TI 2010-074/1
Tinbergen Institute Discussion Paper

Self-Selection and the Power of Incentive Schemes: An Experimental Study

Jana Vyrastekova[1]
Sander Onderstal[2]
Pierre Koning[3]

[1] *Radboud University Nijmegen;*
[2] *University of Amsterdam, and Tinbergen Institute;*
[3] *CPB Netherlands Bureau for Economic Policy Analysis, The Hague.*

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

Self-selection and the Power of Incentive Schemes: An Experimental Study*

Jana Vyrastekova,[†] Sander Onderstal,[‡] and Pierre Koning[§]

July 29, 2010

Abstract

We examine how self-selection of workers into firms depends on the power of the firms' incentive schemes and how it affects the performance of firms that increase the power of the incentive schemes. In a laboratory experiment, we let subjects choose between (low-powered) team incentives and (high-powered) individual incentives. We observe that subjects exhibiting high trust or reciprocity in the trust game are more likely to choose team incentives. When exposed to individual incentives, workers who chose team incentives perform worse if both the unobservable interdependency between workers and their incentive to cooperate under team incentives are high.

JEL Classification: C91, J33, M52

Key words: Incentive scheme; Self-selection; Laboratory experiment

*Financial support from CPB Netherlands Bureau for Economic Policy and the Dutch National Science Foundation (NWO-VENI 051.14.023 and NWO-VICI 453.03.606) is gratefully acknowledged. For valuable discussions and comments, we would like to thank Marcel Canoy, Robert Dur, Jan Potters, Arthur Schram, Jeroen van de Ven, Frans van Winden, participants a seminar at the University of Amsterdam, the Conference on "Incentives and Work Motivation in the Public Sector" 2005 at Erasmus University Rotterdam, and the 2006 EALE conference in Prague. The usual disclaimer applies.

[†]Radboud University Nijmegen, Nijmegen School of Management, Thomas van Aquinostraat 5, 6525 GD Nijmegen, The Netherlands, j.vyrastekova@fm.ru.nl

[‡]Corresponding author. University of Amsterdam, Amsterdam School of Economics, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands, onderstal@uva.nl

[§]CPB Netherlands Bureau for Economic Policy Analysis, P.O. Box 80510, 2508 GM The Hague, The Netherlands, pwck@cpb.nl

1 Introduction

A reorganisation in a firm or the restructuring of the public sector often involves proposals to increase the power of employees' incentive schemes. Recent examples of such proposals in the public sector are with respect to teachers, employment offices, and medical practices (see Burgess and Ratto (2003) for a survey). Indeed, there is some empirical evidence that high-powered incentive schemes improve workers' performance. For instance, Lazear (2000) studies the behaviour of 3,000 employees of a car glass company that changed the compensation method from an hourly wage to piece-rate pay. This shift could be interpreted as a move from a low-powered incentive scheme to a high-powered one. Lazear observes that the average output per worker increases by 44%. Nalbantian and Schotter (1997) find similar evidence in laboratory experiments. They compare payment schemes with low-powered incentives (such as revenue sharing) and high-powered ones (such as tournaments). They document that high-powered incentive schemes perform much better than low-powered ones.

However, the literature also offers mechanisms which explain why low-powered incentive schemes may perform well. These include intrinsic motivation, positive reinforcement, mutual monitoring among team members, and opportunities for workers to reciprocate among each other within a team (see, e.g., Kandel and Lazear (1992), Gneezy and Rustichini (2000), Minkler (2004), and Canton (2005)). In questionnaires, workers indicate

that they very much appreciate non-pecuniary motivations like intrinsic incentives to perform the job (Frey and Jegen (2001)) as well as a fair relationship with the employer (Fehr and Schmidt (2004)). Indeed, Lavy (2002) (teachers), Knez and Simester (2001) (airlines), and Hamilton et al. (2003) (the garment industry) provide evidence on teams in which (low-powered) team incentives do perform well. More generally, low-powered incentive schemes may outperform high-powered ones when the latter crowd out intrinsic motivation (Frey and Jegen (2001)).

In this paper, we study the effect of increasing the power of employees' incentive schemes in a setting where worker types self-selected into firms before being aware of the change in incentive scheme. Indeed, crowding-out of intrinsic motivation may be amplified when intrinsically motivated workers self-selected into firms with low-powered incentives (Delfgaauw and Dur (2008)). We examine how self-selection of workers into firms depends on the preferences for reciprocity and on the power of the firms' incentive schemes, and how it affects the performance of firms that increase the power of the incentive schemes.

We answer these questions using a laboratory experiment. We use a laboratory experiment to answer our research questions because field data are likely to suffer from measurement and identification problems. Measurement problems may, for instance, arise in connection to workers' output. Even if output is measurable, the researcher has no information on an individual's effort and preferences. Identification problems may occur because it is not straightforward to isolate the effect of the power of the incentive scheme

on a worker's performance from other effects. In contrast, in the laboratory we can observe each subject's effort and measure their preferences, and we can expose the same individual to different payment schemes and observe her reaction, while keeping the rest of the environment constant.

In our experiment, subjects perform two tasks in a simple production game. In Task 1, they produce their own output, while in Task 2, they increase the output of the colleague in their team. Before they perform the tasks, subjects have to choose between two payment schemes: A low-powered one and a high-powered one. Subjects choosing the same scheme are matched together. Under either scheme, their payment is based on both team performance and individual performance. In the high-powered one, individual performance has a higher weight in the payment. For 10 rounds, subjects participate under the scheme they have chosen. Next, all subjects play 10 more rounds in which they are paid according to the high-powered payment scheme. A priori, it is not clear whether the performance will improve for those who had chosen the low-powered incentive scheme. The higher power of the scheme may induce workers to expend more effort in Task 1 because they obtain a higher payment for their own output. However, this effect may be counterbalanced because workers no longer have the incentive to reward their colleague for high effort in Task 1 with high effort in Task 2. In other words, reciprocal behaviour is not triggered by the high-powered incentive scheme because money-maximizing workers exert high effort in Task 1 in the subgame perfect Nash equilibrium. The latter effect

may be strong because in the self-selection process, reciprocators may be more likely to choose to work for a low-powered incentive scheme than money-maximisers. However, also money-maximisers who have trust in others to cooperate in both tasks may choose the low-powered incentive scheme so that they can choose to free-ride in Task 2. To analyse the role of trust and reciprocity, we let subjects submit strategies in the trust game and use these as measures for these characteristics.

Our experimental findings show that trust and reciprocity are important determinants of subjects' sorting behavior. The more a subject trusts and the higher her propensity to reciprocate the more likely she is to choose the low-powered payment scheme. Indeed, we observe reciprocal behaviour in the low-powered incentive scheme in the sense that a subject is more likely to expend high effort in Task 2 if both she and her colleague exert high effort in Task 1. We also find that increasing the power of the incentive scheme has ambiguous effects: Subjects who choose the low-powered incentive scheme increase output when confronted with high-powered incentives if and only if the benefits from (unobservable) cooperation are high, from both the individual's and the team's viewpoint. Self-selection partly explains why workers may perform worse if the power of their incentive scheme is increased.

Self-selection of employees in firms has received some attention in the empirical literature in the past few years. In his field experiment in a car glass company, Lazear (2000) not only observes that the average output per worker went up when the company

increased the power of its workers' incentive scheme, he also finds that the firm was increasingly able to attract productive workers. Hamilton et al. (2003) find the opposite effect in a study in a garment factory: They observe that high-ability workers are more likely to join teams than low-ability workers.

Self-selection into payment schemes has also been studied in the laboratory. Dohmen and Falk (2006), Cadsby et al. (2007), and Eriksson and Villeval (2008) observe that the more productive a subject is the more likely she is to opt for a high-powered payment scheme. In contrast to us, those authors do not examine the effect of increasing the power of the incentive scheme for those who choose the low-powered one (i.e. a fixed payment). Keser and Montmarquette (2009) have a similar set-up as ours, with the differences that (1) they let each subject stay in the same two-player team for the duration of the experiment and (2) by construction, the maximum payoff is higher under team incentives than under individualistic incentives, while in our set-up it is the same. Keser and Montmarquette observe that team incentives are popular, and frequently lead to high output. In our experiment, subjects are randomly rematched after each production game so that trigger strategies cannot explain the instances of cooperation we observe.

The remainder of the paper is organised as follows. In section 2, we discuss the design of the experiment and our hypotheses. Section 3 includes the experimental observations. Section 4 concludes.

2 The experiment

In this section, we describe the design of our experiment and the hypotheses that we wish to test based on the results from the theory.

2.1 Design

In the year 2004, we ran 9 experimental sessions.[1] Altogether, 172 students from Tilburg University participated. Participants were paid for all points they earned in the experiment (on average 13 Euro including a 5 Euro participation fee for a session lasting approximately 1.5 hours). The experiments were fully computerised, programmed, and conducted using z-Tree (Fischbacher (2007)). Upon arrival at a session, participants were randomly seated at computer cubicles which were separated by blinds. During the experiment, communication other than via the computer was prohibited.

Of the 172 subjects, 134 participated in the main design, while 38 entered control sessions (see further below). In each experimental session of the main design, subjects had to make decisions at four subsequent stages:

1. the trust game;

2. a "labour market" in which subjects chose between TEAM and INDI;

3. the production game (10 rounds) in the chosen incentive scheme TEAM or INDI;

[1] Instructions are available at http://www.sanderonderstal.com/Instructions/SelfselectionP2.doc.

4. the production game (10 rounds) in INDI.

Subjects only received instructions for the stage that they were on, and were not informed about the stages to follow. Moreover, no feedback on the trust game was given before the end of the experiment so as not to contaminate further decision making in the main part of the experiment.

Let us discuss the experiment in detail. We used the trust game to measure subjects' reciprocity and trust (i.e. their belief in others' reciprocity). Berg et al. (1995) designed the trust game to mimic a situation in which two players, a sender and a receiver, can profit if trust exists between them. The sender has to decide how much of her 10 point endowment to transfer to the receiver. This money is then tripled and the receiver has to decide how much money (if any) to return to the sender. In the unique subgame perfect Nash equilibrium of the trust game, a money maximizing receiver will return zero, so that a money-maximizing sender will transfer zero. However, senders who expect receivers to be sufficiently reciprocal have the incentive to transfer a strictly positive amount of money. We, therefore, use a subject's action in the role of sender to evaluate her level of trust, and her strategy in the role of receiver to measure her level of reciprocity.

In order to obtain a measure of both trust and reciprocity for each subject, we used a strategy method behind the veil of ignorance.[2] We asked subjects to submit strategies for

[2]Vyrastekova and Onderstal (2010) discuss this design and observe that subjects behave similarly as in the standard trust game design.

both roles of the trust game, i.e. they first decided how much to transfer to the receiver in the role of sender, and then how much to return to the sender for every level of transfer the sender could make. At the very end of the experimental session, we let the computer decide at random which role each subject would play and to whom she would be matched.[3] We paid subjects according to the strategies they submitted for the role that was assigned by the computer. A measure of a subject's trust is how much she transfers in the role of sender (which is an integer between 0 and 10).[4] A measure of a subject's reciprocity is the average fraction she returns in the role of receiver (a number between 0 and 1).

In the second stage of the experiment, we present subjects with a choice between the production games INDI and TEAM. Both are 2-stage games. The two subjects provide input in two subsequent parts. In part 1, each subject independently chooses effort $e \in \{L, H\}$. In Task 2, after observing the effort of the other subject, each subject chooses a reward $r \in \{0, R\}$. Two subjects who opt for the same scheme form a team and interact in the game of their choice. Table 1 below includes the parameters of TEAM and INDI. Four sessions used parametrisation P1 and three sessions parametrisation P2 (see table 2).

We interpret the actions of the subjects as follows. Both are workers in a firm. "$e = H$"

[3] In this way, the behavior of others in the trust game cannot affect the interaction in the main experimental game, being the production game.

[4] We are aware of the fact that the sender's motivation to transfer money in the trust game may go beyond the belief in positive reciprocity (Charness (2004)). For instance, risk and betrayal aversion might affect sender's decision to send money in the trust game (see Schechter (2007) and Bohnet and Zeckhauser (2004) respectively). However, Vyrastekova and Garikipati (2005) provide evidence that transfers are a reasonable measure of belief in positive reciprocity.

Parametrisation P1					
INDI			TEAM		
Part 1	$e = H$	$e = L$	Part 1	$e = H$	$e = L$
$e = H$	6,6	4,4	$e = H$	6,6	0,8
$e = L$	4,4	2,2	$e = L$	8,0	2,2
Part 2	$r = R$	$r = 0$	Part 2	$r = R$	$r = 0$
$r = R$	8,8	-6,14	$r = R$	8,8	-2,10
$r = 0$	14,-6	0,0	$r = 0$	10,-2	0,0
Parametrisation P2					
INDI			TEAM		
Part 1	$e = H$	$e = L$	Part 1	$e = H$	$e = L$
$e = H$	6,6	4,2	$e = H$	6,6	-1,-1
$e = L$	2,4	0,0	$e = L$	-1,-1	0,0
Part 2	$r = R$	$r = 0$	Part 2	$r = R$	$r = 0$
$r = R$	6,6	-6,12	$r = R$	6,6	-1,7
$r = 0$	12,-6	0,0	$r = 0$	7,-1	0,0

Table 1: Experiment payoff matrices of the production games

	Parametrisation P1	Parametrisation P2
Number of sessions	4	3
Number of subjects	78	56

Table 2: Experimental sessions

["$e = L$"] refers to exerting high [low] effort in producing own output, "$r = R$" to rewarding/helping the team mate, which increases the team mate's output, and "$r = 0$" to not doing so.[5] The pay-off for the subjects is higher if both choose $e = H$ [$r = R$] than if both choose $e = L$ [$r = 0$]. The interpretation is that the firm shares the fruits of high output with the workers. The difference between INDI and TEAM is that in INDI,

[5] In the experiment, we used neutral labels: "PULL" instead of $e = H$, "PUSH" instead of $e = L$, "GIVE" instead of $r = R$, and "KEEP" instead of $r = 0$.

a worker is mainly rewarded for her own output. So, her pay-offs are high if she chooses $e = H$ in part 1 and if her team mate chooses $r = R$ in part 2. In contrast, in TEAM, workers are to a large extent rewarded by team output, so relative to INDI, they have more incentives to opt for $e = L$ in part 1 and fewer incentives to choose $r = 0$ in part 2.[6]

In each round, a subject was assigned to another anonymous co-player from among the subjects who chose the same scheme.[7] This matching procedure was known to the subjects.[8] Of the 134 subjects in the main design, 6 could not continue into the production game because an odd number of subjects entered either of the two schemes.

After 10 rounds, we informed subjects that they would play 10 more rounds in which all participated in INDI. We rematched them after each round, but only among those who chose the same scheme in the labour market. Having also finished these 10 rounds,

[6] The parameters are consistent with the following production game. The costs of effort are $c(e_i)$, with $c(L) = 0 < c(H) < H - L$. Cost of reward are $k(r_i)$ where $k(0) = 0$ and $0 < k(R) < R$. An agent's effort raises her own output, while her reward raises the output of her team member (which could be interpreted as helping the other person or being cooperative with her). More precisely, the relationship between the efforts and the rewards of players i and j forming one team and the output o_i of player i is $o_i = e_i + r_j$, $\{i, j\} = \{1, 2\}$.

Player i's payment equals $p_i = \alpha o_i + (1 - \alpha) o_j$, $\{i, j\} = \{1, 2\}$, where α is a measure of the extent to which an agent's own output determines her payment. In the extreme case where $\alpha = 1$, only her own output determines what she gets, while the other extreme $\alpha = \frac{1}{2}$ indicates that the payments are only based on total team performance.

The following table includes the parameters for the two parametrisations:

	H	L	$c(H)$	R	$k(R)$	α_{INDI}	α_{TEAM}
Parametrisation P1	18	2	12	16	8	$\frac{7}{8}$	$\frac{5}{8}$
Parametrisation P2	14	0	8	14	8	$\frac{6}{7}$	$\frac{1}{2}$

[7] When an odd number of subjects chose either scheme, one or two subjects were randomly excluded from continuing in the experiment, so that we could match the subjects into pairs.

[8] Subjects did not know how many others chose the same scheme. Therefore, they could not condition their behavior on the self-selection procedure outcome. In each session, more than four subjects self-selected into each scheme so that we did not observe repeated interaction of a fixed matched pair.

the subjects learned the results of the trust game and their accumulated earnings for the whole experiment, and were paid by us in cash.

In order to evaluate how self-selection affects our results, we ran a control session where we forced all subjects to start with TEAM. We did so with 18 and 20 subjects in P1 and P2 respectively. In both control sessions, the procedures were kept as similar to the other sessions as possible. Subjects submitted the trust game strategies, and then played 10 rounds of the production game under TEAM incentives in the given parametrisation. Afterwards, we informed subjects that they would participate in 10 more rounds and we exposed them to the TEAM scheme in the other parametrisation. We chose this approach so as to let subjects earn approximately the same amount of income as in the other sessions. In our analysis, we only compare TEAM without self-selection in rounds 1 to 10 to TEAM with self-selection (rounds 1 to 10 as well).

2.2 Hypotheses

In this subsection, we formulate the hypotheses that we wish to test using experimental data. In order to do so, let us have a closer look at the pay-off tables in table 1. First, note that for money-maximizing subjects, the 2-stage games have a unique subgame perfect Nash equilibrium (SPNE). In the SPNE of INDI [TEAM], both team members choose $e = H$ $[e = L]$ in part 1 and $r = 0$ in part 2. It is readily verified that in the SPNE a subject's pay-off is higher in INDI than in TEAM. Therefore, a money-maximizing

subject will always choose INDI. Moreover, the power of incentives in INDI is higher than in TEAM. The reason is that in task 1 of INDI, $e = H$ is the SPNE strategy, in contrast to TEAM, while both choose the same action in task 2.

However, TEAM may yield a better outcome than INDI if the population contains sufficiently many reciprocal subjects, i.e. subjects who wish to cooperate as long as their team mate does so as well. A reciprocator may play the following "tit-for-tat" strategy in TEAM: She starts off by choosing $e = H$ in part 1, and she continues to cooperate by choosing $r = R$ in part 2 if and only if the other team member chooses $e = H$ part 1 as well. Note that in TEAM, a money maximizing subject (i.e., non-reciprocal) may "free-ride" on a reciprocal subject by choosing $e = H$ in part 1 and $r = 0$ in part 2. Observe that the money maximiser obtains a higher pay-off in TEAM than in INDI if she meets a reciprocator who plays the above "tit-for-tat" strategy.

So, the most important determinant for subjects to choose TEAM seems to be their level of trust. Those who trust the other subject to reciprocate have an incentive to enter TEAM. Others perfer INDI.

Hypothesis 1a: Labor market. Subjects with a high trust level are more likely to choose TEAM than those with a low one.

Hypothesis 1b: Actions. Subjects who choose INDI play $e = H$ in Task 1 and $r = 0$ in Task 2. Reciprocal subjects who choose TEAM play $e = H$ in Task 1 and, if the

other does so as well, chooses $r = R$ in Task 2 and $r = 0$ otherwise. Non-reciprocal subjects who choose TEAM play $e = H$ in Task 1 and $r = 0$ in Task 2.

Next, we spell out the hypothesis most relevant for policy: Does a team perform better the higher the power of the incentive scheme? If we assume an increasing relationship between the subjects' payoffs and the profits of the firm, the answer is no. The reason is that subjects will only choose a scheme if they expect the payments under this scheme to be higher than under the other scheme.

Hypothesis 2: The power of incentives. Subjects who choose to play TEAM in rounds 1-10 produce more in rounds 1-10 than in rounds 11-20 (in which they play INDI).

We may find support for hypothesis 2 if reciprocal subjects self-selected in TEAM. Indeed, if sufficiently many subjects choose $e = H$ in Task 1 and $r = R$ in task 2, they perform worse in INDI.

Finally, we address the question to which extent self-selection contributes to TEAM outperforming INDI - if it does. Subjects may select TEAM to express their willingness to provide high effort without explicit monetary incentives. It is especially relevant when we think of firms operating under team-based incentives. The incentive scheme is usually known to the workers entering the firm beforehand, i.e., it is one of the factors upon which they select the firm. If self-selection matters then it generates a reason for the firm to stick to TEAM incentives. We can evaluate the role of self-selection for the success of

team-based incentives by comparing the behaviour of subjects in TEAM who self-selected this scheme, to subjects who were forced to play TEAM. If more cooperation is found in the former group, then self-selection is responsible for at least part of the success of TEAM.

Hypothesis 3: Self-selection. Subjects who select TEAM in the main design (in which they can choose between TEAM and INDI) are more likely to choose $e = H$ in task 1 and $r = R$ in task 2 than those who participate in the control design (in which all play TEAM).

Hypothesis 3 will not be rejected if the initial sorting of subjects into TEAM and INDI results in higher payoffs for subjects in TEAM. The driving force may be that those who choose TEAM trust more and are more reciprocal than those choosing INDI. If all subjects are forced to play TEAM, less reciprocity may be realised, so that the output in TEAM ends up being lower than it would be under sorting. Note that all hypotheses will be rejected if the population only contains money-maximizing subjects.

3 Data analysis

In this section, we address the results from our experiment in the light of the three hypotheses formulated above. We find support for the first hypothesis for parametrisation P1, and all hypotheses for P2. Before discussing these observations in more detail, we

wish to note that the findings in our trust game set-up do no differ substantially from the observations in Berg et al.'s (1995) standard trust game (quoted in brackets). In our experiment, senders transfer, on average, 51% (52%) of their endowment to receivers, while 9% (6%) of them send nothing. Receivers return on average 36% (30%) of the sender's transfer.

3.1 Labour market and actions

A substantial fraction of subjects chooses TEAM in both parametrisations: 34% in P1 and 44% in P2. This finding allows us to reject the hypothesis that all select INDI, which is the optimal choice if all play according to the SPNE. Figures 1 and 2 present subjects' choices in the trust game conditional on the payment scheme they selected. From these figures, it becomes clear that those who choose TEAM trust more and are more reciprocal. Subjects selecting INDI send on average 4.6 points while those selecting TEAM send significantly more, 5.5 points (Mann-Whitney U test, $p = 0.049$). Subjects selecting TEAM return a fraction that leaves senders' investment profitable (i.e. more than one third of the received number of points), unlike subjects selecting INDI who return less than senders sent to them.

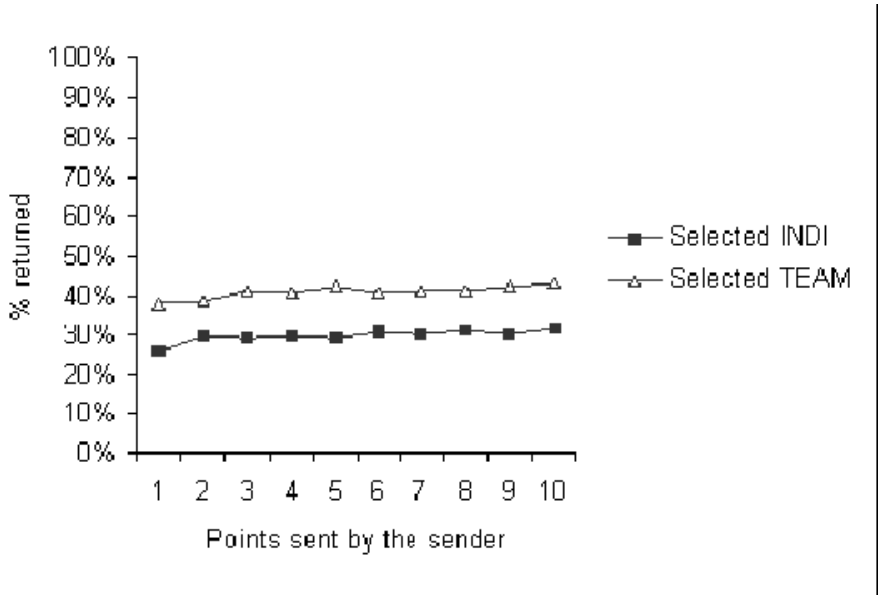

Figure 1: Reciprocity by subjects who selected INDI and TEAM.

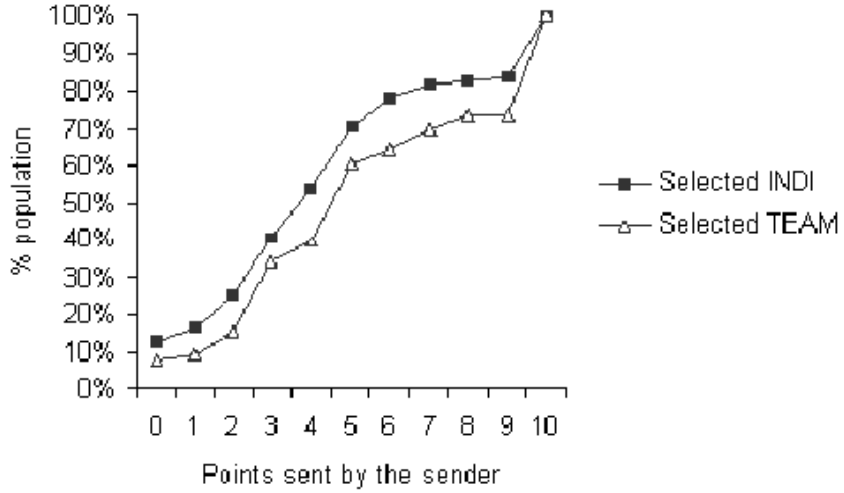

Figure 2: Distribution of "trust" by subjects who selected INDI and TEAM.

Table 3 includes Probit estimates for the choice of TEAM. The outcomes show that there is a non-linear relationship between trust and reciprocity on one hand, and the

choice of the payment scheme on the other. Both trust and reciprocity have a positive effect on the choice of TEAM (in line with the above non-parametric tests). However, the interaction term is negative and (weakly) significant. Its parameter estimate implies that given values of reciprocity below 0.57, higher trust makes entry into TEAM more likely.

Variable	Coefficient estimate	Standard error
Trust	0.057**	(0.024)
Reciprocity	0.74**	(0.33)
Trust * reciprocity	- 0.10*	(0.052)
P2-dummy	- 0.059	(0.089)
# observations	134	
Log Likelihood	- 85.44	
* [**] indicates significance at the 10% [5%] level.		

Table 3: Probit estimates of choice of TEAM. The coefficients are expressed as marginal effects.

These observations have two implications. First, subjects who exhibit high trust are likely to enter TEAM, in line with hypothesis 1a. Second, TEAM may attract "free-riders": People who believe that sufficiently many other are reciprocators, without having the attention to reciprocate themselves. In other words, these subjects may imitate behaviour of reciprocators in Task 1, but free-ride on them in Task 2. Consequently, the TEAM scheme attracts reciprocal as well as money-maximizing subjects, as long as their trust is sufficiently high.[9]

Observation 1a (Labor market): A nonnegligible fraction of subjects (more than one

[9] We tested for the possibility of multi-collinearity between Trust and Reciprocity, and found this not to be of importance.

third in both parametrisations) selects TEAM. On average, subjects who do so trust more and are more reciprocal than subjects who opt for INDI. However, money-maximisers (subjects with low reciprocity levels) also enter TEAM if their trust-level is sufficiently high.

We now turn to analyzing subjects' strategies under the payment scheme of their choice. From Figure 3, we derive that in Task 1, subjects in INDI choose nearly exclusively $e = H$ (consistent with the SPNE). Also in TEAM, we observe $e = H$: In parametrisation P1 [P2] on average 31% [69%] of subjects choose $e = H$. This observation is somewhat surprising: We expect someone who enters TEAM to always play $e = H$, hoping that her team mate will choose $r = R$ in Task 2. One explanation for this discrepancy is that subjects entered TEAM by mistake. There is indeed support for this suggestion. Subjects choosing $e = L$ in TEAM send significantly less points in the trust game than subjects choosing $e = H$ in the same scheme (5.2 points vs. 6.1 points in P1 and 4.1 points vs. 6.4 points in P2 with p-values of the Mann-Whitney U tests being $p = 0.059$ and $p = 0.000$, respectively). This means that their trust is lower than for subjects entering TEAM who choose $e = H$. However, low trust implies that they should prefer INDI.

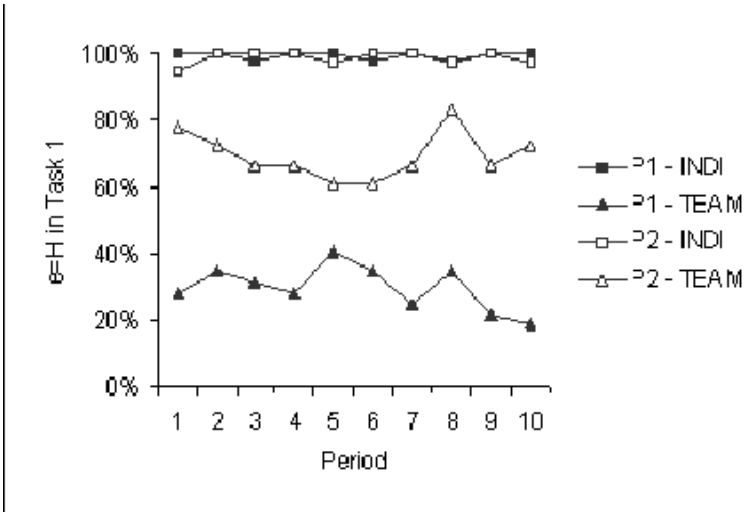

Figure 3: Action $e = H$ in Task 1 for P1 and P2.

Figure 4 indicates that high effort in Task 2 ($r = R$) is much more likely in TEAM than in INDI, especially when both subjects chose high effort in Task 1 ($e = H$). We use a conditional Logit model to investigate whether subjects employ the above "tit-for-tat strategy" in Task 2 (see Table 4 for the estimates). This specification accounts for subject-specific effects because a subject's choices in Tasks 1 and 2 may depend on her individual characteristics. For both P1 and P2, it is apparent that the probability of choosing $r = R$ is highest when both players chose $e = H$ in Task 1 (the coefficients on any other history observed are significant and negative). Moreover, we observe that those who choose $r = R$ in TEAM are more likely to be reciprocators than those who choose $r = 0$. In the trust game, the former return on average 51% [39%] of the sent amount in P1 [P2], while the latter return on average 39% (25%). The difference is significant

in both cases (Mann–Whitney U test, $p = 0.090$ [$p = 0.008$]). So, reciprocators are less inclined to "free ride" in Task 2.

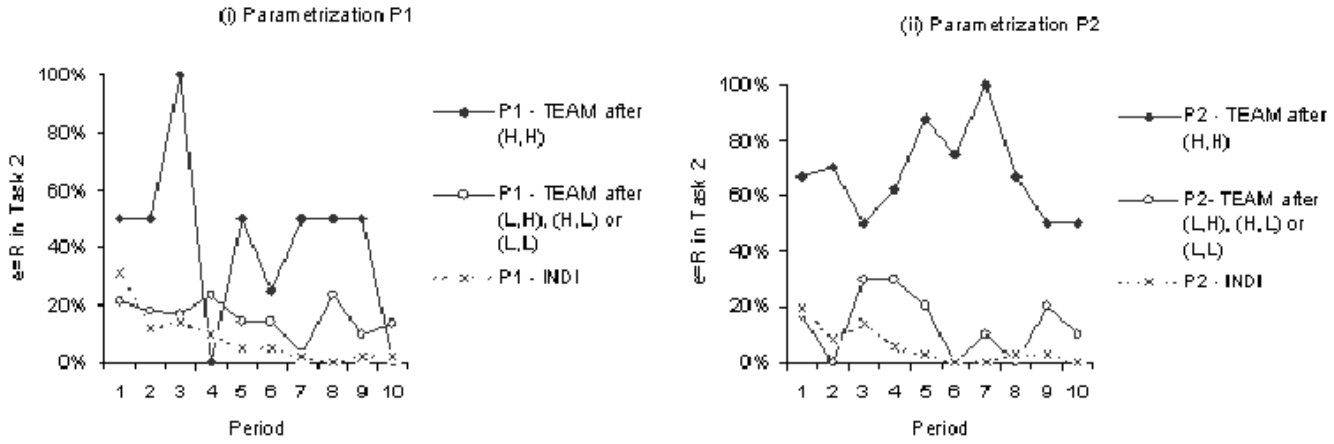

Figure 4: Action $r = R$ in Task 2 for P1 and P2.

Part 1 action		Parametrisation	
Player	Co-player	P1	P2
$e = H$	$e = H$	reference group	
$e = H$	$e = L$	- 1.54 (0.637)***	-3.41 (0.851)***
$e = L$	$e = H$	- 1.63 (0.678)***	- 3.13 (0.885)***
$e = L$	$e = L$	- 1.91 (0.607)***	- 1.98 (0.957)**
Number of observations		190	120
Dropped (no variation)		130	60
Log likelihood		-70.809	-32.249
** [***] indicates significance at the 5% [1%] level.			

Table 4: Conditional Logit model estimates for the probability of $r = R$ in Stage 2 (standard errors between brackets).

Observation 1b (Actions): In INDI, in the far majority of cases, subjects choose $e = H$ and $r = 0$, which is the SPNE. In TEAM, many subjects deviate from the SPNE by choosing $e = H$ or $r = R$. Those who select TEAM are more likely to choose $r = R$ in Task 2 if they observe history ($e = H$, $e = H$) in Task 1 than another history. In TEAM, a subject is more likely to opt for $r = R$ the more reciprocal she is.

3.2 The power of incentives

Is it profitable for the firm to increase the power of the incentive scheme? Table 5 summarizes the performance of INDI and TEAM. It contains the average profits per subject and per round in the self-selected scheme in rounds 1-10 for both parametrisations. In P1 [P2], INDI is significantly more [less] profitable than TEAM for both the firm and the workers. Moreover, in both parametrisations, individuals who self-select into TEAM earn significantly more than the payoff predicted by the Nash equilibrium for this scheme.

	Payoff in SPNE		Average payoff		Mann-Whitney U test
Parametrisation	INDI	TEAM	INDI	TEAM	INDI = TEAM
P1	6	2	6.6 (0.2)	4.6 (0.28)	p = 0.000
P2	6	0	6.3 (0.16)	6.6 (0.39)	p = 0.002

Table 5: Average and Nash equilibrium payoffs per subject per round in the self-selected scheme (rounds 1 to 10) and Mann-Whitney U test (standard errors between brackets).

These observations suggest that the firm can increase output by switching from TEAM

to INDI in parametrisation P1, but not in P2. This is indeed what we find. Figures 5 and 6 display actions chosen in rounds 11 to 20 in Task 1 and Task 2, respectively. They correspond closely to the SPNE.

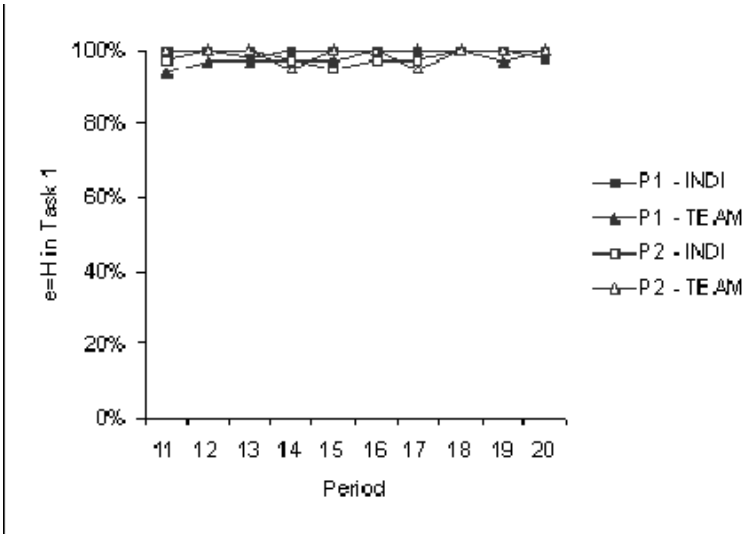

Figure 5: Action $e = H$ in Task 1 for P1 and P2.

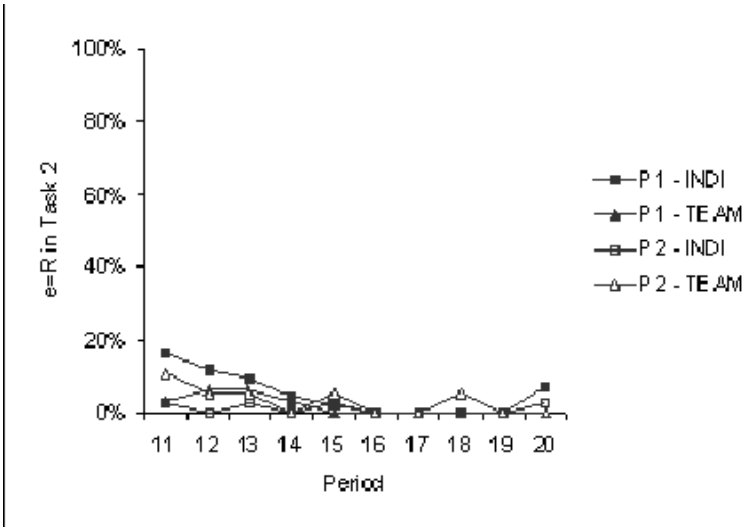

Figure 6: Action $r = R$ in Task 2 for P1 and P2.

Table 6 presents the average payoffs for subjects who chose TEAM in periods 1 to 10 (in which they play TEAM) and periods 11 to 20 (in which they play INDI). The economic performance of the subjects who chose TEAM significantly changes when they are forced to play INDI. However, the change is not unidirectional. In parametrisation P1, we observe a significant increase in output. The opposite is found in parametrisation P2: Performance is worse in INDI than in TEAM. Therefore, increasing the power of the incentive scheme is not profitable in P2. Note that these observations cannot be related to the subject pool composition, because subjects are matched in the same subset of the pool as when they play according to the self-selected scheme.

	Payoff in SPNE		Avg. payoff of those choosing TEAM		Wilcoxon Signed Ranks test
Parametrisation	INDI	TEAM	TEAM (rounds 1-10)	INDI (rounds 11-20)	
P1	6	2	4.6 (0.28)	6.1 (0.11)	p=0.001
P2	6	0	6.6 (0.39)	6.1 (0.18)	p=0.044

Table 6: Average and Nash equilibrium payoffs for 10 rounds in the forced payment scheme TEAM only for subjects who previously self-selected into INDI.

Observation 2 (The power of incentives): Subjects who choose TEAM, when exposed to INDI, improve performance in parametrisation P1, but not in P2, in which performance is worse.

The explanation behind this observation may be that cooperation is more attractive in P2 than in P1. In P2, subjects lose more points ($6 - 0 = 6$) when they fail to initiate

cooperation than in P1 (where they lose $6 - 2 = 4$ points). At the same time, the incentives to free-ride on a cooperator are higher in P1 than in P2 (2 points vs. 1 point). As a result, we expect subjects to be less successful in sustaining cooperation in P1 than in P2. Therefore, in the case of strong (unobservable) interdependency between workers and strong incentives to cooperate under team incentives, switching from team incentives to individual incentives does not improve performance.

3.3 Self-selection

We have observed that in P2, subject who self-select in TEAM perform worse once they are forced to play INDI. Is the success of TEAM explained by the fact that reciprocal subjects choose TEAM in the labour market? We compare the actions in the production game in Figure 7 (Task 1) and Figure 8 (Task 2) with the control group which was forced to play TEAM. We find that in P1, self-selection has no effect ($p = 0.909$ Mann-Whitney U test). However, in parametrisation P2, subjects earn significantly more in the sessions when they sort themselves into TEAM than when we force them to do so ($p = 0.016$ Mann-Whitney U test). More specifically, those who self-select are significantly more likely to choose $e = H$ in Task 1 (Mann-Whitney U test, $p = 0.012$), and $r = R$ in Task 2 (Mann-Whitney U test, $p = 0.005$).

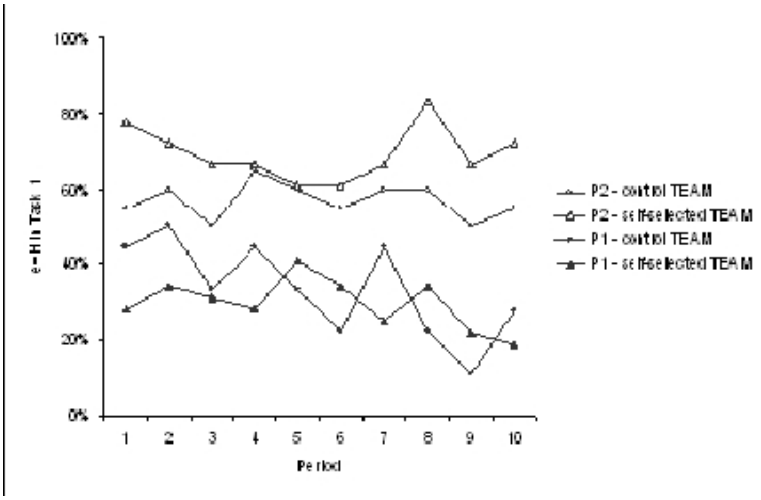

Figure 7: Action $e = H$ in Task 1 for parametrizations P1 and P2 in the self-selected TEAM and in control TEAM sessions.

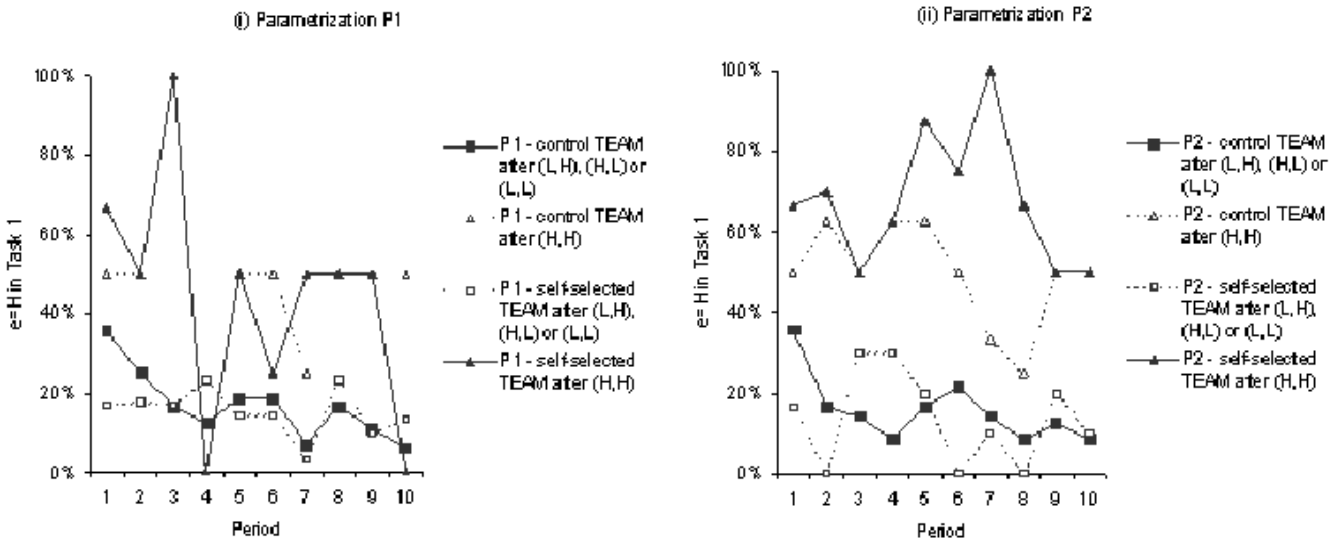

Figure 8: Action $r = R$ in Task 2 for parametrization (i) P1 and (ii) P2 in the self-selected TEAM and in control TEAM sessions.

Observation 3 (Self-selection): The impact of self-selection on subjects' behaviour in TEAM is small in parametrisation P1. In contrast, in P2, subjects in the control treatment (who are forced to play TEAM) perform worse than those who select TEAM in the main design.

Because in P2, performance in TEAM is better if subjects can self-select in this scheme, self-selection contributes to an explanation why TEAM performs better than INDI in P2. In other words, self-selection partly explains why an increase in the power of the incentive scheme may have counterproductive effects.

4 Conclusions

High-powered incentives stand high on the list of policy instruments that should improve the performance of organisations. In this paper, we have addressed the questions (1) which types of workers self-select in firms with low- or high-powered incentive schemes and (2) whether firms perform better if they increase the power of their workers' incentives in situations where workers can self-select. We have answered these questions using a laboratory experiment. We have observed that the more a subject trusts or the more reciprocal she is, the more likely she is to opt for a firm with low-powered incentives. Moreover, production decreased once we confronted subjects who chose low-powered incentives with high-powered incentives if the gains from cooperation are high, from the perspective of both the individual and the team. Self-selection partly explains this finding.

The policy implications of our experiment are as follows. A (public) firm may or may not perform better if the power of workers' incentive schemes is increased, at least in the short run.[10] If the interdependency between workers is high and difficult to observe, and workers' incentive to cooperate is high in the case of a low-powered incentive scheme, a higher-powered incentive scheme may imply worse outcomes. For example, one may think of case managers at employment services, who often have to rely on their mutual expertise to find suitable jobs for their clients. Similar interdependencies may exist between medical doctors and between police officers. In contrast, if the gains from the unobservable interdependencies is low, then introducing high-powered incentives is the preferred choice. Another policy option is to strengthen the advantages of team pay, that is, develop sorting mechanisms to attract reciprocal workers, or strengthen signalling mechanisms within the organisation.

Finally, we wish to note that we used a very cautious design. Our subjects did not have the opportunity to sort themselves into teams endogenously or to build reputation. Free-rider problems could be alleviate if the subjects are able to form groups endogenously on the basis of historical information or repeated interaction. These options remain open for future research.[11]

[10] In the long run, employees who prefer low powered incentives may leave the firm and find jobs that better suit their preferences.

[11] Keser and Montmarquette (2009) is a promising first step.

References

Berg J., J. Dickhaut, and K. McCabe (1995) Trust, Reciprocity, and Social History, *Games and Economic Behavior*, **10**, 122-142.

Bohnet, I., and R. Zeckhauser (2004)Trust, Risk and Betrayal, *Journal of Economic Behavior and Organization*, **55**, 467-484.

Burgess, S., and M. Ratto (2003) The Role of Incentives in the Public Sector: Issues and Evidence, *Oxford Review of Economic Policy*, **19**, 285-300.

Cadsby, C.B., F. Song, and F. Tapon (2007) Sorting and Incentive Effects of Pay for Performance: An Experimental Investigation. *Academy of Management Journal*, **50**, 387-405.

Canton, E. (2005) Power of Incentives in Public Organizations When Employees Are Intrinsically Motivated, *Journal of Institutional and Theoretical Economics*, **4**, 664-680.

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

Delfgaauw, J., and R. Dur (2008) Incentives and Workers' Motivation in the Public Sector, *Economic Journal*, **118**, 171-191.

Dohmen, T., and A. Falk (2006) Performance Pay and Multi-dimensional Sorting: Productivity, Preferences and Gender, IZA Discussion Paper 2001, IZA Bonn.

Eriksson, T. and M.C. Villeval (2008) Performance Pay, Sorting and Social Motivation, *Journal of Economic Behavior and Organization*, **68**: 412-421.

Fehr, E., and Schmidt, K.M. (2004) Fairness and Incentives in a Multi-task Principal-agent Model, *Scandinavian Journal of Economics*, **106**, 453-474.

Fischbacher, U. (2007) z-Tree. Zurich Toolbox for Readymade Economic Experiments, *Experimental Economics*, **10**, 171-178.

Frey, B., and Jegen, R. (2001) Motivation Crowding Theory: A Survey of Empirical Evidence, *Journal of Economic Surveys*, **15**, 589-611.

Gneezy, U., and Rustichini, A. (2000) Pay Enough or Don't Pay at All, *Quarterly Journal of Economics*, **115**, 791-810.

Hamilton, B, J. Nickerson, and H. Owan (2003) Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation, *Journal of Political Economy*, **111**, 465-497.

Kandel, E., and Lazear, E.P. (1992) Peer Pressure and Partnerships, *Journal of Political Economy*, **100**, 801-817.

Keser, C., and Montmarquette, C. (2009) Voluntary versus Enforced Team Effort, working paper, Georg-August-Universität Göttingen.

Knez, M., and Simester, D. (2001) Firm-wide Incentives and Mutual Monitoring at Continental Airlines, *Journal of Labor Economics*, **19**, 743-772.

Lavy, V. (2002) Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement, *Journal of Political Economy*, **110**, 1286-1317.

Lazear, E.P. (2000) Performance Pay and Productivity, *American Economic Review*,

90, 1346-1361.

Minkler, L. (2004) Shirking and Motivations in Firms: Survey Evidence on Worker Attitudes, *International Journal of Industrial Organization*, **22**, 863-884.

Nalbantian, H.R., and Schotter, A. (1997) Productivity under Team Incentives: An Experimental Study, *American Economic Review*, **87**, 314-341.

Schechter, L. (2007) Traditional Trust Measurement and the Risk Confound: An Experiment in Rural Paraguay, *Journal of Economic Behavior and Organization*, **62**, 272-292.

Vyrastekova, J., and Garikipati, S. (2005) Beliefs and Trust: An Experiment, CentER Discussion Paper, Tilburg University.

Vyrastekova, J., and Onderstal, S. (2010) The Trust Game Behind the Veil of Ignorance: A Note on Gender Differences, Tinbergen Institute Discussion Paper.