TI 2009-030/1
Tinbergen Institute Discussion Paper

The Interaction between Explicit and Relational Incentives: An Experiment

Randolph Sloof
Joep Sonnemans

Faculty of Economics and Business, University of Amsterdam, and Tinbergen Institute.

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Most TI discussion papers can be downloaded at http://www.tinbergen.nl.

The interaction between explicit and relational incentives: An experiment

Randolph Sloof[*] and Joep Sonnemans

University of Amsterdam
School of Economics
Roetersstraat 11
1018 WB Amsterdam
the Netherlands

April 12, 2009

Abstract

We consider repeated trust game experiments to study the interplay between explicit and relational incentives. After having gained experience with two payoff variations of the trust game, subjects in the final part explicitly choose which of these two variants to play. Theory predicts that subjects will choose the payoff dominated game (representing a bad explicit contract), because this game better sustains (implicit) relational incentives backed by either reputational or reciprocity considerations. We also explicitly test how game choice is affected by the length of the repeated game.

Keywords: relational contracts, explicit incentives, crowding out, experiments

JEL codes: C91, M52, J41

1 Introduction

Because there will always exist contingencies not covered by the formal contract, any real world contract is necessarily incomplete. This in general precludes reaching full efficiency by using formal contracts alone. Implicit, informal agreements may therefore potentially serve a useful purpose by bringing outcomes closer to first best. One important class of implicit incentives are those provided by so-called *relational contracts*, where the informal agreement is backed by reputational considerations in a repeated (long-term) relationship. In order to understand when and how such relational contracts can be efficiency enhancing, it is important to know exactly how explicit and implicit contracts interact.

In principle two possibilities can be distinguished. First, explicit and implicit incentives may act as complements. In this case better formal contracts increase the scope for implicit contracts. Second, the

[*] Corresponding author. e-mail: r.sloof@uva.nl; phone: +31 (0)20 5255241; fax: +31 (0)20 5254310.

two types of incentives may act as substitutes. In that case, stronger explicit incentives through formal contracts may crowd out the implicit incentives provided by relational contracts and may even reduce efficiency. Some existing theoretical models predict that both situations may indeed occur; see Baker et al. (1994) and Schmidt and Schnitzer (1995). Given the potential interaction between the two types of incentives, a large number of theoretical papers expand on the idea that formal contracts can be chosen optimally as to best facilitate informal relational contracts.[1]

Especially the observation that under certain circumstances formal and relational contracts may be substitutes has received considerable theoretical attention. The intuition behind this somewhat counter-intuitive prediction is that better explicit contracts reduce the worst possible punishment that follows after deviation from the relational contract. Given this improvement in the fall-back outcome, it becomes more difficult to sustain relational contracts as a self-enforcing agreement. The availability of better explicit incentives may thus crowd out these self-enforcing agreements and thereby reduce efficiency. As a result, parties may deliberately write an 'inferior' (i.e., more incomplete) formal contract, in order to sustain the better relational contract (cf. Bernheim and Whinston 1998).

In this paper we present the results from a laboratory experiment designed to test these predictions. Our experiment is based on the two versions of the well-known trust game of Kreps (1990) depicted in Figure 1. In the trust game a worker first decides whether to work according to the letter of the formal contract, or to put in high effort beyond the level that can be contractually enforced. If the worker works according to rule, it is assumed that he does not get a bonus on top of his wage. If, however, the worker puts in higher effort than formally required, the employer decides whether to reward him with a bonus or not. This bonus is discretionary and not part of the formal employment contract. Figures 1a and 1b reflect two different situations. Figure 1a represents the situation of a 'bad' formal contract or, alternatively, a very incomplete explicit contract. If parties strictly behave according to contract, only a moderate fraction of 33% of the maximum available joint surplus is captured. This is not the case in the trust game of Figure 1b, which reflects the situation of a 'good' or less incomplete formal contract. Then parties capture 46% if they behave according to rule.

[Figure 1]

Standard theory predicts that, in a static one-shot relationship, a selfish employer will not pay the bonus. Anticipating this, the worker will just work according to the formal contract. Both parties would thus be best off signing the best formal contract available (cf. Bernheim and Whinston 1998), here represented by game G of Figure 1b. In case of repeated interaction the employer may, however, credibly commit to pay the bonus. In particular, when the trust game is infinitely repeated, trigger strategies may sustain cooperation on the (High effort, Bonus) outcome. For the employer to be indeed willing to keep her implicit promise to pay the bonus, the prospects of future cooperation should be sufficiently valuable. This translates into a sufficiently high discount factor $\delta \geq \frac{12}{24-d}$, with d the payoff the employer gets when the worker works according to rule (and hence no bonus is paid). Note that the higher d, the more difficult it becomes to support cooperation as an equilibrium outcome. So, the better explicit contracts

[1] See e.g. , Baker et al. (2001, 2002), Bernheim and Whinston (1998), Blonski and Spagnolo (2007), Bragelien (2002), Che and Yoo (2001), Demougin and Fabel (2004), Garvey (1995), Halonen (2002), Hart (2001), Itoh and Morita (2006), Kvaloy and Olsen (2006a, 2006b), Levin (2003), MacLeod (2007a), Murdock (2002), Rayo (2007), and Schottner (2007).

are, the *smaller* the scope for relational contracts becomes. The general underlying intuition here is that relational contracts require sufficient additional gains from trade (on top of the gains generated by the formal contract) in order to be self-enforcing; see e.g. MacLeod (2007b) for an insightful discussion.

In our experiment we choose a discount factor equal to $\delta = \frac{2}{3}$. This implies that (theoretically) cooperation can be sustained through a relational contract when the bad explicit contract of Figure 1a applies, but *not* when the good contract of Figure 1b is in place. Subjects are confronted with both situations, so that we can test whether better explicit incentives indeed crowd out relational incentives. Moreover, in the final part of the experiment employers explicitly choose between the bad and the good formal contract.[2] In that way we are able to test whether parties indeed prefer to write an inferior formal contract to better sustain relational incentives.

Relational contracts provide one way in which implicit incentives are given. Another prominent driver of implicit incentives are social preferences. The discussion up till now assumed that parties are selfish and only care about their own material payoffs. But in reality many people are guided by alternative motivations, like e.g. fairness and reciprocity, as well. This may have profound implications for the predicted outcome. In the trust games of Figure 1 for instance, if the employer has a preference to react in kind to the kind choice of the worker to put in high effort, this reciprocal motivation may sustain cooperation even in a static one-shot setting. And because the worker's choice to put in high effort becomes more kind the lower d is, in fact a stronger reciprocal reaction of the employer is predicted (that is, a higher probability of granting the bonus) in game B than in game G. Just as was the case with relational contracts, therefore, parties may prefer the bad explicit contract in order to sustain the better implicit incentives from reciprocal motivations.[3] From that perspective the properties of implicit incentives generated by repeated interaction are similar to those created by reciprocal motivations, see also Fehr and Falk (2002) and Scott (2003, p. 1681).

We also address this issue in our experiment, by explicitly looking at repeated trust games of a given fixed length. In particular, we consider both the case where employer and worker interact for one round only and the one where they interact for exactly three rounds. In the latter treatment, the length of the game equals the expected length of the infinitely repeated game with $\delta = \frac{2}{3}$. By comparing the results with the infinite games then, it can be identified whether cooperation indeed depends on the 'shadow of the future' as standard theory predicts, or merely on the length of the game (Dal-Bo 2005, pp. 1591-1592).

This paper adds to a small but growing empirical literature studying the interaction between explicit and implicit incentives.[4] The distinguishing feature of our experiment lies in the combination of two fac-

[2] Although in theoretical models it is typically assumed that the two parties jointly agree on which contract is implemented, experimentally it is much more attractive to let one of the parties dictate the contract. See Appendix B for a justification why we let the employer choose the stage game.

[3] A number of experiments on the hidden costs of incentives and/or control in (one-shot) trust game like principal-agent settings obtain findings that are in this spirit. Fehr and Rockenbach (2003) and Fehr and List (2004), for instance, introduce a punishment option in the investment game of Berg et al. (1995), where back transfers below a certain threshold can be fined. They observe that when the principal intentionally refrains from installing this punishment option, back transfers are highest. In Falk and Kosfeld (2006) the principal has the option to restrict the agent's effort to a certain minimum amount. Principals who do so in the experiment actually induce lower average effort levels. These findings thus provide support for the crowding out of alternative (social or intrinsic) motivations by explicit incentives / control (see also Gneezy and Rustichini (2000a, 2000b) for empirical evidence on this).

[4] In the words of Brown et al. (2004, p. 775): "...questions regarding the interaction between explicit and implicit incentives have remained largely unexplored empirically. For instance, is it indeed the case that the availability of better explicit incentives makes self-enforcing agreements less likely, as hypothesized by Baker, Gibbons and Murphy (1994) and Schmidt and Schnitzer (1995)?" Or, as Fehr and Schmidt (2007, pp. 180-181) put it, "...we are just beginning to understand

tors. First, to allow for relational contracts that are fully in spirit of the original theoretical contributions by MacLeod and Malcomson (1989), Baker et al. (1994) and Schmidt and Schnitzer (1995), we explicitly consider *infinitely* repeated trust games. Second, within this setting we let parties choose between two different games (after having gained experience with the games in isolation), which represent in reduced form two different types of formal contracts. In that way we directly test whether explicit incentives may indeed crowd out self-enforcing agreements, so that sophisticated parties may set weak explicit incentives in order to better sustain relational contracts.

This paper proceeds as follows. In the next section we discuss the theoretical predictions for the simple (repeated) trust game that we consider. In particular, we illustrate that better explicit incentives may crowd out implicit incentives, which are either backed by reputation in an infinitely repeated game setting, or by reciprocal motivations. Parties may therefore choose a more incomplete contract to facilitate the better implicit contract. Section 3 presents the details of our experimental design and also relates our design to previous experiments. Results are reported in Section 4. The final section summarizes and concludes.

2 Theory

Our experiment is based on the B-game and the G-game as depicted in Figure 1. Both games have the same general decision structure, which is reflected in Figure 2. Player 1 first chooses between Not trust and Trust. If player 1 chooses Trust, player 2 subsequently chooses between Honor and Betray. Payoffs are such that choosing Betray yields player 2 the most in monetary payoffs, whereas Honor corresponds to (costly) rewarding player 1 for his 'kind' choice of Trust. From $f > e > d$ and $c > b > a$ it immediately follows that, if both players are selfish, (Not trust, Betray) is the unique subgame perfect equilibrium (SPE) of the one-shot game. Given $b + d < c + e$, this outcome is inefficient.

[Figure 2]

The trust game of Figure 2 can be interpreted as a reduced form model of explicit and implicit contracts.[5] Player 1 is an employee who chooses an (possibly multi-dimensional) effort level. Part of this effort choice is contractible. Focusing on these explicit incentives only, outcome Not trust results. Effort is not fully contractible though, so explicit contracts do not yield first best. The employee can supplement his contractible effort with additional effort up to the first best level, hoping that the employer (player 2) pays a non-contractible bonus on top of his salary in return. In that way implicit incentives may lead the parties away from the inefficient outcome dictated by the letter of the formal contract. The following two subsections illustrate this, for the two cases where implicit incentives are either backed by reputational considerations or by reciprocal motivations.

the interaction of explicit and implicit incentives, which is a fascinating field for future research."

[5] Building on Schmidt and Schnitzer (1995) and Baker et al. (1994), we discuss in Appendix A a bare bones model of explicit and implicit incentives that underlies this reduced form specification.

2.1 Relational contracts

Suppose the trust game of Figure 2 is infinitely repeated. The efficient outcome (Trust, Honor) can then be sustained as SPE whenever the discount factor δ (with $0 \leq \delta < 1$) is sufficiently high. In particular, suppose the players use grim-trigger strategies in which they start with Trust and Honor, respectively, and continue to choose these actions as long as only actions Trust and Honor have been observed in the past. (Otherwise, players move to the (No trust, Betray) outcome.) Player 2 then has an incentive to keep its promise to honor trust iff:

$$\frac{1}{1-\delta} \cdot e \geq f + \frac{\delta}{1-\delta} \cdot d \Longrightarrow \delta \geq \frac{f-e}{f-d} \equiv \underline{\delta} \tag{1}$$

Note that the cutoff value $\underline{\delta}$ is increasing in parameters d and f and decreasing in e. If the discount factor exceeds $\underline{\delta}$, a *relational contract* exists under which players coordinate on (Trust, Honor) as informal, self-enforcing agreement.

The general implications of better explicit contracts for such relational contracts can easily be illustrated within the simple reduced form setup of Figure 2. Following Schmidt and Schnitzer (1995, p. 198), better explicit incentives have two opposing effects:

1. The worst possible punishment in case of deviation from the relational contract is *less* severe. Here this implies that d increases;
2. A deviation becomes relatively less attractive, because implicit incentives add less surplus. Here this corresponds to a decrease in $f - e$, i.e. either f decreases or e increases (or both).

The first effect leads to an upward pressure on $\underline{\delta}$, reducing the scope for relational contracts. The second effect leads to a lower $\underline{\delta}$ and thus a larger scope for relational incentives. If this second effect dominates, explicit and relational contracts act as complements. But if the first effect dominates, explicit and implicit incentives act as substitutes.[6] In that case the availability of better explicit incentives makes self-enforcing agreements less likely. Parties may then deliberately opt for formal contracts with weak(er) explicit incentives. As already noted in the Introduction, a large number of theoretical papers build on this somewhat counterintuitive prediction.

In our experiment we particularly focus on variations in d. The above analysis then predicts that, the higher d, the less likely it is that player 1 chooses Trust (and player 2 chooses Honor) in the infinitely repeated trust game. Moreover, when given the choice player 2 may explicitly prefer the situation with a low value of d as given by game B in Figure 1a, over a situation in which d is high as reflected by game G in Figure 1b.

2.2 Reciprocity motivations

In the Introduction it was suggested that allowing for social preferences in the one-shot game leads to similar comparative statics predictions as the ones derived above under relational contracts. To illustrate this more formally we make use of the theory of intention-based reciprocity as developed by Rabin (1993)

[6] Clearly, in general the interaction between explicit and implicit incentives may be determined by other effects as well. For instance, in a repeated agency framework Pearce and Stacchetti (1998) derive that explicit incentives (salary) and implicit bonuses are substitutes, in order to smooth the income path of the risk-averse agent over time.

and refined by Dufwenberg and Kirchsteiger (2004). For ease of exposition we assume here that player 1 is selfish and motivated by money maximization only.[7] Player 2 may be motivated by reciprocity though, implying that she is willing to sacrifice in order to reward (punish) player one's good (bad) intentions. In particular, her utility function equals:

$$U_2 = m_2 + Y_2 \cdot \kappa \cdot \lambda \tag{2}$$

In this specification m_2 denotes the monetary payoffs of player 2 and term $Y_2 \cdot \kappa \cdot \lambda$ gives her reciprocity payoffs. Parameter $Y_2 \geq 0$ reflects the reciprocal attitude of player 2. The larger Y_2, the more sensitive to reciprocity she is. Factor κ measures the kindness of player 2 towards player 1. It is positive if she is kind to player 1 and negative if she is unkind to him. Factor λ gives player 2's belief about how kind player 1 is towards her. It is positive when she believes player 1 is kind to her, and negative when she thinks he is unkind. A key characteristic of the model is that a reciprocal player 2 has an incentive to match the sign of her own kindness κ with the sign of the perceived kindness λ of player 1.

In the theory of Dufwenberg and Kirchsteiger (2004), the (perceived) kindness of a particular choice is measured by the difference between what a player actually gives to another player and the average of the minimum and maximum amount she could give him in principle. Now, the larger d in the trust game of Figure 2, the lower the perceived kindness of player 1's choice of Trust is. Formally, λ decreases with d. Player 2 therefore has less incentives to reciprocate in turn by choosing Honor. (Formally, κ is increasing in the probability of choosing Honor.) This in turn reduces player 1's incentives to trust. Similarly, a decrease in $f - e$ has two effects. The direct effect is that the monetary incentives to choose Betray decrease relative to Honor. An indirect effect is that, for given expectations about player 2's reaction, player 1's choice of Trust is perceived as less kind. This follows because the additional gain player 2 in principle can obtain by reneging decreases. In Dufwenberg and Kirchsteiger's specification, the first (direct) effect dominates the second (indirect) effect. The scope for reciprocity thus increases when $f - e$ decreases.

For player 1 to be willing to choose Trust, player 2 should be sufficiently likely to choose Honor. This requires that player 2 is sufficiently motivated by reciprocity. From the formal analysis relegated to Appendix B it follows that this is the case whenever:

$$Y_2 \geq \frac{2 \cdot \underline{\delta}}{(c-a) - \underline{\delta} \cdot (b-a)} \equiv \underline{Y} \tag{3}$$

When this inequality holds, (a selfish) player 1 chooses Trust for sure. The cutoff value $\underline{Y}$ is increasing in $\underline{\delta}$. Hence the scope for reciprocity (as measured by $\underline{Y}$) and the scope for relational contracts (as reflected by $\underline{\delta}$) move together. As a result, similar comparative statics predictions regarding the interaction between explicit and implicit incentives are obtained as in the previous subsection. In particular, (i) the larger payoff d, the less scope there is for intention-based reciprocity, and (ii) the smaller $f - e$, the larger this scope is. Solely focusing on variations in d we thus predict that stronger explicit incentives

[7] The formal analysis in Appendix B starts from the more general assumption that player 1 is motivated by reciprocity as well.

(a higher d) may crowd out implicit incentives based on an informal reciprocity mechanism.[8], [9]

The formal analysis of intention-based reciprocity is already quite involved for the one-shot game (cf. Appendix B). This follows from the fact that the reciprocity payoffs depend on the players' beliefs, so psychological game theory is needed to derive the equilibrium predictions. Matters become even much more complex in a repeated game, because player 2's behavior is then not only guided by her reciprocal motivations towards player 1's past actions (and what player 1 would have done off the equilibrium path), but also by her expectations about player 1's future behavior. An equilibrium analysis of the repeated trust game assuming players have preferences like in (2) is therefore beyond the scope of this paper. But given the co-movement of $\underline{Y}$ and $\underline{\delta}$ as derived above, it seems quite likely that similar comparative statics predictions are obtained when the trust game of Figure 2 is played repeatedly in a row. In fact, it seems reasonable to conjecture that reciprocity and repetition reinforce each other in providing implicit incentives.[10]

Similar remarks apply to the situation where player 2 first fixes the formal contract, i.e. first chooses the version of the trust game that is going to be played (cf. Figure 1). The contract chosen then not only affects the players' subsequent behavior, but the choice of contract itself will be guided by reciprocity motivations as well. This makes the formal analysis much more complex. Although we do not provide a complete equilibrium analysis for this case, in Appendix B it is shown that player 1 is more likely to choose Trust when the bad explicit contract is chosen instead of the good explicit contract.[11] So, also when the bad contract is chosen endogenously (by player 2) rather than exogenously given, it increases the probability of the cooperative outcome.[12]

[8] Also here it holds that in general the interaction between explicit and implicit incentives may be guided by other factors as well. For instance, Hart and Moore (2007, 2008) envisage contracts as providing a reference point for feelings of entitlement. In that way contracts set the benchmark for (negative) reciprocal reactions when a party gets less than the best possible outcome permitted by the contract.

[9] In Sloof et al. (2007) we study the effectiveness of *informational rents* as incentive instrument against holdup. There we also obtain the prediction that explicit incentives (in that setting provided by private information about the actual investment made) may crowd out an informal reciprocity mechanism in place, and we find experimental evidence in line with this prediction.

[10] Some theoretical papers predict that repeated interaction and alternative motivations act as complements. MacLeod (2007a), for instance, incorporates a small preference for honesty in a finitely repeated relationship. Parties can then achieve close to first best by means of a relational contract. The intuition is that in the final period the party with the bargaining power has, given his preference for honesty, some incentive to stick to the implicit agreement (whereas under selfish preferences this party would certainly renege). Therefore, some surplus in the final period can be created through trade, which can be used as a 'stick' to discipline the behavior of this party in the period before. This in turn increases the stick that can be used one period earlier, and so on back to the initial trade period. Murdock (2002) incorporates intrinsic motivation in his analysis of static and relational contracts. He shows that, under certain circumstances, intrinsic motivation harms the profitability of static contracts, but by doing so improves the profitability of long term, relational contracts. His model more generally predicts that intrinsic motivation and relational contracts are complements.

[11] This appendix also shows that the key feature that implementing a bad formal contract favorably affects player 2's honor behavior only applies when player 2 chooses the contract and *not* when player 1 does so. For this reason we let player 2 make the contract choice in our experiment.

[12] Chen (2000) studies a somewhat different contracting setting where parties have a certain tendency to behave trustworthy and keep promises. To save on contracting costs they may therefore opt for an incomplete contract that relies on trustworthy behavior, rather than for a complete contract under which costly arrangements are made to make effort (quality) verifiable. Chen formally shows that an increase in contracting costs may be efficiency enhancing, the intuition being that parties may switch from a complete contract to an incomplete one and thereby save on the socially wasteful contracting costs. A similar result applies in our setting; higher contracting costs may make the good contract relatively less attractive, inducing parties to switch to the more incomplete bad contract (and thereby increasing efficiency).

3 Experimental design and hypotheses

3.1 Treatments and sessions

Our experiment is based on a 3 by 3 treatments design. In all sessions we considered repeated versions of the two trust games depicted in Figure 1. Each repeated trust game consisted of a number of rounds and in each session different repeated games were played after each other in different periods. Within sessions we kept the type of repetition fixed. The first set of sessions considered one-shot trust games only. In these sessions we thus had only one round per period ($l = 1$). In a second set of sessions finitely repeated games with a fixed length of $l = 3$ rounds were studied. The final set of sessions only contained 'infinitely' repeated games with random ending; after each round the probability of a new round equalled $\delta = \frac{2}{3}$. Note that the expected length of these infinite games equals the length of the sessions with $l = 3$; $\frac{1}{1-\delta} = 3$ for $\delta = \frac{2}{3}$.

Within sessions we varied the type of trust game that is considered (either game B or game G) and whether the game is exogenously given or endogenously chosen. Each session consisted of three parts. Table 1 provides an overview of the different sessions. In the first part, the (repeated) trust game to be played was exogenously given, by game B say. In the second part then the other trust game (game G in the example) applied. In the final part the trust game to be played (B or G) was endogenously chosen by player 2. To control for order effects two different orders were considered, starting the session either with game B or with game G. In the repeated game sessions with either $l = 3$ or $\delta = \frac{2}{3}$, each part consisted of 15 periods, i.e. 15 repeated games of expected length 3. Subjects thus (on average) played the trust game 45 times. In order to keep the (average) number of plays per part constant, the number of periods was tripled to 45 in the $l = 1$ sessions with one-shot play.

At the start of the experiment subjects were informed that there were three different parts. Once they had completed the instructions for the first part, they learned their roles. Roles were kept fixed during the whole experiment. At the end of part one subjects obtained an overview of the outcomes of all the games they played in that part. Having studied this overview, subjects continued with the instructions for part 2. These basically informed subjects that only the (payoffs in the) game to be played changes. Upon completion of the second part, subjects again obtained a personal history overview of the outcomes in part 2. Then the instructions for the final part were distributed. Before players' 2 actually had to choose between games, again a personal history overview was presented to them with the outcomes of part one and part two next to each other on the same screen (together with the overall earnings in these two parts). This was done to facilitate the choice between games.

Matching was based on a stranger design. Subjects were informed that they would never be paired with the same other subject in two consecutive periods and control questions were used to explicitly remind subjects of that.[13] Moreover, unknown to the subjects we divided them into two groups that were independently matched if the number of show ups allowed us to do so. In particular, when 16 or more subjects showed up for a session, we formed two independent matching groups.

The experiment was framed neutrally. We referred to Game B as the "Blue structure" and game G as the "Yellow structure". The game trees were printed with a frame of the corresponding color, both

[13] The control questions also reminded the subjects of the fact that in all rounds of a given period, they were matched to the same other subject. Control quesions were included in the instructions for each part. Subjects could only proceed after having answered all control questions correctly.

Table 1: Overview of sessions and treatments

session	length (# rnds)	# subj.	# grps	# periods per part	Part I	Part II	Part III
1	$l = 1$	14	1	45	game B	game G	B versus G
2	$l = 1$	18	2	45	game B	game G	B versus G
3	$l = 1$	20	2	45	game G	game B	G versus B
4	$l = 1$	12	1	45	game G	game B	G versus B
5	$l = 3$	12	1	15	game B	game G	B versus G
6	$l = 3$	20	2	15	game B	game G	B versus G
7	$l = 3$	22	2	15	game G	game B	G versus B
8	$l = 3$	14	1	15	game G	game B	G versus B
9	$\delta = \frac{2}{3}$	18	2	15	game B	game G	B versus G
10	$\delta = \frac{2}{3}$	18	2	15	game B	game G	B versus G
11	$\delta = \frac{2}{3}$	22	2	15	game G	game B	G versus B
12	$\delta = \frac{2}{3}$	22	2	15	game G	game B	G versus B

on the computer screen and on the summary of the instructions handed out on paper. Players 1 and 2 were labelled participant A and B, respectively. A's chose between A-left and A-right and in the latter case B's had to choose between B-left and B-right. The corresponding payoffs to these choices were as in Figure 1.

In the infinite games the lengths of the separate repeated games were drawn in advance. We did so for two reasons. First, the four different sessions that considered the infinite games are then better comparable. Second, in every period the length of the game is then the same for all pairs in the session. If this were not the case, each period would last as long as the length of the longest game, lengthening the duration of the $\delta = \frac{2}{3}$ sessions to a considerable extent. The random draw to determine whether a new round would occur in a given game was visualized on the screen by using a one-armed bandit which started to turn automatically after player 2 made her decision.

Subjects received a show up fee of 7 euros. Their overall earnings equalled the sum of this showup fee and the total number of points they earned in the three different parts together. The conversion rate was set at 50 points for one euro. Subjects on average earned 31.5 euros in around one and a half to two hours the different sessions took. Most participants were undergraduate students in either economics, science or psychology.

3.2 Hypotheses

The main purpose of the experiment is to test the theoretical prediction that better explicit incentives may actually weaken implicit (relational) incentives. Subjects (i.c. players 2) may therefore deliberately choose the inferior explicit contract (game B) in order to facilitate the better implicit contract. As explained in Section 2, these predictions can be backed by standard repeated game like arguments and/or by reciprocal motivations.

First assume subjects are entirely selfish. In that case the single subgame perfect equilibrium of the finitely repeated trust games is (Not trust, Betray) in all games. Therefore, player 2 is expected to prefer game G over game B in the treatments with either $l = 1$ or $l = 3$. The predictions for the infinite games follow from the analysis in Subsection 2.1. For game B the cutoff value for the continuation probability δ equals $\underline{\delta} = \frac{3}{5}$ whereas for game G this cutoff value equals $\underline{\delta} = \frac{12}{17}$. Because $\frac{3}{5} < \delta = \frac{2}{3} < \frac{12}{17}$, cooperation on (Trust, Honor) can be sustained as equilibrium outcome under game B, but not when game G applies. Player 2 may therefore prefer game B over game G in the infinite games. Comparing the infinite games with the finite ones, we expect game B to be chosen more often in the infinite games. Moreover, given that game B applies, player 1 is more likely to choose Trust and player 2 is more likely to choose Honor.

Next assume that subjects have reciprocal motivations. As illustrated in Subsection 2.2, game B allows for a larger scope for intention-based reciprocity than game G does. In particular, it holds that cutoff $\underline{Y}$ is lower when game B applies than when game G applies ($\frac{1}{8} < \frac{2}{13}$). Player 1 is thus more likely to choose Trust under game B and therefore Player 2 may already prefer game B in the finite games.

Overall we expect that, for a given level of repetition, player 2 is more likely to honor trust and player 1 is more likely to trust in game B than in game G. This may induce players 2 to choose game B instead of game G when giving them the opportunity to do so. Moreover, because it seems reasonable to conjecture that reciprocity and repetition complement each other, we also expect that when we move from treatment $l = 1$ to $l = 3$ to $\delta = \frac{2}{3}$, player 2 is more likely to honor trust, player 1 is more likely to trust, and player 2 is more likely to choose game B in part 3.

3.3 Related experiments

Engle-Warnick and Slonim (2004, 2006) compare five round repeated trust games ($l = 5$) with infinitely repeated trust games ($\delta = 0.8$). Their main purpose is to infer the strategies subjects employ in these games and how these strategies evolve over time. One striking observation is that in the infinite games, only one strategy for the first moving player 1 is inferred, viz. grim-trigger. Strategies evolve over time in a best response manner. In the finite games this implies that player 1 increasingly chooses Not trust while in the infinite games players increasingly use strategies that sustain the cooperative outcome (Trust, Honor). As a result, once subjects have gained experience, the infinite games induce a larger level of trust and higher levels of efficiency than the finite games do.[14]

Other experimental studies consider repeated versions of the investment game of Berg et al. (1995). The investment game is just an extension of the (mini) trust game with binary choices reflected in Figure 1 to a situation where both players have multiple available choices. Cochard et al. (2004) compare seven round investment games with one round investment games. Their main finding is that repetition increases trust on average, although there is a clear end round effect. Keser (2002) evaluates two different reputation management mechanisms. A 20 round repetition of the investment game with a stranger design serves as baseline. In two other treatments player 1 rates player 2 once the latter has

[14] Anderhub et al. (2002) study the impact of private information on trusting behavior. The standard finitely repeated trust game is compared with one in which player 2 is privately informed about her type; with probability equal to $\frac{1}{3}$ she is forced to reward player 1's choice to trust. Differences in trusting behavior between the two treatments appear to be very small.

taken her decision, either as ‘positive’, ‘negative’ or ‘neutral’. Under ‘short run reputation’ only the most recent rating becomes available to the next trading partner, under ‘long run reputation’ the full set of ratings becomes available. The two reputation mechanisms induce significantly higher investment levels and prevent that investment decays over time.

van Huyck et al. (1995) vary the order of play in the investment game. They compare a discretion treatment which corresponds to the standard investment game, with a ‘commitment’ treatment where player 2 first commits to a return rate before player 1 invests. In line with theoretical predictions, commitment significantly increases investments and thus efficiency. In van Huyck et al. (2001) a third reputation treatment is considered where subjects play the standard investment game repeatedly in a row (based on a partners design). The main finding is that reputation is an imperfect substitute for commitment, because it is less efficient on average.[15]

Unlike we do, none of the above studies considers subjects’ endogenous choice between different payoff variations of the constituent trust game.[16] Hence these studies have little to say about whether subjects will indeed opt for weak explicit incentives in order to better sustain relational contracts. Other experiments are explicitly designed to study subjects’ actual choice of contract. Fehr et al. (2007), for instance, allow for three different contracts within a gift-exchange setting. In a so-called ‘trust’ contract the employer pays an (high) up-front wage in the hope that the worker will reciprocate with (high) effort in return. Incentive contracts contain a fine as well. This fine is (stochastically) imposed when the worker supplies less effort than desired. Finally, in a bonus contract the employer promises to pay a non-contractible bonus ex post for good performance. In the experiments the bonus contract performs best whereas the trust contract performs worst.[17] A sensible explanation for these findings is that subjects are partly driven by (distributive) fairness concerns.

In the Fehr et al. experiments interactions are one-shot. In contrast, we explicitly study infinitely repeated interactions as well, to allow for relational contracts that are fully in spirit of the ones discussed in the theoretical literature. Given that this makes the experimental game much more complex, we have designed the actual contracts between which subjects can choose in highly reduced form. Instead of having different types of actual contracts as in Fehr et al. (2007), we let subjects choose between the two trust games of Figure 1. By doing so we are able to explore the interaction between explicit and relational incentives in a simple setup.

[15] Falk et al. (1999) and Gächter and Falk (2002) compare one-shot (strangers) and repeated (partners for 10 consecutive rounds) gift exchange treatments. Their main finding is that in a repeated relationship reputation and reciprocity reinforce each other (i.e. are complementary).

[16] In his study of infinitely repeated Prisoners’ dilemma games, Dal-Bo (2005) compares two different payoff matrices. In one of these cooperation can be sustained as equilibrium in the repeated game whereas in the other it cannot (for a continuation probability of one half). The rate of cooperation is indeed higher in the former. Dal Bo does not let subjects choose between different payoff matrices.

[17] Fehr and Schmidt (2007) report an experiment that allows for combined incentive-bonus contracts in the Fehr et al. (2007) setup. Among other things they find that in the combined contracts employers reward high effort levels less generously than in pure bonus contracts; combined contracts thus provide lower implicit incentives (crowding out). One explanation put forward is that the explicit threat of using a sanction may be interpreted as unkind, which triggers a negative reciprocal reaction by the agent. Another one is that the contract offer may have been interpreted as a signal of the principal’s trustworthiness.

4 Results

In this section we first provide a general overview of the aggregate outcomes observed under the two different contracts and the actual contract choices made. Then we more formally test, for each given length of the repeated game, whether the bad explicit contract indeed better facilitates coordination on the cooperative (Trust, Honor) outcome. We also compare the different game lengths and verify whether reciprocity and repetition reinforce each other in providing implicit incentives. The final subsection analyses the choice of contract.

4.1 General overview of aggregate outcomes

Figure 3 provides a percentage wise overview of the observed outcomes in the two games. Here all outcomes of different rounds and periods are bunched together, so the overview can only be used to paint a rough picture. Because sessions and matching groups are equally balanced over the different orders of the (exogenous) treatments (cf. Table 1), we have pooled the data from sessions that differ only in the order of treatments.

Some observations are immediate from Figure 3. First, the exogenous treatments lead by and large to the same distribution of outcomes. For these treatments there is no indication that the bad contract induces a larger fraction of cooperative outcomes than the good contract does; the percentages belonging to the (Trust, Honor) outcome are very similar across games, see the lower segments in the bars belonging to the exogenous treatments. Second, in the endogenous treatments we do observe that implementation of the bad contract instead of the good contract makes it more likely that outcome (Trust, Honor) is observed, independent of the length of the game. Not only is player 1 more likely to choose Trust, conditional on him making this choice player 2 is also more likely to honor this trust. Third, in line with theoretical predictions the infinite games induce much more cooperation than the finite games do.

[Figure 3]

Actual contract choices in the endogenous treatments are reported within parentheses below the corresponding bars in Figure 3. For all different game lengths the bad contract is chosen in a non-neglible fraction of the cases. Somewhat surprisingly, the bad contract is chosen more often in the one-shot games than in the repeated games.

Although inconclusive, the overview of aggregate outcomes in this subsection suggests that the bad contract may indeed make cooperation more likely, but only when this contract is endogenously chosen by the subjects themselves. This may in turn explain why players 2 do opt for the bad explicit contract in a non-negligible fraction of cases.

4.2 Impact of formal contract on implicit incentives

In this subsection we more formally test whether the bad contract better facilitates cooperation. We do so for each game length ($l = 1$, $l = 3$ or $\delta = \frac{2}{3}$) separately and focus on the outcomes of the individual stage games. For the $l = 1$ treatments this simply corresponds to the single (stage) game played in each match. In the two other treatments we both look at the first stage game of each match (i.e. the stage game played in round one of a given period) and, in a separate analysis, at the final stage game of each match (the stage game played in the last round of a given period).

Honor versus betray. We first look at player 2's decision whether to honor trust or not, given that player 1 decided to trust. Table 2 reports the estimates of random effect probit models of the probability that player 2 chooses Honor in response to Trust (with the id's of player 2 as the clustering variable). In all specifications we include, besides a constant, a 0/1-dummy indicating whether or not the bad contract applies in that period and another one indicating whether or not the contract is endogenously chosen. Also the interaction between these two dummies is included. To take account of learning, we include a time trend labelled "match number" as well, which simply reflects the number of matches player 2 has been involved in up till now (including the present match).

For the infinite games we find that subjects are indeed more likely to honor trust when the bad contract applies; the game B dummy is positive and highly significant in both specifications. In the finite games honor behavior is independent of the type of game in the treatments where the game is exogenously given. But when the type of contract is endogenously chosen, players 2 are significantly more likely to honor (the first occurrence of) trust when the bad contract applies. Overall we thus obtain some evidence that player 2 is more likely to honor trust when game B applies.

Trusting behavior. Table 3 reports random effect estimates of the probability that player 1 chooses trust (with now the id's of player 1 as the clustering variable). Here the pattern is more clear. In the exogenous treatments there is no difference in trusting behavior under the two contracts. This follows from the fact that the game B dummy is insignificant in all specifications. In contrast, the interaction term with the endogeneity dummy is highly significant in all columns, indicating that player 1 is more likely to trust when the bad contract is endogenously chosen over the good contract. Taken together, we thus find strong evidence that the bad contract facilitates trust only if this contract is endogenously chosen. Another observation that can be made from Table 3 is that trusting behavior decays over time; the time trend 'match number' is significantly negative in all cases.

Cooperation rates. Tables 2 and 3 consider the individual choices of players 1 and 2. Whether cooperation on (Trust, Honor) occurs, depends on both their decisions. To take account of the interdependencies between subjects that are (repeatedly) matched, we analyze these joint decisions at the matching group level. As explained in the previous section, we formed two independent matching groups when at least 16 subjects showed up for a session. In total, we have 6 matching groups in the sessions with $l = 1$ and $l = 3$. For $\delta = \frac{2}{3}$ we have 8 matching groups (cf. Table 1).

For each matching group we calculate the fraction of cooperative outcomes (Trust, Honor) observed under the two contracts. We do so separately for the case where the contract is exogenously given and the one where the contract is endogenously chosen. This yields fractions at the matching group level, which we compare by means of Wilcoxon signrank tests for matched pairs. Table 4 reports the outcomes of the performed tests, together with the average fractions observed (where averages are taken over the group fractions). Besides focusing on first round and last round outcomes only, we also calculate these fractions based on the observed outcomes in all rounds of the repeated games.

Table 2: Random effects probit estimations of player 2 choosing Honor

	(1)	(2)	(3)	(4)	(5)
	$l = 1$	$l = 3$	$l = 3$	$\delta = \frac{2}{3}$	$\delta = \frac{2}{3}$
		first round	last round	first round	last round
game B	0.147	0.077	-0.088	0.364***	0.283***
	(0.107)	(0.117)	(0.311)	(0.134)	(0.103)
endo	-0.095	0.836***	0.106	0.094	0.748***
	(0.170)	(0.190)	(0.679)	(0.221)	(0.176)
game B×endo	0.768***	1.092***	0.531	-0.092	-0.165
	(0.196)	(0.247)	(0.657)	(0.323)	(0.222)
match number	-0.004**	-0.086***	0.002	-0.002	-0.031***
	(0.002)	(0.007)	(0.017)	(0.007)	(0.006)
constant	-0.611	2.334***	-1.253***	1.940***	1.104***
	(0.387)	(0.209)	(0.344)	(0.288)	(0.168)
Log L	-666.754	-534.156	-89.804	-381.423	-626.719
N	1930	1282	209	1711	1293
n (# of clusters)	32	34	33	40	40
rho	0.787***	0.424***	0.585***	0.670***	0.377***
LR-chi2	38.165***	269.253***	2.471	8.887*	32.875***

Remark: Standard errors in parentheses. ***/**/* indicates significance at the 1/5/10% level. Rho gives the proportion of overall variance contributed by the panel-level component; its significance is based on a likelihood ratio test that rho=0. LR-chi2 reports the test statistic from testing that all coefficients (except the constant) are zero.

Table 3: Random effects probit estimations of player 1 choosing Trust

	(1)	(2)	(3)	(4)	(5)
	$l = 1$	$l = 3$	$l = 3$	$\delta = \frac{2}{3}$	$\delta = \frac{2}{3}$
		first round	last round	first round	last round
game B	0.075	0.162	0.126	0.095	0.051
	(0.064)	(0.141)	(0.116)	(0.221)	(0.082)
endo	0.366***	0.211	0.672***	-0.609*	-0.187
	(0.105)	(0.197)	(0.222)	(0.316)	(0.131)
game B×endo	0.921***	0.997***	0.601**	0.790**	0.755***
	(0.112)	(0.260)	(0.239)	(0.362)	(0.167)
match number	-0.008***	-0.060***	-0.062***	-0.021*	-0.014***
	(0.001)	(0.008)	(0.007)	(0.011)	(0.004)
constant	0.120	3.109***	-0.430**	3.594***	0.942***
	(0.303)	(0.350)	(0.173)	(0.406)	(0.106)
Log L	-1609.814	-423.412	-469.230	-206.384	-1000.359
N	4320	1530	1530	1800	1800
n (# of clusters)	32	34	34	40	40
rho	0.731***	0.707***	0.379***	0.685***	0.122***
LR-chi2	189.987***	175.571***	148.861***	54.260***	72.221***

Remark: Standard errors in parentheses. $^{***}/^{**}/^{*}$ indicates significance at the 1/5/10% level. Rho gives the proportion of overall variance contributed by the panel-level component; its significance is based on a likelihood ratio test that rho=0. LR-chi2 reports the test statistic from testing that all coefficients (except the constant) are zero.

Table 4: Fractions of cooperative outcome by game length and game

			Exo	Endo	Exo vs. Endo
$l = 1$	first round	Bad	0.199	0.261	0.600
	$(n = 6)$	Good	0.181	0.164	0.600
		B vs. G	0.527	0.249	
$l = 3$	first round	Bad	0.686	0.728	0.463
	$(n = 6)$	Good	0.683	0.330	0.028
		B vs. G	0.917	0.028	
$l = 3$	last round	Bad	0.038	0.041	0.833
	$(n = 6)$	Good	0.041	0.022	0.140
		B vs. G	0.833	0.829	
$l = 3$	all rounds	Bad	0.332	0.324	0.917
	$(n = 6)$	Good	0.321	0.132	0.028
		B vs. G	0.753	0.028	
$\delta = \frac{2}{3}$	first round	Bad	0.884	0.912	0.311
	$(n = 8)$	Good	0.853	0.741	0.050
		B vs. G	0.263	0.018	
$\delta = \frac{2}{3}$	last round	Bad	0.582	0.679	0.128
	$(n = 8)$	Good	0.540	0.397	0.012
		B vs. G	0.575	0.018	
$\delta = \frac{2}{3}$	all rounds	Bad	0.624	0.717	0.091
	$(n = 8)$	Good	0.578	0.434	0.012
		B vs. G	0.107	0.018	

Remark: The rows 'B vs. G' report the p-values of Wilcoxon signrank tests for matched pairs, comparing the B-game with the G-game. The column 'Exo vs. Endo' reports the p-values of Wilcoxon signrank tests for matched pairs, comparing the exogenous treatment with the endogenous treatment (for a given game). All tests are based on group level data, with n giving the number of groups.

When the formal contract is exogenously given, cooperation rates do not differ between the two contracts, independent of the length of the game. This follows because all "B vs. G" comparisons are insignificant. This does not apply when contracts are endogenously chosen though. Then cooperation is more likely under the bad contract than under the good contract. Differences are not significant in the one-shot games, but are significant for the first round of the repeated games. Here the bad contract thus stimulates subjects to start off more cooperatively. In the final round of the three-round games ($l = 3$), differences are insignificant and cooperation is hardly ever observed under both contracts. For the infinite games we still observe a significant difference between the bad contract and the good contract even in the final round. The cooperation rates calculated over all rounds in the repeated games are also significantly higher when the bad contract is endogenously chosen, both for $l = 3$ and for $\delta = \frac{2}{3}$.

The exogenous versus endogenous comparison in the final column of Table 4 yields a consistent pattern. Cooperation rates under the bad contract typically increase when this contract is endogenously chosen rather than exogenously given, although the differences observed are insignificant. In contrast, cooperation under the good contract declines when this contract is endogenously chosen. Here the differences are typically significant.

Overall we conclude from the results reported in Tables 2 through 4 that the bad contract indeed better facilitates cooperation, but only does so when this contract is endogenously chosen. In that case player 1 is significantly more likely to choose trust and player 2 is more likely to honor. This significantly increases the probability of a cooperative outcome. The flip side of this finding is that better explicit incentives may actually weaken implicit (relational) incentives.

Length of the game. Up till now we focused on how changes in the underlying stage game affect the amount of cooperation. We next turn to the impact of the length of the repeated game. When we compare the cooperation rates reported in Table 4 across different game lengths (for a given stage game), we observe that cooperation in the first round increases with the length of the game.[18] However, cooperation rates in the final round of the repeated games are much lower than those in the first round. In the $l = 3$ games, cooperation even unravels completely towards the end of the game. Here the cooperation fractions in the final round are also lower than those in the $l = 1$ treatments, although differences are typically insignificant according to a ranksum test at the 5%-level. In the infinite games cooperation decays over time as well, but it does so at a much slower pace. Cooperation rates in the final round are still significantly higher than those observed in the one-shot games.[19] Comparing cooperation fractions calculated over all rounds (which for $l = 1$ just equal those for the first round), differences between the $l = 1$ and $l = 3$ treatments are always insignificant whereas they are always significantly higher in the infinite games (ranksum tests, 5% level). In line with (standard) theoretical predictions, therefore, the 'shadow of the future' in the infinite games significantly increases cooperation (cf. Dal-Bo 2005).

The observation from Table 4 that only in the repeated games the (endogenously chosen) bad contract leads to significantly more cooperation than the good contract does, suggests that repetition and reciprocity reinforce each other in providing implicit incentives. Indeed, the difference in average overall

[18] When we compare (first round) cooperation fractions accross different game lengths by means of Mann-Whitney ranksum tests, five out of six comparisons are significant at the 5%-level in the exogenous treatments (the single exception being the difference between $l = 3$ and $\delta = \frac{2}{3}$ when the good contract applies). In the endogenous treatments four out of six differences appear significant at the 5%-level.

[19] Using ranksum tests, all four comparisons between $l = 1$ and the last round of $\delta = \frac{2}{3}$ yield significant differences at the 5% level.

cooperation rates between the two contracts increases with the length of the game: from 0.097, to 0.192 to 0.283.[20] Although the differences in these differences are substantial, formal ranksum tests (again performed at the matching group level) reveal that they are insignificant.[21] A plausible explanation here is the relatively few group observations that we have. We thus only obtain weak evidence that reciprocity and repetition complement each other.

4.3 Choice of contract

Theory predicts that the bad contract will be chosen more often in the infinite games than in the finite games. The aggregate overview provided in Figure 3 already indicates that this prediction is not borne out in the data. On average the bad contract is chosen in about 40% of the cases when a one-shot game is played and in around 25% of the cases when a repeated game is played. A more formal analysis is presented in Table 5, which reports random effects probit estimates of the probability that player 2 opts for the bad contract. Apart from two treatment dummies for the two versions of the repeated games ('three' for the three-round repeated games and 'infinite' for the infinite games), we also include a 0/1 dummy that indicates whether the games are presented in the order 'first B, then G' or not. Learning effects are accounted for by incorporating the number of game choices player 2 has made so far. This yields the estimates in the first column of Table 5.

In line with the percentages reported above the repeated games lower the probability that the bad contract is chosen, although the two dummies for the repeated games appear insignificant at conventional levels. The remaining order dummy is highly significant, indicating that when subjects start off with the exogenously given bad contract, they are more likely to choose this contract later on. The bad contract is also more likely to be chosen, the more experience players 2 have with making the contract choice.

Because significant order effects are found, the second column verifies whether these interact with the length of the game. This appears not to be the case; the two interaction terms are insignificant. The significant order effect most likely reflects the different experiences subjects have with the two contracts in isolation. Tables 2 and 3 have namely shown that cooperative behavior decays over time. Subjects therefore probably have relatively better experiences with the first contract they are confronted with. To investigate this, the third column adds the payoff difference player 2 experienced between the two contracts in the exogenous treatments. In particular, $\Pi_B - \Pi_G$ reflects the difference between player 2's overall payoffs under the exogenously given bad contract and his overall payoffs under the exogenously given good contract. By including this payoff difference the order dummy becomes insignificant. This is not surprising, given that the two are highly correlated ($\rho = 0.615$). The final column only includes the payoff difference and leaves the order dummy out. This reveals that the better relative experience player 2 has with the bad contract, the more easily he is induced to choose this contract in the final part of the experiment. Note that by including $\Pi_B - \Pi_G$ the infinite game dummy becomes significant.

Contract choices and profits. In order to better understand actual contract choices, we ran (random effects) regressions of the average profit player 2 earns, using the same set of regressors as in Tables 2

[20] These numbers follow from the cooperation rates calculated over all rounds, focusing on the case where the contract is endogenously chosen (cf. column 'Endo' in Table 4). For $l = 1$ we have $0.261 - 0.164 = 0.097$, for $l = 3$ the difference equals $0.324 - 0.132 = 0.192$ and for $\delta = \frac{2}{3}$ we obtain $0.717 - 0.434 = 0.283$.

[21] For the $l = 1$ versus $l = 3$ comparison we obtain $p = 0.200$, for $l = 1$ vs. $\delta = \frac{2}{3}$ we get $p = 0.116$ and for $l = 3$ vs. $\delta = \frac{2}{3}$ we have $p = 0.153$.

Table 5: Random effects probit estimations of player 2 choosing the bad contract (game B)

	(1)	(2)	(3)	(4)
constant	-1.475***	-1.492***	-1.011*	-0.517
	(0.307)	(0.365)	(0.539)	(0.371)
three	-0.520	0.015	-0.557	-0.567
	(0.444)	(1.462)	(0.444)	(0.451)
infinite	-0.691	-1.210	-0.840*	-0.948**
	(0.428)	(1.436)	(0.454)	(0.456)
BG-order	1.002***	1.018*	0.619	
	(0.362)	(0.589)	(0.506)	
game choice #	0.008**	0.008**	0.008**	0.008**
	(0.004)	(0.004)	(0.004)	(0.004)
order×three		-0.354		
		(0.897)		
order×infinite		0.351		
		(0.868)		
$\Pi_B - \Pi_G$			0.004	0.007***
			(0.004)	(0.003)
Log L	-837.121	-836.840	-836.503	-837.248
N	2550	2550	2550	2550
n (# of clusters)	106	106	106	106
rho	0.764***	0.763***	0.761***	0.763***
LR-chi2	14.425***	14.987**	15.660***	14.170***

Remark: Standard errors in parentheses. ***/**/* indicates significance at the 1/5/10% level. Rho gives the proportion of overall variance contributed by the panel-level component; its significance is based on a likelihood ratio test that rho=0. LR-chi2 reports the test statistic from testing that all coefficients (except the constant) are zero.

and 3. For briefness, here only the main findings are discussed (Appendix C reports the actual estimates obtained). For the case where the contract is exogenously given, player 2 earns significantly less under the bad contract, irrespective of the length of the game. This does not apply when the contract is endogenously chosen. Then in the one-shot games ($l = 1$) player 2 actually earns significantly more when he opts for the bad contract, while for the repeated games no profit differences between the two types of contracts are found.

These findings provide a partial explanation for actual contract choices. In the exogenous treatments the good contract typically yields player 2 more than the bad contract does. Based on these experiences, player 2 may be inclined to choose the good contract in the final part. If he chooses the bad contract anyway, he actually does significantly better in the one shot games but not in the repeated games. This may explain why player 2 is more likely to choose the bad contract under spot interaction than under repeated interaction.

Betrayal rates. A final concern is why choosing the bad contract is profitable for player 2 in the one-shot games but not in the repeated games. This is somewhat puzzling, given that the differences in cooperation rates between the two contracts are not significant when $l = 1$, but are significant for $l = 3$ and $\delta = \frac{2}{3}$ (cf. Table 4). Cooperation is not the only way to make money for player 2, however. Within a round he earns the most when the (Trust, Betray) outcome pertains. In the one shot games the bad contract may thus be attractive, because it increases the frequency that this betrayal outcome is observed. The precise fractions of the (Trust, Betray) outcome in the various treatments are reported in Appendix C. Here we again focus on the main findings.

Like with cooperation rates, the betrayal fractions do not differ between the two contracts when the contract is exogenously given. But when the contract is endogenously chosen and the game is one-shot, the betrayal outcome occurs more often under the bad contract than under the good contract. For the finitely repeated games there is by and large no difference while for the infinite games (Trust, Betray) occurs significantly more often under the good contract. This makes the bad contract relatively more attractive under spot interaction and relatively less attractive under repeated interaction.

A tentative explanation for these findings runs as follows. Choosing the bad contract increases the probability that player 1 chooses Trust (cf. Table 3), probably because this is interpreted as a kind choice that signals player 2's trustworthiness. Once player 1 chooses to trust, however, player 2 can in the one shot games costlessly betray without affecting future outcomes. This is not the case in the repeated games. After betrayal player 1 typically does not trust anymore and player 2 is stuck with a low payoff in the remainder of the repeated game. Future trust thus requires player 2 to honor. Roughly put, future trust not only requires choosing the bad contract, but also to honor in the current stage game. In the one shot games this is not needed and choosing the bad contract in the next period (where player 2 is matched to a different player 1) suffices.

5 Conclusion

A large theoretical literature exists that studies the interplay between formal and relational contracts in providing incentives. An important and intriguing insight that can be obtained from these studies is that formal contracts that are suboptimal in themselves – because better contracts exist when formal incentives are studied in isolation – may actually be conducive for sustaining relational incentives. Parties may therefore deliberately opt for inferior (more 'incomplete') formal contracts in order to support better implicit contracts.

In this paper we report the results from a laboratory experiment that tests whether the availability of better explicit incentives indeed makes self-enforcing relational contracts less likely. A distinguishing mark of our experiment is that we consider infinitely repeated trust games in some of our treatments and let subjects choose between two different payoff variations of the constituent trust game. These payoff variations represent in reduced form two different types of formal contract, viz. a bad explicit contract and a good explicit contract. Besides infinitely repeated games, we also look at repeated games that lasts for three rounds and at one-shot games.

In line with theoretical predictions, our results indicate that cooperation on the (Trust, Honor) outcome is more likely when the bad contract applies. However, this appears only to be the case when the bad contract is endogenously chosen by the subjects themselves. In the treatments where the contracts are exogenously given we do not observe significant differences between the two contracts. By comparing the different game lengths we also find some evidence that repetition and reciprocity complement each other in providing informal relational incentives.

The finding that choosing the bad contract facilitates cooperation rationalizes that subjects choose this contract in a substantial fraction of cases. However, subjects are more likely to choose the bad contract in the one-shot games than in the repeated games, although especially in the latter the bad contract is helpful in boosting cooperative behavior. A tentative explanation here is that in our experiment the trustee chooses the contract. Opting for the bad contract then increases trust in the current round (i.e. stage game). In the one shot games this is all that is needed, because in the next round the trustee is matched to a different trustor. In the repeated games, however, the trustee should honor current round trust as well in order to stimulate future trust. This makes the bad contract relatively more attractive under spot interaction.

Appendix A: A simple model

In this section we present a simple model in the spirit of Baker, Gibbons, and Murphy (1994) and Schmidt and Schnitzer (1995) to justify the reduced form game setup of Figure 1.

Consider a principal-agent relationship between a firm and a worker. Suppose the worker's contribution to firm value equals $\pi(e) = e$, with e the effort level chosen by the worker. The disultility of effort the worker bears is $C(e)$, with $C(e)$ increasing and convex. If the worker does not work for the firm, both parties get their outside options payoffs; w_a for the worker and π_a for the firm. The overall net surplus created by the worker's effort thus equals $S(e) \equiv e - C(e) - w_a - \pi_a$. This surplus is maximized by effort level e_1 that satisfies $C'(e_1) = 1$. We assume that there are gains from trade, i.e. $S(e_1) > 0$. This implies that e_1 reflects the efficient effort level. Unfortunately, however, due to contractual incompleteness only effort levels up to e_0 (with $e_0 < e_1$) can be contracted upon.

In a one-shot relationship the predicted outcome depends on the sign of $S(e_0)$. If $S(e_0) > 0$ the parties are predicted to sign a contract specifying that the worker receives some wage w_0 for supplying the maximum enforceable effort level e_0. The value of w_0 depends on the bargaining power of the two parties involved. In case $S(e_0) < 0$ trade is expected not to occur.

When the game is infinitely repeated, multiple equilibria exist. We are particularly interested in the combination of an explicit contract that enforces $e_{com} \leq e_0$ (for a contract wage w_{com}) and an implicit contract which asks the worker to supplement effort up to the efficient level (i.e. put in additional effort of $e_1 - e_{com}$) in return for some non-contractible bonus β. We focus on trigger-like strategies, i.e. the parties stick to the implicit agreement if they both always did so in the past. But as soon as one of the parties deviates, from that point onwards the one-shot equilibrium is played forever after. From Proposition 4 in Schmidt and Schnitzer (1995) it then follows that such a combined contract can be made self-enforcing if and only if:

$$C(e_1) - C(e_{com}) \leq \frac{\delta}{1-\delta} \cdot [S(e_1) - \max\{0, S(e_0)\}] \tag{A1}$$

Here δ (with $0 \leq \delta < 1$) denotes the common discount factor. The intuitive idea is that, once one of the parties reneges on the implicit contract, the equilibrium of the one-shot game is played. That is, the parties renegotiate their original contract specifying (e_{com}, w_{com}) either into one that stipulates (e_0, w_0) (when $S(e_0) > 0$), or into no trade at all (in case $S(e_0) < 0$). This fall back equilibrium yields a joint surplus of $\max\{0, S(e_0)\}$ per period. The loss in joint surplus thus equals $S(e_1) - \max\{0, S(e_0)\}$ per period and the r.h.s. in the above inequality just gives the net present value of all these future losses. The l.h.s. represents the worker's short term gain of deviating from the implicit contract. Note that this gain is lower the higher e_{com} is. Hence, the best combined contract to choose in order to facilitate cooperation specifies $e_{com} = e_0$; this gives the weakest incentive to deviate and thus the weakest constraint on δ.

For $e_{com} = e_0$ inequality $(A1)$ can be rewritten as:

$$\delta \geq \frac{C(e_1) - C(e_0)}{(C(e_1) - C(e_0)) + [S(e_1) - \max\{0, S(e_0)\}]} \equiv \underline{\delta} \tag{A2}$$

In case $\delta \geq \underline{\delta}$ a combined explicit-implicit contract $(e_0, e_1, w_{com}, \beta)$ exists that supports the efficient effort level e_1 as equilibrium outcome. Lower bound $\underline{\delta}$ increases with e_0 when $S(e_0) > 0$ and decreases with e_0 if $S(e_0) < 0$. We next consider the latter two cases separately, to illustrate that the reduced form trust game can capture the essence of the above model.

The case $S(e_0) > 0$. Consider the combined contract $(e_0, e_1, w_{com}, \beta)$. The worker has an incentive to stick to this contract (by choosing e_1 rather than e_0) whenever:

$$\frac{1}{1-\delta} \cdot [w_{com} + \beta - C(e_1)] \geq w_{com} - C(e_0) + \frac{\delta}{1-\delta} \cdot (w_0 - C(e_0))$$

The final term on the r.h.s. follows because once the worker has deviated, the one-shot equilibrium contract (e_0, w_0) arises (which is based on explicit incentives only). Similarly so, the firm has an incentive to pay the promised bonus β after e_1 whenever:

$$\frac{1}{1-\delta} \cdot [e_1 - w_{com} - \beta] \geq e_1 - w_{com} + \frac{\delta}{1-\delta} \cdot (e_0 - w_0)$$

Rewriting these two inequalities, the combined contract is self-enforcing iff:

$$\underline{\beta} \equiv C(e_1) - C(e_0) + \delta\,(w_0 - w_{com}) \leq \beta \leq \delta\,(e_1 - e_0) + \delta\,(w_0 - w_{com}) \equiv \overline{\beta}$$

Note that term $\delta\,(w_0 - w_{com})$ appears in both the lower bound $\underline{\beta}$ and the upper bound $\overline{\beta}$. Hence the combined contract wage w_{com} is immaterial as instrument for self-enforcement and thus can be independently used for distributive purposes (cf. Proposition 1(b) in Schmidt and Schnitzer (1995)). Also note that lower bound $\underline{\beta}$ is decreasing in e_0. Intuitively, the better explicit contracts are, the lower the bonus needed to induce the worker to supply the extra effort $e_1 - e_0$.

Any combined contract with $\beta' < \underline{\beta}$ is not self-enforcing and outcome-equivalent to contracts specifying $\beta'' > \overline{\beta}$. We therefore focus on contracts with $\beta \geq \underline{\beta}$ only. For these contracts the worker never has an incentive to renege given that the firm sticks to the combined contract. Effectively, accepting the combined contract $(e_0, e_1, w_{com}, \beta)$ and then deviating to e_0 is not a relevant option for the worker. Only two relevant choices remain: (i) accepting and working according to the formal contract (e_0, w_0) based on explicit incentives only, or (ii) accepting and working according to the combined contract $(e_0, e_1, w_{com}, \beta)$. In turn, the relevant choice for the firm is whether to pay the non-contractible bonus β when the combined contract applies. Figure A1 depicts the corresponding game tree.

[Figure A1]

In this game the firm has an incentive to keep its promise to pay bonus β if:[22]

$$\delta \geq \frac{\beta}{(e_1 - w_{com}) - (e_0 - w_0)}$$

[22] For the worker to prefer the combined contract it is required that $w_{com} - C(e_1) + \beta \geq w_0 - C(e_0)$, i.e. $\beta \geq C(e_1) - C(e_0) + (w_0 - w_{com})$. Given $\beta \geq \underline{\beta}$ this condition is certainly satisfied when $w_0 \leq w_{com}$. Because w_{com} does not affect self-enforcement but distribution only, we assume that $w_0 \leq w_{com}$ and that remaining distributional issues are solved by lump sum payments at the contracting stage.

For a fixed bonus level β, the r.h.s. is increasing in e_0.[23] This upward pressure results from the fact that the worst possible punishment in case of deviation is less severe when better explicit contracts are available. However, also the minimum required bonus $\underline{\beta}$ decreases with e_0, so the overall impact depends on which effect dominates. Plugging in $\beta = \underline{\beta}$ and rewriting yields the requirement:

$$\delta \geq \frac{C(e_1) - C(e_0)}{(e_1 - e_0)} \equiv \underline{\delta}$$

This corresponds with $(A2)$ for $S(e_0) > 0$. Given that $C(e)$ is convex it follows that $\frac{\partial \underline{\delta}}{\partial e_0} \geq 0$. Hence, better explicit incentives lead to worse implicit incentives in this case, because the improvement in the fall back after reneging outweighs the decrease in the minimally required bonus $\underline{\beta}$.

The case $S(e_0) < 0$. The analysis here is similar to the one above. The no-reneging conditions for the worker and the firm now equal respectively:

$$\frac{1}{1-\delta} \cdot [w_{com} + \beta - C(e_1)] \geq w_{com} - C(e_0) + \frac{\delta}{1-\delta} \cdot w_a$$

$$\frac{1}{1-\delta} \cdot [e_1 - w_{com} - \beta] \geq e_1 - w_{com} + \frac{\delta}{1-\delta} \cdot \pi_a$$

Therefore, the combined contract $(e_0, e_1, w_{com}, \beta)$ is self-enforcing iff:

$$\underline{\beta} \equiv C(e_1) - C(e_0) + \delta\left(C(e_0) + w_a - w_{com}\right) \leq \beta \leq \delta(e_1 - w_{com} - \pi_a) \equiv \overline{\beta}$$

Focusing on contracts that satisfy $\beta \geq \underline{\beta}$, the worker effectively chooses between (i) not working for the firm, or (ii) accepting and working according to the combined contract. Figure A2 gives the appropriate game tree for this case.

[Figure A2]

The condition for the firm to pay the bonus is now given by:[24]

$$\delta \geq \frac{\beta}{(e_1 - w_{com} - \pi_a)}$$

Because e_0 does not affect the fall back outcome, there is no upward pressure on the requirement on δ in this case. Only the effect of a decrease in the minimum bonus $\underline{\beta}$ remains. For $\beta = \underline{\beta}$ the above requirement reduces to:

$$\delta \geq \frac{C(e_1) - C(e_0)}{(e_1 - C(e_0)) - (\pi_a + w_a)} \equiv \underline{\delta}$$

This corresponds with $(A2)$ for $S(e_0) < 0$. From $C(e)$ convex it follows that $\frac{\partial \underline{\delta}}{\partial e_0} \leq 0$. In this case better explicit incentives improve implicit incentives and thus are complements, the intuition being that the gain from deviation is smaller when the formal contract fixes a higher value of e_0.

[23] When e_0 increases, the corresponding wage $w_0(e_0)$ will increase as well. Assuming that the firm can capture at least some part of the additional surplus, $e_0 - w_0(e_0)$ will be increasing in e_0.

[24] The worker prefers the combined contract if $w_{com} - C(e_1) + \beta \geq w_a$, i.e. $\beta \geq C(e_1) + (w_a - w_{com})$. Given $\beta \geq \underline{\beta}$ this condition is certainly satisfied when $w_a + C(e_0) \leq w_{com}$. We assume that w_{com} is chosen such that it satisfies this constraint.

Appendix B: Formal analysis under reciprocity motivations

In this appendix we formally analyze the one-shot trust game depicted in Figure 2, assuming that players are guided by intention-based reciprocity like in Dufwenberg and Kirchsteiger (2004). That is, player j's utility is given by:

$$U_j = \pi_j + Y_j \cdot \kappa_{jk} \cdot \lambda_{jkj}$$

Here π_j gives player j's monetary payoffs and parameter $Y_j \geq 0$ her reciprocal attitude. Term κ_{jk} represents j's kindness to k; it is positive when j is kind to k and negative if j is unkind to k. Factor λ_{jkj} gives j's belief about how kind player k is to her, i.e. the kindness of player k as perceived by j.

Let h denotes the probability that player 2 chooses Honor. The following lemma provides the equilibrium value of h.

Lemma 1. In equilibrium player 2 chooses Honor with probability:[25]

$$h^e = \left\lfloor \frac{f-d}{f-e} - \frac{2}{Y_2 \cdot (c-a)} \right\rceil_0^1 \equiv \psi(d)$$

Proof. Suppose player 1 chooses Trust. Her equitable payoff then equals $\frac{a+c}{2}$. Player's 2 kindness equals what player 1 gets compared to this payoff, i.e. $\kappa_{21} = h \cdot c + (1-h) \cdot a - \frac{a+c}{2} = \left(h - \frac{1}{2}\right) \cdot (c-a)$. Similarly so, $\lambda_{212} = h'' \cdot e + (1-h'') \cdot f - \frac{1}{2}\left[h'' \cdot e + (1-h'') \cdot f + d\right]$. Here h'' denotes player 2's (second order) beliefs about what player 1 beliefs about h. Rewriting this we get $\lambda_{212} = \frac{1}{2}\left[(f-d) - (f-e) \cdot h''\right]$. Overall utility for player 2 thus becomes:

$$U_2 = f - h \cdot (f-e) + Y_2 \cdot \left(\left(h - \frac{1}{2}\right) \cdot (c-a)\right) \cdot \frac{1}{2}\left[(f-d) - (f-e) \cdot h''\right]$$

From $\frac{\partial U_2}{\partial h} = 0$ we obtain $-(f-e) + Y_2 \cdot (c-a) \cdot \frac{1}{2}\left[(f-d) - (f-e) \cdot h''\right] = 0$. In equilibrium necessarily $h'' = h^e$, yielding the result. ■

Note that when $Y_2 = 0$ we have $h^e = 0$ necessarily. A selfish player 2 will never honor trust. The next lemma gives the equilibrium probability q^e with which player 1 chooses Trust.

Lemma 2. Player 1's equilibrium behavior can be characterized as follows:[26]

$$\begin{aligned}
(i) \quad &: \quad h^e < \min\left\{\frac{b-a}{c-a}, \frac{1}{2}\right\} : q^e = 0 \\
(iia) \quad &: \quad \frac{1}{2} < h^e < \frac{b-a}{c-a} : \ q^e = 0 \\
(iib) \quad &: \quad \frac{b-a}{c-a} < h^e < \frac{1}{2} : q^e = \left\lceil \frac{(c-a) \cdot h^e - (b-a)}{Y_1 \cdot ((f-d) - (f-e) \cdot h^e) \cdot \left[\left(\frac{1}{2} - h^e\right) \cdot (c-a)\right]} \right\rceil^1 \\
(iii) \quad &: \quad h^e > \max\left\{\frac{b-a}{c-a}, \frac{1}{2}\right\} : q^e = 1
\end{aligned}$$

[25] This notation means that $\lfloor x \rceil_0^1$ equals 0 for $x < 0$, x for $0 \leq x \leq 1$, and 1 for $x > 1$.

[26] In case (iia) actually multiple equilibria may exist side by side. We report the single equilibrium that always exists when h^e is within the given bounds for this case.

Proof. We first show that player 1's equilibrium behavior is determined by the following derivative (with q the probability of choosing Trust):

$$\begin{aligned}\frac{\partial U_1}{\partial q} &= -(b-a)+(c-a)\cdot h'+ \\ &\quad Y_1\cdot((f-d)-(f-e)\cdot h')\cdot q''\left[\left(h'-\frac{1}{2}\right)\cdot(c-a)\right]\end{aligned}$$

Player 1 believes that player 2 chooses Honor with probability h'. The equitable payoff for player 2 thus equals $\pi_2^e = \frac{1}{2}\left[d+f\cdot(1-h')+e\cdot h'\right]$. Therefore, $\kappa_{12} = (1-q)\cdot d+q\cdot[f\cdot(1-h')+e\cdot h']-\pi_2^e$. This reduces to $\kappa_{12} = \left(q-\frac{1}{2}\right)\cdot[(f-d)-(f-e)\cdot h']$. Likewise, $\lambda_{121} = (1-q'')\cdot b+q''\cdot(a\cdot(1-h')+c\cdot h')-\frac{1}{2}\left[2\left(1-q''\right)\cdot b+q''\cdot(a+c)\right] = q''\cdot\left(h'-\frac{1}{2}\right)\cdot(c-a)$. By $U_1 = (1-q)\cdot b+q\cdot[(1-h')\cdot a+h'\cdot c]+Y_1\cdot\kappa_{12}\cdot\lambda_{121}$ and taking the derivative yields the expression for $\frac{\partial U_1}{\partial q}$.

Now, if $h' < \min\left\{\frac{b-a}{c-a},\frac{1}{2}\right\}$, it holds that $\frac{\partial U_1}{\partial q} < 0$ necessarily. Therefore $q^e = 0$. Similarly, for $h' > \max\left\{\frac{b-a}{c-a},\frac{1}{2}\right\}$, we have $\frac{\partial U_1}{\partial q} > 0$ and thus $q^e = 1$. In the in between case $\min\left\{\frac{b-a}{c-a},\frac{1}{2}\right\} < h' < \max\left\{\frac{b-a}{c-a},\frac{1}{2}\right\}$ the two types of payoffs give opposing incentives. First assume $\frac{b-a}{c-a} < h^e < \frac{1}{2}$. In this case player 1 prefers $q = 1$ on the basis of monetary payoffs only and $q = 0$ on the basis of his reciprocity payoffs. Balancing these two forces the equilibrium value q^e results; q^e is just the value of q'' that solves $\frac{\partial U_1}{\partial q} = 0$. Note that here q^e is weakly increasing in h^e. Second, when $\frac{1}{2} < h^e < \frac{b-a}{c-a}$ player 1 prefers $q = 0$ on the basis of monetary payoffs. Here $q^e = 0$ is always an equilibrium, because for $q'' = 0$ the second term in $\frac{\partial U_1}{\partial q}$ vanishes and thus $\frac{\partial U_1}{\partial q} < 0$ (given $h^e < \frac{b-a}{c-a}$). When $Y_1\cdot((f-d)-(f-e)\cdot h^e)\left[\left(h^e-\frac{1}{2}\right)\cdot(c-a)\right] > (b-a)-(c-a)\cdot h^e$ also $q^e = 1$ and $q^e = \frac{(b-a)-(c-a)\cdot h^e}{Y_1\cdot((f-d)-(f-e)\cdot h^e)\cdot\left[\left(h^e-\frac{1}{2}\right)\cdot(c-a)\right]}$ exist at the same time. ■

Lemma 2 can be understood as follows. Player 1 is guided by both monetary and reciprocity payoffs. The relevant cutoff for h^e is $\frac{b-a}{c-a}$ when monetary payoffs are considered in isolation. If h^e is larger than this cutoff, player 1 prefers Trust on the basis of monetary payoffs. Otherwise he chooses Not trust. Likewise, in regard to reciprocity payoffs the relevant cutoff for h^e is $\frac{1}{2}$.

Now, in case (i) both monetary payoffs and reciprocity payoffs induce player 1 to choose Not trust. Similarly so, if h^e exceeds the given upper bound of case (iii), player 1 prefers Trust on the basis of both types of payoffs. If h^e falls in between, however, monetary payoffs and reciprocity payoffs are in conflict. In case (iia) actually multiple equilibria may exist side by side for some subset of parameters. The lemma reports the single equilibrium that always exists in this case. In case (iib) a unique equilibrium results. In this equilibrium player 1 is more likely to Trust (i.e. a higher q^e) the more probable it is that player 2 chooses Honor (i.e. the higher is h^e).

Note that when $Y_2 = 0$, necessarily $h^e = q^e = 0$. For (Trust,Honor) to be an equilibrium outcome it is thus necessary that player 2 has sufficiently strong reciprocal motivations. $Y_1 > 0$ is not needed for this. Taking Lemma 1 and Lemma 2 together we obtain the following proposition, which justifies expression (3) in the main text.

Proposition 1. Both the probability that player 1 chooses Trust (q^e) and the probability that player 2 honors trust (h^e) are weakly decreasing in both d and $(f-e)$. When player 1 is selfish ($Y_1 = 0$), he chooses Trust for sure iff $Y_2 \geq \frac{2\cdot\underline{\delta}}{(c-a)-\underline{\delta}\cdot(b-a)} \equiv \underline{Y}$.

Next we analyze an extended game where, in a first stage, one of the players chooses between a "bad" and a "good" explicit contract. In the former case the payoffs after Not trust equal (b_b, d_b) whereas in the latter case these are (b_g, d_g), with $b_b \leq b_g$ and $d_b < d_g$. After the contract has been chosen, the game continues with first player 1's Not trust/Trust choice and subsequently player 2's Betray/Honor choice. We first explore whether the initial contract choice affects player 2's propensity to honor trust. Lemma 3 below shows that this is the case only when player 2 chooses the contract.

Lemma 3. (i) Suppose player 1 makes the contract choice $j \in \{b, g\}$. Then player 2 chooses Honor after Trust with probability $\psi(d_g)$,[27] *independent* of the actual contract chosen. (ii) If player 2 chooses contract $j \in \{b, g\}$, then he chooses Honor with probability $\psi(d_j)$. Hence the probability of honor is weakly higher when the bad contract is chosen; $h_g^e = \psi(d_g) \leq \psi(d_b) = h_b^e$.

Proof. (i) In this case player 1 makes two choices in a row. Player 2's reaction to Trust may in principle depend on which contract has been chosen. Let h_b (h_g) denote the probability that player 2 chooses Honor after player 1's combined choice for the bad contract (good contract) and Trust. We show that player 2's equilibrium reaction always follows from Lemma 1 with $d = d_g$, i.e. $h_b^e = h_g^e = \psi(d_g)$.

Suppose player 1 chooses contract $j \in \{b, g\}$ and subsequently chooses Trust. The minimum player 2 can then give to player 1 is a whereas the maximum is c. Hence the equitable payoff for 1 equals $\pi_1^e = \frac{a+c}{2}$. We obtain $\kappa_{21}^j = \left(h_j - \frac{1}{2}\right) \cdot (c - a)$ for contract $j \in \{b, g\}$ chosen. With respect to perceived kindness, the equitable payoff for player 2 after history $(j, Trust)$ equals $\pi_2^e = \frac{1}{2}[d_g + \max\{h_b'' \cdot e + (1 - h_b'') \cdot f, h_g'' \cdot e + (1 - h_g'') \cdot f\}]$. Note that here always d_g enters as the relevant minimum payoff, independent of the actual contract j chosen. This holds because player's 1 choice for $(j = b, Not\ trust)$ is Pareto-dominated by $(j = g, Not\ trust)$ and therefore has no impact on the equitable payoff of player 2 (cf. Dufwenberg and Kirchsteiger (2004, p. 276)). After history $(j, Trust)$ player 2's perception about the kindness of player 1 thus equals $\lambda_{212}^j = h_j'' \cdot e + (1 - h_j'') \cdot f - \pi_2^e$. Suppose $h_b'' > h_g''$. Then $\lambda_{212}^b < \lambda_{212}^g$ and hence from the proof of Lemma 1:

$$\frac{\partial U_2}{\partial h_b} = -(f - e) + Y_2 \cdot (c - a) \cdot \lambda_{212}^b < -(f - e) + Y_2 \cdot (c - a) \cdot \lambda_{212}^g = \frac{\partial U_2}{\partial h_g}$$

But this implies $h_b \leq h_g$, contradicting $h_b'' > h_g''$ under correct beliefs. A similar contradiction follows from assuming $h_b'' < h_g''$. Hence the equilibrium probability of choosing Honor is independent of contract choice. This gives $\lambda_{212}^b = \lambda_{212}^g = \frac{1}{2}[(f - d_g) - (f - e) \cdot h'']$ and the equilibrium value for h follows from Lemma 1 with $d = d_g$; $h_b^e = h_g^e = \psi(d_g)$.

(ii) Suppose player 2 chooses contract $j \in \{b, g\}$ and player 1 chooses Trust. From part (i) we again have $\kappa_{21}^j = \left(h_j - \frac{1}{2}\right) \cdot (c - a)$ for contract $j \in \{b, g\}$ chosen. But the equitable payoff for player 2 after history $(j, Trust)$ now equals $\pi_{2j}^e = \frac{1}{2}\left[\left(h_j'' \cdot e + (1 - h_j'') \cdot f\right) + d_j\right]$ for $j \in \{b, g\}$. This follows because, in calculating the perceived kindness of player 1 towards 2 at node $(j, Trust)$, the contract as chosen by 2 should be taken as given. Hence $\lambda_{212}^j = \frac{1}{2}\left[(f - d_j) - (f - e) \cdot h_j''\right]$ at this node. The equilibrium value for h_j then directly follows from Lemma 1; $h_j^e = \psi(d_j)$. ■

Lemma 3 reveals that when player 1 chooses the contract, the existence of the bad contract does not impact the equilibrium probability of honoring trust, even not when the bad contract is actually chosen.

[27] The function $\psi(\cdot)$ has been defined in Lemma 1.

The intuitive idea is that choosing the bad contract followed by Not trust, is a Pareto-dominated choice for player 1. Payoffs (b_b, d_b) therefore should be discarded when determining the perceived kindness of player 1 choosing Trust (cf. Dufwenberg and Kirchsteiger (2004, p. 276)). This does not apply when player 2 makes the contract choice. In that case payoffs (b_b, d_b) become relevant for determining the (perceived) kindness of player 1's choice to trust. Trust is considered more kind when the bad contract applies, because the relative gain for player 2 is then larger. Player 2 is therefore more likely to honor trust. Overall, the key feature that implementing a bad formal contract may favorably affect player 2's propensity to honor trust thus only applies when player 2 chooses the contract.

A complete equilibrium analysis of the extended game with contract choice is difficult, because player 1's choice between Trust and Not trust will not only be guided by player 2's anticipated reaction, but also by reciprocal motivations towards either 2's actual contract choice (when 2 has chosen the contract), or what 2 would have done would player 1 have chosen the other contract. This makes calculating the equilibrium probabilities q_j^e (for $j \in \{b, g\}$) of player 1 choosing Trust much more involved than in Lemma 2. But, as indicated by Lemma 3, the difference in induced honor behavior of player 2 is key. For simplicity we therefore focus on the case in which player 1 is not motivated by reciprocity at all (i.e. $Y_1 = 0$). Our final lemma then indicates that (only) when player 2 chooses the contract, he can increase the probability of a cooperative outcome by implementing the bad contract.

Lemma 4. Suppose $Y_1 = 0$. (i) If player 1 makes the contract choice, then the outcome corresponds with the case where the bad contract is absent. (ii) If player 2 makes the contract choice, then the probability of player 1 choosing Trust is weakly higher when the bad contract is chosen.

Proof. (i) Note that in this case $h_b^e = h_g^e = \psi(d_g)$. Choosing the good contract thus always yields player 1 weakly more in monetary terms. When $\psi(d_g) > \frac{b_g - a}{c-a}$ player 1 chooses Trust and her contract choice is irrelevant for the outcome, in case $\psi(d_g) < \frac{b_g - a}{c-a}$ player 1 chooses ($j = g$, *Not trust*). This corresponds with the outcome where the bad contract is absent.

(ii) Given contract $j \in \{b, g\}$, player 1 chooses Trust whenever $h_j = \psi(d_j) > \frac{b_j - a}{c-a}$. Given $h_b \geq h_g$ and $b_b \leq b_g$, this inequality is more easily satisfied for $j = b$. ■

Appendix C: Profit regressions and betrayal rates

In this appendix we report the details of the profit regressions and the across treatment comparisons of betrayal rates referred to in Section 4.3.

Table C-1 reports random effects regression estimates of the average profit player 2 earns. Besides two treatment dummies indicating the type of contract that applies and whether this contract is endogenously chosen (together with their interaction), a time trend is incorporated that counts the number of matches player 2 has been involved in so far. For the repeated games average profits are simply calculated over the multiple rounds within a period. For the one-shot game we estimate two specifications. The first one is based on the (single-round) period payoffs only, see the first column in Table **??**. In a second specification we group three consecutive periods into a single "pseudo-period" and calculate averages over the three periods. In that way average payoffs are calculated over three stage games, just like (on average) in the repeated games.

Table C-1: Random effects regressions of player's 2 average profits

	(1)	(2)	(3)	(4)
	$l = 1$	$l = 1$	$l = 3$	$\delta = \frac{2}{3}$
	1 period	3 periods		
game B	-1.377***	-1.377***	-1.233***	-0.553***
	(0.249)	(0.241)	(0.196)	(0.174)
endo	1.770***	1.901***	0.899***	-0.803***
	(0.423)	(0.412)	(0.322)	(0.289)
game B×endo	2.494***	2.083***	0.990**	0.640*
	(0.477)	(0.461)	(0.398)	(0.359)
match number	-0.030***	-0.090***	-0.106***	0.003
	(0.005)	(0.013)	(0.011)	(0.009)
constant	13.258***	13.273***	13.694***	13.363***
	(0.690)	(0.689)	(0.289)	(0.242)
Overall R^2	0.019	0.037	0.093	0.009
N	4320	1440	1530	1800
n (# of clusters)	32	32	34	40
rho	0.224	0.481	0.120	0.095
Wald-chi2	98.025***	99.108***	176.424***	18.476***

Remark: Standard errors in parentheses. ***/**/* indicates significance at the 1/5/10% level. Rho gives the proportion of overall variance contributed by the panel-level component. Wald-chi2 reports the test statistic from testing that all coefficients (except the constant) are zero.

The game B dummy and its interaction with the endogenous treatment dummy are of main interest. The former is always significantly negative, the latter significantly positive. The game B dummy and the interaction term are jointly significant only for the one shot games ($l = 1$); when we test the hypothesis that GameB + gameB×endo equals zero by means of a chi-square test, we obtain a p-value of $p = 0.01$ in the first column and $p = 0.07$ in the second.

Table C-2: Fractions of (Trust, Betray) outcome by game length and game

			Exo	Endo	Exo vs. Endo
$l = 1$	first round	Bad	0.242	0.304	0.173
	$(n = 6)$	Good	0.231	0.210	0.173
		B vs. G	0.753	0.116	
$l = 3$	first round	Bad	0.217	0.138	0.345
	$(n = 6)$	Good	0.199	0.337	0.249
		B vs. G	0.917	0.075	
$l = 3$	last round	Bad	0.133	0.086	0.116
	$(n = 6)$	Good	0.121	0.027	0.046
		B vs. G	0.917	0.116	
$l = 3$	all rounds	Bad	0.209	0.173	0.173
	$(n = 6)$	Good	0.205	0.162	0.116
		B vs. G	0.917	0.753	
$\delta = \frac{2}{3}$	first round	Bad	0.090	0.027	0.018
	$(n = 8)$	Good	0.121	0.136	0.623
		B vs. G	0.362	0.063	
$\delta = \frac{2}{3}$	last round	Bad	0.186	0.120	0.128
	$(n = 8)$	Good	0.191	0.166	0.093
		B vs. G	0.779	0.043	
$\delta = \frac{2}{3}$	all rounds	Bad	0.181	0.109	0.043
	$(n = 8)$	Good	0.197	0.190	0.575
		B vs. G	0.263	0.018	

Remark: The rows 'B vs. G' report the p-values of Wilcoxon signrank tests for matched pairs, comparing the B-game with the G-game. The column 'Exo vs. Endo' reports the p-values of Wilcoxon signrank tests for matched pairs, comparing the exogenous treatment with the endogenous treatment (for a given game). All tests are based on group level data, with n giving the number of groups.

References

Anderhub, V., D. Engelmann, and W. Guth (2002). An experimental study of the repeated trust game with incomplete information. *Journal of Economic Behavior and Organization 48*, 197–216.

Baker, G., R. Gibbons, and K. Murphy (1994). Subjective performance measures in optimal incentive contracts. *Quarterly Journal of Economics 109*, 1125–1156.

Baker, G., R. Gibbons, and K. Murphy (2001). Bringing the market inside the firm? *American Economic Review 91*, 212–218.

Baker, G., R. Gibbons, and K. Murphy (2002). Relational contracts and the theory of the firm. *Quarterly Journal of Economics 116*, 39–84.

Berg, J., J. Dickhaut, and K. McCabe (1995). Trust, reciprocity, and social history. *Games and Economic Behavior 10*, 122–142.

Bernheim, B. and M. Whinston (1998). Incomplete contracts and strategic ambiguity. *American Economic Review 88*, 902–932.

Blonski, M. and G. Spagnolo (2007). Relational efficient property rights.

Bragelien, I. (2002). Asset ownership and relational contracts. Working Paper, Norwegian School of Economics and Business Administration.

Brown, M., A. Falk, and E. Fehr (2004). Relational contracts and the nature of market interactions. *Econometrica 72*, 747–780.

Che, Y.-K. and S.-W. Yoo (2001). Optimal incentives for teams. *American Economic Review 91*, 525–541.

Chen, Y. (2000). Promises, trust, and contracts. *Journal of Law, Economics and Organization 16*, 209–232.

Cochard, F., P. Van, and M. Willinger (2004). Trusting behavior in a repeated investment game. *Journal of Economic Behavior and Organization 55*, 31–44.

Dal-Bo, P. (2005). Cooperation under the shadow of the future: Experimental evidence from infinitely repeated games. *American Economic Review 95*, 1591–1604.

Demougin, D. and O. Fabel (2004). The determinants of salary and bonus for rank and file employees.

Dufwenberg, M. and G. Kirchsteiger (2004). A theory of sequential reciprocity. *Games and Economic Behavior 47*, 268–298.

Engle-Warnick, J. and R. Slonim (2004). The evolution of strategies in a repeated trust game. *Journal of Economic Behavior and Organization 55*, 553–573.

Falk, A., S. Gächter, and J. Kovacs (1999). Intrinsic motivation and extrinsic incentives in a repeated game with incomplete contracts. *Journal of Economic Psychology 20*, 251–284.

Falk, A. and M. Kosfeld (2006). The hidden costs of control. *American Economic Review 96*, 1611–1630.

Fehr, E. and A. Falk (2002). Psychological foundations of incentives. *European Economic Review 46*, 687–724.

Fehr, E., A. Klein, and K. Schmidt (2007). Fairness and contract design. *Econometrica 75*, 121–154.

Fehr, E. and J. List (2004). The hidden costs and returns of incentives - trust and trustworthiness among ceos. *Journal of the European Economic Association 2*, 743–771.

Fehr, E. and B. Rockenbach (2003). Detrimental effects of sanctions on human altruism. *Nature 422*, 137–140.

Fehr, E. and K. Schmidt (2007). Adding a stick to the carrot? the interaction of bonuses and fines. *American Economic Review 97*, 177–181.

Gächter, S. and A. Falk (2002). Reputation and reciprocity: Consequences for the labour relation. *Scandanavian Journal of Economics 104*, 1–26.

Garvey, G. (1995). Why reputation favors joint ventures over vertical and horizontal integration: A simple model. *Journal of Economic Behavior and Organization 28*, 387–397.

Gneezy, U. and A. Rustichini (2000a). A fine is a price. *Journal of Legal Studies 29*, 1–17.

Gneezy, U. and A. Rustichini (2000b). Pay enough or don't pay at all. *Quarterly Journal of Economics 115*, 791–810.

Halonen, M. (2002). Reputation and the allocation of ownership. *Economic Journal 112*, 539–558.

Hart, O. (2001). Norms and the theory of the firm. *University of Pennsylvania Law Review 149*, 1701–1715.

Hart, O. and J. Moore (2007). Incomplete contracts and ownership: Some new thoughts. *American Economic Review 97*, 182–186.

Hart, O. and J. Moore (2008). Contracts as reference points. *Quarterly Journal of Economics 123*, 1–48.

Itoh, H. and H. Morita (2006). Formal contracts, relational contracts, and the holdup problem. CESifo Working Paper No. 1786.

Keser, C. (2002). trust and reputation building in e-commerce. Working Paper, CIRANO.

Kreps, D. (1990). Corporate culture and economic theory. In J. Alt and K. Shepsle (Eds.), *Perspectives on Positive Political Economy*. Cambridge University Press.

Kvaloy, O. and T. Olsen (2006a). Endogenous verifiability and relational contracting. Working paper.

Kvaloy, O. and T. Olsen (2006b). Team incentives in relational employment contracts. *Journal of Labor Economics 24*, 139–169.

Levin, J. (2003). Relational incentive contracts. *American Economic Review 93*, 835–857.

MacLeod, B. and J. Malcomson (1989). Implicit contracts, incentive compatibility, and unvoluntary unemployment. *Econometrica 57*, 447–480.

MacLeod, W. (2007a). Can contract theory explain social preferences? *American Economic Review 97*, 187–192.

MacLeod, W. (2007b). Reputations, relationships, and contract enforcement. *Journal of Economic Literature 45*, 595–628.

Murdock, K. (2002). Intrinsic motivation and optimal incentive contracts. *Rand Journal of Economics 33*, 650–671.

Pearce, D. and E. Stacchetti (1998). The interaction of implicit and explicit contracts in repeated agency. *Games and Economic Behavior 23*, 75–96.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review 83*, 1280–1302.

Rayo, L. (2007). Relational incentives and moral hazard in teams. *Review of Economic Studies 74*, 937–963.

Schmidt, K. and M. Schnitzer (1995). The interaction of explicit and implicit contracts. *Economics Letters 48*, 193–199.

Schottner, A. (2007). Relational contracts, multitasking, and job design. *Journal of Law, Economics and Organization 24*, 138–162.

Scott, R. (2003). A theory of self-enforcing indefinite agreements. *Columbia Law Review 103*, 1641–1699.

Sloof, R., H. Oosterbeek, and J. Sonnemans (2007). Does making specific investments unobservable boost investment incentives? *Journal of Economics and Management Strategy 16*, 911–942.

van Huyck, J. B., R. C. Battalio, and M. F. Walters (1995). Commitment versus discretion in the peasant-dictator game. *Games and Economic Behavior 10*, 143–170.

van Huyck, J. B., R. C. Battalio, and M. F. Walters (2001). Is reputation a substitute for commitment in the peasant-dictator game? Working Paper.

Figure 1a: “Bad” explicit contract (game B)

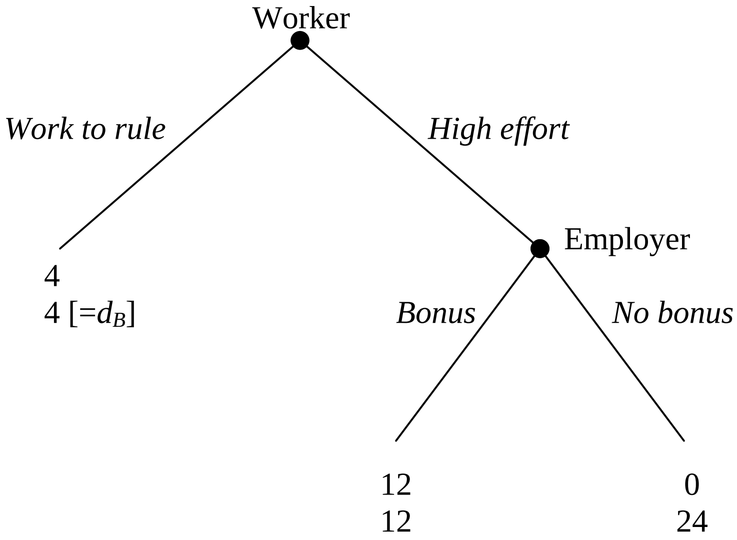


Figure 1b: “Good” explicit contract (game G)

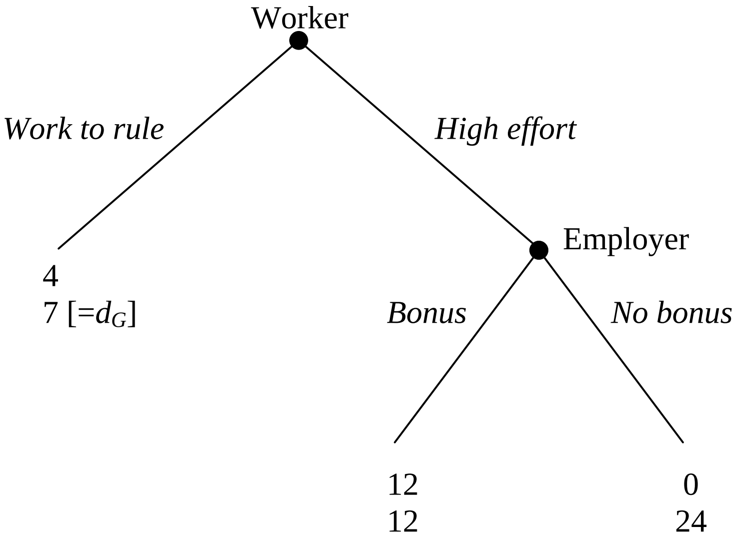
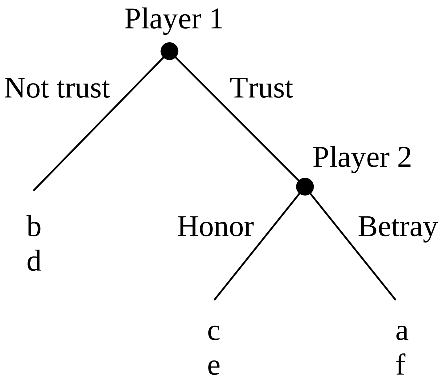


Figure 2: The trust game



with c ≥ b > a, f > e > d and c+e > max{b+d,a+f}.

Figure 3: Percentages of outcomes by game length and contract (and contract choice)

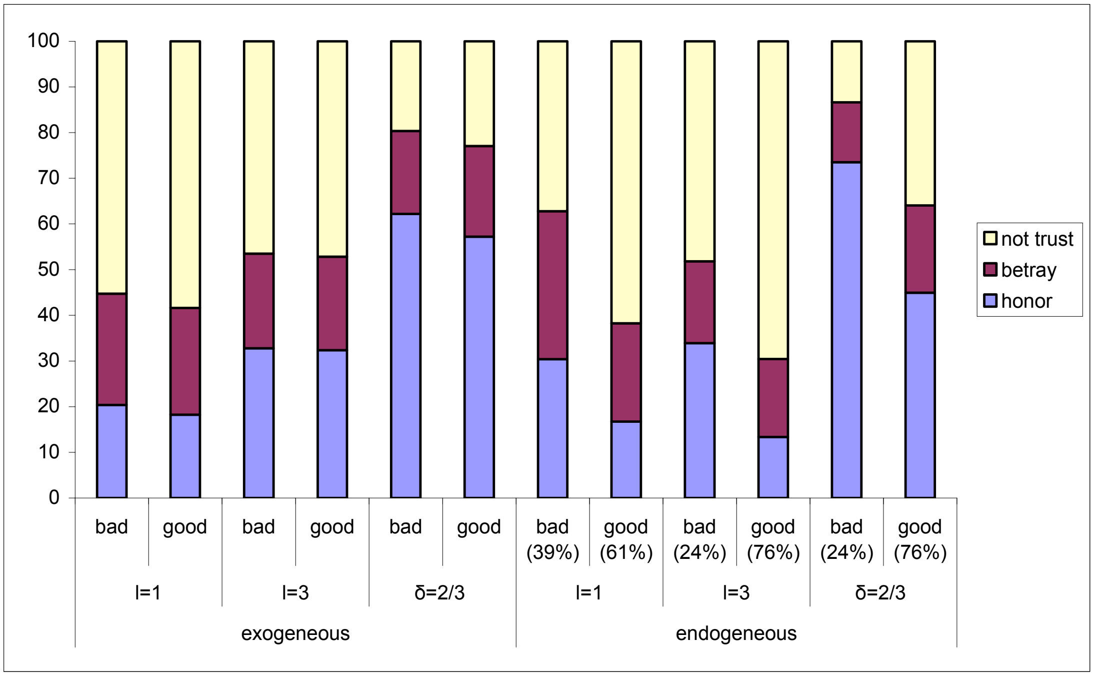


Figure A1: Underlying game when $S(e_0) > 0$

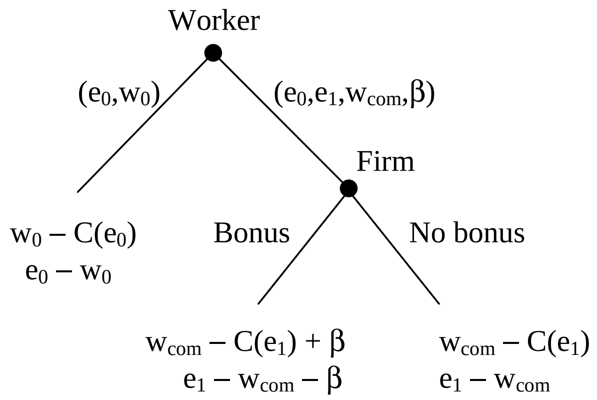
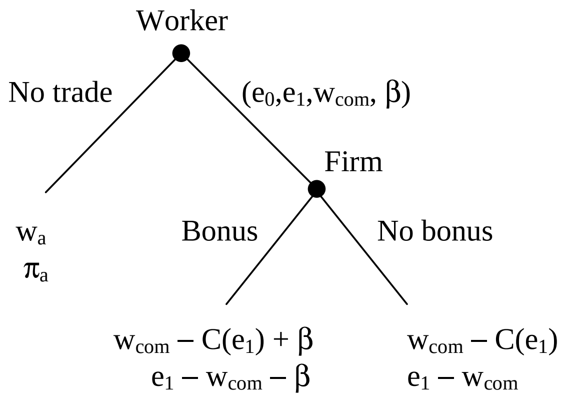


Figure A2: Underlying game when $S(e_0) < 0$


Worker
No trade
(e_0,e_1,w_{com}, β)
Firm
w_a
π_a
Bonus
No bonus
$w_{com} - C(e_1) + \beta$
$e_1 - w_{com} - \beta$
$w_{com} - C(e_1)$
$e_1 - w_{com}$