIELY (2004): "Effort for Payment - a Tale of Two Markets," *Psychological Science*, 15, 787-793.
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.
LAZEAR, E. P. (2000): "Performance Pay and Productivity," *The American Economic Review*, 90, 1346-1361.
LEVINE, D. K. (1998): "Modeling Altruism and Spitefulness in Experiments," *Review of Economic Dynamics*, 1 3, 593-622.
MACLEOD, W. B. (2007): "Reputations, Relationships, and Contract Enforcement," *Journal of Economic Literature*, 45, 595-628.
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.

SHEARER, B. S. (2004): "Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment," *Review of Economic Studies*, 71, 513-534.
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.
SLIWKA, D. (2007): "Trust as a Signal of a Social Norm and the Hidden Costs of Incentive Schemes," *The American Economic Review*, 97, 999-1012.
SOBEL, J. (2005): "Interdependent Preferences and Reciprocity," *Journal of Economic Literature*, 43, 392-436.
WILLIAMSON, O. (1985): *The Economic Institutions of Capitalism*. New York: Free Press.
WOOLDRIDGE, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press.