Team production benefits from a permanent fear of exclusion

Kopányi-Peuker, A.; Offerman, T.; Sloof, R.

TI 2015-067/VII
Tinbergen Institute Discussion Paper

Team Production benefits from a Permanent Fear of Exclusion

Anita Kopányi-Peuker
Theo Offerman
Randolph Sloof

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

Team production benefits from a permanent fear of exclusion

Anita Kopányi-Peuker, Theo Offerman and Randolph Sloof*

May 27, 2015

Abstract

One acclaimed role of managers is to monitor workers in team production processes and discipline them through the threat of terminating them from the team (Alchian and Demsetz, 1972). We extend a standard weakest link experiment with a manager that can decide to replace some of her team members at a cost. The amount of contractual commitment ('termination possibilities') and the precision of the manager's monitoring information serve as treatment variables. Our results show that the fear of exclusion has a profound effect on team performance even if workers are imperfectly monitored; the most flexible contract induces the highest output while the one with no firing possibilities leads to the lowest production. However, once the fear is eliminated for some workers, because permanent workers cannot be fired after a probation phase, effort levels steadily decrease.

JEL classification: C72, C92, M51, M55
Keywords: team-production, weakest-link game, exclusion, probation, experiment.
Acknowledgements: We thank for suggestions of audiences at 2013 Florence Workshop on Behavioral and Experimental Economics, at the CCC-meeting 2013 (Amsterdam), at the 5th Anniversary of Thurgau Experimental Economics Meeting (Stein am Rhein, Switzerland), at the 7th Maastricht Behavioral and Experimental Economics Symposium, and at the European ESA meeting (2014) in Prague. Financial support of NWO is gratefully acknowledged.

*CREED, University of Amsterdam and Tinbergen Institute, Roetersstraat 11, 1018WB Amsterdam, The Netherlands. A.G.Kopanyi-Peuker@uva.nl; T.J.S.Offerman@uva.nl; R.Sloof@uva.nl

1 Introduction

In practice many team-production processes have weakest link characteristics. Examples include the construction of a new building, an operation performed by a surgical team and the preparation of an airplane for take-off. In the latter case, the plane cannot take off before the baggage is loaded, the catering has replenished the pantry, the crew has arrived, all passengers are seated and the plane is refueled and checked. The slowest component of these will determine when the plane is ready to take off. The teams engaging in a task with weakest-link aspects are usually quite successful in real life. Buildings rarely collapse, patients who are operated usually do not die and most planes are ready for take-off according to schedule. This picture contrasts sharply with the results from minimum-effort games that are used to model team production with weakest link characteristics in the laboratory. In the experiments, team-production typically fails unless the team consists of very few members.[1]

In this paper, we design a series of experiments to pursue two goals. First, we test if the fear of exclusion can account for (part of) the difference in team performance in the laboratory and the field.[2] In contrast to the minimum-effort experiments where the fear of exclusion is absent, team members in the field face the fear of being fired if their performance slackens. Thus, we address a key hypothesis formulated by Alchian and Demsetz (1972) who argue that an important reason for why firms need managers is that efficient team production is facilitated if someone specializes in monitoring workers and excludes those whose performance falls behind (see also Jensen and Meckling, 1976). Second, we investigate if the fear of exclusion needs to be permanently maintained in labor relations. In some contracts workers are effectively protected from firing after they have survived a probation phase. The fear of exclusion may encourage workers to perform well in their probation phase. In fact, it may even facilitate coordination on the efficient equilibrium in the minimum-effort game. An open question that we address is if workers continue to perform well after the probation phase has ended.

In agreement with the labor applications that motivate our research, we include a manager in the minimum-effort game. The manager monitors a team of six workers and benefits from the production in the same way as workers do, but she does not participate herself in the production process. Instead, the manager has the possibility to replace some workers in her team (with the fired workers becoming unemployed). In the experiments we vary two aspects of the game: (i) the extent to which workers are protected by contracts, that is, they can be fired every round, never, or they can only be fired during their probation

[1] Van Huyck et al. (1990) first studied the minimum-effort game and showed that high effort levels were only sustainable with a fixed group of 2, but not with random pairs, and neither with groups of 14-16 members. Subsequent research confirmed these earlier findings and showed that groups converged to the worst equilibrium unless they consisted of only 2 or 3 members (e.g. Knez and Camerer, 1994; Chaudhuri et al., 2001; Weber, 2006). Devetag and Ortmann (2007) provide a survey.

[2] Obviously, other (complementary) explanations for this difference may exist as well, including strong social norms and peer pressure. In this paper we focus on the fear of exclusion.

phase, and (ii) how well the manager is informed of the workers' performance. Regarding the latter variable, workers' effort levels are either perturbed by noise, or not. This allows us to explore the robustness of the results. If noise is involved in the production process, managers cannot trace back the exact effort workers exert in order to have a high output. This feature may weaken a potential positive effect of the possibility to exclude team members. For instance, with noise there is a danger that managers judge workers too quickly, and do not sufficiently account for the possibility that a worker is affected by bad luck.[3] More generally, Alchian and Demsetz (1972, p. 786) conjecture that "...the cost of managing team inputs increases if the productivity of a team member is difficult to correlate with his behavior." Our treatment variation in noise allows us to explore the similar notion that the value-added of having a monitoring manager decreases when monitoring information becomes less precise.

In the experiments, we find strong and significant support for the hypothesis that the fear of exclusion efficiently encourages workers to perform well. When managers have perfect discretion, workers tend to anticipate from the start that they cannot afford to slacken. This contrasts strongly to the case where workers are protected by long-term contracts. There, workers' performance gradually deteriorates over the experiment. Interestingly, the result carries over to the case where workers' productivity is affected by good or bad luck. Also here, workers perform much better when the manager has the power to discipline workers. When firing is allowed, managers need to use their stick. The mere threat of being fired is not enough for workers to behave well.

Our results for the probation treatments suggest that the fear of exclusion needs to be a permanent fixture in the labor market. Workers start with high effort levels in their probation phase. Once some of the workers become permanently employed and cannot be fired any longer, the team's performance deteriorates. Especially novice permanent workers substantially reduce their effort upon becoming permanent, a finding similar in spirit to the Peter principle (cf. Lazear, 2004). Overall, workers do not perform significantly better when they receive probation contracts than when they are completely protected by long-term contracts.[4]

Our paper contributes to three stands of literature. First of all, some studies investigated other ways to improve efficiency in the minimum-effort game. Weber (2006) let small groups play the minimum effort game, and then added new group members. If they were aware of the group's previous performance, newcomers often adhered to the norm already existing in the smaller group. More recently Salmon and Weber (2011) investigated how adding lower-performing

[3]In the context of a financial asset market, Gneezy and Potters (1997) find that agents perform worse when they are offered the possibility to evaluate their portfolio continuously.

[4]Ichino and Riphahn (2005) provide empirical field evidence about the effect of probation on worker behavior. They measure absenteeism in a large Italian bank both during and after probation, and find significantly higher absence rates once workers are fully protected. To interpret their data they rely on a standard principal − single agent framework in which workers (ceteris paribus) benefit from exerting less effort. In our team production setting, workers benefit from higher output and coordinating their effort, so even if the fear of exclusion is lifted they may in principle still want to exert the same amount of effort.

members to a high-performing group affects performance. Without restrictions on entering the high-performing group, high performance could not be maintained after growth. However, if restrictions were introduced (e.g. only one person could enter in a round), efficiency was preserved in larger groups, too.

Secondly, some studies investigated the possibility of exclusion from a team or endogenous group formation in public good games (rather than minimum effort games). Cinyabuguma et al. (2005) and Maier-Rigaud et al. (2010) started with a larger group and allowed group members to vote to exclude their fellow group members for the remainder of the given part (or for a subsequent game as in Masclet, 2003). Charness and Yang (2010), Ahn et al. (2008) and Page et al. (2005) had smaller groups and let them endogenously decide about group-formation, viz. by forced or voluntary exit and mergers in the first two, and by ranking others in the latter. Both mechanisms helped group members to contribute higher amounts than in the baseline treatment where no decision could be made about group members. Güth et al. (2007) introduced a leader who could exclude players from the public good game. In contrast to our setting, this leader was a group member who contributed before the others did. Güth et al. showed that leaders increase contributions in the public good game, especially when they were endowed with the power to exclude.

A third line of studies investigated the possibility of exclusion or endogenous group formation in the minimum-effort game. In the design of Croson et al. (2015), exclusion occurred automatically, and the player(s) contributing the least was (were) excluded from the team-production. Simple exclusion did not significantly raise contributions, but when exclusion was paired with redistribution of the excluded players' contribution, contributions significantly increased compared to the baseline treatment. Riedl et al. (2011) studied coordination in large groups in a weakest-link network game. In their experiment subjects chose with whom they wanted to interact. An interaction only took place with mutual consent. The more people a subject interacted with, the higher the potential benefits were but also the higher the strategic uncertainty. Riedl et al. show that endogenous group formation helps players to coordinate more efficiently.

To the best of our knowledge, we are the first to investigate the fear for a manager's power to fire in a labor market context. Another novel feature of our paper is that we investigate if the disciplining effect of the fear of exclusion is eroded when managers observe workers' performance only imperfectly. Most importantly, perhaps, none of the previous experimental studies addresses the question what happens if the danger of exclusion disappears when workers have survived the probation phase.

The remainder of the paper is organized as follows. Section 2 introduces our game. Section 3 describes the experimental design. Section 4 presents the experimental results, and Section 5 concludes.

2 The game

We consider a team-production setting with 6 team members (workers) who play the same game repeatedly. Workers simultaneously choose an effort level (e_i - integer between 1 and 9), which determines their productivity (p_i). If there is no noise in the game, workers' productivity equals their chosen effort level. With noise, productivities are the perturbed effort choices. For each worker, there is an independent random noise term which together with the effort choice determines his productivity: $p_i = e_i + \varepsilon_i$. Productivities are always in the ± 2 range of the chosen effort (and between 1 and 9). If the chosen effort level e_i is between 3 and 7, the productivity is symmetric around e_i; p_i equals the chosen effort with 50% probability, the chosen effort ± 1 both with 22.5% probability and the chosen effort ± 2 with 2.5% probability each. Since productivities cannot be higher than 9 or lower than 1, on the edges the probability distribution is not symmetric around the chosen effort level. The probability mass which would fall outside the feasible range is shifted to the nearest possible productivity (so either to 1 or to 9).[5]

The minimum of the individual productivities determines the output of the firm, with the restriction that output is never higher than 7. We imposed this restriction to allow workers to reach the highest possible output with probability 1 even in the noise case (that is, if everybody chooses an effort level of 9).[6] The production function is thus given by:

$$Q = \min\{p_1, ..., p_6, 7\}$$

Effort is costly for the workers with marginal costs equal to 10. Naturally, workers also benefit from higher output. In particular, a worker's payoff from the production process is the following:

$$\pi_i = 20 \cdot Q - 10 \cdot e_i + 50$$

Besides the workers, there is a manager who also benefits from production. The manager does not choose an effort level, but benefits from the output in the same way as workers do without bearing the costs of effort. In some of our treatments, the manager can decide to replace at most 3 workers by unemployed people. If she decides doing so, she bears a firing cost per fired worker. The manager's payoff is given by:

$$\pi_m = 20 \cdot Q + 50 - 20 \cdot n_f$$

where n_f is the number of fired workers. In some instances, the manager's firing ability is limited. If the manager cannot fire any worker, $n_f = 0$ automatically in the payoff function above.

[5] For example: if the chosen effort level is 2, then the productivity is 1 with 25% probability, 2 with 50%, 3 with 22.5%, and 4 with 2.5%.

[6] Note that noise can cause severe decrease in output. If everybody chooses the same effort level, then the chance that the output will be two units lower than the effort levels is almost 15% (since the probability of at least 1 worker having a noise term equal to -2 is $1 - (1 - 0.025)^6 \approx 0.141$). The output will be one unit lower with 68.1% probability.

Table 1: Overview of the treatments

Treatment	Noise	Firing	# of subjects per group	# of groups
FS	no	every round	10	6
NS	yes	every round	10	6
FL	no	not possible	7	6
NL	yes	not possible	7	6
FP	no	during probation	10	6
NP	yes	during probation	10	6
FM	no	every 3^{rd} round	10	6
NM	yes	every 3^{rd} round	10	6

Unemployed people do not participate in the production process: they neither choose an effort level, nor benefit from production. Instead, they receive an unemployment benefit of 30 and are inactive. They can only get hired if the manager decides to fire somebody. If a worker is fired, he becomes unemployed.

We consider three types of contract in the above-mentioned team-production game. Under Spot contract, managers can replace workers in every round. Under Longterm contract, workers are protected by contracts and the manager cannot replace anybody. Under Probation contract, workers are on probation in the first 5 rounds of their working phase, and they can be fired only during these rounds. After that the manager cannot fire them any more. We consider two production processes: one with Noise and the other without noise (Full information). This leads to six different treatments: FS, NS, FL, NL, FP and NP, where F (N) stands for Full information (Noise), and S, L, P stands for Spot, Longterm, Probation, respectively. An overview of the treatments is presented in Table 1.[7]

Managers are better informed than other labor market participants. When workers have chosen their effort levels, managers observe both the team's output and the individual productivities. However, other subjects can only observe the output of the team. Thus, in treatments where the manager can fire workers, she can condition her choice on workers' productivity levels. Note that without noise, productivities simply equal the chosen effort levels.

Next, we informally discuss the equilibria of our game. We do not intend to provide a comprehensive equilibrium analysis; the upshot of the discussion will be that, given the multiplicity of equilibria, standard game theoretic arguments provide little guidance which treatment effects to expect.

First consider the single round weakest link (stage) game without a manager. With Full information all effort profiles in which workers choose the same effort level but not higher than 7 can be supported as (symmetric) Nash equilibrium, as in a standard minimum-effort game. Equilibrium refinements may potentially restrict this set of (symmetric) equilibria. Such refinements are often based on (Pareto) efficiency and risk considerations. The equilibrium in which

[7]Besides these 6 treatments, we considered also two treatments with a Medium contract length. In these treatments, managers can only fire in every third round. This contract is discussed in Appendix C.

every worker chooses 7 is payoff dominant, but also contains much strategic uncertainty. Harsanyi and Selten (1988)'s notion of risk dominance cannot be readily generalized to games with more than two players. However, as argued by Goeree and Holt (2005), the related concept of maximizing the "potential" of the game (cf. Monderer and Shapley, 1996) can be applied and selects the secure equilibrium in which all workers choose minimum effort. This is in line with the empirical observation that in experiments with larger groups (of 5 and above), workers rapidly converge to the worst possible equilibrium. When noise is introduced in the stage game, only the secure Nash equilibrium remains.[8]

In our experiment subjects play the stage game repeatedly with an indefinite end. Still ignoring the manager for the moment, under Full information then again any symmetric effort profile with effort levels below 8 constitutes an equilibrium. Under Noise all symmetric effort profiles except effort levels 2 and 3 can be supported for a sufficiently high discount factor / continuation probability δ using trigger like strategies, where workers revert to minimum effort forever after any detectable deviation (see Appendix A.2 for a calculation of the corresponding threshold discount factors). The intuition here is that for effort levels of 2 and 3 downward deviations cannot be identified from bad luck, so a trigger strategy that only punishes clear cut deviations does not exist.

We finally consider equilibria in which the manager plays an active role. The manager has an interest to keep up high output and thus that the workers coordinate on a high effort level. She may naturally pursue this with a threshold strategy and fire workers whose productivity levels fall below the threshold. As discussed in Appendix A.3, any threshold Q^* above minimum effort can be supported as equilibrium under Full information. Firing then does not occur on the equilibrium path. The picture is different when the manager observes the effort levels of the workers imperfectly. With Noise, firing may even occur on an equilibrium path supporting high effort. The occasional firing of unlucky workers may serve the purpose of clearly communicating the manager's intolerance of slacking performance.[9] Yet again multiple thresholds can be supported.

Due to the multiplicity of equilibria standard theory does not make clear cut predictions about the impact of our treatment variations. However, our experimental design allows us to explore some qualitative arguments. With full information and the power to exclude, the manager's stick may help to reduce the strategic uncertainty and allow workers to coordinate on the efficient equilibrium. One thus would expect that, under Full information, Spot contracting works better in sustaining high team output than Longterm contracting does (where the threat of exclusion is absent). With Full information there will be no need to fire after workers have learned to anticipate the manager's threshold (firing) strategy. The question is if the manager's stick remains equally pow-

[8]This can be shown by calculating the expected loss and gain from deviating to a lower effort level. From any level exceeding 1, this results in an increase in expected profit. On the other hand, unilaterally deviating to a higher level when others choose 1 deteriorates expected payoffs. The calculations are presented in Appendix A.1.

[9]Similar results are reported in the cartel literature (see Green and Porter, 1984) where cartels may punish their members in case of bad luck.

erful if she observes effort levels imperfectly. She then may need to continue firing workers every now and then, to avoid that slackening workers can pass themselves off as hard workers just being unlucky. At the same time, workers may be demotivated if the manager does not appropriately take the imprecision into account and fires workers too often. In addition, our experimental design allows us to investigate if workers are disciplined after they are fired. That is, if they get a new chance to work, do they enhance their effort compared to before they were fired?

A further important question is if workers continue to perform well after the manager has given up her stick. Under Probation, workers can potentially learn to choose high effort levels, which they might be able to maintain once their probation phase is over. The manager's firing strategy may endogenously create overlapping generations in the group, with some workers still on probation whereas others cannot be fired any more. Those who are on probation might then work hard in order to avoid firing, which can trigger permanent workers also to work hard. Probation may then not only serve to motivate the worker in question, but also indirectly his permanent team mates.

3 Experimental design and procedures

The experiment was conducted in the CREED lab at the University of Amsterdam. In total, 444 subjects participated in overall 20 sessions. We collected data for 6 independent groups for each treatment. Every subject participated in at most one session. Participants were mainly undergraduate students from different fields (e.g. economics, law, psychology). The average session lasted about 75-90 minutes, and subjects earned on average 20.5 euros. During the experiment, subjects earned points which were converted to euros at the end of the experiment. Participants received 1 eurocent for each 1.6 points. Earnings were paid privately in cash at the end of the experiment. The experiment was computerized, and programmed in php. Instructions were given on computer screens, and subjects' questions were answered privately.

At the start of the experiment, subjects were assigned to a group that consisted of 10, or in the Longterm treatments, 7 people. In an average session, two or three groups were formed at the same time. Subjects did not know which other subjects were assigned to their group, but they knew that the group remained constant throughout the experiment. In the first part of the experiment, subjects were informed that the advantageous role of the manager was assigned on the basis of how well they performed in the secretary problem (Seale and Rapoport, 1997). Subjects had to hire a "secretary" out of the 25 possible (imaginary) applicants. Each applicant had an integer quality independently drawn from U[0,420]. Subjects did not know the range of qualities, and once they rejected an applicant, they could not reconsider this decision. If a subject chose to reject the first 24 applicants, he had to hire the last one. In each group, the subject who hired the best applicant became the manager and kept this role throughout the whole experiment.[10]

In the second part of the experiment subjects played the modified minimum-effort game presented in Section 2. They were informed that there would be at least 25, and at most 40 rounds of the game. In practice they played 30 rounds of the game.

In each session only one of the eight treatments discussed in the previous section was played (see Table 1). Because in the Longterm treatments firing was impossible we did not have unemployed subjects there. We decided to keep the managers in the Longterm treatments. They were inactive but otherwise earned a payoff in the same way as in the treatments where firing was allowed. We kept the manager to make the treatments with and without firing more similar.

During the experiment, each worker had a fixed worker ID while employed and their productivities was communicated to the manager by these IDs. Then, if the manager was allowed to fire workers, the manager decided who to replace by using this worker ID. Once a worker was fired, he lost his ID. When he was rehired, he got a new ID which might have been different from the old one. By doing so, each worker could have a fresh start after being rehired. Furthermore, managers could not decide who to hire from the unemployed subject pool. To those who started the experiment in the role of being unemployed we paid a bonus by a lottery. With 50% probability the subject received an additional bonus of 120 points over the unemployment benefit (in all initial unemployment rounds). We decided to implement this bonus to remove the manager's incentive to fire just to make sure everybody got the opportunity to work.

In each treatment a history screen was continuously available for every subject. This screen contained information about past production in the own labor market. For each previous round, managers could observe workers' productivities, the output, and their own firing decision. Workers' productivities were displayed by different colors in the history screen to make the history better tractable. Workers and unemployed subjects could see their own productivity level, the output, and the manager's firing decision. Examples of the history screen and instructions for the second part for the Spot contract treatment with Noise (NS) can be found in Appendix B.

4 Results

In Section 4.1 we focus on worker behavior by comparing effort decisions across treatments. In Section 4.2 we study managers' firing decisions and how these drive workers' effort choices. Finally, Section 4.3 deals with efficiency. All non-parametric tests (unless otherwise stated) are carried out at the matching group level, with two-sided ranksum tests for treatment differences, and Wilcoxon signed-rank tests for differences over time.

[10]We only used the secretary problem to assign the role of manager on the basis of performance. In the data analysis we ignore these data and we focus on the main game of the experiment.

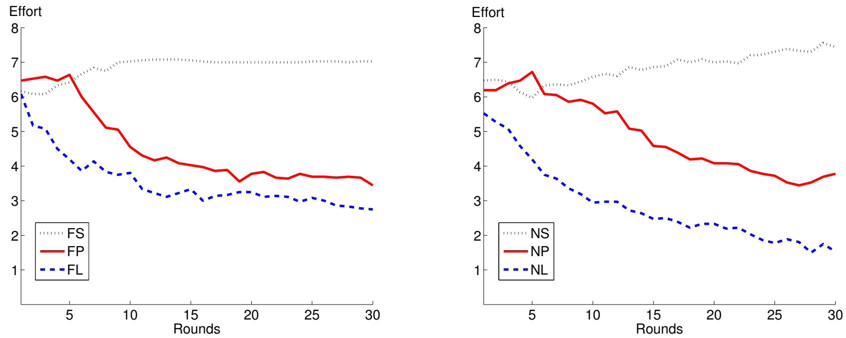


Figure 1: Average effort over time under full information (left panel) and under noise (right panel)

Table 2: Average effort level

Treatment	First half (rounds 1-15)	Second half (rounds 16-30)	First vs. second half
FS	6.71	7.01	0.03**
NS	6.49	7.19	0.03**
FL	4.04	3.03	0.05**
NL	3.69	2.02	0.03**
FP	5.32	3.72	0.09*
NP	5.83	3.93	0.05**
FS vs. FL	0.04**	0.01***	
FS vs. FP	0.26	0.10*	
FL vs. FP	0.23	0.69	
NS vs. NL	0.01**	0.00***	
NS vs. NP	0.23	0.01***	
NL vs. NP	0.02**	0.15	

Notes: ***: significant at 1% level, **: significant at 5% level according to two-sided ranksum test with $n = 6$ for the treatment differences, and Wilcoxon-test for the differences over time.

4.1 Effort decisions

Figure 1 shows the average effort levels over time for the different treatments. In the Longterm treatments, the average effort decreases over time just like in previous experiments with a standard minimum-effort game. The possibility to fire offered in the Spot treatments profoundly improves the team performance. With Spot contracts, average effort increases over time, and is substantially higher than in the Longterm contracts. The Probation contract increases effort level a bit compared to the Longterm contract, but to a much lesser extent than the Spot contract.

Table 2 reports the extent to which the treatments differ systematically. Effort in the Spot contracts is always significantly higher than in the Longterm contracts. The Spot contract also significantly outperforms the Probation contract in the second half of the experiment. Effort levels in the Probation contracts are not significantly different from the effort levels in the Longterm contract (except for the Noise case in the first half of the experiment).[11]

There are differences not only in average effort choices across treatments, but also in the extent to which very high effort levels of 8 or 9 are chosen. In the second part of the FS treatment, in 1% of the cases subjects choose an effort level of 8, and they never choose 9. In contrast, these numbers are 36% and 2% in NS, respectively. Note however, that in case of Noise, the most efficient equilibrium in the one-shot game is that all workers choose an effort level of 8. Thus, in this treatment a substantial fraction of workers coordinates on the most efficient equilibrium. The difference in the percentage of subjects choosing the effort level 8 in FS and NS is significant at the 1% level (p=0.006). Additionally, subjects choose significantly less often 8 (4%) and 9 (0.2%) under NP than under NS in the second half of the experiment (p=0.01 and p=0.06, respectively). There are no further differences in choosing these high effort levels between the other treatments.

We also compared first round effort levels across different contract types, keeping the production process fixed. With ranksum tests on individuals we found that the first round effort levels in the NL treatment are significantly lower than those in the NS treatment ($p = 0.00$). There are no further differences across treatments. These findings show that, with imperfect monitoring, workers are already affected in the first period by the threat of firing under Spot contract, whereas under the other contracts they only start to react on firing if it indeed happens in their group.

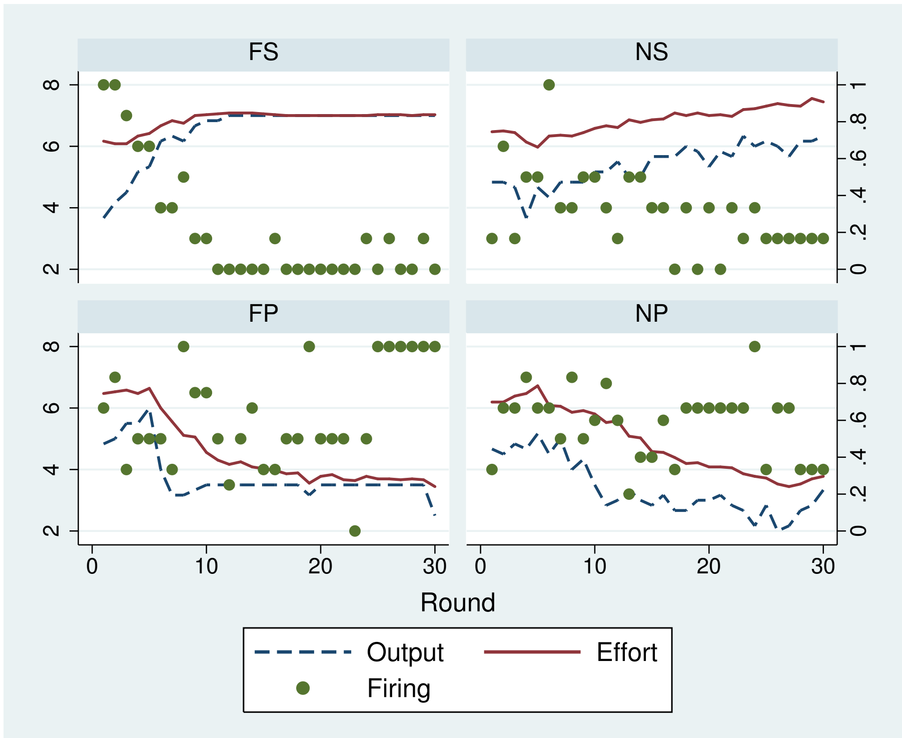
During the production process, effort levels translate into productivities and output realizes. In the experiment, output has about the same trend as effort levels have. In the second half of the experiment, output is significantly higher at the 1%-level in the Spot treatments than in the Longterm and Probation treatments (between the latter two there is no significant difference). Thus always having the firing possibility not only increased workers' effort levels, but also the production of the team.

To sum up, effort levels are significantly higher if workers face the threat of being fired. However, once the threat of being fired ceases under the Probation contract, workers' performance deteriorates and their efforts are not significantly higher than under the Longterm contract.

4.2 Firing: causes and consequences

In the experiment, managers regularly fire bad-performing workers if they have the power to do so. Figure 2 shows the firing decisions for the Spot and Probation contracts over time. Output and effort levels are measured on the left axis, while the frequency of firing in a given round is measured on the right axis. The figure shows that firing decisions differ substantially between treatments. All managers in FS fire at least one worker in the early rounds. This results in a

[11] For a given contract type, the difference in effort choice across Full information and Noise is never significant ($p > 0.30$ for all cases).



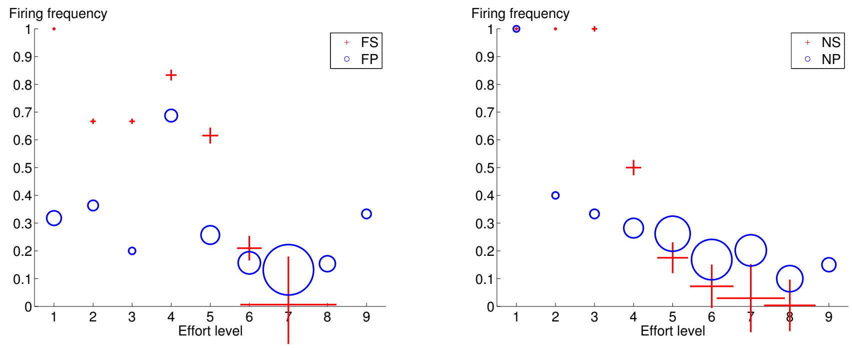
Notes: The left axis presents the output and effort levels while the right axis shows the frequency of firing. Here we only consider whether firing happens in a round, not the exact number of workers who got fired. In case of Probation the frequency of firing is measured over those cases when firing was still possible, that is, there were still workers on probation.

Figure 2: Firing decisions, effort and output over time in the different treatments

steadily increasing average effort level and output, which diminishes the need to fire. After round 10 hardly any firing occurs. Overall managers fire in 21% of all the possibilities.

In treatment NS, even though early outputs are not at the maximum, managers initially make less use of the opportunity to fire. This results in a less steep increase in the effort choices and output. Nevertheless, over all rounds firing occurs somewhat more frequently when there is noise in the productivity levels: in 32% of all possibilities a worker is fired. In both treatments firing decreases significantly from the first half to the second half ($p_{FS} = 0.03$, $p_{NS} = 0.05$). In the second half of the experiment, managers in treatment FS fire significantly less often than managers in treatment NS ($p = 0.02$). This finding is in line with the intuition that occasional firing is needed with imperfect monitoring, even when worker behavior has more or less stabilized.

Managers are far less effective in using the firing tool under the Probation contract than under the Spot contract. As will be discussed in more detail below, team performance especially deteriorates if some workers get a perma-



Notes: The size of the markers is proportional to the number of observations in each case. Under Probation the frequency of firing is measured over those cases when firing was still possible for that given effort level.

Figure 3: Firing decisions as a function of the effort

nent status. In FP, managers fire in 57% of the actual firing possibilities (i.e. when there are still workers on probation). However, by giving workers permanent status managers lose about 50% of the firing possibilities. These firing possibilities are reduced to a lesser extent in NP (27%) where managers keep their workers on probation more often. Here, managers fire at approximately the same rate as in FP (58%) if they have the possibility to do so. In both treatments firing decreases from the first to the second half of the experiment.

We next investigate the factors that make a manager decide to fire a worker. Not surprisingly effort levels are highly correlated with the firing decision. Figure 3 shows how effort levels relate to the probability of firing in the different treatments. The higher the effort level is, the less likely it becomes that the worker is fired. Most managers fire workers with the lowest productivity in the team. In treatment FS, only 2 managers make one "mistake" by firing someone with a high effort level when the output is low. In treatment NS, only 1 manager makes the same mistake once. Managers' strategies under Spot contracting are broadly consistent with a threshold strategy, which means that they only fire if the output is below some threshold. In the Full information case the average threshold level is between 5 and 6, while in the Noise case the threshold is between 4 and 5.[12] So managers tend to fire in FS when the output is below 6 and in NS when it is below 5. The difference in managers' threshold levels is significant between the two treatments (p=0.025). Thus, managers are more lenient when workers' productivities are distorted by noise. They then take into account that workers' productivities might be driven downwards by bad luck. Managers' threshold levels appear to be quite stable over time.

The case is not that simple under the Probation contract, because there sometimes managers simply do not have the possibility to fire the worst per-

[12] We determined the threshold levels by checking the number of mistakes managers make if they fire according to a given threshold level. A manager's threshold is defined as the level that produces the fewest mistakes.

Table 3: Regression results for individual effort as dependent variable

Dependent variable: Effort	Longterm	Spot	Probation
round	-0.05 (0.00)***	0.01 (0.00)***	-0.02 (0.00)***
lagged output	0.74 (0.03)***	0.36 (0.02)***	0.50 (0.02)***
noise	0.01 (0.16)	0.36 (0.08)***	0.78 (0.17)***
firing in previous round	-	0.39 (0.06)***	0.66 (0.12)***
firing in previous round*noise	-	-0.13 (0.06)**	-0.44 (0.15)***
newly hired	-	-0.01 (0.06)	-0.27 (0.10)***
being permanent	-	-	-1.26 (0.12)***
permanent co-worker in group	-	-	0.04 (0.11)
constant	2.17 (0.16)***	4.19 (0.12)***	3.40 (0.19)***
Number of panels	72	108	104
Avg # of observations per panel	29	19.3	12.5

Notes: ***: significant at the 1% level, and **: significant at the 5% level. Std errors are in parentheses after the coefficients. Panels are the individuals. Due to firing, the sample is unbalanced. Under Probation we exclude observations where firing is not possible anymore.

forming worker (in total 30 and 27 workers out of 54 got permanent status in FP and NP, respectively). Moreover, they may want to fire well performing workers just to avoid that they become permanent and cannot be fired anymore. Some managers indeed follow this strategy of keeping at least part of their workforce purposefully on probation.[13] The behavior of the other managers is broadly consistent with a threshold strategy, in which either a fixed threshold is used as under the Spot contract, or a "dynamic" one that starts as a fixed threshold but is adapted downwards to the productivity level of the worst-performing *permanent* worker once there are permanent workers in place who exert less effort than the original fixed threshold. A potential rationale for the latter strategy is that, if permanent workers are largely unaffected by firing of others and constitute the weakest link, costly firing of those still on probation serves little purpose.

We next turn to the question of how firing drives workers' effort choices. Table 3 presents the results of random effects panel regressions with individual effort levels as the dependent variable. We study each of the three contracts in a separate regression. The Longterm contract, under which firing is not possible, is included as benchmark. Independent variables here are a time trend 'round', team output in the previous round, and a dummy indicating whether monitoring is noisy or not (1: Noise, 0: Full information). In line with Figure 1 we observe that effort levels significantly decrease over time. Moreover, we do not find differences in effort levels under Full information and Noise.

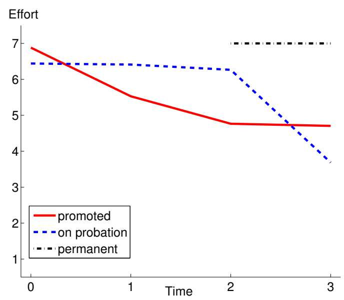
In the regression for the Spot contract we include some additional independent variables: a dummy indicating whether there was any firing in the previous round, the interaction of this dummy with the noise dummy, and a dummy equal to one only if the worker is newly hired in that particular round. Again in line

[13]Under Full information there is one manager who successfully keeps everybody on probation. Under Noise, there are two managers keeping an eye on workers' probation phase. A third manager has 5 permanent workers already in round 8 and keeps firing the remaining worker (however this is not efficient).

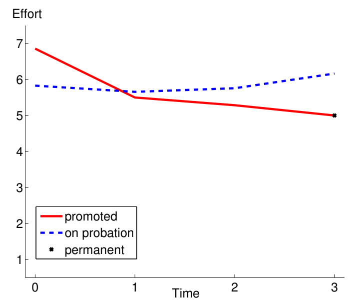
with Figure 1 effort levels now significantly increase over time. The explanatory variable of particular interest is “firing in the previous round”. The dismissal of another worker appears to have a significantly positive effect on a worker’s effort choice. The interaction term reveals that this effect is significantly smaller for Noise than for Full information. As noted in Section 2, with noise firing may occur on an equilibrium path supporting high effort. The significantly negative interaction term is thus in line with the intuition that, when monitoring is imperfect and low productivity can also be due to bad luck, occasionally firing may be needed just to avoid that workers slacken off. Furthermore, we find that newly hired workers do not choose different effort levels than those already in the team. This suggests that firing has a true disciplining effect on the current work force and that the increase in average effort is not mainly due to sorting (i.e. above average workers entering and replacing low performers in the team). Finally, we observe that workers choose significantly higher effort levels when there is noise in the productivity levels than when there is not (this also holds true for the Probation contracts, see Table 3). Workers thus try to compensate the effect of noise in their productivity by choosing higher effort levels.

The final column in Table 3 concerns the Probation contracts. Here we include two additional dummies: one indicating whether or not the worker is permanent himself, and another one indicating whether there is at least one permanent co-worker in the group. Being a permanent worker oneself significantly decreases effort levels, while having a permanent co-worker in the group does not significantly changes effort levels. The effect of being permanent oneself is a much stronger force. Under Probation contract newly hired workers choose significantly lower effort levels than those already in the team. This might be the case because of the overall decrease in effort levels. The finding is again in line with a true disciplinary effect of firing, rather than sorting.

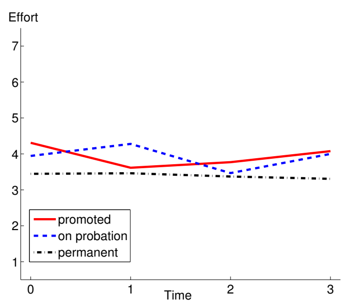
The regression results for the probation contract suggest there is a ‘structural break’ directly after a worker gets tenure. Indeed, workers who survive their probation phase immediately reduce their effort level on average by 1.1 units in FP and by 1.8 units in NP in the first round after being permanent compared to the last round of probation. This decrease is significant at the 5%-level for both Full information and Noise. The decrease in the now permanent workers’ effort levels also induces a smaller decrease in effort levels of workers still on probation. The latter choose on average higher effort levels than permanent workers to avoid firing. These observations are illustrated in Figure 4. The figure consists of four panels; the two panels on the left belong to the Full information case, while the two on the right correspond to the Noise case. In each panel, time 0 on the horizontal axis refers to the last round in which a given worker works under probation. This worker is granted permanent status (‘promoted’) at the end of that round, so from time 1 onwards this worker is permanent. The two upper panels focus on the case where the worker in question is the first one to get permanent status. Each panel depicts both the average effort level of the now permanent worker and of all the other, non-permanent



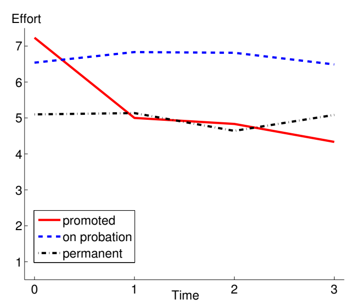
(a) First permanent workers - FP



(b) First permanent workers - NP



(c) Further permanent workers - FP



(d) Further permanent workers - NP

Notes: At time 0 some workers get promoted to a permanent status. The upper figures correspond to cases when the promoted workers get permanent status as first in their group (with 17 observations in FP, and 14 observations in NP), whereas the lower figures correspond to the case when there are already permanent workers in the group when workers are promoted (with 13 observations in both FP and NP). The lines are average efforts based on a strict partition of the six team members.

Figure 4: Effort decisions when getting promoted

workers.[14] The two bottom panels consider the situation where there are already other permanent workers around at time 0. These panels include a third line for the whole 4 rounds with the average effort level of these other permanent workers. All figures display a strict partition of the 6 workers of the team per round.

Subjects' changed behavior after having received permanent status as displayed in Figure 4, together with the manager's strategy of whom to promote, suggests a potential reason why the probation contract does not work well. Managers typically give permanent status to those who choose above average effort during probation, i.e. those who appear to be exemplary workers. These effort choices are likely to be driven by two motives: (i) an incentive not to get fired and (ii) an incentive to match the expected minimum of the other team mates

[14]Note that new permanent workers can also arrive to the group in round 2 or 3 if they get promoted just one or two rounds later than the worker in question. This happens relatively rarely (so the averages for 'permanent' in Figure 4a and 4b are based on a few observations), and in FP only in groups where the group managed to maintain high effort levels.

Table 4: Average relative effort before firing and after rehiring

Treatment	Before firing	After rehiring	After-Before	p-value for differences
FS (6)	0.74	0.98	0.24	0.03**
NS (6)	0.79	0.96	0.17	0.03**
FP (4)	0.91	1.06	0.15	0.14
NP (5)	0.97	1.04	0.07	0.14

Notes: **: significant at 5% level according to two-sided Wilcoxon-test for the differences over time. The numbers in the brackets represent the number of observations in that given treatment. Averages are based on matching group averages.

(as choosing a higher effort level does not increase monetary pay in the given round). Consider the first worker to get permanent status. Given the manager's strategy to promote (presumably) hard workers, this is likely one with an above average motive not to get fired (or an above average, perhaps inflated belief of what it takes to get promoted). After obtaining permanent status the firing incentive vanishes for him and only the matching incentive remains. Immediately after being the first to receive promotion, subjects thus should reduce their effort more (or increase less) than those still on probation. This "regression to the mean" is what we observe. The first workers to get permanent status are indeed those who have higher than average effort in the round just before. Yet due to their larger reduction in effort, they turn from exemplary workers into slackers; i.e. those who receive permanent status first immediately become the weakest link. This resembles Lazear's (2004) explanation for the "Peter principle", i.e. the empirical observation that individuals perform worse after being promoted.[15] Once there is a permanent worker in the team, the next one to get promoted has an incentive to follow suit after getting permanent status and copy the work morale of the other permanent people in place. The lower panels in Figure 4 nicely illustrate this.

Finally, we take a quick look at the behavior of rehired workers. For each worker we calculated their relative effort compared to the other group members in each round right before firing and in the first round after rehiring. Table 4 shows these average relative efforts based on matching group averages. Non-parametric tests reveal that in both Spot treatments workers significantly increase their relative effort level after rehiring, whereas there is no significant difference between relative efforts before firing and after rehiring under the Probation contract. This is another illustration that in the Probation treatments managers are much less effective with firing than in the Spot treatments.

[15] To paraphrase Lazear (2004, p. S146) by replacing his notion of ability with work motivation: "Individuals who are promoted are promoted in part because they are likely to have high permanent work morale [ability], but also because the transitory component of their motivation [ability] is high."

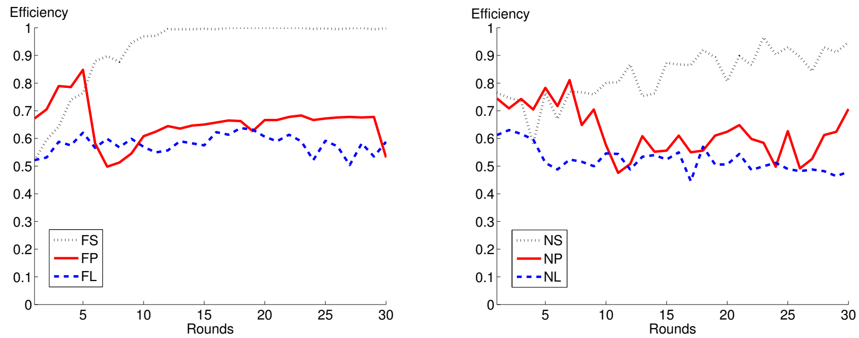


Figure 5: Relative efficiency over time under Full information (left panel) and under Noise (right panel)

4.3 Efficiency

In Section 4.1 we have observed that teams managed to perform better in terms of effort levels in the Spot treatments than in the Probation or Longterm treatments. Here we investigate if this result extends to efficiency. Do managers use their power efficiently under Spot contracts? Or do they spend too much on firing? Is efficiency harmed because workers choose different effort levels? To answer these questions we calculate the relative efficiency of the groups, including actual firing costs of the managers.

The appropriate benchmark to gauge welfare depends on the production process. In case of Full information, we calculate the total earnings of the 6 workers and the manager if all workers choose 7 and the manager does not fire. This leaves the team with total earnings of 910. In case of Noise, we calculate the expected payoffs. It can be shown that the expected payoffs for the workers are the highest if everybody chooses 8 (see Table 9 in Appendix A.2). In this case the expected welfare is 830.27. To calculate the teams' relative efficiency, we divide the actual group earnings by these benchmark levels.

Figure 5 presents these relative efficiency levels over time per treatment. The figure shows that relative efficiency is higher in the Spot treatments, and in fact, it is almost always 1 for FS after round 12. Furthermore, efficiency is increasing in the first half of the experiment for the Spot treatments, but is stable in the Longterm and Probation treatments. In these treatments teams only earn about 50-60% of the possible highest earnings. This stability seems to contradict the development of effort levels over time (Figure 1). Note however, that it is not necessarily the case that efficiency declines as effort levels decline, because workers start to coordinate better (on a lower effort level) and there is much less wasted effort.

Table 5 presents the average efficiency levels (with and without the manager) for the first and the second half of the experiment, as well as the comparison between treatments. The Spot contract yields significantly higher efficiency levels than the Longterm and Probation contracts do for each production process

Table 5: Average relative efficiency

Treatment	First half (rounds 1-15)	Second half (rounds 16-30)	First vs. second half
FS	0.85 (0.86)	0.998 (0.999)	0.03** (0.03**)
NS	0.76 (0.75)	0.90 (0.88)	0.03** (0.04**)
FL	0.57 (0.57)	0.59 (0.60)	0.25 (0.12)
NL	0.54 (0.54)	0.50 (0.52)	0.35 (0.92)
FP	0.65 (0.66)	0.66 (0.68)	0.92 (0.60)
NP	0.66 (0.66)	0.59 (0.60)	0.17 (0.12)
FS vs. FL	0.05* (0.05*)	0.03** (0.02**)	
FS vs. FP	0.26 (0.34)	0.03** (0.02**)	
FL vs. FP	0.63 (0.52)	0.75 (0.75)	
NS vs. NL	0.02** (0.01**)	0.00*** (0.00***)	
NS vs. NP	0.20 (0.23)	0.01*** (0.01***)	
NL vs. NP	0.05** (0.04**)	0.42 (0.34)	
FS vs. NS	0.20 (0.08*)	0.00*** (0.00***)	
FL vs. NL	1.00 (0.87)	1.00 (0.87)	
FP vs. NP	0.75 (0.43)	0.42 (1.00)	

Notes: ***: significant at 1% level, **: significant at 5% level, *: significant at 10% level according to two-sided ranksum test with $n = 6$ for the treatment differences, and Wilcoxon-test for the differences over time. The numbers in parentheses represent average relative efficiency for workers only (thus ignoring the manager) and their test results.

in the second half of the experiment. Furthermore, the Probation contract is significantly more efficient than the Longterm contract in the first half of the experiment under Noise. This is due to the fact that in the former workers are still on probation in the beginning and they want to avoid firing. The same pattern can also be observed in the Full information case, but there efficiency starts to decline earlier under the Probation contract. As a result, there are no significant differences between the Longterm and Probation contracts.

Comparing the Full information and the Noise case, we observe that efficiency is the same for the Longterm and Probation contracts, but is higher in FS than in NS in the second half of the experiment. To disentangle the effect of bad luck and workers' behavior, we calculate the 'hypothetical' relative efficiency for the Noise case in which we have taken out the effect of noise. That is, we calculate the output as if it had resulted from workers' effort levels, not from their productivities. The hypothetical efficiency of treatment NS in the second half is 0.997, which is not significantly different from the actual efficiency in FS.

If we exclude the manager's surplus from the efficiency measure, almost the same picture emerges (see Table 5). That is, workers are harmed by noise in their productivity. More importantly, keeping the production process fixed, they gain a lot from Spot contracts. The possibility to always fire thus not only benefits managers, but also the workers themselves.

5 Conclusion

In this paper we investigated the role of the fear of exclusion in team production with weakest-link characteristics. In an experimental labor market, we varied the extent to which managers were allowed to fire workers and the quality of the information they had about workers' performance.

Our design allowed us to address two main themes. First, we identified the profound role that the fear of exclusion may have in the labor market. When the manager was perfectly informed about workers' performance and had the discretion to fire workers in every single round, workers were disciplined swiftly and teams performed efficiently. In stark contrast, when workers were completely protected from being fired, their performance gradually deteriorated and overall performance was substantially worse. Interestingly, the role of the fear of exclusion did not diminish when managers had only imperfect information of the workers' performance.

Second, our research allows us to address how important it is to maintain the fear of exclusion. Can managers afford (i.e. without harming team performance) to abandon the possibility to fire workers after the workers have been disciplined, as regularly happens in practice when workers receive a permanent position after they have successfully passed a probation phase? Our results show that the team's performance tends to steadily deteriorate after some of its members have been assigned a permanent contract. Especially these members immediately reduce their effort upon getting a permanent status, an observation similar in spirit to the well-known Peter principle. We think that our results highlight an important drawback of probation contracts.

Appendix A Equilibria

A.1 Stage game with noise

In this appendix we derive the symmetric (pure strategy) equilibrium of the one-shot game in the Noise case. With just a single round there is no role for the manager, so we focus on worker behavior only. During the analysis we will use the following notation: $P_e(q|e_i)$ is the probability that output q occurs if everybody else chooses an effort e and player i chooses e_i. $q_{e_i|e}$ is the output when player i chooses an effort level e_i and everybody else chooses e. Player i's expected payoff of choosing the same effort level (e) as the others is $20Eq_{e|e} - 10e + 50$. If he deviates to a lower effort level e_i, his expected payoff is $20Eq_{e_i|e} - 10e_i + 50$. The gain is $10(e - e_i)$, while the loss equals $20(Eq_{e|e} - Eq_{e_i|e})$. It is worth to deviate if the gain is higher than the loss: $10(e - e_i) > 20(Eq_{e|e} - Eq_{e_i|e})$.

First consider deviations to an effort level one unit lower than everybody else. In this case the gain is exactly the effort cost, 10. To determine the expected losses we will make use of two tables that list the probabilities of the different output levels for given effort choices. Table 6 shows the probability

Table 6: Probability density of output if everybody chooses the same effort

Effort	Output						
	7	6	5	4	3	2	1
9	1	-	-	-	-	-	-
8	0.859	0.141	-	-	-	-	-
7	0.178	0.681	0.141	-	-	-	-
6	0.000	0.178	0.681	0.141	-	-	-
5	0.000	0.000	0.178	0.681	0.141	-	-
4	-	0.000	0.000	0.178	0.681	0.141	-
3	-	-	0.000	0.000	0.178	0.681	0.141
2	-	-	-	0.000	0.000	0.178	0.822
1	-	-	-	-	0.000	0.000	0.999

Notes: The probabilities are rounded to three digits. The most left numbers in each row from effort 5 downwards are 2.441E-10. The one next to it is 2.441E-4.

Table 7: Probability density of output if one player chooses an effort level one lower than the others

Effort	Output						
	7	6	5	4	3	2	1
9	0.975	0.025	-	-	-	-	-
8	0.661	0.314	0.025	-	-	-	-
7	0.059	0.601	0.314	0.025	-	-	-
6	0.000	0.059	0.601	0.314	0.025	-	-
5	-	0.000	0.059	0.601	0.314	0.025	-
4	-	-	0.000	0.059	0.601	0.314	0.025
3	-	-	-	0.000	0.059	0.601	0.339
2	-	-	-	-	0.000	0.059	0.941

Notes: The probabilities are rounded to three digits. The most left numbers in each row from effort 6 downwards are 2.441E-5.

density of output if everybody chooses the same effort.[16] Table 7 shows the probability density of output if all but one worker choose the same effort level and the remaining worker an effort level one unit below the others.[17] Using these two tables, we calculate expected losses from deviating to a lower effort level. Table 8 presents these losses.[18]

From Table 8 we observe that the expected loss of deviating one unit down-

[16] To illustrate where the numbers in the table come from, consider an effort level in the middle, e.g. $e = 5$. Let p be the probability of a noise equal ± 1 ($p = 0.225$) and q be the probability of ± 2 ($q = 0.025$). If everybody chooses 5, then $P_5(3|5) = 1 - (1-q)^6$, $P_5(4|5) = (1-q)^6 - (1-p-q)^6$, $P_5(5|5) = (1-p-q)^6 - (p+q)^6$, $P_5(6|5) = (p+q)^6 - q^6$ and $P_5(7|5) = q^6$. The calculations are similar for effort levels 4 and 3. For effort levels above 5 and below 3, the probability mass that would fall outside the range of 1-7 is shifted to either 1 (for effort 1 and 2), or to 7 (for effort 6, 7, 8, 9).

[17] The calculations here are similar to the previous ones. Just to illustrate, consider e.g. the case where all but one chooses 6 and the remaining worker chooses 5. In this case $P_6(3|5) = q$, $P_6(4|5) = 1 - (1-q)^5 * (1-p-q) - q$, $P_6(5|5) = (1-q)^5 * (1-p-q) - (p+q) * (1-p-q)^5$, $P_6(6|5) = (p+q) * (1-p-q)^5 - q * (p+q)^5$ and $P_6(7|5) = q * (p+q)^5$.

[18] The calculations in this table are also straightforward. As an illustration, we present the loss from deviation from 6 to 5: $\text{loss}_{6,5} = 20 \cdot (P_6(7|6) \cdot 7 + P_6(6|6) \cdot 6 + P_6(5|6) \cdot 5 + P_6(4|6) \cdot 4 - P_6(7|5) \cdot 7 - P_6(6|5) \cdot 6 - P_6(5|5) \cdot 5 - P_6(4|5) \cdot 4 - P_6(3|5) \cdot 3) = 6.84$.

Table 8: Expected loss from deviating to one unit below the others

Others' effort level							
9	8	7	6	5	4	3	2
0.5	4.46	6.84	6.84	6.84	6.84	6.34	2.38

wards always fall short of 7 and thus deviating is always beneficial, because the gain equals 10 in reduced effort costs. Thus, the only candidate for a symmetric equilibrium is where everybody chooses minimum effort. Now suppose that everybody chooses an effort level of 1 and one worker considers an upward deviation. In this case he loses 10 in increased effort costs, but gains by the increase in the expected output. Note however, that it is not possible to increase output above 3 by unilateral deviation. Deviations to effort levels of 6 and above are therefore dominated by deviating to 5. Furthermore note that even though the deviator increases the probability of higher output, this increase is very small. If everybody else chooses 1, the probability of an output equal to 3 cannot be higher than $q^5 = 9\text{E-}9$ and the probability of an output equal to 2 cannot be higher than $(p+q)^5 - q^5 = 9.77\text{E-}4$. Therefore, the maximum expected output, which happens when the deviator chooses an effort level of at least 5, is $3 \cdot q^5 + 2 \cdot ((p+q)^5 - q^5) + 1 \cdot (1 - (p+q)^5) = 1.00098$. The expected output if everybody chooses 1 equals 1.00024. Thus the increase in expected output (and the expected gain) is not substantial compared to the increased costs from the deviation. This means that deviating to a higher effort level is not beneficial for the workers. Hence the only symmetric equilibrium of this game is when all workers choose an effort level equal to 1.

A.2 Repeated game

In this section we show that both under Full information and under Noise almost any symmetric effort profile can be supported as equilibrium in the infinitely repeated game for high enough discount factor (even without the manager).

First note that under Full information all symmetric effort profiles with $e \leq 7$ are equilibrium profiles. Thus, adhering to a strategy to always play the same effort level (below 8) while everybody else does the same is a Nash equilibrium, regardless of the discount factor.[19]

Now let us turn to the Noise case. It is again trivial that if everybody always chooses an effort level of 1, then this is an equilibrium (because it is just an infinite repetition of the Nash equilibrium of the stage game). It can be shown that effort levels higher than 3 can also be sustained in the infinitely repeated game with a trigger strategy which punishes deviation forever. In particular, we consider trigger strategies in which players always choose the same effort level e, but if an output inconsistent with this effort level is observed, then they switch to effort level 1 forever after. In Table 9 we summarize, for all effort

[19]Note that since workers never observe effort levels, there is no equilibrium in which they would choose effort levels 8 and 9. These effort levels are always dominated by 7 as the output can be 7 at most.

Table 9: Expected payoff of given effort levels for one round, and the threshold discount factor under Noise

	Effort levels (e)							
	9	8	7	6	5	4	3	2
e	50	57.181	50.746	40.746	30.746	20.746	10.746	3.564
$e-1$	59.5	58.905	53.78	43.803	33.823	23.842	14.372	11.187
$e-2$	64.5	49.94	48.36	39.024	29.024	19.524	14.524	-
δ^*	0.905	0.824	0.749	0.799	0.856	0.92	-	-

levels from 2 to 9, the expected payoff of choosing the same effort level as the others, and of one and two units below. Note that always choosing effort level 1 gives an expected payoff of 10.005 in one round (here we exclude the constant 50). The probability that players enter the punishment phase depends on the effort level of the deviator (except for the case of effort levels 2 and 3). If the deviator chooses an effort level one unit below the others, this probability is 0.025. However, if he chooses an effort level two units below the others, it equals 0.25. The lower effort he chooses, the higher the chance of punishment is. It can be shown that the expected payoff of the deviation is decreasing as the effort level of the deviator decreases.[20]

Table 9 also contains the threshold discount factors for each effort level. If players have a higher discount factor than the threshold, the given effort level is sustainable in equilibrium with the above-mentioned trigger strategy.[21]

For our parameters there is no discount factor for which effort levels 2 and 3 are sustainable in equilibrium. In these cases there are no strategies in which only deviation can be punished, but players might also be punished in case of bad luck. Since the expected payoff of playing 1 forever is not much worse (in fact it is better than playing 2 forever) than the original effort levels, punishment is not harsh and it cannot deter players from deviation.

A.3 Firing using threshold strategy

We finally consider equilibria in which the manager plays an active role. Suppose the manager chooses a threshold level Q^* and fires workers who have a productivity lower than this threshold level. For simplicity we assume in our analysis that there is no restriction on the number of workers being fired in a round and that workers are not re-hired back later, but they become unemployed forever after being fired. Furthermore, we assume that workers choose minimum effort as soon as a deviation has occurred (e.g. if output is inconsistent with the threshold but the manager did not fire, or a worker deviates from the equilibrium effort level such that the resulting output is incompatible

[20] The only exception here are effort levels 9 and 3, where the best deviation for one round is effort levels 7 and 1, respectively.

[21] To calculate these threshold levels, we use the classical game theoretic approach and calculate the discounted sum of expected payoffs from adhering to the strategy and from the best deviation. Then we require that adhering is better for the worker than deviating from the strategy.

with equilibrium behavior), and that managers do not fire anymore (all workers including newly hired ones use the same strategy).

Under Full information, any threshold $2 \leq Q^* \leq 9$ can be supported as equilibrium threshold. To see this, suppose the manager uses threshold Q^*. From the discussion in Appendix A.2 it immediately follows that all workers choosing the same effort level e^* weakly above Q^* is an equilibrium. For effort levels 8 and 9 a sufficiently high discount factor is required.[22] With all workers choosing e^* using a threshold weakly below that is a best response for the manager. If she would deviate to a strictly higher threshold, she would have to fire all workers and (given the workers' trigger strategies) from the next round onwards the minimum effort outcome would always result.

Given the assumed trigger strategies, firing does not occur in (a symmetric pure strategy) equilibrium. Suppose all workers choose the same effort level $e < Q^*$ and are fired at the end of the round on the supposed equilibrium path. In that round the manager then earns $20e - 120$, as she has to bear 120 in firing costs. With $Q \leq 7$ this equals 20 at most. Hence on the supposed equilibrium path the manager would earn 20 per round at most. Not firing anybody gives strictly more in the current round (as the manager saves on firing costs) and yields 20 in each subsequent round where all workers choose minimum effort. Deviating thus yields strictly more and firing is not an equilibrium response.

Under Noise, even if workers choose an effort higher than the threshold, they might be fired due to bad luck. Note, however, that choosing an effort level 2 units higher than Q^* (with the threshold below 8 in this case) will ensure staying in the game for the next round. After a deviation workers will then either be immediately fired, turn to minimum effort from the subsequent round onwards or, if the luck component compensates appropriately for the deviation such that it goes undetected, continue with equilibrium behavior. The punishment is thus more severe than without the manager (as being unemployed yields strictly less than the minimum effort), implying that effort levels can be sustained in equilibrium for lower discount factors than we have in Table 9. In this case, the manager's strategy is also optimal (since she does not need to fire), so for high enough discount factors $e^* \geq Q^* + 2$ can be supported in equilibrium.

Similar to the Full information case, no equilibrium exists in which all workers are fired for sure at the end of the round (i.e. in which the manager employs threshold Q^* while the workers coordinate on an effort level $e < Q^* - 2$). Yet other equilibria exist in which workers are fired occasionally on the equilibrium path. To illustrate, suppose workers coordinate on an effort level e in the $Q^* - 2$ to $Q^* + 2$ range. In that case there are two possible deviations workers might consider. Deviating downwards will give them an instant benefit of reduced effort costs, but increases the chance of being immediately fired (note that if the manager does not deviate from her strategy, workers never enter the punish-

[22] First note that 8 and 9 can only be supported in equilibrium if the threshold is equal to the effort level. Players can always beneficially deviate if the threshold is strictly below 8 or 9 as choosing 7 or 8 does not change the output, but gives an instant benefit of the effort cost. If $Q^* = 8$, then $e^* = 8$ is supported in equilibrium if $\delta \geq \frac{1}{9}$. For $Q^* = 9$ and $e^* = 9$ we need that workers' discount factor is at least $\frac{2}{9}$.

ment phase because a possible deviation from a worker is either undetected or punished by the manager with firing). Deviating upwards decreases the round earnings, but also decreases the probability of being terminated from the team. It depends on the exact value of Q^* and the workers' common effort level e which deviation is the best one, and then on the discount factor whether this optimal deviation is attractive. If not and the chosen effort level is optimal for the workers, than we subsequently need to verify whether the manager behaves optimally with firing in equilibrium. This again depends on the exact value of Q^* and common effort level e. Note, however, that the lower the chosen effort is compared to Q^*, the more likely the manager has to fire, and the more beneficial a deviation to no firing becomes. By calculating the expected payoff of firing and no firing, we can also determine the discount factor for the manager with which firing is sustainable in equilibrium.[23]

Under Full information the same equilibria emerge under the Probation contract. Under Noise, the equilibria change because we need to take into account the likelihood that a worker becomes permanent. If everybody becomes a permanent worker, then we essentially are in the situation where there is no manager. Thus, the set of equilibria under the Probation constract is larger compared to the Longterm case, but smaller than in the Spot contract case.

Appendix B Instructions

In this section, we present the instructions for the second part of the experiment for the most difficult treatment, treatment NS. The other instructions are similar, and are available upon request from the authors.

[23]To illustrate this, consider $Q^* = 5$ (as in NS) and $e^* = 6$, and workers' behavior. If everybody chooses 6, then each worker is fired with probability $q = 0.025$. First of all note, that the two best deviations can only be deviating either to 5 or to 7 (see Table 9 for lower deviations and the explanation in Section A.1 for higher deviations). Adhering to effort level 6 when everybody plays the equlibrium strategy is better than deviating to 5 when $\delta_{\text{worker}} \geq 0.189$. Furthermore, deviating to 7 is not beneficial, when $\delta_{\text{worker}} \leq 0.866$. The intuition behind this is that by choosing 7, a worker can avoid firing for sure. So a very patient worker is better off by 7 as the instant loss is not that substantial compared to the loss from being fired ($E\pi_{6|6} = 90.746$, $E\pi_{7|6} = 82.264$ and $u = 30$). Thus if $0.189 \leq \delta_{\text{worker}} \leq 0.866$, then workers do not want to deviate from $e^* = 6$. Considering the manager, she never wants to fire if all productivities are at least Q^*, as it is costly, and the workers would punish her. Firing a worker with $p_i \geq Q^*$ if there exists a worker with $p_j < Q^*$ is costly again, and does not change anything (so it is not a beneficial deviation). The only possible deviation is to fire nobody when she sees $p_i = 4$ (firing *some but not all* workers with $p_i = 4$ is costly, and results in entering the punishment phase). If the manager adheres to the firing strategy, she needs to fire 6 workers in the worst case scenario, and afterwards workers still choose effort level 6, and the manager needs to fire in expectations $6 * q = 0.15$ workers per round. By deviating to no firing, she can save the firing cost, but she enters the punishment phase. Firing is supported in equlibrium if $\delta_{\text{manager}} \geq 0.607$ To summarize, $Q^* = 5$ and $e^* = 6$ are supported in equilibrium if $0.189 \leq \delta_{\text{worker}} \leq 0.866$ and $\delta_{\text{manager}} \geq 0.607$. Firing occurs in each round with probability 0.141, and the manager fires 0.15 workers per round in expectations.

INSTRUCTIONS PAGE 1

In part 2, you again will earn points, which will be transferred into money according to the same exchange rate of part 1: each 160 points will be exchanged for 1 euro. When everyone has finished reading the instructions of part 2 and before the experiment starts, you will receive a handout with a summary of these instructions. At the start of part 2, you will receive a starting capital of 600 points. You will not have to pay back this starting capital. In addition, you will earn points based on your decisions in combination with the decisions of the other participants in your labor market.

Part 2 consists of at least 25 rounds and at most 40 rounds. Your labor market of 10 participants will not change during the experiment. Each labor market contains 1 manager, 6 workers and 3 unemployed participants. The manager is the participant who earned the highest payoff in part 1; this person will keep this role during the entire experiment. At the start of the experiment, the other participants will learn whether they are a worker or unemployed. Initially, these roles are randomly assigned.

SUMMARY

In each round, the manager and the workers form a firm. Each worker has to choose an effort level which determines his or her productivity. The productivities of the six workers together determine the output of the firm. In every round, the manager can decide whether to fire any of the workers. Workers who are fired become unemployed and are automatically replaced by unemployed people. Thus, unemployed participants may become workers, and vice versa. In each round the earnings of the workers and the manager are determined by the effort levels and the output, whereas the unemployed are passive and receive an unemployment benefit. It is important that all participants read and understand the instructions for ALL roles (not only the part pertaining to the own role).

INSTRUCTIONS PAGE 2

WORKERS

At the beginning of each round, workers choose an effort level without knowing what the other workers choose. The effort level can be any integer from 1 to 9 (that is, 1, 2, 3, ... 9). A worker's productivity depends on effort and on chance. Productivity always lies within two units of the effort level chosen. It is most likely that productivity equals effort, and a one unit deviation is more likely than a deviation of two units. The following table gives the probability of a given productivity after a given effort level:

	Productivity								
Effort	1	2	3	4	5	6	7	8	9
1	75%	22.5%	2.5%						
2	25%	50%	22.5%	2.5%					
3	2.5%	22.5%	50%	22.5%	2.5%				
4		2.5%	22.5%	50%	22.5%	2.5%			
5			2.5%	22.5%	50%	22.5%	2.5%		
6				2.5%	22.5%	50%	22.5%	2.5%	
7					2.5%	22.5%	50%	22.5%	2.5%
8						2.5%	22.5%	50%	25%
9							2.5%	22.5%	75%

For example, if the chosen effort level is 4, then there is 2.5% chance that the worker's productivity is 2, 22.5% chance that it is 3, 50% chance that it is 4, 22.5% chance that it is 5, and 2.5% chance that it is 6.

In short, a worker's productivity cannot be lower than the chosen effort-2, and cannot be higher than the chosen effort+2. Moreover, productivity always lies between 1 and 9.

The output of the firm equals the minimum of the productivities of all workers. However, the output will never be higher than 7.

Workers benefit from output in the following way: for each level of output, a worker receives 20 points. For example, if one worker's productivity equals 3 while the other productivities are higher, the output will be 3, and each worker will receive 3*20=60 points. However, if all productivities are at least 8, the output of the firm will still be 7 and each worker receives 7*20=140 points.

Effort is costly for the workers. The higher the chosen effort level is, the more costly it is. For each level of effort the worker will incur a cost of 10 points.

Besides the output benefits and effort costs, workers also receive 50 points in every round.

In total, the payoff of the worker is determined by the output of the firm and the chosen effort level in the following way:

$$\text{Payoff worker} = 50 + 20 * \text{output} - 10 * \text{chosen effort}$$

MANAGER

In every round, the manager observes each worker's productivity. The manager does not choose an effort level. Instead, in every round, the manager can decide whether he or she wants to fire at most 3 workers. Workers that are fired are automatically replaced by unemployed participants. Firing workers is costly for the manager. For each worker fired, the manager incurs a cost of 20. If the manager fires someone, that worker becomes unemployed from the next round, and is automatically replaced by an unemployed person. The manager cannot choose which unemployed person is hired.

The manager benefits from the output of the firm in the same way as the workers do, and he / she also receives an additional fixed amount of 50 points in every round. The manager does not pay an effort cost. Instead, he or she may incur firing costs if he or she decides to fire any worker(s). In total, the manager's payoff is given by:

Payoff manager = 50 + 20 * output - 20 * number of fired worker

UNEMPLOYED

Unemployed participants have the possibility to observe the process in the firm. That is, they can observe the firm's output and the firing decisions of the manager. Only when an unemployed participant is hired, he or she will start making decisions, like the other workers in the firm. If a worker becomes unemployed, there is a possibility that he or she is hired again in a later round. Whether unemployed participants are hired is completely decided by the manager.

Unemployed participants receive an unemployment benefit.

Payoff unemployed person = 30

Unemployed participants receive an additional payoff if they have never been fired before. That is, the participants who start the experiment as unemployed will receive this additional payoff in every round until they are hired. The additional payoff is determined by a lottery. With probability 50%, the unemployed person receives an additional 120 points. With probability 50%, the unemployed person gets no additional points for the lottery. This additional payoff comes on top of the unemployment benefit of 30 points.

INSTRUCTIONS PAGE 3

OUTPUT AND PAYOFFS

For the workers, the highest possible payoff results if all workers have a productivity level of 7. In this case, each worker earns a payoff equal to 50 + 20*7 - 10*chosen effort. Notice that if other workers have high productivity levels of 7, it may not be in the best interest of the worker to choose a low effort level of 1, because then the firm's output will be diminished. With such an effort level, her or his productivity level is most likely to be 1, which would result in a total payoff of only 50 + 20*1 - 10*1 = 60.

Notice, however, that if other workers have a low productivity level of for instance 1, the worker cannot unilaterally enhance output by choosing a high effort level. (Remember, output is determined by the minimum of the productivity levels). In this case the worker's payoff in the round will be highest if he or she chooses an effort level of 1.

Because the manager does not have a cost for exerting effort, the manager is always better off if the effort levels of the workers and the output are high.

INFORMATION

After all workers have chosen their effort level, everybody in the labor market is informed about the output and only the manager is informed about the productivity of each individual worker. After that the manager decides whether he or she wants to fire at most 3 workers. Everybody in the labor market is informed of the manager's decision.

Each worker has an ID (an integer from 1 to 6) which is fixed during the time span he or she is employed. However, if the worker is fired, the newly hired worker gets his or her ID. Since worker ID's are fixed only for the spell of employment, it can happen that a worker gets a different ID when he or she is rehired. A worker loses his or her ID as soon as he or she is fired.

HISTORY OVERVIEW

On the lower part of the screen, a history screen will be provided. This history screen is different for the different roles. Managers can see the individual productivities by worker ID, firm output, and his / her own firing decision (how many workers and who gets fired, if applicable) in his / her labor market for each round. Workers and unemployed participants can only see their own effort level if they happen to be employed in that round (so not that of other workers), firm output and the manager's firing decision (how many workers and who gets fired, if applicable) in their labor market for each round. One row contains information about one round. The history screen updates twice in a round: after the workers' decision and after the manager's decision.

In the history screen for the manager, workers' productivities are denoted by three different colors. In green, you can see the productivities of the recent (active) workers in the firm. The two shades of blue denote the productivities of fired workers. Each worker's productivities across rounds are denoted by the same color as long as he or she is employed. The color changes, if the worker with that ID is fired.

Below you find examples of the history screens. The observations are sorted descending by round, so you can find the most recent round always at the top.

History screen for the manager:

Round	Productivity of						Output	Number of fired workers	ID's of fired workers
	Worker 1	Worker 2	Worker 3	Worker 4	Worker 5	Worker 6			
4	6	7	7	4	5	4	4	2	2,5
3	4	8	5	6	5	5	4	0	n.a.
2	5	4	7	8	3	4	3	1	2
1	3	4	2	4	3	6	2	3	1,3,6

History screen for workers and unemployed:

Round	Own productivity	Output	Number of fired workers	ID's of fired workers
4	7	4	2	2,5
3	8	4	0	n.a.
2	unemployed	3	1	2
1	6	2	3	1,3,6

On the next screens you will be requested to answer some control questions. Please answer these questions now.

Appendix C Results for the Medium contract

The opposite results we find for the Longterm and Spot contracts confirm our conjecture that the possibility of firing can enhance team-production and efficiency substantially. A natural question to ask is how workers perform in an environment that is in between these two extremes. In real life, contracts vary from full commitment (our Longterm case) to no commitment (our Spot variant), and there are intermediate cases where workers are partially protected by contracts. Either contracts are for a shorter term, or they can be renegotiated (see e.g. Salanié, 1997, p. 162-163). Even if contracts are for the very short term, it may be too costly for managers to seriously consider the possibility of firing every day. The Probation contract discussed in the main text represents one realistic in between case. In this appendix we look at an alternative intermediate situation where the manager can fire only every third round (labeled 'Medium' contract).

Compared to Spot, the Medium contract has more limited firing possibilities, yet it shares the important feature that workers are never fully secure. A priori one would thus expect that the Medium contract performs more similar to the Spot contract and will be more efficient than the Probation contract (where firing possibilities are limited in another way). In fact, one might even conjecture that the Medium contract could improve on the Spot contract in case monitoring is imprecise. With Noise managers might judge workers too quickly under the Spot contract and take insufficient account of the effect of bad luck.[24] Attribution error (Reeder and Spores, 1983) can further aggravate this problem. In a labor market, a manager makes an attribution error if she

[24] Abreu et al. (1991) examined theoretically the effect of different interval lengths to act in

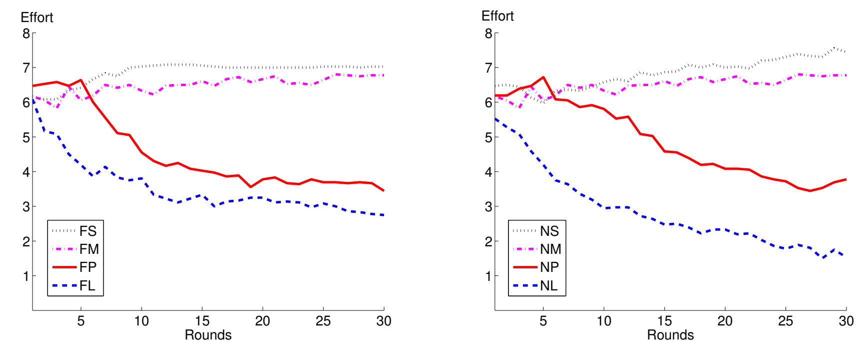


Figure 6: Average effort level over time for the Full information (left panel) and the Noise case (right panel)

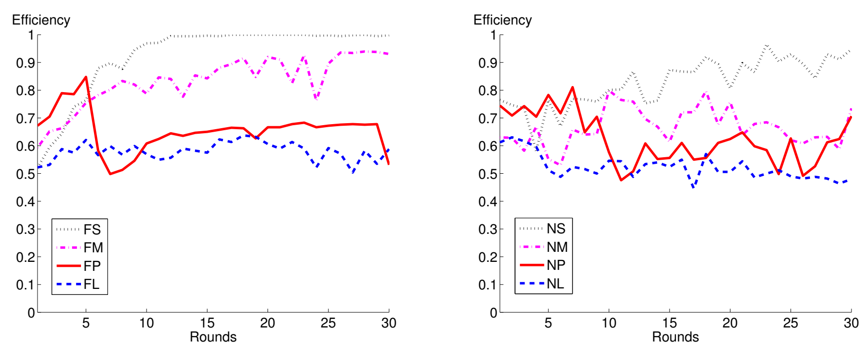


Figure 7: Average efficiency level over time for the Full information (left panel) and the Noise case (right panel)

downplays the possibility that a worker's bad performance may be caused by bad luck. By observing workers for more rounds before having the possibility to fire, the manager has better information about the worker when she decides about dismissing him.

The results of the Medium treatments are closest to the ones we obtained for the Spot contracts. Figure 6 presents the average effort levels of the Medium contract over time, together with the benchmarks provided by the other treatments. For both Full information and Noise, the Spot contract outperforms the Medium contract, though this difference is only significant under Noise. Furthermore, the Medium contract significantly outperforms the Longterm contract and it is insignificantly better than Probation.[25]

games with imperfect monitoring. Bigoni et al. (2011) found experimental evidence that in a 2x2 Cournot game, collusion is harmed with high or low flexibility but not with intermediate flexibility.

[25] This result might be due to the following: in FP, there is a huge variation in effort levels across matching groups. Two groups maintained average effort levels around 7, whereas three

Table 10: Average effort and efficiency level for the Medium contract

Treatment	First half (rounds 1-15)	Second half (rounds 16-30)	First vs. second half
Panel A: Effort level			
FM	6.32	6.66	0.05**
NM	5.68	5.34	0.75
FS	6.71	7.01	0.03**
NS	6.49	7.19	0.03**
FL	4.04	3.03	0.05**
NL	3.69	2.02	0.03**
FP	5.32	3.72	0.09*
NP	5.83	3.93	0.05**
FS vs. FM	0.34	0.10	
FM vs. FL	0.05*	0.02**	
FM vs. FP	0.34	0.20	
NS vs. NM	0.08*	0.02**	
NM vs. NL	0.03**	0.01**	
NM vs. NP	0.63	0.20	
FM vs. NM	0.15	0.15	
Panel B: Efficiency level			
FM	0.77	0.90	0.05**
NM	0.66	0.68	0.46
FS	0.85	0.998	0.03**
NS	0.76	0.90	0.03**
FL	0.57	0.59	0.25
NL	0.54	0.50	0.35
FP	0.65	0.66	0.92
NP	0.66	0.59	0.17
FS vs. FM	0.63	0.14	
FM vs. FL	0.11	0.04**	
FM vs. FL	0.42	0.11	
NS vs. NM	0.08*	0.05*	
NM vs. NL	0.04**	0.05*	
NM vs. NP	0.75	0.52	
FM vs. NM	0.15	0.02**	

Notes: **: significant at 5% level, *: significant at 10% level according to two-sided ranksum test with $n = 6$ for the treatment differences and Wilcoxon-test for the differences over time. For ease of comparison we repeat the effort levels and efficiency for the other contract types.

A very similar picture emerges for the comparison of efficiency levels. Figure 7 presents the efficiency levels of all four contract types over time, separately for the treatments with Full information and the treatments with Noise. Efficiency levels of the Medium contracts are in between those of Spot and Longterm contracts in each case. Again efficiency levels of the Medium contract are closer to the levels of the Spot contract than of the Longterm (and under Full information the Probation) contract, but the patterns is not as dramatic as for effort. Table 10 presents the average effort and efficiency levels together with the test results for the comparisons across treatments.

In sum, the Medium contract is also an efficient device to increase team-

groups had an average around 1, and the remaining group was in between.

production in this game. Workers choose considerably higher effort levels than in case of no firing possibilities. Yet we find little evidence that reduced firing possibilities may even help the manager to overcome a potential bias to fire too quickly; the Spot contract still outperforms the Medium contract in terms of team production and efficiency.

References

Abreu, D., Milgrom, P., and Pearce, D. (1991). Information and Timing in Repeated Partnerships. *Econometrica*, 59(6):1713–1733.

Ahn, T.-K., Isaac, R., and Salmon, T. C. (2008). Endogenous Group Formation. *Journal of Public Economic Theory*, 10(2):171–194.

Alchian, A. A. and Demsetz, H. (1972). Production, information costs, and economic organization. *The American Economic Review*, 62(5):777–795.

Bigoni, M., Potters, J., and Spagnolo, G. (2011). Flexibility and Cooperation with Imperfect Monitoring. *Working paper.*

Charness, G. and Yang, C.-L. (2010). Public goods provision with voting for exclusion, exit, and mergers: an experiment. *Unpublished Manuscript.*

Chaudhuri, A., Schotter, A., and Sopher, B. (2001). Talking Ourselves to Efficiency: Coordination in Inter-Generational Minimum Games with Private, Almost Common and Common Knowledge of Advice. *CV Starr Center for Applied Economics*, No. 2001-11.

Cinyabuguma, M., Page, T., and Putterman, L. (2005). Cooperation under the threat of expulsion in a public goods experiment. *Journal of Public Economics*, 89(8):1421–1435.

Croson, R., Fatas, E., Neugebauer, T., and Morales, A. J. (2015). Excludability: A laboratory study on forced ranking in team production. *Journal of Economic Behavior & Organization*, 114:13–26.

Devetag, G. and Ortmann, A. (2007). When and why? A critical survey on coordination failure in the laboratory. *Experimental Economics*, 10(3):331–344.

Gneezy, U. and Potters, J. (1997). An Experiment on Risk Taking and Evaluation Periods. *The Quarterly Journal of Economics*, 112(2):631–645.

Goeree, J. K. and Holt, C. A. (2005). An experimental study of costly coordination. *Games and Economic Behavior*, 51(2):349–364.

Green, E. J. and Porter, R. H. (1984). Noncooperative Collusion under Imperfect Price Information. *Econometrica*, 52(1):87–100.

Güth, W., Levati, M. V., Sutter, M., and Van Der Heijden, E. (2007). Leading by example with and without exclusion power in voluntary contribution experiments. *Journal of Public Economics*, 91(5-6):1023–1042.

Harsanyi, J. C. and Selten, R. (1988). *A general theory of equilibrium selection in games*. The MIT Press.

Ichino, A. and Riphahn, R. T. (2005). The effect of employment protection on worker effort: Absenteeism during and after probation. *Journal of the European Economic Association*, 3(1):120–143.

Jensen, M. C. and Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4):305–360.

Knez, M. and Camerer, C. (1994). Creating Expectational Assets in the Laboratory: Coordination in 'Weakest-Link' Games. *Strategic Management Journal*, 15(SI):101–119.

Lazear, E. P. (2004). The peter principle: A theory of decline. *Journal of Political Economy*, 112(S1):S141–S163.

Maier-Rigaud, F. P., Martinsson, P., and Staffiero, G. (2010). Ostracism and the Provision of a Public Good: Experimental Evidence. *Journal of Economic Behavior & Organization*, 73(3):387–395.

Masclet, D. (2003). Ostracism in work teams: a public good experiment. *International Journal of Manpower*, 24(7):867–887.

Monderer, D. and Shapley, L. S. (1996). Potential Games.

Page, T., Putterman, L., and Unel, B. (2005). Voluntary association in public goods experiments: Reciprocity, mimicry and efficiency. *The Economic Journal*, 115(506):1032–1053.

Reeder, G. D. and Spores, J. M. (1983). The attribution of morality. *Journal of Personality and Social Psychology*, 44(4):736–745.

Riedl, A., Rohde, I. M. T., and Strobel, M. (2011). Efficient Coordination in Weakest-Link Games. *CESifo Working Paper Series*, No. 3685.

Salanié, B. (1997). *The Economics of Contracts: A Primer.* The MIT Press.

Salmon, T. C. and Weber, R. A. (2011). Maintaining Efficiency While Integrating Entrants From Lower-Performing Environments: An Experimental Study. *University of Zurich, Department of Economics, Working Paper*, No. 35.

Seale, D. A. and Rapoport, A. (1997). Sequential Decision Making with Relative Ranks: An Experimental Investigation of the "Secretary Problem". *Organizational Behavior and Human Decision Processes*, 69(3):221–236.

Van Huyck, J. B., Battalio, R. C., and Beil, R. O. (1990). Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure. *The American Economic Review*, 80(1):234–248.

Weber, R. A. (2006). Managing Growth to Achieve Efficient Coordination in Large Groups. *The American Economic Review*, 96(1):114–126.