



Institute for Empirical Research in Economics  
University of Zurich

Working Paper Series  
ISSN 1424-0459

---

Working Paper No. 34

**Do Incentive Contracts Undermine Voluntary  
Cooperation?**

Ernst Fehr and Simon Gächter

April 2002

---

# Do incentive contracts undermine voluntary cooperation?<sup>\*</sup>

ERNST FEHR<sup>\*\*</sup>

*University of Zürich and CEPR*

SIMON GÄCHTER<sup>\*\*\*</sup>

*University of St. Gallen and CESifo*

**April 2002**

**Abstract:** In this paper we provide experimental evidence indicating that incentive contracts may undermine voluntary cooperation. This suggests that explicit incentives may have costly side effects that have been largely neglected by economists. In our experiments the undermining effect is so strong that the incentive contracts are less efficient than contracts without any incentives. Buyers, who are in the role of principals, nonetheless, prefer the incentive contracts because they allow them to appropriate a much larger share of the (smaller) total surplus and are, hence, more profitable for them. The undermining of voluntary cooperation through incentives is, in principle, consistent with models of inequity aversion and reciprocity. Additional experiments show, however, that the reduction of voluntary cooperation through incentives is partly due to a framing effect. If the incentive is framed as a price deduction the reduction of voluntary cooperation is much stronger compared to a situation where the incentive is framed as a bonus paid on top of a base price.

**JEL-Classification:** J41, C91, D64

**Keywords:** Incentive contracts, reciprocity, incomplete contracts, voluntary cooperation, experiments.

---

<sup>\*</sup> This paper is part of the EU-TMR Research Network ENDEAR (FMRX-CT98-0238). Support from the Swiss National Science Foundation under project number 1214-051000.97 and from the MacArthur Foundation Network on Economic Environments and the Evolution of Individual Preferences and Social Norms is gratefully acknowledged. Gary Charness, Bruno Frey, Werner Güth, Patrick Legros, George Loewenstein, Sendhil Mullainathan, Axel Ockenfels, Andrew Schotter, Robert Slonim, and George Wu, and the participants at the Verein für Socialpolitik in Koblenz, the GEW-meeting in Meissen, the ESA conference in Lake Tahoe, the IAREP meeting in Belgirate, the Francqui conference on Contract Theory in Brussels, the First Asian Conference on Experimental Business Economics in Hong Kong and the ASSA meetings in New Orleans, as well as seminar participants in Saarbrücken and Berlin provided helpful comments. We are particularly indebted to Armin Falk and Beatrice Zanella for many discussions and support.

<sup>\*\*</sup> University of Zurich, Institute for Empirical Economic Research, Blümlisalpstrasse 10, CH-8006 Zurich. Phone: 0041-1-6343709. Fax: 0041-1-6344907. E-mail: [efehr@iew.unizh.ch](mailto:efehr@iew.unizh.ch).

<sup>\*\*\*</sup> University of St. Gallen, FEW-HSG, Varnbuelstrasse 14, CH-9000 St. Gallen. Phone: 0041-71-224-2535. Fax: 0041-71-224-2302. E-mail: [simon.gaechter@unisg.ch](mailto:simon.gaechter@unisg.ch).

## 1. INTRODUCTION

Economic theories almost always rely in some way on the effectiveness of performance incentives. From the economist's viewpoint the question why certain actions are undertaken is closely related to the question which incentives the decision-makers are facing. There can thus be little doubt that the effectiveness and the limits of performance incentives belong to the fundamental questions in economics. In this paper we examine experimentally how buyers in a buyer-seller relationship use the available incentives and to what extent the provision of performance incentives undermines the sellers' propensity to cooperate voluntarily.

Our results show that incentive contracts may undermine voluntary cooperation. In our experiments the reduction in voluntary cooperation is so strong that contracts providing explicit performance incentives are on average less efficient than fixed-price contracts that do not provide any performance incentives at all. This result contradicts the standard economic model predicting that the incentive contract will be more efficient than the contract without incentives.<sup>1</sup> In spite of the efficiency loss the buyers, as principals, have a material advantage from the provision of incentives because the incentive contracts allow them to reap a much larger share of the smaller surplus. Thus, the redistributive effect of incentive contracts is sufficiently strong to over-compensate the induced efficiency loss.

We also show that the undermining of voluntary cooperation is, in principle, consistent with models of inequity aversion and reciprocity that assume fully rational agents (e.g., Fehr and Schmidt 1999, Bolton and Ockenfels 2000). However, additional experiments indicate that a significant part of the undermining effect is due to the negative framing of the explicit incentive. If the incentive is framed as a price deduction in case of malfeasance, the sellers are much less willing to cooperate voluntarily compared to a situation where a bonus is paid on top of a base compensation in case of appropriate performance. This difference in the behavior of sellers across frames occurs despite the fact that – in economic terms – the price-deduction-frame is identical to the bonus-frame. Therefore, most rational choice models of reciprocity and inequity aversion are not consistent with the framing effect.<sup>2</sup>

---

<sup>1</sup> For our purposes we define the standard economic model by two assumptions. First, individuals are fully rational and second, the individuals' objective function is defined in quite a narrow way in the sense that only income and the costs of effort matter. For seminal papers on the standard approach see, e.g., Holmström (1979), Grossman and Hart (1983), Hart and Holmström (1987) and Jewitt (1988). Interesting overviews about the empirical evidence are provided in Gibbons (1997) and Prendergast (1999).

<sup>2</sup> The potential exemption is the model by Dufwenberg and Kirchsteiger (1999), which generically exhibits a large number of equilibria in sequential games like the one considered in this paper. This

It seems that by framing incentives positively or negatively, perceptions of kindness and hostility, which are crucial for reciprocal responses, can be manipulated. The framing effect may, therefore, be due to a shift in the reference point that provides the basis for judging actions as hostile or kind. Different frames may, however, also affect the players' beliefs, which, in case of multiple equilibria, may induce the players to play a different equilibrium.

In our view these results are important because voluntary cooperation is a characteristic feature of many real world contexts. Whenever agents have discretion over the intensity or the type of activity they perform, voluntary cooperation is very valuable for principals. In case of the employment relationship, for instance, the relevance of voluntary cooperation is neatly confirmed by the extensive study of Bewley (1995, 1999) who reports the following: "Managers claim that workers have so many opportunities to take advantage of employers that it is not wise to depend on coercion and financial incentives alone as motivators" (Bewley 1995, p. 252). Voluntary cooperation is also often important in markets for complex production inputs like customized software. In these markets it is typically not possible to specify all the potential contingencies in a contract so that the buyers depend on the sellers' voluntary cooperation.

In the standard principal-agent approach the agents' objective function is, in general, assumed to be increasing and strictly concave in income, decreasing in effort and additively separable in the two arguments. This model provides the benchmark for our definition of voluntary cooperation: If agents determine their effort by maximizing the objective function assumed in the standard approach they do not cooperate voluntarily. If, in contrast, they work harder we speak of voluntary cooperation. Voluntary cooperation may have different sources: Agents may be simply irrational or they may have an intrinsic preference for the activity. Agents may also work harder because they exhibit inequity aversion or a preference for reciprocity. Another possibility is that agents may want to gain the social approval of relevant reference agents by providing excess effort relative to their narrow material self-interest. The neglect of these sources of voluntary cooperation in the standard approach is no big problem as long as these factors do not interact with the incentives provided.<sup>3</sup> Yet, as

---

means that, for a given incentive contract, multiple quality levels are consistent with equilibrium behavior of the sellers. If framing affects the selection of equilibria our framing result could be consistent with this model.

<sup>3</sup> Occasionally voluntary cooperation has been taken into account. Holmström and Milgrom (1991) assume, e.g., that employee's are willing to provide positive effort levels in the absence of any performance incentives. This is a critical assumption for their conclusion that in a multi-tasking environment principals may do better if agents face no incentives at all.

our experiments suggest, explicit incentives may interact in important ways with inequity aversion, bounded rationality and reciprocity motives. In addition, they are likely to interact with social approval and, perhaps, also with task-specific intrinsic motivation.<sup>4</sup>

There is a large literature in social psychology, which suggests that material rewards for a pleasant activity may decrease the intrinsic motivation to perform this activity (see, e.g. Deci and Ryan 1985).<sup>5</sup> Frey (1993) and, in particular, Frey (1997) discuss the relevance of this literature to economics. Recently, there has been a debate regarding the validity of the empirical claims put forward in this literature (see, e.g., Cameron and Pierce 1994, Eisenberger and Cameron 1996, Deci, Koestner and Ryan (1999), Kunz and Pfaff 2002, Fehr and Falk 2002). In the concluding section we compare our results with the experiments and claims made in the social psychology literature. The most important difference concerns the fact that in our experiments the incentives are set by the principals while in the social psychology experiments the agents face predetermined incentives that have been set by the experimenter. This means that this type of experiment cannot capture motivational forces like inequity aversion or reciprocity arising from the interaction between principals and agents.

The remainder of this paper is organized as follows. In the next section we will describe the experimental design and the experimental procedures. In Section 3 we derive the behavioral predictions. Section 4 presents the results. In Section 5 we summarize our findings and provide a concluding discussion.

## 2. EXPERIMENTAL DESIGN AND PROCEDURES

The basic feature of our design is the comparison of two treatments in the context of a “gift exchange game” as developed by Fehr, Kirchsteiger and Riedl (1993). The first treatment, providing the baseline, is the so-called “Trust-Treatment” (henceforth TT). Its main purpose is to establish the extent of voluntary cooperation in the absence of performance incentives for cooperation. The second treatment is the “Incentive-Treatment” (henceforth IT). The IT is identical to the TT, with the exception that performance incentives for contractual compliance are introduced. To examine the reasons for the undermining of voluntary cooperation in the IT we have conducted an additional treatment, which we call Bonus Treatment (henceforth BT). We will

---

<sup>4</sup> In Fehr and Falk (2002) one of us deals more explicitly with these interactions.

<sup>5</sup> Recently, Bénabou and Tirole (1999) developed a model that neatly formalises the crowding out of intrinsic motivation by extrinsic incentives.

present and discuss the BT in Section 4. In the current section we will describe the TT and the IT and the common procedural details that were implemented in all treatments.

### 2.1 The Trust-Treatment

The constituent game was a version of the “gift exchange game”, which comprises three stages. In the first stage buyers make a contract offer, which consists of a fixed price  $p$  and a *desired* quality level  $\hat{q}$ . In the second stage sellers can decide whether they want to accept one of the available contract offers made by the buyers. The third stage is entered only if sellers accept a contract offer. In the third stage they then choose their quality level. They are not committed by the buyer’s desired quality level  $\hat{q}$ , i.e., they may choose  $q \neq \hat{q}$ . In the experiments, a buyer’s profit is given by:

$$(1) \quad \pi = \begin{cases} vq - p & \text{if the contract is accepted;} \\ 0 & \text{if no contract is concluded.} \end{cases}$$

$vq$  is the return for the buyer as a function of the seller’s actual quality  $q$ . In the experiments  $v = 100$  and prices  $p$  have to obey  $0 \leq p \leq 100$ . Actual and desired quality levels have to be elements of the set  $\{0.1, 0.2, \dots, 1\}$ .

The seller’s payoff in the experiment is given by

$$(2) \quad u = \begin{cases} p - c(q) & \text{if the contract is accepted;} \\ 0 & \text{if no contract is concluded.} \end{cases}$$

Quality  $q$  causes disutility  $c(q)$ . In the experiments  $c(q)$  was determined according to Table 1.

TABLE 1  
*Quality levels and costs of quality for the seller in TT and IT.*

$q, \hat{q}$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$c(\hat{q}), c(q)$	0	1	2	4	6	8	10	12	15	18

## 2.2 The Incentive-Treatment

The basic difference between the IT and the TT is the buyers' possibility of punishing sellers in the IT whose quality choice falls short of  $\hat{q}$ , *provided* the seller's shirking can be verified. In particular, buyers' contract offers in the IT also specify – in addition to  $p$  and  $\hat{q}$  – a fine  $f$  that has to be paid to the buyer in case that  $q < \hat{q}$  can be verified. Shirking can be verified with probability  $0 < s < 1$ . With probability  $1 - s$  shirking cannot be verified and, hence, the buyer is committed to pay  $p$ . Therefore, a buyer's (expected) payoff in the IT is

$$(3) \quad \pi = \begin{cases} vq - p & \text{if the contract is accepted and } q \geq \hat{q}; \\ vq - p + sf & \text{if the contract is accepted and } q < \hat{q}; \\ 0 & \text{if no contract is concluded.} \end{cases}$$

A seller who does not shirk earns the contractually agreed-upon price and has to bear the quality costs, i.e. her utility is  $p - c(q)$ . This is also the utility in case of unverifiable shirking. In case of verifiable shirking the price is reduced by the fine  $f$ . Therefore, a seller's (expected) payoff is given by

$$(4) \quad u = \begin{cases} p - c(q) & \text{if the contract is accepted and } q \geq \hat{q} \\ p - c(q) - sf & \text{if the contract is accepted and } q < \hat{q}. \\ 0 & \text{if no contract is concluded.} \end{cases}$$

In the experiment the probability with which shirking could be verified was  $s = 1/3$ . The fine had to obey the restriction  $0 \leq f \leq 13$ .<sup>6</sup> All the other parameters and restrictions were the same as in the TT. In both treatments all the players knew their own payoff functions and the payoff functions of the trading partners, i.e.,  $v$ ,  $c(q)$  as given in Table 1,  $s$  and the feasible values of  $p$  and  $f$  were known to all the players.

Both in the TT and the IT the buyers' marginal return from an increase in  $q$  is 10 while, according to Table 1, the marginal cost of  $q$  is always much lower. There are thus high efficiency gains from higher levels of  $q$  and  $q = 10$  is the first best level of  $q$ .<sup>7</sup>

---

<sup>6</sup> In reality there are often limits to principals' sanctioning possibilities. Some of these limitations are due to legal regulations, norms, or collective bargaining agreements. It is also possible that these limits arise endogenously because the monitoring technology is not perfect or there is a problem of "moral hazard" on the principals' side. In addition, agents may also face liquidity constraints prohibiting large fines.

<sup>7</sup> One referee asked what would happen if the efficiency gains from higher  $q$  were lower. There is evidence that experimental subjects care positively for efficiency in non-strategic settings where one

### *2.3 Procedures and common features*

After their arrival all the subjects were randomly allocated their roles as buyers and sellers, respectively. All subjects kept their role during the whole experiment. Every subject participated in one treatment only. The experiments were conducted manually. After the subjects had been assigned their roles, sellers and buyers were sent to different rooms where they were seated remote from each other. Subjects first had to read their instructions (see Appendix) and then they had to answer a set of hypothetical questions to test their understanding of payoff calculations. The experiments started only after all the subjects had given correct answers to all the questions. Most subjects were able to answer the hypothetical questions in less than 10 minutes. Occasionally it happened that one or two subjects in a session needed more than 10 minutes.

In both treatments there were six subjects in the role of a buyer and eight subjects in the role of a seller. At the beginning buyers first had to decide privately on a contract offer and to enter it into a decision sheet. In the IT, after buyers had made their choice, they had to roll a six-sided die, which determined whether shirking, in case it occurred, could be verified. In particular, if the numbers 1 and 2 turned up, the seller's quality choice could be verified, i.e., in case of shirking the buyer could collect the specified fine  $f$ . If a number between 3 to 6 turned up, shirking was not verifiable. After all the buyers had made their choice (and, in the IT, rolled the die) all contract offers were made public by writing them on the blackboard in the buyers' room. Then the buyers' contract offers were transmitted to the sellers' room, where they were also written on the blackboard. The sellers could then choose – in a random order – among the available contract offers. Each buyer could only employ one seller and each seller could only accept one job. Hence, there was an excess supply of sellers. Both buyers and sellers knew the exact number of players on each side of the market. The reason for this was that we wanted to prevent buyers from offering high prices simply because they feared the rejection of their offers.<sup>8</sup>

---

player can unilaterally dictate the whole allocation of material payoffs (Engelmann and Strobl 2002). In strategic settings, however, efficiency considerations do not seem to affect players' social preferences. Fehr and List (2002) have conducted trust games in which the efficiency of the back-transfer of the second-mover varied. In the baseline condition the second-mover had to spend \$1 to increase the first-mover's payoff by \$1. In the high-efficiency condition the first-movers' payoff increased by \$3 for every \$1 the second mover spent on the back-transfer. Yet, in both conditions the second-movers spent the same amount on back-transfers indicating the absence of an efficiency motive.

<sup>8</sup> After we had conducted our experiments we became aware of the work of Brandts and Charness (2001) who report the remarkable result that prices and quality in gift exchange markets are the same regardless of whether the number of sellers is below or above the number of buyers. This indicates that competition has little or no impact on outcomes in gift exchange markets.



Sellers who accepted a contract had to determine their *actual* quality level. They had to insert their choice into a decision sheet. In the IT they had to make this decision before knowing whether their quality choice could be verified. However, immediately after they had made their choice they were *privately* informed about verification, i.e., whether they would have to pay the fine in case of shirking.<sup>9</sup> After a seller had determined his quality, the experimenter informed the respective buyer (and nobody else) of the actual quality level. Thus quality levels were private information of the two parties involved in a trade. This setting rules out any group effects regarding the choice of quality levels. At the end of the third stage buyers and sellers had to calculate their payoffs. After payoffs had been calculated a new market was opened. To allow for learning and to test the robustness of decisions, we had twelve periods (called “trading days”) in which the above-described market was operative. All the players in the market knew this.

In the experiment we avoided potential value-laden terms. Therefore, the fine was described in a neutral way as a “potential price deduction”. The instructions in the Appendix provide further details on the framing of terms.

Before we put forward our behavioral predictions we want to stress that our experimental procedures ensured that nobody was ever informed of the trading partner’s identity. Thus, a player who concluded a contract only knew that this contract was concluded with somebody on the other side of the market. Players did not know the identification numbers of their trading partners nor did they know their current trading partner’s past behavior. It was thus completely impossible for a player to gain an individual reputation.<sup>10</sup> Note also that buyers could not make offers to specific sellers. They only could make offers to the market, i.e., the whole group of sellers, and then individual sellers could accept one of the available offers in a random order. Thus, even if sellers would have had the opportunity to gain an individual reputation for, e.g., being a high quality seller, it would have been impossible for them to gain from this reputation because it was impossible for the buyers to select their trading partner. All these experimental procedures were implemented to enhance the one-shot nature of a trading day.

---

<sup>9</sup> We also informed the non-shirkers whether they would have had to pay the fine if they had shirked. This was done to enhance the credibility of our random verification procedure. Technically, sellers were informed about verification by checking the appropriate box on their decision sheets (see instructions in the Appendix). Thus, since *all* the sellers were informed about verification they could not identify the shirkers among them.

<sup>10</sup> In this regard it is also important that the offers were written on the blackboard in a *random order*. We implemented a random order to prevent possible identification of buyers. Therefore, it was not possible that individual buyers could gain a reputation for particular offers.

In total we conducted 13 sessions with 8 buyers and 6 sellers each. In the IT we ran four and in the TT five sessions. In addition, we conducted 4 sessions in the Bonus Treatment (BT) to examine framing effects. In total 182 undergraduates from the universities in Zurich participated in our experiments. All the participants were recruited from a large database to minimize the likelihood that they know each other.<sup>11</sup> In principle, losses were possible in our experiments. Therefore, we endowed all subjects with an additional amount of 9 Swiss Francs. All losses had to be covered by this endowment and the earnings made during the experiments. Subjects were paid in cash immediately after the experiments. An experiment lasted between 2 and 2.5 hours. Subjects earned on average 50 Swiss Francs (about \$36) including a show-up fee of 15 Swiss Francs. The hourly earnings of Swiss students for temporary jobs are roughly 20 – 25 Swiss Francs.

### 3. BEHAVIORAL PREDICTIONS

#### 3.1 *Behavior in the Trust Treatment*

Due to the one-shot nature of our game, a rational seller whose preferences are given by (2) will always choose the minimal quality  $q^{min}$  because higher quality levels are costly and yield no return. Consequently, the buyer will offer the minimal price that is necessary to induce the seller to accept the contract, i.e., the buyer offers  $p = 1$ .<sup>12</sup>

However, the situation changes substantially if sufficiently many sellers have a propensity to cooperate voluntarily in response to generous contract offers. In our experiments a major reason for voluntary cooperation is likely to be found in subjects' "social preferences". A considerable number of studies indicate that a substantial fraction of experimental subjects – across a wide variety of different subject pools – does exhibit social preferences (for surveys see, e.g., Fehr and Schmidt 2000, Sobel 2001). A person exhibits social preferences if the person does not only care about the material resources allocated to her but also cares about the material resources allocated to relevant reference agents. In the context of our experiments social preferences imply, for example, that the actions of the sellers are not only affected by

---

<sup>11</sup> A post-experimental questionnaire confirmed that indeed most subjects had never met another participant before.

<sup>12</sup> Price offers had to be integers. Therefore, if sellers have selfish preferences a price offer of 1 can be an equilibrium outcome. Of course, price offers of zero can also be an equilibrium outcome. In the following we always concentrate on the strict equilibrium in which sellers are strictly better off when they accept the contract.

the performance incentives they themselves face but also by the perceived fairness of the payoff allocation resulting from the contract offer.

Quite an extensive literature has emerged discussing the different motivational sources of social preferences. The literature is based on the development of formal models of inequity aversion, reciprocity and spitefulness and on a large set of experiments that isolate the different motivational sources for non-selfish behaviors. Rabin (1993), Dufwenberg and Kirchsteiger (1999) and Falk and Fischbacher (1999) have developed reciprocity models while Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) developed models of inequity aversion. A reciprocal individual responds to actions that are perceived to be kind in a kind manner, and to actions that are perceived to be hostile in a hostile manner. Whether an action is perceived as kind or hostile depends on the fairness or unfairness of the consequences *and* the intention associated with the action. The fairness of the intention, in turn, is determined by the equitability of the payoff distribution, relative to the set of feasible payoff distributions, caused by the action. Inequity aversion models, in contrast, do not depend explicitly on the perceived fairness or unfairness of the intentions but they also rely on the perceived fairness of the actions' consequences.

Intuitively it is easy to see that reciprocity or inequity aversion models predict that sellers voluntarily cooperate, i.e., choose  $q > q^{min}$ , in response to generous contract offers. Reciprocal sellers infer kind intentions from generous offers and, therefore, they reciprocate (see, e.g., Dufwenberg and Kirchsteiger 2000). Inequity averse sellers are willing to put forward non-minimal quality levels in response to generous prices because by raising the quality level they can implement outcomes that they perceive as more equitable. To illustrate this let us consider a simplified version of the model by Fehr and Schmidt (1999). In the two-player case the utility function of an inequity averse (i.e., fair) player is given by

$$(5) \quad U_i(x) = x_i - \alpha_i \max \{x_j - x_i, 0\} - \beta_i \max \{x_i - x_j, 0\},$$

$i \in \{1, 2\}$ ,  $i \neq j$ , where  $x=(x_i, x_j)$  denotes the vector of monetary payoffs and  $\alpha_i \geq \beta_i$ ,  $0 \leq \beta_i < 1$ . The term weighted with  $\alpha_i$  measures the utility loss that stems from inequality to  $i$ 's disadvantage, while the term weighted with  $\beta_i$  measures the loss from advantageous inequality.<sup>13</sup> For convenience we use a grossly simplified version of this

---

<sup>13</sup> In principle, inequity aversion is not the same as inequality aversion. For laboratory experiments inequality aversion may, however, be a good first approximation of inequity aversion. The subjects typically enter the laboratory as equals, they don't know anything about each other, and they are allocated to different roles in the experiment at random. Thus, it seems natural to assume that the egalitarian outcome constitutes a salient reference point.

model because the purpose of the subsequent deliberations is not to provide precise quantitative predictions but to illustrate some important qualitative regularities arising from inequity averse preferences.<sup>14</sup> We assume that fair subjects exhibit  $\alpha_i \geq \beta_i = 0.5$ . In a dictator game subjects with such preferences are willing to give 50 percent of the money to the recipient whereas when being in the responder role in an ultimatum game they reject offers that give them less than 25 percent of the available money. We also assume that – at the quality stage of the experiment – the only reference player for the seller is his buyer. This allows us to use the two-person utility function above to study the quality behavior of the sellers. We believe that this assumption approximates the subjects’ perception of the situation well because they do not know the quality that is provided in other trades. In addition, once a contract has been concluded the buyer has clearly a salient position for a seller in our experiment.

Once a fair seller with the above preferences has accepted a contract in the TT he will, for sufficiently high prices, choose a quality level such that  $\pi = 100q - p = p - c(q) = u$  holds.<sup>15</sup> Differentiation of this equation shows that quality responds to a price change according to  $dq/dp = 2/[100+c'(q)] > 0$ . Thus, by raising the seller’s price sufficiently such that  $u > \pi$  holds at  $q^{min}$ , the seller is induced to choose a payoff equalizing quality level, which we denote by  $q^e$ , and the higher the price the higher will be  $q^e$ . An immediate consequence of this is that if the fraction of fair sellers is sufficiently high, even purely selfish buyers have an incentive to pay generous prices.

Previous experimental research has shown that in games that are similar to our TT quality levels are indeed positively affected by prices and that – for certain price ranges – the buyers can raise their profits by raising prices (for overviews see, e.g., Fehr and Schmidt 2000, Sobel 2001 or Gächter and Fehr 2001).<sup>16</sup>

### 3.2 Behavior in the Incentive Treatment

Analogous to the TT we first discuss the behavioral predictions for the case that all the buyers and sellers are selfish and rational. A selfish seller will accept a contract in the IT if the participation constraint

$$(6) \quad p \geq c(q^*)$$

---

<sup>14</sup> To generate more precise quantitative predictions at the individual level it would, e.g., be necessary that  $\beta$  varies with the amount of advantageous inequality. If  $\beta$  is constant the agent either prefers full equality or does not care for equality at all.

<sup>15</sup> Of course, if the price is so low that for all feasible quality levels  $\pi > u$  holds the seller chooses  $q^{min}$ .

<sup>16</sup> These results also hold in high-stake environments. In Fehr and Tougareva (1996) subjects earned, on average, the monetary income of ten weeks in a gift exchange experiment similar to the TT presented in this paper.

holds. In (6)  $q^*$  denotes the quality level that maximizes the seller's material payoff. A rational and selfish risk-neutral seller who has accepted a contract will perform at the desired quality level (i.e.,  $q = \hat{q}$ ) if the No-Shirking Condition (NSC)

$$(7) \quad sf \geq c(\hat{q})$$

is satisfied.<sup>17</sup> The seller will perform at the desired level if the expected fine  $sf$  for shirking is larger than the costs of contractual compliance, which are given by  $c(\hat{q}) - c(q^{min}) = c(\hat{q})$  because  $c(q^{min}) = 0$  (see Table 1). It is easy to check that with our parameterization of  $s = 1/3$  and  $f \leq 13$  quality costs of 4.33 can be enforced. Hence, according to the quality cost function in Table 1, the largest quality level that can be enforced by an incentive compatible contract, which we denote by  $\hat{q}^*$ , is given by  $\hat{q}^* = 0.4$ . More generally, the best reply quality choice  $q^*$  of a risk-neutral seller who is rational and selfish is given by:

$$(8) \quad q^* = \begin{cases} q^{min} & \text{for all } f, \hat{q} \text{ that obey } (1/3)f < c(\hat{q}) \\ \hat{q} & \text{for all } f, \hat{q} \text{ that obey } (1/3)f \geq c(\hat{q}). \end{cases}$$

Since the marginal revenue of quality is 10 while the marginal cost of quality is always strictly below 10, a profit-maximizing buyer who keeps the seller at her reservation utility, will always prefer the highest enforceable quality level  $\hat{q}^* = 0.4$ .

We can summarize the above discussion for the IT in the following behavioral prediction: If all the subjects are selfish and rational we will observe that sellers accept all contracts obeying  $p \geq c(q^*)$  and choose the quality level  $q^*$  according to (8). Buyers, therefore, offer the contract  $[p = c(\hat{q}^*) + 1 = 5; f = f^{max} = 13; \hat{q} = \hat{q}^* = 0.4]$ .<sup>18</sup> Note that if this prediction is met  $q - q^*$  will be zero, i.e., there will be no voluntary cooperation. Moreover, sellers' behavior then will not respond to variations in the price offered so that it is not possible to induce positive voluntary cooperation ( $q - q^* > 0$ ) by offering a generous price. In addition, self-interested sellers will not reduce their quality below  $q^*$  in response to low prices.

---

<sup>17</sup> To derive our benchmark predictions we assume that subjects in our experiment are risk neutral because, as Rabin (2000) has shown, assuming risk aversion over the typical stake levels in laboratory experiments is highly implausible. Even small degrees of risk aversion over typical experimental stakes imply an absurd degree of risk aversion over large stakes.

<sup>18</sup> There is another, non-strict subgame perfect equilibrium, in which  $p = 4, f = 13, \hat{q}^* = 0.4$  holds and where the seller accepts the contract and chooses  $q^* = 0.4$ . For empirical purposes the slight difference between these two equilibria is negligible.

The situation again is substantially different if we account for the presence of reciprocal or inequity averse sellers. To keep the exposition simple we again concentrate on the case of inequity averse sellers as described in the previous subsection. If an inequity averse seller faces a contract  $(p, \hat{q}, f)$  such that  $\hat{q}$  is below  $q^e$ , the seller will never shirk. Instead, the seller will choose  $q^e > \hat{q}$  (see Figure 1). This means that for all contracts in which  $\hat{q} < q^e$  an inequity averse seller exhibits the same behavior as in the TT. In these contracts the fine does not affect the seller's response because the seller does not consider performing below  $\hat{q}$ . The situation changes, however, if  $\hat{q}$  is above  $q^e$ . In this case an inequity averse seller would, in principle, prefer to reduce the quality to  $q^e$  but if he does so he faces the threat of being fined.

An inequity averse seller who decides to shirk may end up in one of two situations. With probability  $s$  he will be caught shirking and will have to pay the fine  $f$ . If the seller knew for sure that he would end up in this situation he would choose a quality level such that the payoff difference between the buyer's payoff  $100q - p + f$  and his own payoff  $p - c(q) - f$  is removed. In Figure 1 we denote this quality by  $q^f$ . With probability  $1 - s$  shirking will not be verified and the seller does not have to pay a fine. If the seller knew for sure that he would end up in this situation he would choose the quality level  $q^e$  (see Figure 1). Thus, in case of shirking, the seller's quality will be between  $q^f$  and  $q^e$ .<sup>19</sup> Since ex-ante the seller does not know in which situation he will end up his utility from shirking  $U^s$  is given by

$$(9) \quad U^s = s(p - c(q) - f - \alpha[100q - p + f - (p - c(q) - f)]) + (1-s)(p - c(q) - \beta[p - c(q) - (100q - p)]).$$

We can examine the conditions under which a shirking seller prefers to provide  $q^f$ . Differentiation of  $U^s$  with respect to quality gives

$$\partial U^s / \partial q = -c'(q) - (100 + c'(q))(s\alpha - (1-s)\beta)$$

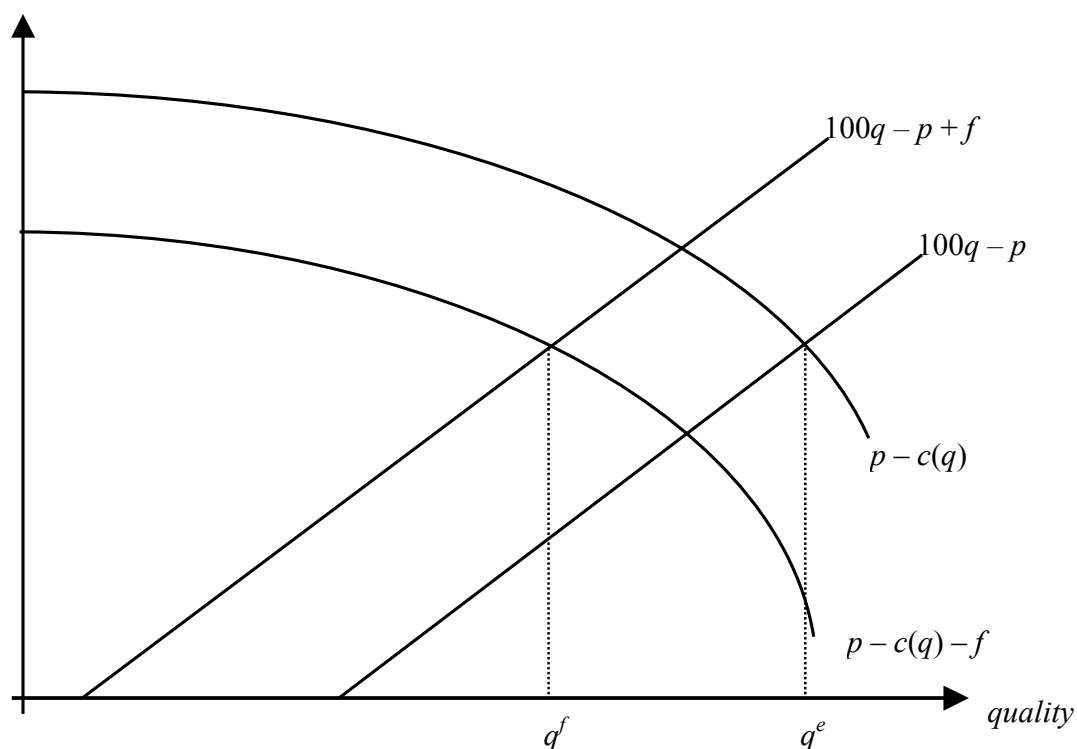
It is easy to see that this expression is always negative if  $s\alpha - (1-s)\beta$  is positive. When this is the case the utility from shirking is maximized at  $q^f$ . Thus, if the seller's inequity aversion is sufficiently biased against disadvantageous inequality (i.e.,  $\alpha \geq 2\beta$  for  $s = 1/3$ ) a shirking seller prefers to reduce the quality provided below the egalitarian quality  $q^e$  that would be chosen if there were no fine or if the buyer demanded a quality at or below  $q^e$ . In other words, once the seller has decided to shirk

---

<sup>19</sup> If at  $q^{min}$  the buyer's payoff,  $100q - p + f$ , is still bigger than the seller's payoff,  $p - c(q) - f$ , the seller chooses  $q$  such that  $q^{min} \leq q \leq q^e$  holds.

the existence of the fine has a counterproductive effect because the shirking seller chooses a lower quality with the fine than without the fine. This is, of course, not the only effect of the fine because the fine also affects whether the seller decides to shirk.

FIGURE 1  
Inequity aversion and quality choice



To see this in more detail we write the utility from non-shirking,  $U^{ns}$ , as

$$(10) \quad U^{ns} = p - c(\hat{q}) - \alpha[100\hat{q} - p - (p - c(\hat{q}))].$$

The term in parenthesis measures the difference between the buyer's and the seller's payoff at  $\hat{q}$ . The seller prefers to perform at  $\hat{q}$  if  $U^{ns} \geq U^s$  holds. Using (9) and (10) the non-shirking condition can be written as

$$(11) \quad s\alpha[100q - p + f - (p - c(q) - f)] + (1-s)\beta[p - c(q) - (100q - p)] \geq \\ \geq c(\hat{q}) - c(q) - sf + \alpha[100\hat{q} - p - (p - c(\hat{q}))]$$

Intuitively, the right hand side of (11) measures the pecuniary and the non-pecuniary gains from shirking:  $c(\hat{q}) - c(q) - sf$  measures the monetary gains from shirking while  $\alpha[100\hat{q} - p - (p - c(\hat{q}))]$  measures the non-pecuniary inequality costs associated with the proposed contract. By shirking the seller can avoid these non-pecuniary costs. The left-hand side of (11) measures the non-pecuniary inequality costs that arise from shirking. Note that if the seller chooses  $q^f$  the first term on the left side vanishes while if the seller chooses  $q^e$  the second term on the left side vanishes. Condition (11) shows that, by decreasing the monetary gains from shirking, the fine deters shirking. In addition, (11) indicates that the inequality associated with the contract is also a decisive shirking determinant. The more unequal the payoff distribution of a proposed contract the more likely it is that the seller shirks. Finally, the non-shirking condition also highlights the important role of  $\hat{q}$  for the shirking decision. While in the TT  $\hat{q}$  represents only cheap talk, in the IT a higher level of  $\hat{q}$  makes it more likely that the non-shirking condition is violated. In case that  $\hat{q}$  exceeds  $q^e$  a rise in  $\hat{q}$  increases the monetary gains as well as the non-pecuniary gains from shirking.

The previous discussion shows that in the presence of inequity aversion the incentive contract may not necessarily do better than a pure trust contract – in fact it may even do worse. However, inequity aversion is not the only, and perhaps not even the most important, reason why the incentive contract may do worse. Incentive contracts may also cause negative side effects if reciprocity motives alone are present. Remember that reciprocity means that sellers respond in a hostile manner to actions that reveal a hostile intention. In our view the fining of sellers may reveal hostile intentions for two reasons.<sup>20</sup> First, the fine per se may be perceived as hostile. Second, threatening to fine a seller is an indication of distrust. To the extent to which trusting actions are perceived as kind and distrusting actions as hostile, a fine will be perceived as a hostile act. Whatever the true reason for perceiving somebody's intention as hostile, if the sellers perceive the fine as a hostile act they may no longer be willing to put forward the same quality compared to a situation with no fine.

---

<sup>20</sup> To our knowledge there is no formal model of reciprocity that captures these two reasons explicitly.



## 4. RESULTS

In the following we first describe the results for the TT and the IT. Then we move on to discussing the BT. In the TT buyers offered 357 contracts in total, of which all but one were accepted. In the IT the total number of offered contracts was 287 and 7 of them (2.4 percent) were rejected. Before we provide a detailed statistical analysis of subjects' voluntary cooperation behavior we present the evidence at a more descriptive level. Our first result provides insights into the average contract offered by the buyers in the TT and the IT.

**Result 1:** *In the TT buyers offer, on average, higher prices and demand higher quality levels than in the IT. The average fine in the IT is close to the maximum fine.*

Support for R1 is provided in Table 2. Table 2 summarizes the average behavior over all the periods in the IT and the TT and compares them with the predictions of the self-interest model for both treatments. It shows that the average price offered in the TT is considerably higher than in the IT. The same regularity is exhibited by median prices. The median price is more than two times larger in the TT than in the IT. Table 2 also provides information on the average and median fines. Note that the median fine is exactly the maximal fine of  $f = 13$ . In fact, buyers in the IT imposed the maximum fine in 69 percent of all cases and only in 4 contracts they chose  $f = 0$ . It is, therefore, not surprising that the average fine is close to the maximal fine. Table 2 also indicates that  $\hat{q}$  is considerably larger in the TT. On average,  $\hat{q} = 0.65$  in the TT while buyers desire only  $\hat{q} = 0.49$  in the IT.

TABLE 2

*Contracts and actual quality levels in the Trust and the Incentive Treatment*

	Trust-Treatment (TT) N = 356			Incentive-Treatment (IT) N = 280			
Predictions:	( $p^*=1, q^*=0.1$ )			( $p^*=5, f^*=13, \hat{q}^*=0.4, q^*=0.4$ )			
Actual:	$p$	$\hat{q}$	$q$	$p$	$f$	$\hat{q}$	$q$
Mean	30.2	0.65	0.37	19.7	11.7	0.49	0.27
Median	34.0	0.70	0.30	15.0	13.0	0.50	0.30
Std.dev.	17.6	0.27	0.26	12.6	2.6	0.17	0.17

The frequent use of maximal fines together with the considerably lower prices in the IT conveys the impression that buyers in the IT relied to a large extent on the “stick”

and much less on the “carrot” compared to the TT. It is, therefore interesting to see how this strategy is related to actual quality levels and to the sellers’ willingness to cooperate voluntarily.<sup>21</sup>

**Result 2:** *Quality and voluntary cooperation are lower in the IT compared with the TT. This is due to the following reasons: (i) A fraction of sellers shirks in the IT even when the No-Shirking Condition (NSC) is met. (ii) Voluntary cooperation ( $q - q^* > 0$ ) vanishes almost completely for incentive compatible contracts in the IT. (iii) If the NSC is violated in the IT sellers choose the minimum quality in the vast majority of cases. In the TT quality levels above the minimum are provided in the majority of cases. (iv) In the TT voluntary cooperation responds strongly to the price level while in the IT voluntary cooperation does no longer respond to the price level.*

A first indication for R2 is given in Table 2, which shows that the mean actual quality level is lower in the IT. Further support for R2 is provided by Figure 2, which presents the distributions of actual quality levels. The figure shows that the distribution of quality has considerably more mass at lower levels in the IT. This is also confirmed by a Kolmogorov-Smirnov-test ( $prob < 0.001$ ). In particular, in the IT only 7.5 percent of quality levels are above 0.5, whereas in the TT quality levels above 0.5 account for about 27 percent of all quality choices. The distribution of quality choices in the IT shows two peaks – one at  $q^{min} = 0.1$  and the other at  $q = 0.4$ . In contrast, in the TT there is only one mode at  $q^{min}$ . However, roughly 70 percent of all quality choices in the TT are at non-minimal levels.

Table 3 sheds light on the reasons for the higher quality levels in the TT. The table separates actual choices along two dimensions: (i) Does actual quality deviate from the best reply quality  $q^*$ , i.e., to what extent is there excess quality or under-performance relative to the self-interest model? (ii) Does the contract meet the NSC  $sf \geq c(\hat{q})$ ?

The first interesting result in Table 3 concerns quality behavior in cases where the NSC is met. The final column of the table informs us that the condition was met in 41.3 percent (115 out of 280) of all the contracts in the IT. The table also indicates that in 16.5 percent of these cases (19 out of 115) sellers shirked. In the vast majority of these cases sellers chose  $q^{min}$ . Moreover, there were only two cases of a seller providing excess quality when the NSC was met. This means that for incentive compatible contracts voluntary cooperation is almost completely removed. These

---

<sup>21</sup> In the following the terms “No-Shirking Condition (NSC)” and “incentive compatible contracts” relate to inequality (7),  $sf \geq c(\hat{q})$ , in the text.

regularities also mean that, on average, sellers in the IT provide less than the best reply quality when they face incentive compatible contracts. This result is also confirmed by Figure 3 which shows that the average quality for incentive compatible contracts (see IT-IC-graph) is generally below  $q^*$ .

FIGURE 2  
Distribution of sellers' actual quality choices

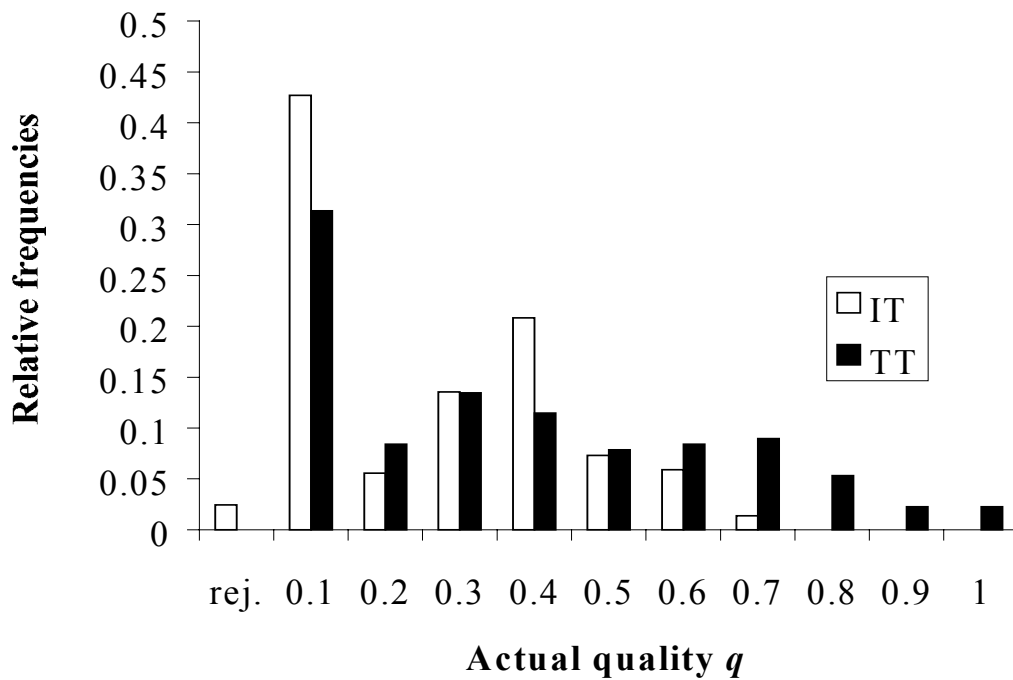


TABLE 3

*Sellers' deviation from the best reply quality level  $q^*$  in the IT and the TT*

		$q < q^*$	$q = q^*$	$q > q^*$	Sum
Treatments:		N	N	N	N
IT	no	-	102	63	165
	<i>NSC (7) met?</i>				
	yes	19	94	2	115
Sum		19	196	65	280
TT			112	244	356

**Note:** Entries are number of cases in the respective category.  $q^*$  denotes the best reply quality choice according to (8).

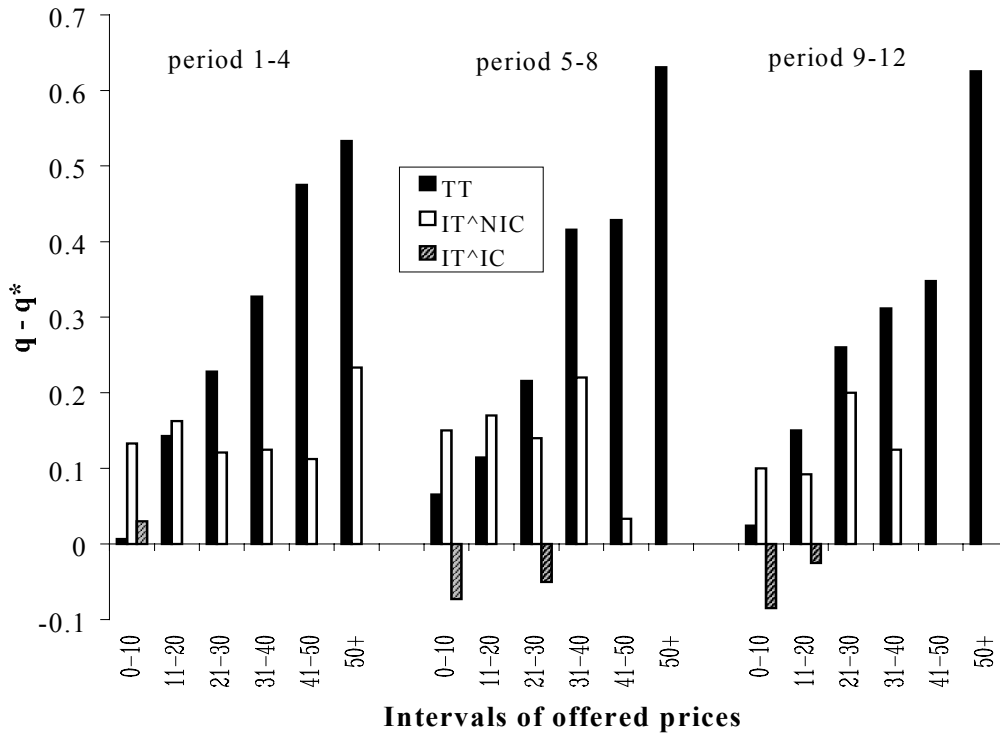
How did the sellers respond to non-incentive compatible contracts? In 62 percent of those cases in which the NSC was violated in the IT (102 out of 165 cases), quality choices conform to the best reply level  $q^* = 0.1$ , i.e., sellers shirk fully. Comparing this result with the quality behavior in the TT we observe a big difference because in 69 percent of all the contracts in the TT (244 out of 356) sellers provide quality above the minimum. The fact that in 62 percent of the non-incentive compatible IT-contracts sellers deliver the minimal quality, while in 69 percent of the TT-contracts they provide quality above the minimal level, is quite remarkable. From the viewpoint of the standard model sellers should exhibit identical behavior in these two situations. This result is, however, consistent with inequity aversion and notions of reciprocity. For instance, once inequity averse sellers have decided to shirk they are fined with positive probability. This will, in general, induce them to perform less than what they would be willing to provide in the TT. Thus, the considerably lower quality levels for non-incentive compatible contracts in the IT can be taken as an indication that voluntary cooperation is undermined by the use of fines.

Figure 3 provides further support for this interpretation. The figure shows that in the TT voluntary cooperation is high and depends positively on the price level. Moreover, this pattern of voluntary cooperation prevails in all three time intervals depicted in Figure 3 indicating stability over time.<sup>22</sup> This contrasts sharply with the pattern of voluntary cooperation observed for non-incentive compatible contracts in the IT (see graph IT<sup>NIC</sup> in Figure 3). Except at very low price levels voluntary cooperation is generally lower in the IT and it does not respond positively to the price.

---

<sup>22</sup> See Gächter and Falk (2002) for a recent experiment, which examined, among other things, the impact of learning and experience in a twenty times repeated experiment (with changing opponents) that was akin to our TT. In their experiment learning and experience also did not diminish voluntary cooperation over time.

FIGURE 3  
Voluntary cooperation ( $q - q^*$ ) as a function of prices



The graphical results depicted in Figure 3 are also confirmed by more detailed econometric analyses of voluntary cooperation in the TT and the IT. For the TT we regressed  $(q - q^*)$  on a constant, prices and the desired quality level. Our regressions take into account that the dependent variable is censored at zero and the reported standard errors are based on the fact that only observations across sessions are independent while observations within a given session are dependent. Table 4 presents the results of our TT-regressions. The table shows that in each of the sub-periods for which we conducted the regression as well as for the regression with the data from all periods price increases are associated with highly significant increases in voluntary cooperation. Recall that according to the Fehr-Schmidt model the desired quality level represents merely cheap talk, i.e., it should not affect actual quality choices. It turns out that in the first 4 periods  $\hat{q}$  does affect actual quality positively but in the remaining periods  $\hat{q}$  is no longer significant. This pattern of voluntary cooperation contrasts sharply with the pattern observed in the IT.

**TABLE 4**

*Response of voluntary cooperation to offered prices in the Trust Treatment*

	Dependent variable: $q - q^*$							
	all periods		periods 1-4		periods 5-8		periods 9-12	
	coef. (rob. s.e.)	z- value	coef. (rob. s.e.)	z- value	coef. (rob. s.e.)	z- value	coef. (rob. s.e.)	z- value
constant	-0.285*** (0.0243)	-11.75	-0.319*** (0.0843)	-3.79	-0.181*** (0.0478)	-3.78	-0.399*** (0.0892)	-4.47
$p_{it}$	0.012*** (0.0014)	8.93	0.010*** (0.001)	10.61	0.012*** (0.0009)	12.90	0.014*** (0.0027)	5.31
$\hat{q}_{it}$	0.199*** (0.076)	2.63	0.336*** (0.1149)	2.93	0.115 (0.1192)	0.97	0.187 (0.2077)	0.90
	N = 356 LL = -94.27 Wald $\chi^2(2)=389.9$ p = 0.000		N = 119 LL = -24.73 Wald $\chi^2(2)=280.8$ p = 0.000		N = 119 LL = -20.78 Wald $\chi^2(2)=313.1$ p = 0.000		N = 118 LL = -42.74 Wald $\chi^2(2)= 69.8$ p = 0.000	

**Note:** The estimation procedure is censored regression with robust standard errors.  $p_{it}$  denotes the price that worker  $i$  received in period  $t$ .  $\hat{q}_{it}$  denotes the desired quality level worker  $i$  faced in period  $t$ . Robust standard errors are in parentheses. \*\*\*, \*\*, \* denote significance at the 1-, 5- and 10-percent level, respectively.

The results of our econometric analysis of voluntary cooperation for the IT are reported in Table 5 where we regressed  $q - q^*$  on a constant and a dummy IC for incentive compatible contracts. We also interacted the IC-dummy and a dummy for non-incentive compatible contracts (NIC) with prices  $p_{it}$  and desired quality levels  $\hat{q}_{it}$ . Most importantly, the table shows that  $p_{it}$  interacted with NIC is close to zero and never significant. This indicates that for non-incentive compatible contracts prices have no effect on voluntary cooperation. Prices have a small significant impact for incentive compatible contracts – the higher the price the lower is on average the sellers' shirking. A further interesting result is that the desired quality level has a *negative* impact on voluntary cooperation. Recall that in the TT  $\hat{q}$  had a positive but insignificant impact between periods 5 and 12 indicating that  $\hat{q}$  represents cheap talk. However, for the incentive compatible contracts in the IT  $\hat{q}$  has a large negative impact on voluntary cooperation during these periods. In addition,  $\hat{q}$  also has a negatively significant impact in non-incentive compatible contracts for the data of all periods as well as for the data in the final three periods.

TABLE 5  
*Response of voluntary cooperation ( $q - q^* > 0$ ) and under-performance ( $q - q^* < 0$ ) to  
prices in the Incentive Treatment*

	Dependent variable: $q - q^*$							
	all periods		periods 1-4		periods 5-8		periods 9-12	
	coef. (rob. s.e.)	z- value	coef. (rob. s.e.)	z- value	coef. (rob. s.e.)	z- value	coef. (rob. s.e.)	z- value
constant	0.276*** (0.0587)	4.70	0.235** (0.0939)	2.51	0.305*** (0.0967)	3.15	0.358*** (0.1104)	3.24
$IC$	-0.130* (0.0771)	-1.69	-0.080 (0.1599)	-0.50	-0.142 (0.1248)	-1.13	-0.216 (0.1393)	-1.55
$p_{it} \times IC$	0.005*** (0.0019)	2.73	-0.001 (0.0012)	-0.96	0.008** (0.0039)	2.06	0.009* (0.0047)	1.93
$p_{it} \times NIC$	0.002 (0.0012)	1.28	0.001 (0.0016)	0.63	0.001 (0.0035)	0.33	0.004 (0.0034)	1.03
$\hat{q}_{it} \times IC$	-0.765*** (0.193)	-3.96	-0.372 (0.3103)	-1.20	-0.994*** (0.339)	-2.94	-0.876** (0.3818)	-2.29
$\hat{q}_{it} \times NIC$	-0.289** (0.1054)	-2.74	-0.190 (0.1554)	-1.22	-0.308 (0.2522)	-1.22	-0.528*** (0.1840)	-2.87
	N = 280 LL = 68.63 Wald $\chi^2(5)=102.7$ p = 0.000		N = 95 LL = 27.8 Wald $\chi^2(5)=41.3$ p = 0.000		N = 90 LL = 18.76 Wald $\chi^2(5)=37.8$ p = 0.000		N = 95 LL = 26.49 Wald $\chi^2(5)=33.45$ p = 0.000	

**Note:** The estimation procedure is censored regression with robust standard errors.  $IC$  denotes a dummy for the incentive compatible contracts.  $p_{it} \times IC$  is an interaction variable of the  $IC$ -dummy with prices.  $p_{it} \times NIC$  is an interaction variable of  $p_{it}$  with the dummy  $NIC$ , which denotes non-incentive compatible contracts.  $\hat{q}_{it} \times IC$  ( $\hat{q}_{it} \times NIC$ ) is an interaction between  $\hat{q}$  and the dummy  $IC$  ( $NIC$ ). Rejected contract offers are excluded. Robust standard errors are in parentheses. \*\*\*, \*\*, \* denote significance at the 1-, 5- and 10-percent level, respectively.

The negative impact of  $\hat{q}$  on voluntary cooperation in the IT is not compatible with the standard model. If sellers are rational and self-interested they provide the desired quality in case of incentive compatible contracts and they fully shirk in case of non-incentive compatible contracts. In the context of the standard model a higher  $\hat{q}$  may render a contract non-incentive compatible if the higher value of  $\hat{q}$  violates the no-shirking condition  $sf \geq c(\hat{q})$ . But within the class of incentive compatible contracts and within the class of non-incentive compatible contracts, higher values of  $\hat{q}$  will not affect voluntary cooperation and underperformance. Yet, if sellers are, e.g., inequity averse then a higher desired quality level may well affect voluntary cooperation and underperformance. This is so because, in addition to the pecuniary costs of shirking, variations in  $\hat{q}$  also affect the perceived fairness of the contract offer in the IT. As our discussion of the NSC (11) for inequity averse sellers has shown a higher level of  $\hat{q}$

will increase the non-pecuniary gains from shirking. Thus, even if a rise in  $\hat{q}$  leaves the NSC for selfish sellers (i.e.  $sf \geq c(\hat{q})$ ) intact the rise may nevertheless violate the NSC for inequity averse sellers because it is also associated with an increase in the non-pecuniary gains from shirking. This shows that in case that the NSC for selfish sellers is met inequity averse sellers may increase underperformance in response to a rise in  $\hat{q}$ . Likewise, it can be shown that if the NSC for selfish sellers,  $sf \geq c(\hat{q})$ , is already violated an inequity averse seller may reduce voluntary cooperation in response to an increase in  $\hat{q}$ . To see this assume that  $sf \geq c(\hat{q})$  is violated and that  $\hat{q}$  is initially below  $q^e$  – inducing the inequity averse seller to provide  $q = q^e$  initially. Assume further that the rise in  $\hat{q}$  renders  $\hat{q}$  larger than  $q^e$ . From our discussion in Section 3.2 we know that inequity averse sellers may well respond to this by reducing quality below  $q^e$ .

In view of R2 it is interesting to what extent the reduction of voluntary cooperation reduces the total surplus  $S$ , defined as the total material payoffs from a contract, in the IT relative to the TT. Before we present this result it is worthwhile to emphasize that the standard model implies that  $S$  is more than two times higher in the IT than in the TT. However, in stark contrast to this hypothesis we can report the following result:

**Result 3:** *The total surplus  $S$  is on average higher in the TT than in the IT. This holds irrespective of whether we compare the TT-contracts with incentive compatible or with non-incentive compatible IT-contracts. The profit for the buyers is highest for incentive compatible contracts, second highest for TT-contracts and lowest for non-incentive compatible IT-contracts.*

R3 basically means that the incentive opportunities in the IT allow buyers to increase their profits relative to the TT but that this is associated with an efficiency loss. Support for R3 is presented in Table 6, which compares the actual with the predicted surplus and shows buyers' and sellers' average profits from a contract. Note first that in the TT the realized surplus equals 33 while it is only 21.5 for non-incentive compatible contracts and 27.5 for incentive compatible contracts. Thus, relative to the surplus predicted by the standard model for the TT we observe a more efficient outcome in the TT that is due to sellers' voluntary cooperation. To a lesser extent this is also true for those IT-contracts that are not incentive compatible. However, actual efficiency is lower than the predicted efficiency for incentive compatible contracts.



TABLE 6

*Average profits and average total gains per contract.*

	Trust-Treatment	Incentive-Treatment	
		NSC violated	NSC holds
Predicted surplus $S^*$ ( <i>according to the standard model</i> )	$vq^{min} - c(q^{min}) = 10$	$vq^{min} - c(q^{min}) = 10$	$v\hat{q}^* - c(\hat{q}^*) = 36$
Realized Surplus $S$	33.0	21.5	27.5
Profit Buyer	7.1	1.9	18.2
Profit Seller	25.9	19.6	9.2

Table 6 also shows how the use of incentive compatible contracts in the IT allows the buyers to vastly change the distribution of the gains from trade. In the TT buyers reap only 21.5 percent (7.1/33) of the total surplus while in the IT they receive 66 percent (18.2/27.5) of the smaller total surplus when they propose incentive compatible contracts. The table also indicates that incentive compatible contracts are much more profitable for the buyers than non-incentive compatible IT-contracts. In view of these large profit differences, it is surprising that 59 percent of all the contracts in the IT are *not* incentive compatible. This figure hides, however, a strongly increasing time trend in the share of incentive compatible contracts. While in period 1 only 17 percent of all IT-contracts were incentive compatible, the share of incentive compatible contracts rises to 63 percent in the final two periods.

In our view R2 and R3 are quite remarkable because they indicate that in the absence of incentives the outcome may well be more efficient than predicted by the standard approach while the use of incentives may trigger factors that decrease efficiency relative to the standard prediction. Taken together this may render the complete absence of incentives, like in our TT, more efficient. It is, therefore, quite important to have a better understanding of the behavioral forces behind R2 and R3. From R1 we already know that buyers in the IT are considerably less generous than buyers in the TT. This suggests the possibility that the driving force behind R2 and R3 is given by buyers' reluctance to make generous offers in the IT. Perhaps, the possibility of fining sellers in the IT has seduced buyers to use only the "stick" and to forget about eliciting quality by the "carrot" of generous offers. Our next result shows, however, that the undermining of voluntary cooperation in the IT is not caused by less generous buyers in the IT relative to the TT.

**Result 4:** The reduction of voluntary cooperation in the IT was not caused by low price offers. Rather, low price offers were the response to the reduction of the sellers' voluntary cooperation in the IT.

A first piece of evidence in favor of R4 is given in Figure 3 and Tables 4 and 5. The figure and the tables show that, for given prices levels, the amount of voluntary cooperation is generally lower in the IT compared with the TT. Thus, even if the buyers would have liked to induce high levels of voluntary cooperation in the IT by offering generous prices they would not have been able to achieve their goal.

Figures 4, 5A and 5B provide further support for R4. Figure 4 shows that in period 1 the average price is even higher in the IT than in the TT. Yet, despite this feature voluntary cooperation is considerably lower in the IT as indicated by Figures 5A and 5B. Figure 5A shows that in period 1 of the IT the sellers exhibited voluntary cooperation in only 5 of the 24 cases while in 19 cases their choices coincided with the best reply quality. This differs sharply from the quality choices in the TT. Figure 5B shows that in period 1 of the TT only 6 of the 30 quality choices were at the best reply level of  $q^{min}$  while in 24 of the 30 cases the sellers exhibited voluntary cooperation.

FIGURE 4  
Evolution of prices in Trust and Incentive Treatment

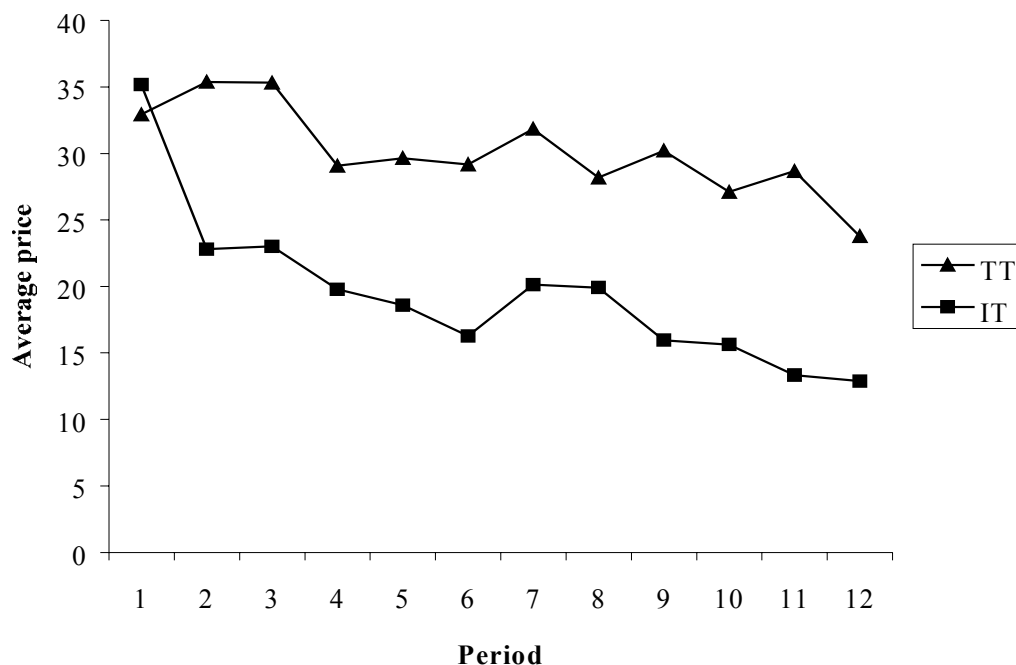


FIGURE 5A  
 Voluntary cooperation in period 1 of the Incentive Treatment<sup>23</sup>

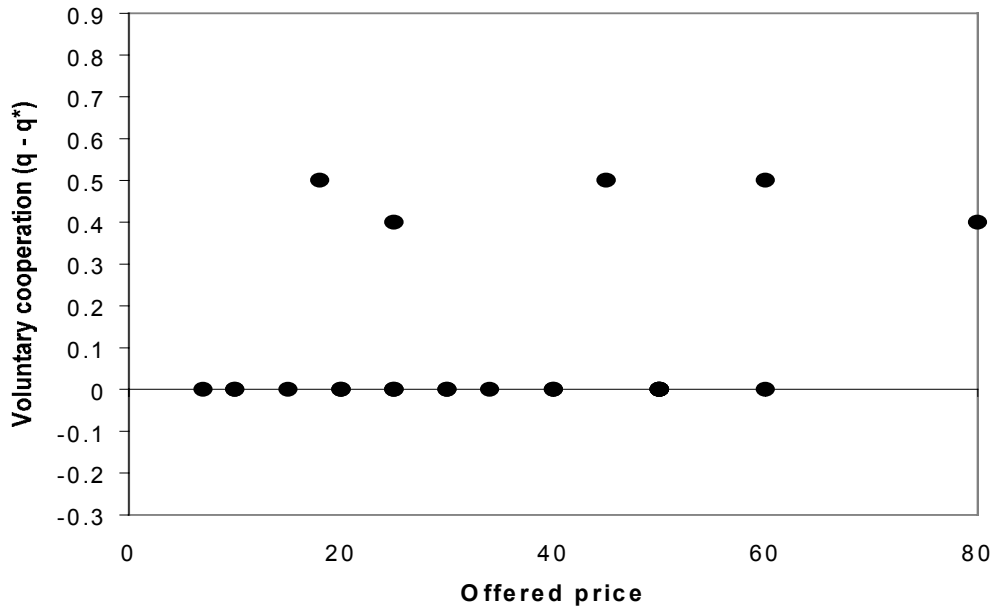
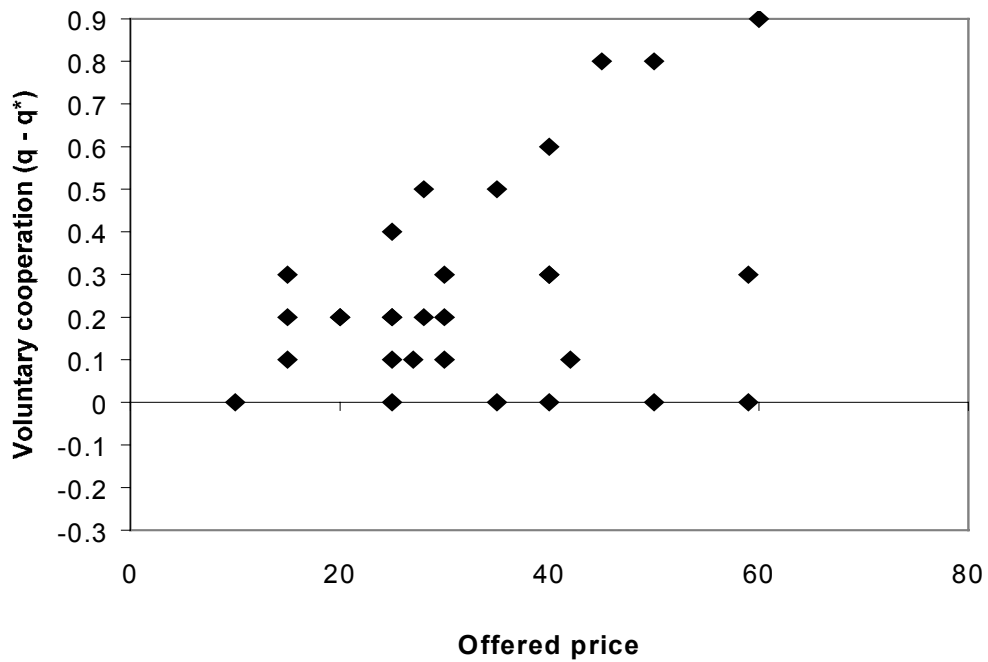


FIGURE 5B  
 Voluntary cooperation in period 1 of the Trust Treatment



<sup>23</sup> Due to identical observations not all observations with  $q = q^*$  are visible in Figure 5A. Only 4 of the 24 contract offers were incentive compatible, i.e., in 20 of the 24 cases  $q^{min}$  was the best reply quality in the IT.

The substantial reduction of voluntary cooperation in the IT is also associated with a significantly lower absolute quality level relative to the TT ( $prob = .020$ , Mann Whitney test, two-tailed). On average, the first period quality in the TT is 0.36 whereas in the IT it is only 0.23. Thus, despite the fact that in period 1 prices in the IT were slightly more generous than in the TT the sellers were less willing to cooperate voluntarily and to provide quality in the IT. This low level of voluntary cooperation in period 1 of the IT caused losses for most of the buyers. As a consequence, the buyers substantially cut back price offers in period 2 of the IT (see Figure 4). From period 2 onwards the price levels are always much lower in the IT compared with the TT but this reduction in prices in the IT is plausibly interpreted as a response to the breakdown of voluntary cooperation that occurred already in period 1 of the IT.

If the low price offers of the buyers were not the cause but the consequence of the low levels of voluntary cooperation in the IT – what was then the cause for the breakdown of voluntary cooperation? In Section 3.2 we discussed several potential reasons that are related to inequity aversion and to notions of reciprocity. One potentially important reason is that in the TT  $\hat{q}$  is merely cheap talk while in the IT  $\hat{q}$  can affect the sellers' quality choices negatively. In the TT it does not matter for an inequity averse seller whether  $\hat{q}$  is low or high – the seller can always choose the egalitarian quality level  $q^e$  without having to face any negative consequences in case that the buyer demanded more than  $q^e$ . In the IT, however, an inequity averse seller who faces  $\hat{q} > q^e$  has the following problem. If he provides  $\hat{q}$  he incurs non-pecuniary costs from the unfair offer. If he wants to correct the unfairness by providing less than  $\hat{q}$  he faces the threat of being punished, which will in general induce him to choose quality levels even below  $q^e$  because the pending punishment will change the payoff distribution unfavorably for the seller. A necessary condition for the effectiveness of this mechanism is that the sellers in the IT did face a large number of contracts with  $\hat{q} > q^e$ . It turns out that this was indeed the case. In 84 percent of all contract offers in the IT  $\hat{q} > q^e$  prevailed. Interestingly, in 76 percent of the contract offers in the TT  $\hat{q}$  also exceeded  $q^e$ . The fact, that in the TT voluntary cooperation prevailed despite the existence of a large majority of offers with  $\hat{q} > q^e$  lends support to the argument that unfair offers are more likely to reduce voluntary cooperation in the IT because the threat of being fined causes further disadvantageous changes in the payoff distribution.

The above argument is, however, not the only possible reason for the breakdown of voluntary cooperation in the IT. We already mentioned in Section 3.2 that fines designed to elicit high quality levels may be perceived as hostile per se. In addition, explicit incentives may be perceived as a signal of distrust and to the extent to which

distrust is perceived as unkind, reciprocally motivated sellers respond to the fine with unkind quality levels. We also conjectured that the perception of the incentive device as kind or unkind could be subject to framing effects. If the same incentive is framed in a positive manner it may perhaps be possible to prevent the reduction of voluntary cooperation. To test this conjecture we have conducted the so-called Bonus Treatment (BT).

The BT is just a reframing of the IT. The only difference between the IT and the BT is that in the IT a shirking seller has to pay a fine  $f$  if caught shirking while in the BT a shirking seller is not paid the bonus  $b$  if caught shirking. While in the IT a contract consists of  $(p, \hat{q}, f)$  the buyers in the BT stipulate contracts  $(p_o, \hat{q}, b)$  where  $p_o$  denotes the base price and  $b$  obeys  $0 \leq b \leq 13$ . Note that the bonus in the BT obeys the same constraints as the fine in the IT. Likewise, the probability that the bonus is not paid out in case of shirking is identical to the probability that the fine has to be paid. Thus, while in the IT the seller faces the threat of having to pay the fine  $f$  with probability  $s$ , in the BT the seller faces the isomorphic threat of not earning the bonus  $b$  with probability  $s$ .<sup>24</sup> A seller in the BT who faces a contract  $(p_o, \hat{q}, b)$  is exactly in the same situation as a seller in the IT who faces a contract  $(p, \hat{q}, f)$  if  $p = p_o + b$  and  $f = b$ . As a consequence the standard model, models of inequity aversion and most reciprocity models predict that sellers who face identical contracts across conditions should behave identically. This is so because all these models rely on rational behavior and predict a unique quality response to any given contract. Thus, rational agents in the BT who face – in economic terms – the same contract as in the IT should behave identically in the two situations. Yet, as our next result shows there are large behavioral differences across IT and BT.

**Result 5:** *Voluntary cooperation is much lower in the IT compared to the BT. Yet, in the BT voluntary cooperation is lower than in the TT.*

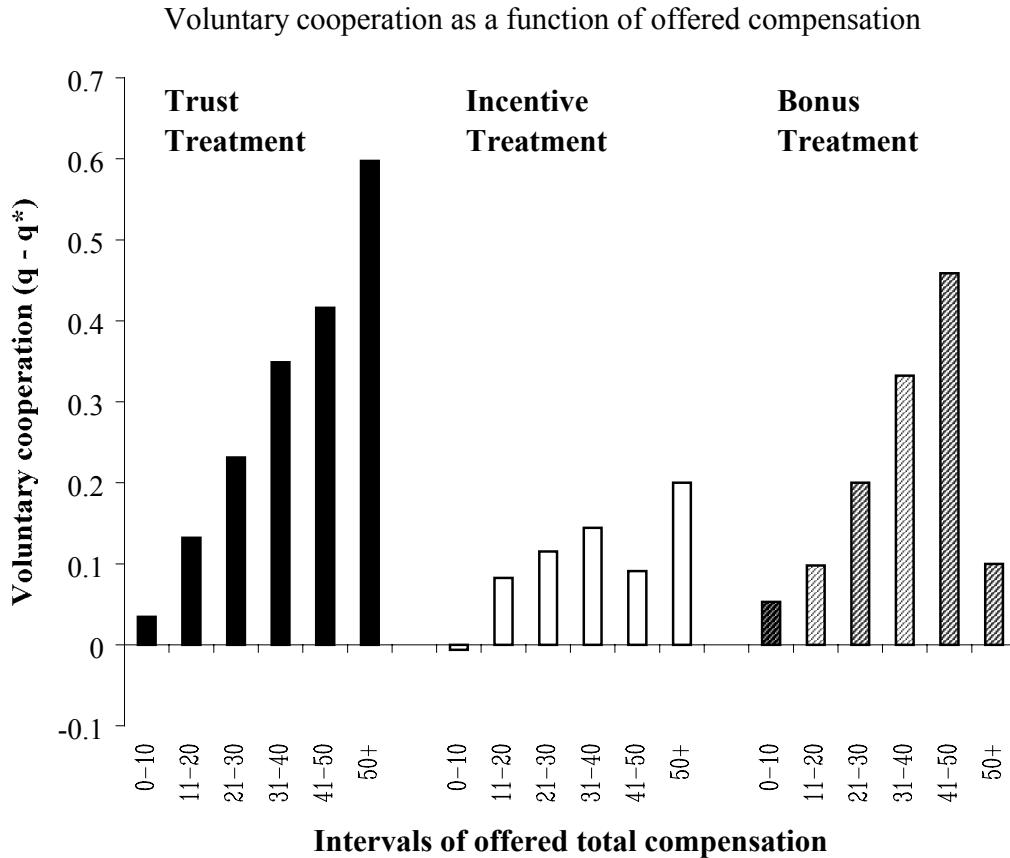
We provide support for R5 in Figure 6 and in Table 7. Figure 6 shows voluntary cooperation as a function of the offered total compensation. In the TT and the IT the offered total compensation in case of  $q = \hat{q}$  was equal to the price. In the BT the offered total compensation for  $q = \hat{q}$  was equal to  $p_o + b$ . Figure 6 indicates that voluntary cooperation is highest in the TT, somewhat lower in the BT and lowest in

---

<sup>24</sup> Recall that in the experimental instructions for the IT we did not use the term fine. We used instead the terms “potential price deduction” from the agreed upon price. In the instructions for the BT we used the term “potential supplementary price” that was paid in addition to the agreed upon price.

the IT. Moreover, the figure also suggests that while voluntary cooperation does not respond to total compensation in the IT, it responds positively in the TT and the BT.

FIGURE 6



The impression conveyed by Figure 6 is confirmed by a comparison of regression results in Tables 4 and 5 with those in Table 7. In Table 7 we present the regressions for the BT. In contrast to the regressions for the IT (see Table 5) the coefficient for  $p_{it} \times NIC$  is significantly positive indicating that the compensation offered in case of non-incentive compatible contracts has a positive impact on voluntary cooperation in the BT. This effect prevails in all three time periods considered as well as for the overall regression in Table 7. However, a comparison of the coefficient for the compensation for non-incentive compatible contracts,  $p_{it} \times NIC$ , in Table 7 with the slope for  $p_{it}$  in the TT (see Table 4) indicates that the slope in the BT is generally flatter. The null hypothesis of equal slopes can be rejected at the  $p = .024$  level of significance indicating that increases in monetary compensation led to larger increases in voluntary cooperation in the TT.

TABLE 7  
*Response of voluntary cooperation ( $q - q^* > 0$ ) and under-performance ( $q - q^* < 0$ )  
to the monetary compensation ( $p_o + b$ ) in the Bonus Treatment*

	Dependent variable: $q - q^*$							
	all periods		periods 1-4		periods 5-8		periods 9-12	
	coef. (rob. s.e.)	z- value	coef. (rob. s.e.)	z- value	coef. (rob. s.e.)	z- value	coef. (rob. s.e.)	z- value
constant	0.025 (0.0426)	0.58	0.030 (0.0887)	0.34	0.047 (0.083)	0.56	-0.028 (0.0600)	-0.46
$IC$	0.102 (0.07223)	1.41	0.159 (0.1312)	1.21	-0.041 (0.1421)	-0.29	0.129 (0.1199)	1.08
$p_{it} \times IC$	0.002 (0.0028)	0.62	0.002 (0.0037)	0.63	0.010 (0.0060)	1.62	0.044 (0.0473)	0.93
$p_{it} \times NIC$	0.008*** (0.0017)	4.95	0.007** (0.0026)	2.53	0.010*** (0.0027)	3.82	0.010*** (0.0024)	4.19
$\hat{q}_{it} \times IC$	-0.782** (0.3500)	-2.23	-1.295** (0.5759)	-2.25	-0.496 (0.3874)	-1.28	-2.100 (2.0701)	-1.01
$\hat{q}_{it} \times NIC$	-0.041 (0.0889)	-0.46	-0.012 (0.1670)	-0.07	-0.106 (0.1612)	-0.66	-0.029 (0.1233)	0.23
	N = 284 LL = -2.65 Wald $\chi^2(5)=58.3$ p = 0.000		N = 94 LL = -8.17 Wald $\chi^2(5)=24.9$ p = 0.000		N = 95 LL = 0.25 Wald $\chi^2(5)= 33.8$ p = 0.000		N = 95 LL = 10.15 Wald $\chi^2(5)=23.1$ p = 0.000	

**Note:** The estimation procedure is censored regression with robust standard errors.  $IC$  denotes a dummy for the non-incentive compatible contracts and  $p_{it}$  measures the total compensation ( $p_o + b$ ) in case of  $q = \hat{q}$ .  $p_{it} \times IC$  is an interaction variable of the  $IC$  dummy with  $p_{it}$ .  $p_{it} \times NIC$  is an interaction variable of  $p_{it}$  with a dummy  $NIC$ , which denotes non-incentive compatible contracts.  $\hat{q}_{it} \times IC$  ( $\hat{q}_{it} \times NIC$ ) measures the interaction between  $\hat{q}$  and the dummy  $IC$  ( $NIC$ ). Rejected contract offers are excluded. Robust standard errors are in parentheses. \*\*\*, \*\*, \* denote significance at the 1-, 5- and 10-percent level, respectively.

The considerably higher levels of voluntary cooperation in the BT relative to the IT indicate that the framing of incentives matters. There are, in our view, two potential explanations for the framing effect. One explanation relies on a combination of the notion of reciprocity with a specific aspect of bounded rationality.<sup>25</sup> The other explanation relies on the existence of multiple equilibria in the Dufwenberg-Kirchsteiger model of reciprocity. Recall that the definition of reciprocity depends on the notion of kind and hostile actions. What is perceived as kind or hostile depends, in turn, inevitably on a neutral reference point. The presence of a framing effect is consistent with the view that the reference point, which provides the basis for

<sup>25</sup> Selten (1998) also argues that there are interactions between bounded rationality and reciprocity.

categorizing an action as kind or hostile, can be manipulated by the framing of the incentive. In the negative frame the total compensation in case of non-shirking seems to provide a natural reference point and the fine focuses attention on the fact that something will be taken away in case of verified shirking. In the positive frame the base compensation is the natural reference point and the bonus focuses attention on the fact that something will be given if the desired quality is provided. It may well be that “taking away something” is perceived as less friendly than “not giving something” even if the total compensation is identical.

A prominent feature of the Dufwenberg-Kirchsteiger model is that in sequential games like ours the model typically exhibits a host of equilibria.<sup>26</sup> Some of these equilibria are characterized by mutual kindness whereas others are characterized by mutual hostility. In principle, it is therefore possible that the framing of the incentive affects the players’ beliefs and as a consequence the sellers choose different quality levels in the different frames. If, for example, the fine changes the players beliefs such that they coordinate on an equilibrium with mutual hostility whereas the bonus induces beliefs such that they coordinate on an equilibrium with mutual kindness a framing effect arises.

The above interpretations are, of course, ex-post rationalizations of the evidence that require further thought and experiments. To examine the validity of the second interpretation it would, in particular, be necessary to measure the players’ beliefs. Having said this, it is also worthwhile to point out that an important part of the decrease in voluntary cooperation is consistent with models of inequity aversion because in the BT voluntary cooperation is significantly lower than in the TT. This means that the framing effect explains only part of the difference in voluntary cooperation across the TT and the IT.

## 5. SUMMARY AND DISCUSSION

This paper shows that voluntary cooperation may indeed be undermined by incentive contracts. If the buyers can fine the sellers for malfeasance they almost universally propose contracts with fines and pay, on average, considerably lower prices compared to a situation where no fines are available. The use of fines causes, however, a substantial reduction in voluntary cooperation and efficiency. Even if the buyers combine the threat of fining with the payment of relatively high prices the sellers’

---

<sup>26</sup> For instance, in the ultimatum game and in the sequential prisoner’s dilemma there are a large number of equilibria.



voluntary cooperation remains low. The payment of generous prices makes thus no sense in the Incentive Treatment whereas in the Trust Treatment the buyers can induce increases in voluntary cooperation by increasing their price offers.

The undermining of voluntary cooperation through explicit fines is incompatible with the standard principal-agent model that assumes self-interested agents. It is, in principle, consistent with rational choice models of inequity aversion. If sellers are inequity averse and the buyers demand quality levels above the egalitarian quality  $q^e$  the sellers incur non-pecuniary inequity costs if they provide the demanded quality level. However, if they want to avoid the inequity costs by providing less than the demanded quality they are fined with positive probability. This causes a further disadvantageous change in the payoff distribution. As a consequence, inequity averse sellers will, in general, provide quality below  $q^e$  when they face a threat of being fined while in the absence of this threat they are willing to provide  $q^e$ .

Further experiments show, however, that an interpretation of the observed counterproductive incentive effect in terms of inequity aversion cannot tell the whole story. If the available incentive is framed as a bonus that is paid in addition to a base compensation voluntary cooperation is, *ceteris paribus*, significantly higher than in a situation in which the incentive is framed as a price deduction. In the bonus frame voluntary cooperation again responds positively to the price level. This framing effect is compatible with the view that perceptions of kindness and hostility, which are crucial for reciprocal quality responses, can be manipulated by the framing of incentives. The framing effect may, therefore, be due to a shift in the reference point that provides the basis for judging actions as hostile or kind. The framing effect is, however, also consistent with the Dufwenberg-Kirchsteiger model of reciprocity. This model typically exhibits multiple equilibria and to the extent to which different frames coordinate players on different equilibria a framing effect occurs.

Our experiment and our conclusions differ in several important ways from the social psychology literature that aims at testing the undermining of intrinsic motivation through monetary rewards (Cameron and Pierce 1994, Deci, Koestner and Ryan 1999).<sup>27</sup> First, and perhaps most importantly, in the psychological experiments no principals are present, i.e., the experimental subjects are always confronted with an incentive that is exogenously set by the experimenter. However, as our discussion in Section 3.2 suggests, in our context variations in voluntary cooperation are the result of variations in the perceived fairness of the payoff distribution between a principal and an agent and variations in the perceived fairness of the principal's action. Thus,

---

<sup>27</sup> For a critical evaluation of this literature from the behavioral economics viewpoint see Fehr and Falk (2002).

the psychological forces stemming from the relation between principal and agent are absent in the experiments on intrinsic motivation. Instead, the subjects in these experiments are engaged in an activity that is assumed to be intrinsically pleasurable.

Second, on the basis of the psychological experiments it is not possible to judge the economic importance of the undermining of intrinsic motivation because the value of the output of the activity is not determined. This is relevant insofar as the incentive may affect intrinsic motivation negatively but this negative effect may be overcompensated by the positive effect of extrinsic motivation on quality so that overall efficiency is nonetheless enhanced by the incentive.

Third, the proponents of an undermining effect claim that even flat payments reduce intrinsic motivation (Deci, Koestner and Ryan 1999). This is highly implausible in our context where fairness concerns play a prominent role. In fact, in our experiments increases in the price are never associated with decreases in average cooperation or average quality. Instead price increases in the TT and the BT are associated with increases in voluntary cooperation. Even for the incentive compatible contracts in the IT we observe a positive effect of price increases because higher prices decrease the under-performance of sellers. This demonstrates that the forces contributing to voluntary cooperation in our experiments are quite distinct from the forces assumed to be operative in the psychological experiments. As a consequence, the mechanisms responsible for counter-productive incentive effects are likely to be quite different, too.

What are the implications of our results for contract and incentive theory? Although in our view the results certainly do not mean that performance incentives will, in general, impair efficiency we believe that the undermining of voluntary cooperation is a possibility that should be taken seriously. There are now several other papers showing similar effects. In Gneezy and Rustichini (2000a) it is shown that the introduction of price incentives for parents who are late picking up their children from the kindergarten increases the frequency of late-coming parents. Gneezy and Rustichini (2000b) show that monetary rewards for pupils who collect money from households for social purposes may decrease the amount of money collected. Fehr and Falk (2002) show that the results by Gneezy and Rustichini can be neatly interpreted in terms of interactions between social approval and explicit performance incentives. Bohnet, Frey and Huck (2001) conducted a contract enforcement experiment in which the probability of fining sellers for the breach of a contract is varied. It turns out that the propensity to obey the contract is non-monotonic in this probability. At low and high probabilities contracts are less likely to be broken relative to a situation with an intermediate probability of being fined. Fehr and Rockenbach (2000) conducted a

Trust experiment with and without punishment opportunities for the principal. Their results indicate that if the principal – although he has the opportunity to punish the agent for malfeasance – voluntarily refrains from threatening to punish the agent, voluntary cooperation by the agents is enhanced.

Although these results do not yet give us clear and sharp answers why and under which conditions voluntary cooperation will be undermined they suggest that explicit incentives may indeed have counterproductive incentive effects that cannot be accounted for by traditional principal-agent models that assume completely selfish preferences. However, regardless of the impact of framing, there seem to be incentive contracts that leave voluntary cooperation intact. For instance, Anderhub, Gächter and Königstein (2000) observe a lot of voluntary cooperation when agents face fair revenue-sharing contracts. There is also evidence that implicit performance incentives generated by repeated interactions have powerful *positive* effects on quality (see Gächter and Falk 2002). Thus, an important question for future research is to identify the conditions under which performance incentives backfire. This requires, among other things, a much better understanding of the interaction between psychological and economic forces in principal-agent relations.

## REFERENCES

- ANDERHUB, V., GÄCHTER, S. and KÖNIGSTEIN, M. (2000), “Efficient Contracting and Fair Play in a Simple Principal-Agent Experiment”, forthcoming in: *Experimental Economics*.
- BÉNABOU, R., and TIROLE, J. (1999), “Self-Confidence and Social Interactions”, (Mimeo, University of Toulouse).
- BEWLEY, T. (1995), “A Depressed Labor Market as Explained by Participants”, *American Economic Review, Papers and Proceedings*, **85**, 250-254.
- BEWLEY, T. (1999), *Why Wages Don't Fall During a Recession* (Cambridge: Harvard University Press).
- BOHNET, I., FREY, B. and HUCK, S. (2001), “More Order with Less Law: On Contract Enforcement, Trust, and Crowding”, *American Political Science Review* **95**, 131-144.
- BOLTON, G. and OCKENFELS, A. (2000), “ERC–A Theory of Equity, Reciprocity, and Competition”, *American Economic Review* **90**, 166-193.
- BRANDTS, J. and CHARNES, G. (2001), “Do Market Conditions Affect Gift Exchange? Evidence from Experimental Markets with Excess Supply and Excess Demand”, Working Paper, Institute de Analisis Economico, Barcelona.
- DECI, E. and RYAN, R. (1985), *Intrinsic Motivation and Self-Determination in Human Behavior* (New York: Plenum Press).

- DECI, E., KOESTNER, R., and RYAN, R. (1999), "A Meta-analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation", *Psychological Bulletin* **125**, 627-668.
- DUFWENBERG, M. and KIRCHSTEIGER, G. (1998), "A Theory of Sequential Reciprocity", (Working paper, Stockholm University).
- DUFWENBERG, M. and KIRCHSTEIGER, G. (2000), "Reciprocity and Wage Undercutting", *European Economic Review* **44**, 1069-1078.
- ENGELMANN, D. and STROBEL, M. (2002), "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments", Working Paper, Humboldt University Berlin.
- EISENBERGER R. and CAMERON J. (1996), "Detrimental Effects of Reward. Reality or Myth?", *American Psychologist*, **51**, 1153-1166.
- FALK, A. and FISCHBACHER, U. (1999), "A Theory of Reciprocity", (Working Paper No. 6, University of Zürich).
- FEHR, E. and FALK, A. (2002), "Psychological Foundations of Incentives", forthcoming in: *European Economic Review*.
- FEHR, E. and LIST, J. (2002), "The Role of the Efficiency Motive in Strategic Interactions – An Experiment with CEOs and Students", mimeo, University of Zürich.
- FEHR, E. and ROCKENBACH, B. (2000), "Incentives and Intentions - The Hidden Rewards of Economic Incentives" (Mimeo, University of Zurich).
- FEHR, E. and SCHMIDT, K. (1999), "A Theory of Fairness, Competition and Cooperation", *Quarterly Journal of Economics*, **114**, 817-868.
- FEHR, E. and SCHMIDT, K. (2000), "Theories of Fairness and Reciprocity – Evidence and Economic Applications", forthcoming in: M. Dewatripont, L.P. Hansen, S. Turnovski, *Advances in Economic Theory, Eighth World Congress of the Econometric Society*, Cambridge: Cambridge University Press.
- FEHR, E. and TOUGAREVA, E. (1996), "Do High Stakes Remove Reciprocal Fairness?—Evidence from Russia", (Discussion paper, University of Zurich).
- FREY, B. (1993), "Shirking or work morale? The impact of regulating", *European Economic Review*, **37**, 1523-1532.
- FREY, B. (1997), *Not Just For the Money. An Economic Theory of Personal Motivation* (Cheltenham: Edward Elgar).
- GÄCHTER, S. and FALK A. (2002), "Reputation and Reciprocity: Consequences for the Labour Relation", *Scandinavian Journal of Economics* **104**, 1-26.
- GÄCHTER, S. and FEHR, E. (2001), "Fairness in the Labour Market. A Survey of Experimental Results", forthcoming in: BOLLE, F. and M. LEHMANN-WAFFENSCHMIDT, M. (eds.): *Surveys in Experimental Economics. Bargaining, Cooperation and Election Stock Markets*. Physica Verlag 2001.
- GIBBONS, R. (1997), "Incentives and Careers in Organizations", in: KREPS, D. and WALLIS, K. (ed.) *Advances in economics and econometrics: theory and applications, Seventh World Congress Vol. II*, (Cambridge: Cambridge University Press).
- GNEEZY, U. and RUSTICHINI, A. (2000), "Pay enough or don't pay at all", *Quarterly Journal of Economics* **115**, 791-810.
- GNEEZY, U. and RUSTICHINI, A. (2000), "A Fine Is a Price", *Journal of Legal Studies* **XXIX**, 1-17.

- GROSSMAN, S., and HART, O. (1983), "An Analysis of the Principal-Agent Problem", *Econometrica*, **51**, 7-45.
- HART, O., and HOLMSTRÖM, B. (1987), "The theory of contracts", in: BEWLEY, T. (ed.) *Advances in Economic Theory, 5<sup>th</sup> World Congress* (Cambridge: Cambridge University Press).
- HOLMSTRÖM, B. (1979), "Moral Hazard and Observability", *Bell Journal of Economics*, **10**, 74-91.
- HOLMSTRÖM, B., and MILGROM, P. (1991), "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design", *Journal of Law, Economics, and Organization*, **7**, 24-52.
- JEWITT, I. (1988), "Justifying the First-Order Approach to Principal-Agent Problems", *Econometrica*, **56**, 1177-1190.
- KUNZ, A. and PFAFF, D. (2001), "Agency Theory, Performance Evaluation, and the Hypothetical Construct of Intrinsic Motivation" (Mimeo, Institute for Accounting and Controlling, University of Zurich).
- LEVINE, D. (1998), "Modeling Altruism and Spitefulness in Experiments", *Review of Economic Dynamics* **1**, 593-622.
- PRENDERGAST, C. (1999), "The Provision of Incentives in Firms", *Journal of Economic Literature*, **XXXVII**, 7 – 63.
- RABIN, M. (2000), "Risk Aversion and Expected-Utility Theory: A Calibration Theorem", *Econometrica* **68**, 1281-1292.
- SELTEN, R. (1998), "Features of experimentally observed bounded rationality", *European Economic Review* **42**, 413-436.
- SOBEL J. (2001), "Interdependent Preferences and Reciprocity", mimeo, University of California, San Diego.

## APPENDIX: INSTRUCTIONS

*These instructions were originally written in German. Here we present the instructions of our Incentive-Treatment. The instructions of the Trust-Treatment were identical, except for those differences in design that are peculiar to the Incentive-Treatment.*

### **Introductory Remarks** (*identical for buyers and sellers*)

The experiment, in which you participate today, is part of a research project that is funded by various research funds. Its purpose is to study decision making in markets. Your income in this experiment consists of Fr. 15.- for your show-up and all payment that you will earn during the experiment according to your decisions and those of other participants. During the experiment your income will be calculated in points where **1 point = 8 Rappen**. Calculated in points the show-up fee of Fr. 15 amounts to 187.5 points. In addition you will receive an endowment of 112.5 points, which implies that **in total** you will have **300 points** at your disposal to cover losses that may occur during the experiment. **However, with your own decisions you can always prevent losses with certainty. At the end of the experiment all points, which you earned during the experiment will be summed up, exchanged into Swiss Francs, and paid out to you in cash immediately.**

First we would like to ask you to read these instructions carefully, and then to answer the control questions. After all participants have correctly answered all questions, we will start with the experiment, in which you will need the 12 decision sheets and the leaflet which have been handed out to you together with these instructions. **Please notice that all written information that you receive from us, is for your private use only. You are not allowed to transmit any information to other participants of this experiment. It is also prohibited to communicate with the other participants. Otherwise we would have to break off the experiment. If you have questions, please ask us.**

### **General Information** (*identical for buyers and sellers*)

- In this experiment there are buyers and sellers. The experiment comprises 12 trading periods.
- Each trading day consists of two stages. At the **first stage** each buyer decides on an offer, which contains the conditions under which the buyer is prepared to buy a commodity from a seller. Such an offer consists of a price, a desired quality and a potential price deduction. There are ten possible quality levels.
- At the **second stage** a random mechanism determines an order according to which the sellers can choose among the available offers. No seller is obliged to accept a bid, and no buyer is forced to make an offer. All sellers who have accepted an offer, have to decide which quality they will actually deliver. Choosing a quality level entails costs for the seller. After the seller has determined the “actual quality” the respective buyer will be informed about it.
- In principle the seller can choose a quality level that is higher, equal, or lower than the desired quality. If the actual quality is lower than the desired quality, then the potential price deduction specified in the contract is due with a probability of 33.3 percent. Hence, with a lower than the desired quality, the specified price deduction is **on average** due in one out of three cases. The seller will only learn after his actual quality choice whether the

price deduction is due or not. A trading day is over after all sellers who have accepted an offer, have determined their actual quality and all bargaining partners have been informed whether there is a price deduction.

- There are more sellers than buyers. All sellers and buyers know this. Each seller (or buyer, resp.) can only sell (buy) one unit per trading day. In the following you will find an exact description of the stages, i.e., which decisions are possible, and an exact description of how payoffs are calculated.

### Detailed Information for Buyers

In the market a certain commodity is traded and each seller sells the same commodity. Each seller can sell to each buyer and each buyer can buy from each seller. The market is organized as follows. We open the market for a trading day and you will receive from us **100 points** for each commodity that you buy. This amount is the same for all buyers. Each buyer and each seller knows that you will receive 100 points per unit of the commodity. You now have the possibility to make an offer. **An offer consists of a price, a desired quality and a potential price deduction.** For making an offer the following rules hold:

1. Per trading day you are only allowed to make one offer. You are not obliged to make an offer.
2. Concerning the "**desired quality**" the following holds: There are **ten possible quality levels**, from which you can choose your desired quality. The lowest quality is 0.1 and the highest one is 1. Below, the impact of the seller's delivered quality on your payoff will be described in more detail. It holds true that your payoff in points is the higher the higher the delivered quality is. A quality choice entails **quality costs** for the seller. **On the leaflet you will find the table with all feasible quality levels and the associated quality costs! All sellers have the same cost schedule.**
3. The price can **at most be 100** and has **at least to cover the seller's quality cost**. For example, if you ask for the quality level 0.3, you have to offer at least a price of 2, for a quality of 0.3 entails costs of 2 units for the seller. Prices have to be in **integers**. In summary, for the determination of the price the following rule holds:

$$100 \geq \text{price} \geq \text{quality costs.}$$

4. The potential price deduction has not to be lower than 0 and larger than 13:

$$13 \geq \text{potential price deduction} \geq 0.$$

**The potential price deduction has to be stated in integers.**

If you have decided on a price, a potential price deduction and a desired quality, please insert them in the box "**proposed offer**" on your **decision sheet**.

**After** you have made your offer, you have to roll a six-sided die. Rolling the die determines whether the price deduction is due in case of an under-provision of the desired quality. The following rule hold: If the numbers **1 or 2** show up (with a probability of 1/3) the price deduction is exacted in case of an under-provision; if the numbers **3, 4, 5, or 6** show up, the price deduction will not be implemented. Please insert the result in the box "**Price deduction due in case of under-provision?**" on your **decision sheet**.

Your offer will be written on the blackboard and transmitted to the sellers. In the sellers' room all offers on a trading day will be written on the blackboard in a **random order**. Moreover, on each trading day a random mechanism determines the order according to which sellers are allowed to choose among the offers. **The sellers will not learn which buyer has made which offer and you as a buyer will not learn which seller has accepted which offer.** Each seller can only accept one offer per trading day. The sellers cannot make counteroffers.

**After** a seller has accepted an offer, he determines the “actual quality” of the commodity, i.e., he chooses a quality level from the quality levels mentioned on the leaflet. Hence, the sellers can choose among exactly the same quality levels as you can. As already mentioned, for the seller choosing a quality entails quality costs.

When a seller determines the “actual quality” of the sold commodity, he does not know whether in case the actual quality falls short of the desired quality the price deduction will be implemented or not. Hence, the seller does not know the numbers that showed up in your throwing of the die. The seller only knows that an under-provision leads to a price deduction with a probability of 1/3.

Each seller personally and completely anonymously decides on the actual quality of which only you will be informed (i.e., it will be inserted in the row “actual quality” on your decision sheet). You will not learn the identity of the seller. Hence, no other buyer and no other seller will learn about the actual quality choice of “your” seller.

If the seller has made an actual quality choice that falls short of your desired quality and if the price deduction is due (i.e., if the die numbers are 1 or 2) then you only have to pay the seller the offered price minus the price deduction. If the price deduction is not due, you have to pay the offered price.

You now have all necessary information to calculate your own payoff as well as the payoff of “your” seller. This ends a trading day and the next one starts. In total there will be 12 trading days during which you can earn money.

### **The Calculation of Buyer’s Payoffs at the End of a Trading Day** (*for sellers this sheet was adapted accordingly but was otherwise identical*)

At the end of a trading day there are the following possibilities:

1. If you have not made an offer, or if your offer has not been accepted by a seller, you have not bought a unit and your **payoff is 0 points**.
2. Your offer has been accepted and the seller’s *actual* quality conforms to or is higher than your desired quality. In this case your payoff and the payoff of your seller is (in points):

<b>Your payoff</b>	=	$100 \times \text{actual quality} - \text{price}$
<b>Seller’s payoff</b>	=	$\text{Price} - \text{quality costs}$

3. Your offer has been accepted, but the seller has chosen a lower than your desired quality.

(a) The price deduction cannot be exacted, because the die numbers 3, 4, 5 or 6 showed up. In this case your payoff and the payoff of your seller is:

<b>Your payoff</b>	=	$100 \times \text{actual quality} - \text{price}$
<b>Seller’s payoff</b>	=	$\text{Price} - \text{quality costs}$

(b) The die numbers 1 and 2 showed up, i.e., the price deduction can be implemented. In this case your payoff and the payoff of your seller is:

<b>Your payoff</b>	=	$100 \times \text{actual quality} - \text{price} + \text{price deduction}$
<b>Seller’s payoff</b>	=	$\text{Price} - \text{quality costs} - \text{price deduction}$



Each seller and each buyer is informed about the details of this payoff calculation. Hence, “your” seller can calculate your payoff in points and you can calculate the payoff of “your” seller. Do you have any questions?

**Buyers’ Control Questionnaire** (*for sellers the control questionnaire was adapted accordingly but was otherwise identical*)

1. You have not made an offer. What is your payoff?
2. (*not asked with sellers*) You would like to demand an effort level of 0.7.
  - (a) What is the highest feasible price?
  - (b) What is the lowest price that you have to offer?
  - (c) What is the maximal potential payoff deduction?
3. You have made the following offer:
 

**Price = 40**  
**Desired quality = 0.8**  
**Potential price deduction = 10**

  - (a) The actual quality of “your” seller is 0.8 and die numbers 1 and 2 showed up.  
 What is your payoff? What is the payoff of “your” seller?
  - (b) The actual quality of your seller is 0.2 and die numbers 1 and 2 showed up.  
 What is your payoff? What is the payoff of “your” seller?
  - (c) The actual quality of your seller is 0.2 and die number 5 showed up.  
 What is your payoff? What is the payoff of “your” seller?
4. You have made the following offer:
 

**Price = 15**  
**Desired quality = 0.9**  
**Potential price deduction = 13**

  - (a) The actual quality of “your” seller is 0.9.  
 What is your payoff? What is the payoff of “your” seller?
  - (b) The actual quality of your seller is 0.1 and die numbers 1 and 2 showed up.  
 What is your payoff? What is the payoff of “your” seller?
  - (c) The actual quality of your seller is 0.1 and die number 3 showed up.  
 What is your payoff? What is the payoff of “your” seller?

---

**Leaflet** (*identical for buyers and sellers*)

Points for buyer per unit bought: 100

For making an offer the following rules hold:

$100 \geq \text{price} \geq \text{quality costs}$

$13 \geq \text{potential price deduction} \geq 0$

**Quality and cost of quality for the seller:**

quality	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
cost of quality	0	1	2	4	6	8	10	12	15	18

---

**Buyers' decision sheet** (1 per period; 12 sheets in total)

*(The decision sheet of sellers was adapted accordingly but was otherwise identical)*

<b>Proposed offer</b>	
Price (p)	
Desired quality	
Potential price deduction (s)	
<b>Price deduction due in case of under-provision?</b> (Please mark the appropriate box)	
Die numbers 1, 2	Die numbers 3, 4, 5, 6
<input type="radio"/> Yes	<input type="radio"/> No
<b>Seller's actual quality (q)</b>	
<b>Payoff of the concluded deal</b>	
<b>No offer made</b>	
Payoff =	
<b>If the desired quality has been delivered or exceeded</b>	
Your payoff $100 \times q - p =$	Payoff seller $p - \text{costs of quality} =$
<b>If the actual quality fell short of the desired quality</b>	
<b>Price deduction is due</b>	
Your payoff $100 \times q - p + s =$	Payoff Seller $p - c(q) - s =$
<b>Price deduction is <u>not</u> due</b>	
Your payoff $100 \times q - p =$	Payoff Seller $p - c(q) =$

### Detailed Information for Sellers

In the market a certain commodity is traded and each seller sells the same commodity. Each seller can sell to each buyer and each buyer can buy from each seller.

Each buyer receives from us on each trading day 100 points, which he can use for buying a commodity. All buyers and sellers know this.

The organization of the market is as follows: We open the market for one trading day. First, without communicating with other buyers, each buyer can make an offer. **An offer consists of a price, a desired quality and a potential price deduction.** There are ten possible quality levels from which the buyer, and you as a seller, respectively, can choose. The lowest quality is 0.1 and the highest quality is 1. The impact of the quality of the delivered good on the payoffs will below be described in more detail. In general, however, it holds true that a high quality increases your cost and the payoff of the buyer. On the leaflet you will find a table with all possible quality levels and the associated quality costs of your quality choice.

The next step after all buyers had the opportunity to make an offer, is the transmission of the offers to this room where the offers will be written on the blackboard in a random order. You will not learn which buyer made which offer. Now a random device determines the order according to which you as a seller can choose among the offers. We implement this as follows. You have to draw one out of 8 cards that are numbered from 1 to 8. The seller who picks card no. 1 is the first who has the opportunity to pick an offer; the seller who draws card no. 2 is the second to pick an offer, and so on. You will make your choice as follows: When it is your turn to make a choice, you state your **seller number** and your chosen offer.

On a trading day you can accept only one offer. You are not obliged to accept an offer. **The buyers will not be informed which offer you have accepted; the buyer will only know, whether their offer has been accepted or not.** Your chosen offer will be deleted from the blackboard and it is the turn of the next seller to make a choice among the remaining offers.

If you have accepted an offer, we ask you to insert on your **decision sheet** the price, the desired quality and the potential price deduction into the box “**Accepted Offer**”. Now you have to decide which quality level you will deliver. As already mentioned, the choice of a quality level is associated with quality costs that you have to bear. **On the leaflet you will find the table with the feasible quality levels and the associated costs for you! Both the buyers and the sellers know this table.** Please insert your actual quality level on your decision sheet in the row “**Actual Quality**”. No other seller will be informed about your quality choice. **We therefore ask you not to talk about your “actual quality”.** Each buyer is only informed about the “actual quality” of “his” seller. Moreover, the buyer will **not** learn the identity of “his” seller. Hence, the anonymity of your quality choice is completely secured.

Whether the potential price deduction is due or not, depends on your quality choice and on chance. If you have delivered or exceeded the desired quality you will receive the accepted price in any case and the price deduction is **not** exacted. If, however, your actual quality fell short of the desired quality, the price deduction may be implemented. Whether the price deduction is due, depends on the result of the following procedure: “your” buyer rolls a six-sided die. If the numbers 1 and 2 turn up, the price deduction will be implemented. If the numbers 3, 4, 5, or 6 show up, the price deduction will not be implemented. We will indicate whether the price deduction – in case of an under-provision of quality – is due or not, by checking the respective box on your decision sheet. You will receive this information after you have determined your actual quality level.

You now have all necessary information to calculate your own payoff as well as the payoff of “your” seller. This ends a trading day and the next one starts. In total there will be 12 trading days during which you can earn money.