Team production benefits from a permanent fear of exclusion

Anita Kopányi-Peuker, Theo Offerman and Randolph Sloof*

December 22, 2017

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. We extend a standard weakest link experiment with a manager who can decide to replace some workers at a cost. We address two main questions: (i) Does the fear of exclusion need to be a permanent element of contractual agreements? (ii) Are the results robust to the introduction of noise in workers' productivity? We find that the fear of exclusion strongly encourages cooperation among workers, but it does not generate the trust needed for cooperation once the fear of exclusion is lifted. That is, once some workers receive a permanent contract, effort levels steadily decrease. The results are robust to the introduction of noise in the link between effort and productivity.

JEL classification: C72, C92, M51, M55
Keywords: team-production, weakest-link game, exclusion, probation, experiment.
Acknowledgements: We thank for suggestions of audiences at the 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 (grant number: 406-11-017).

*University of Amsterdam and Tinbergen Institute, Roetersstraat 11, 1018WB Amsterdam, The Netherlands. A.G.Kopanyi-Peuker@uva.nl (corresponding author); 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 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. We think that the fear of exclusion is key in explaining (part of) the difference in team performance in the laboratory and the field.[2]

In this paper, we design a series of experiments in which we vary the role of a manager with discretionary power. The experiments allow us to pursue two main goals. First, and most importantly, 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. It is not clear though whether the fear of exclusion is effective in creating the trust that is needed for continued cooperation after the fear of exclusion is lifted. An open question that we address is if workers continue to perform well after the probation phase has ended in a setting where maintaining high effort levels is in everybody's best interest.[3]

A second goal is to investigate if the results are robust to the introduction of

[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., 2009; Weber, 2006). Devetag and Ortmann (2007) provide a survey.

[2] Alchian and Demsetz (1972) 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).

[3] 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 vanished they may in principle still want to exert the same amount of effort.

noise in the link between effort and productivity. In practice, workers' performances will be affected by luck. This feature may erode a potential positive effect of the possibility to exclude team members. One effect of noisy performance may be that cooperative equilibria are no longer supported in an equilibrium of the stage game. In the noisy minimum effort game that we study, deviating down from any symmetric profile is a best response in the stage game. Another behavioral possibility is that with noise there is a danger that managers judge workers too quickly, and do not sufficiently account for the possibility that a worker's performance is affected by bad luck. 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 possibility that the added value of having a monitoring manager decreases when performance is noisy. Thus, the introduction of noise may make it harder for managers to motivate the workers. To the best of our knowledge, we are the first to address the fear of exclusion in the presence of noisy performance.

Exclusion may facilitate cooperation in team production through two mechanisms. First, the possibility of exclusion may simply boost performance because workers who consider to shirk refrain from doing so because they fear to be fired which is costly to them. This is an incentive effect. The other potential mechanism is that there is heterogeneity in workers' attitudes, and the possibility to exclude badly performing workers allows a manager to create a homogenous team that consists of members with a cooperative attitude. This would be the selection effect. Notice that a comparison of the performance of workers who survived a probation phase and workers who are subject to an ongoing threat of being fired allows us to identify the incentive effect. If a positive effect of exclusion is realized through selection only, performance should not differ between the two groups once the right type of workers is selected. A selection effect is revealed by a comparison of the performance of workers who survived probation and the performance of workers who never face a threat of exclusion.[4]

In agreement with the labor applications that motivate our research, we include a manager in our experimental 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,

[4]We do not disentangle the pure monetary effect of incentives and their symbolic effect. That is, in our experiment, we cannot determine if the incentive effect is a response to the monetary effect of losing one's job, or a response to an informal reprimand of the manager. In previous work on the voluntary contribution mechanism, Masclet et al. (2003) find that informal sanctions can be effective in raising cooperation, although their effect in the longer term is not as large as what can be accomplished with monetary sanctions. Dugar (2010) finds that non-monetary disapproval is very effective in the minimum-effort game, while non-monetary approval does not increase effort. Riedl et al. (2016) have a treatment where exclusion is only symbolic in the sense that the exclusion does not affect the payoff of the person excluded. In that treatment, subjects coordinate on the highest effort level.

and (ii) how well the worker's productivity level reflects the worker's effort level. Regarding the latter dimension, a worker's productivity is either equal to her effort level, or it is equal to her effort level plus a noise term.

In real life, the extent to which workers are contractually protected varies from no commitment under spot contracting to full commitment in case of a tenured position. In intermediate cases workers are only partially protected by contracts. One particularly relevant in between case concerns probation contracts, where the relationship essentially moves from no commitment during the probation phase to full commitment after having obtained tenure. Another realistic in between case are short term contracts that do not last as long as the potential relationship. In our experiment we consider all four possibilities: workers can be fired either every round (Spot), every third round (Medium), only during their probation phase (Probation), or never (Longterm).

Compared to Spot, our Medium contract has more limited firing possibilities, yet it shares the important feature that workers are never 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 when productivity is noisy. In that case managers might judge workers too quickly under the Spot contract and take insufficient account of the effect of noise.[5] Attribution error (Reeder and Spores, 1983) can further aggravate this problem if managers downplay 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. Together the four contracts also allow us to investigate whether it is primarily the possibility of future exclusion that disciplines workers, or rather the frequency of firing opportunities.

In the experiments, we find that the fear of exclusion efficiently encourages workers to perform well. When managers have perfect discretion to fire workers, workers tend to anticipate from the start that they cannot afford to slacken. This contrasts strongly to the opposite case where workers are fully protected by long-term contracts. There, workers' performance gradually deteriorates over the experiment. The Medium contract performs slightly worse than the Spot contract, but significantly better than the Longterm contract. Interestingly, these results carry over to the case where workers' productivity is affected by good or bad luck. Also then workers perform much better when the manager has the power to discipline workers.

Regarding our first main question, our results for the Probation treatments suggest that the fear of exclusion needs to be a permanent fixture in the labor

[5] Abreu et al. (1991) examined theoretically the effect of different interval lengths to act in 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.In the context of a financial asset market, Gneezy and Potters (1997) found that agents perform worse when they are offered the possibility to evaluate their portfolio continuously.

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. This finding is similar in spirit to the so-called "Peter principle", i.e. the empirical observation that individuals perform worse after being promoted (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. Thus, our data provide clear support for the incentive effect of exclusion, while there is only modest support for the effect of selection.

The remainder of this paper is organized as follows. The next section provides a review of related experiments. Section 3 introduces our game. In Section 4 we describe the experimental design and in Section 5 we presents our experimental results. Section 6 concludes.

2 Related literature

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 (2017) investigated how adding lower-performing 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. 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 (2014), 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.[6]

[6] Solda and Villeval (2017) studies the impact of exclusion on the behavior of the excluded members after their reintegration. They find that exclusion has a positive effect on cooperation after reintegration when reintegration is quick (instead of slow).

Some other studies investigated the possibility of exclusion or endogenous group formation in the minimum-effort game. Closest to our paper is the study of Croson et al. (2015) who study the role of exclusion in the minimum-effort game, the voluntary contribution game and the best-shot game. An important difference is that in their study exclusion from the consumption of the team's product occurs automatically while in our study it results from the decision of a manager. The managers in our experiment do much better than the automatic mechanism in the minimum-effort game of Croson et al. (2015). In their mechanism, all team members who provide less effort than the maximum of the group are automatically excluded from consuming the team's production. In their treatment on the weak-link mechanism without redistribution (WLM-EX), i.e. the treatment that is closest to our spot contract treatment without noise, this mechanism leads to the perverse result that all team members tend to choose the lowest effort level, such that no one is excluded from the team's production, while no one benefits from the presence of the mechanism either.[7] In contrast, the workers in our experiment know that they may get fired if they work badly, even if all the other workers provide low effort as well, and they perform very well. Another important difference is that in Croson et al. exclusion from consumption was only for one round, while in our game it was for multiple rounds. Thus, the presence of a human manager with stronger discretionary power may have a better effect than a simple automatic exclusion rule. Naturally, there are also many minor differences between their and our experiment, so it is not sure that a difference in performance must be attributed to a difference in the exclusion mechanism. For instance, they considered a finitely repeated game with a surprise restart, their teams consisted of four members, and there are some differences in the payoff parameters.

Riedl et al. (2016) 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.[8]

Finally, the possibility of endogenously choosing contractual parties was already investigated in labor market settings with individual production. Brown et al. (2004) and Brown et al. (2012) examined how high effort can be maintained without a third-party enforcement of contracts with an excess number of workers or firms, respectively. In this research, firms hired workers endogenously, and the fear of getting no contract was effective in both settings. Other studies further investigated the fear of exclusion in gift-exchange games, and found that

[7] Croson et al. (2015) find that the automatic exclusion of subjects who exert less effort than the maximum is an effective mechanism to enhance effort in the voluntary contribution mechanism and the best-shot mechanism. In addition, when the losses of the excluded members are redistributed to the members who are not excluded, performance is also improved in the minimum-effort game.

[8] Corgnet et al. (2015) also investigate firing in a team production setting, In their setup individuals' performances do not have any effect on others' earnings, except for possibly the manager.

flexible contracts can enhance effort levels (see e.g. Berninghaus et al., 2013 and Bernard et al., 2016). Furthermore, Falk et al. (2015) also considers the possible adverse effect of probation, but in a gift exchange setting. They find a similar result to ours that workers decrease effort levels once they become permanently employed. Note that in our setting this result is more striking because, in contrast to the gift exchange game, if the others work hard, it is always a best response to work hard as well independently of being permanently employed or not.

Overall, compared to the existing experimental literature a novel feature of our paper is that we investigate if the disciplining effect of the fear of exclusion is eroded when productivity is noisy. In addition, a contribution of practical importance is that we address the question what happens in team-production with weakest link characteristics if the danger of exclusion disappears when workers have survived the probation phase.

3 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).[9]

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).[10] This feature reflects the possibility that in organizations workers work harder just to avoid the possibility that they are unintentionally regarded as shirker

[9] 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%.

[10] 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.

when they face bad luck. 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.

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 four types of contract in the above-mentioned team-production game. Under Spot contract, managers can replace workers in every round. In contrast, under Longterm contract workers are fully protected by contracts and the manager cannot replace anybody. The two remaining contracts represent in between cases. Under Medium contract the manager can fire workers only every 3^{rd} round, so only in rounds 3, 6, 9 et cetera. Finally, 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 anymore. The length of the arbitration period is to some extent arbitrary. We opted for a 5 round probation period for practical reasons. We did not want the experiment to last too long. At the same time, we wanted to accomplish that workers stayed together for a while before all members got a permanent position, and we wanted to have sufficient observations of what happened to a group after everyone got a permanent position.[11]

In order to investigate whether results are robust to the introduction of noise in the link between effort and productivity, we consider two different production processes. In both production processes, managers observe workers' productivity. In the one case, there is no noise, and productivity equals effort. We refer

[11] In the experiment, among the groups where all workers became permanent workers, this happened on average at round 14, while the experiment lasted 30 rounds, which means that our goal was accomplished.

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
FM	no	every 3^{rd} round	10	6
NM	yes	every 3^{rd} round	10	6
FP	no	during probation	10	6
NP	yes	during probation	10	6
FL	no	not possible	7	6
NL	yes	not possible	7	6

Notes: A group consisted of 6 workers, 1 manager, and (if firing was possible) 3 unemployed persons.

to this case as 'Full information', because managers can perfectly infer workers' efforts. In the other case, productivity is noisy. In this case, managers are only imprecisely informed of workers' efforts. We refer to this case as 'Noise'. We chose the Noise and Full Information versions of the game because we think that these are the most interesting situations to compare. That is, in theoretical modelling of team production processes and experimental tests thereof it is often assumed that effort translates perfectly into productivity and productivity is what the manager observes. In practice, workers' performances are affected by good or bad luck, and the manager cannot distinguish between the amount of effort of a worker and luck.[12] This leads to eight different treatments: FS, NS, FM, NM, FP, NP, FL and NL, where F (N) stands for Full information (Noise), and S, M, P, L stands for Spot, Medium, Probation, Longterm, respectively. An overview of the treatments is presented in Table 1.

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

[12]Notice that our design does not allow us to separate the role that noisy information plays from the role of noisy productivity. To identify the role of noisy information per se, it is interesting to compare our noisy treatments with treatments where productivity is noisy but the manager is completely informed of both effort and luck.

Table 2: Equilibrium predictions ignoring the manager

	Full information	Noise
one-shot	all symmetric effort profiles with effort below 8	all workers choose effort level 1
repeated	all symmetric effort profiles with effort below 8	all symmetric effort profiles except effort levels 2 and 3

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, that is, the only Nash equilibrium is when all workers choose an effort level 1.[13] Note however, that the most efficient outcome in this case would be if all effort levels are 8.

In our experiment subjects play the stage game repeatedly with an indefinite end. In the game without a manager, under Full information 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. Table 2 summarizes the different predictions without the manager.

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, assuming firing is always possible any threshold P^* above minimum effort can be supported as equilibrium under Full information. Firing then does not occur on the equilibrium path. The picture might be different when productivity is noisy. With Noise, we can construct equilibria in which firing never occurs on the equilibrium path with sufficiently high effort levels, and other equilibria where firing may even occur on an equilibrium path

[13] 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.

supporting high effort. The occasional firing of unlucky workers may serve the purpose of clearly communicating the manager's intolerance of slacking performance.[14] 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. A manager may be more effective in facilitating cooperative outcomes because the manager can identify and target badly performing members individually. Without a manager specializing in monitoring individual team members, the cooperative outcome can also be supported with trigger strategies, but punishments would affect all team members, not only the badly performing ones. In addition, without exclusion, the badly performing members stay in the team, and may engage in costly retaliation.

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, both Spot and Medium contracting work better in sustaining high team output than Longterm contracting does (where the threat of future 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 powerful if performance is noisy. 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 become demotivated if the manager does not appropriately take the noise into account and fires workers too often or too quickly. The Medium contract may potentially mitigate the latter problem relative to Spot, because managers observe workers for more rounds before having the possibility to fire, and thus potentially outpeform Spot in case of noise. 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 question of main interest is whether 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.

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

4 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. Subjects had to answer some control questions correctly before they could proceed to the experiment.

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 terms of the roles that were assigned to the subjects and the choices that they made, we chose to provide them with the context that corresponds to the situation that motivates our research. Given that in previous experiments subjects have sometimes been found to respond to the framing of the experiment, we prefer to have the framing that corresponds to the situation that we investigate, so that we can be more confident that the results apply to (at least) the situation that we have in mind. Loewenstein (1999) provides a more comprehensive discussion of this issue.

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.[15]

In the second part of the experiment subjects played the modified minimum-effort game presented in Section 3. 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. We chose this quasi indefinitely repeated game procedure because we thought that for our purposes a finitely repeated game with known end round was less suitable. In such cases there may be an artificial end-effect which may make the comparison across treatments more complicated (if there is an interaction effect with the treatment). We do not believe that it is possible

[15] 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.

to have a literally indefinitely repeated game in the lab (because after some time the probability of coninuation must become small when security closes the lab and the building). Another probably more important drawback of a design with a random continuation rule is that treatments may not be balanced in the number of rounds that are played.

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 productivity 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.

5 Results

In Section 5.1 we focus on worker behavior by comparing effort decisions across treatments. In Section 5.2 we subsequently study managers' firing decisions and how these drive workers' effort choices. All reported non-parametric tests are (unless stated otherwise) carried out at the matching group level, with two-sided ranksum tests for treatment differences and Wilcoxon signed-rank tests for differences over time between the first half and the second half of the experiment.

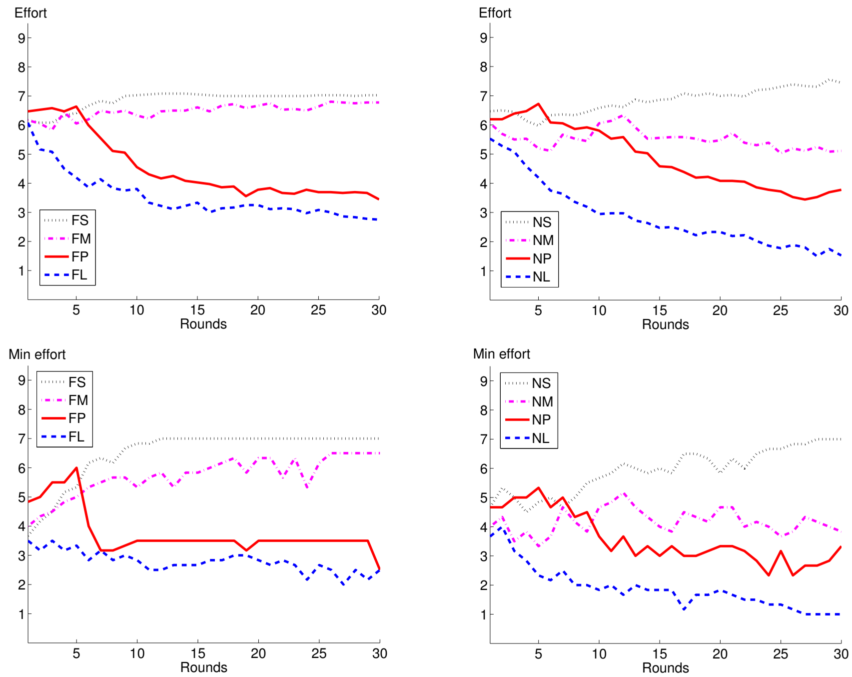


Figure 1: Average effort level (top panels) and minimum effort level (bottom panels) over time for the Full information (left panels) and the Noise case (right panels)

5.1 Effort decisions

Figure 1 shows the average effort levels and the average minimum effort levels over time for the different treatments. In the Longterm treatments, both average and minimum effort decrease over time, just like in previous experiments with a standard minimum-effort game. The possibility to fire offered in the Spot and Medium treatments profoundly improves team performance. The Probation contract increases effort levels slightly compared to the Longterm contract, but to a much lesser extent than either the Spot or the Medium contract do. Because minimum effort follows the same pattern as average effort, always having the possibility of future exclusion not only increases workers' effort levels, but also the output of the team.[16]

Table 3 reflects the extent to which the treatments differ systematically, by reporting the results from our non-parametric tests on matching group averages.[17] Effort levels under the Spot contract are significantly higher than under

[16]Similar patterns are thus also observed if we consider efficiency. These results are relegated to Appendix C.

[17]Our main conclusions continue to hold if we apply a Bonferroni multiple hypotheses t-test. In that case significant differences are similar as in Table 3, with the exception that NM is

Table 3: Average effort level

Panel A - average effort levels			
Treatment	First half (rounds 1-15)	Second half (rounds 16-30)	First vs. second half
FS	6.71	7.01	0.03**
FM	6.32	6.66	0.05**
FP	5.32	3.72	0.09*
FL	4.04	3.03	0.05**
NS	6.49	7.19	0.03**
NM	5.68	5.34	0.75
NP	5.83	3.93	0.05**
NL	3.69	2.02	0.03**

Panel B - comparison across treatments

	FS	FM	FP	FL		NS	NM	NP	NL
FS		0.34	0.26	0.04**	NS		0.08*	0.23	0.01**
FM	0.10		0.34	0.05*	NM	0.02**		0.63	0.03**
FP	0.10*	0.20		0.23	NP	0.01***	0.20		0.02**
FL	0.01***	0.02**	0.69		NL	0.00***	0.01**	0.15	

Notes: ***: significant at 1% level, **: significant at 5%, *: 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 ($n = 6$). In Panel B the above diagonal depicts differences in the first half, and below diagonal the differences in the second half.

the Longterm contract, both under Full information and under Noise. The Spot contract also significantly outperforms the Probation contract in the second half of the experiment, and the Medium contract under Noise. The latter also outperforms the Longterm contract under both productivity structures. A priori one could expect that the Medium contract performs best because it protects the manager against an attribution error of firing workers too quickly when they do not sufficiently appreciate the role of noise in workers' productivities. As we can see, however, we do not find evidence that the reduced firing opportunities under Medium actually help the manager to sustain higher effort levels than in Spot. Effort levels under the Probation contract are not significantly different from those under the Longterm contract (except for the Noise case in the first half of the experiment).[18] Together these results suggest that especially the possibility of future exclusion disciplines workers, rather than the exact frequency of firing opportunities. This holds irrespective of whether or not productivity is noisy. Indeed, noise does not seem to have a systematic upward or downward shift on effort levels; for each of the four different contracts the difference in

not significantly different from NS, and only in the second half from NL.

[18] Eyeballing Figure 1, a remarkable difference is in the effort levels of the Probation and Longterm contracts with Noise. The difference is significant in the first half but ceases to be significant in the second half of the experiment. The latter may be due to the conservative testing procedure using averages per matching group as data. The same test with individual efforts as data is significant in the second half ($p = 0.00$). As we explain below, this effect is mainly driven by the managers who do not assign a permanent position to all workers. Once all workers in a group have a permanent position in the Probation treatments, the differences in effort levels between the Probation treatments and the Longterm treatments are negligible.

effort levels between Full information and Noise is never significant ($p > 0.14$ for all cases).

The previous non-parametric tests only compare whether or not two effort distributions are shifted relative to each other (by looking at the equality of average ranks), and thus essentially whether effort levels are systematically higher in one of the two distributions. Even if these tests are insignificant, the distributions can still differ in their spread though. Figure 2 therefore displays the distribution of effort levels over time for the different treatments. The panels on the left reveal that under Full information subjects often either coordinate on high or low effort levels, intermediate effort levels are less frequent. Compared to FS and FM, we observe more diverse effort levels under FP and FL because groups are more heterogeneous. Under Noise (right panels), effort levels take more intermediate values as well. Thus, even though there are no differences in average effort between Full information and Noise, the distribution of effort levels is not the same. Of particular interest is 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 equal 36% and 2% in NS, respectively. Note however, that in case of Noise, the most efficient choice 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 outcome.

We also compared first round effort levels across different contract types, keeping the link between effort and technology 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 ($p = 0.005$). There are no further differences across treatments. These findings show that, with noisy productivity workers are already affected in the first round 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.

To assess the role that the incentive and selection mechanisms play, we consider the workers in those groups in the probation treatments where all members have become permanent workers, and compare their effort levels with those of the workers in the longterm treatments. Notice that in both cases an incentive effect of the fear of exclusion is absent, because no worker can be fired. Thus, this comparison provides an estimate of the selection effect.[19] Figure 3 displays

[19]Our experiment was not purposely (and thus not perfectly) designed to identify the selection effect. One concern is that the round in which all members are permanent for the first time differs across groups in the probation treatments. Another potential concern in this analysis is that the assessment of a selection effect is obfuscated by the fact that in the probation treatments subjects have a larger incentive to cooperate in the early rounds of the experiment, because the prospect of a permanent contract is a larger stimulus to behave well than for instance the fear for a temporary loss of a position in the team in the spot treatments. This potential confound does not seem to play a role in our data, however. Comparing subjects' effort levels in the Probation and Spot treatments in the first 5 rounds (where no worker in any probation group had been assigned a permanent position), the p-value for the ranksum test under Full information is 0.26, and under Noise is 0.81 (using matching group averages as data).

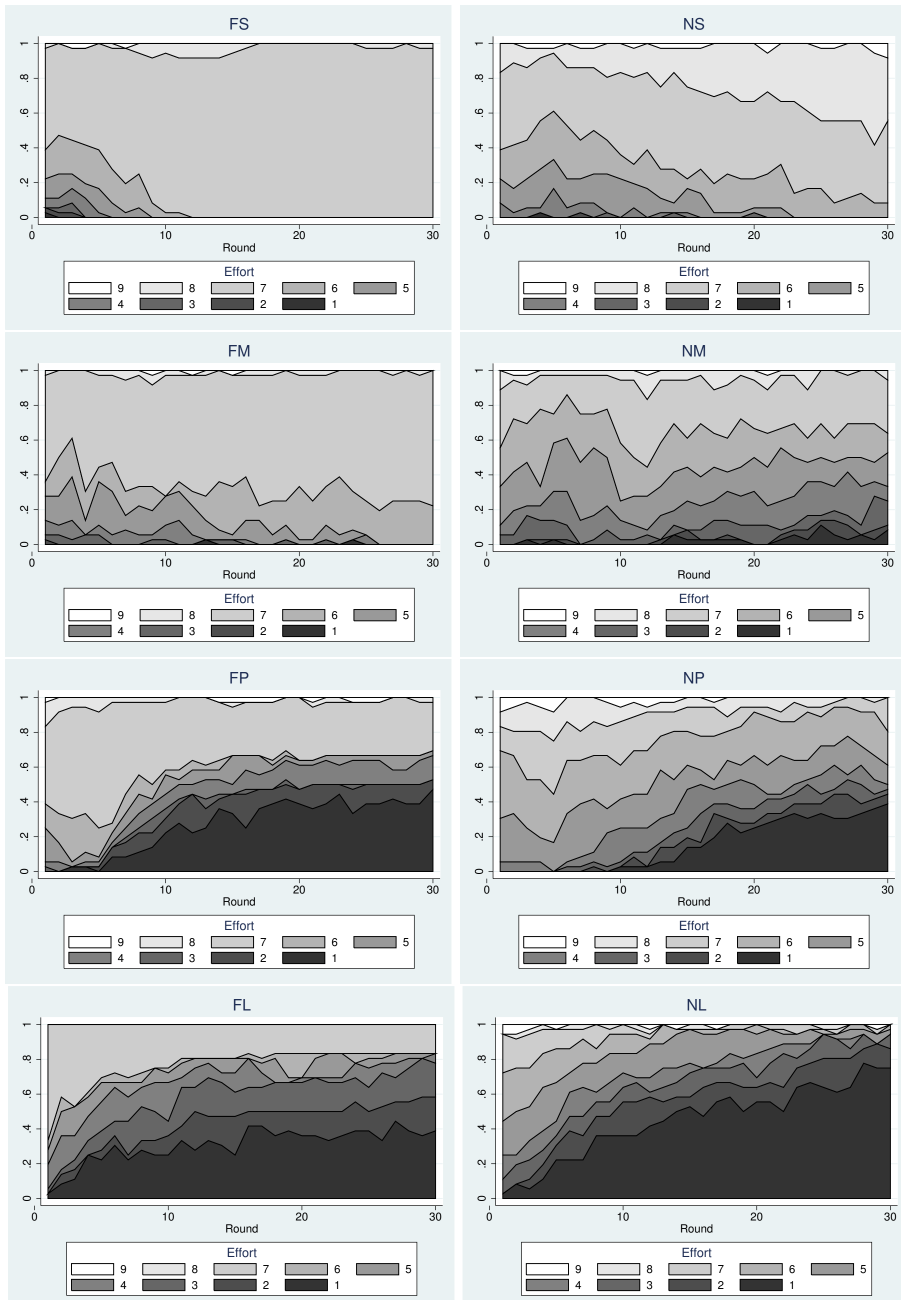
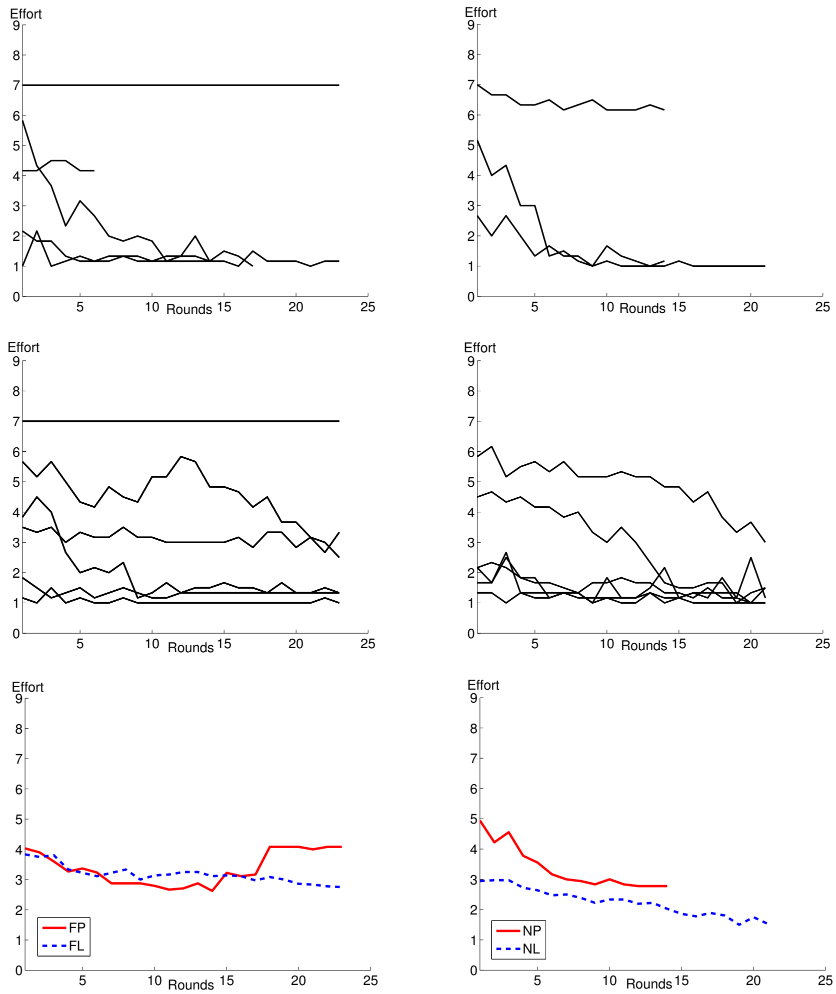


Figure 2: Frequency of different effort levels over time under full information (left panel) and under noise (right panel)



Notes: In the Probation treatments (FP top left, and NP top right panel), the average effort per group is displayed for those periods where all group members were permanent workers. In the Longterm treatments (FL middle left, and NL middle right panel), the average effort is displayed from the periods where all group members are permanent workers in at least one Probation group from the same productivity technology (bottom two panels).

Figure 3: Average effort level over time under full information (left panel) and under noise (right panel) once all group members are permanent workers

these time-series, where we only consider those rounds where all group members are permanent workers. The corresponding series of the longterm treatments use the earliest round in which the first group only consisted of permanent workers as round 1. The figure shows that there are no substantial differences between the series. Combined with the fact that we do observe a substantial difference

Table 4: Firing frequencies

Treatment	0	1	2	3	# poss. firing rounds	% firing rounds used	total # of firings
FS	142	29	6	3	180	21%	50
NS	124	49	7		180	31%	63
FM	41	13	4	2	60	32%	27
NM	23	23	13	1	60	62%	52
FP	42	35	14	6	97	57%	81
NP	55	46	25	5	131	58%	111

Notes: The columns labelled 0, 1, 2, 3 shows the absolute number of rounds in which the manager fired 0, 1, 2, or 3 workers. With (0) denoting the column labelled 0 etc., we have: # poss. firing rnds = (0) + (1) + (2) + (3), overall # of firings = (1) + 2 ∗ (2) + 3 ∗ (3).

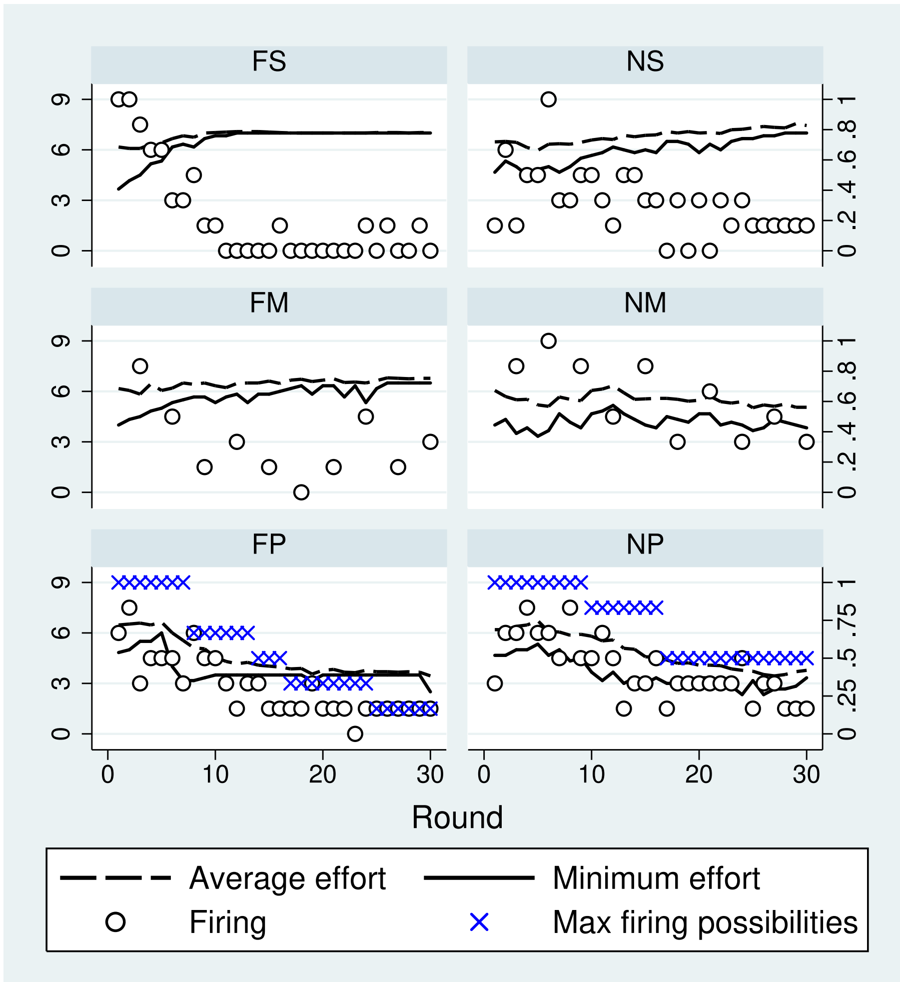
in effort between the spot treatments and the longterm treatments, where both the incentive and the selection mechanism may play a role, we conclude that the fear of exclusion operates primarily through the incentive mechanism instead of the selection mechanism.

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. These results are robust to the introduction of noise. The data reveal a substantial role for the incentive mechanism and a minor role of the selection mechanism.

5.2 Firing: causes and consequences

In the experiment, managers regularly fire bad-performing workers if they have the power to do so. The left part of Table 4 reports the frequency distribution over number of workers fired for the rounds in which firing is possible. In treatments FS and NS firing is possible in every round, so with 6 groups and 30 rounds we have 180 possible firing rounds in total. Treatments FM and NM allow for firing in one third of the rounds, while for FP and NP this depends on the managers' choices; if all workers are permanent, there are no possible firing rounds left. The penultimate column reports the percentage of possible firing rounds in which firing actually takes place, the final column gives the absolute number of workers that are fired. Figure 4 complements the information in Table 4 by displaying the firing decisions over time in relation to how (minimum) effort evolves. Average and minimum effort levels are measured on the left axis, while the frequency of firing in a given round – i.e., the fraction of managers (out of 6) firing at least 1 worker – is measured on the right axis. For the FP and NP treatments the figure also depicts the fraction of managers that still have firing possibilities left.

Both the table and the figure show that firing decisions differ substantially between treatments. All managers in FS fire at least one worker in the early rounds (cf. Figure 4). This results in a steadily increasing average effort level and output, which diminishes the need to fire. After round 10 hardly any firing



Notes: The left axis presents the average effort and minimum 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 the managers that still have firing possibilities is also depicted.

Figure 4: Firing decisions, average and minimum effort over time

occurs. 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 both treatments firing decreases from the first half to the second half. 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 noisy productivity, even when worker behavior has more or less stabilized. A similar pattern in firings can be observed for the Medium contract. There, firing also occurs more often in the Noise than in the Full treatment.

Managers are far less effective in using the firing tool under the Probation contract than under the Spot and Medium contracts. In both FP and NP effort levels and output decrease over time. As will be discussed in more detail below, team performance especially deteriorates if some workers get a permanent status. Under Probation some managers fire simply to prevent workers from getting permanent status. In NP, managers fire at approximately the same rate as in FP if they have the possibility to do so. Even though workers receive permanent status, managers still fire more often under Probation than under the Spot contracts, both in terms of overall number of workers fired and in terms of firing possibilities used (see Table 4).

We next investigate the factors that make a manager decide to fire a worker. In particular, we investigate if managers use a threshold strategy of firing those workers whose productivity is below some threshold, and if managers change such a threshold during the experiment. Under the Spot and Medium contracts, most managers's behavior is broadly consistent with a threshold strategy. 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.[20] The difference in managers' threshold levels is weakly significant between the two treatments ($p = 0.06$). 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.[21]

The managers' firing strategy under the Probation contract is different, as per design there managers sometimes simply do not have the possibility to fire the worst performing worker. In total 30 and 27 workers out of 54 got permanent status in FP and NP respectively. Moreover, managers may want to fire well performing workers just to avoid that they become permanent. Some managers indeed follow this strategy of keeping at least part of their workforce purposefully on probation. In the Full treatment, there is one manager who successfully keeps everybody on probation (and by doing so, the group coordinates on an output of 7). In the Noise treatment, there are two managers keeping an eye on workers' probation phase. A third manager has 5 permanent

[20]We determined the threshold levels by checking the number of mistakes managers make if they fire according to a given presumed threshold level for the productivity. For example, if the presumed threshold is 5, the manager would fire workers if the productivity is below 5. However, he makes a mistake if he fires someone when the productivity is 6, or if he does not fire anybody when the productivity is 4. A manager's actual threshold is defined as the level that produces the fewest mistakes.

[21]For the Medium contract, it is harder to determine a manager's threshold, because it can depend on a worker's productivities in the last three rounds. Overall, managers' firing strategies do not seem to differ between the Spot and Medium contracts. The average relative effort of fired workers (compared to non-fired workers, based on matching group averages) equals 0.74 and 0.79 in treatments FS and NS respectively, with comparable relative effort levels of 0.72 and 0.75 for FM and NM.

Table 5: Regression results for individual effort as dependent variable

Dependent variable: Effort	Longterm	Spot	Medium	Probation
round	-0.05 (0.01)***	0.01 (0.01)**	-0.02 (0.01)	-0.03 (0.01)***
output_{t-1}	0.66 (0.13)***	0.33 (0.05)***	0.41 (0.13)***	0.44 (0.08)***
firing_{t-1}	-	0.40 (0.13)**	0.46 (0.23)*	0.60 (0.21)**
additional firing_{t-1}	-	-0.11 (0.07)	-0.06 (0.11)	0.08 (0.09)
firing_{t-1}*noise	-	-0.18 (0.11)	-0.15 (0.19)	-0.45 (0.28)
newly hired	-	0.18 (0.11)	-0.16 (0.19)	-0.22 (0.05)***
being permanent	-	-	-	-1.43 (0.38)***
permanent co-worker	-	-	-	0.10 (0.28)
constant	2.45 (0.36)***	4.66 (0.30)***	4.29 (0.45)***	4.24 (0.41)***
Number of panels	72	108	106	104
Avg # of obs per panel	29	19.3	19.7	12.5

Notes: ***: significant at the 1% level, and **: significant at the 5% level. Std errors are based on clustering at the group level, and are reported 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.

workers already in round 8 and keeps firing the remaining worker (however this is not efficient). 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 5 presents the results of a fixed effects panel regressions with individual effort levels as the dependent variable. To take account of the fact that individual efforts within a team are correlated, standard errors are based on clustering at the group level. We study each of the four contracts in a separate regression. The Longterm contract, under which firing is not possible, is included as a benchmark. Independent variables here are a time trend 'round', and team output in the previous round (denoted 'output_{t-1}'). In line with Figure 1 we observe that effort levels significantly decrease over time.

In the regressions for the Spot and Medium contracts we add the following independent variables: a dummy indicating whether there was any firing in the previous round, the interaction of this dummy with a noise dummy (1: Noise, 0: Full information), a variable measuring the additional number of firings in the previous round on top of the first one (i.e. equal to 1 if there were 2 firings, 2 if there were 3, and 0 otherwise), and a dummy equal to 1 only if the worker is newly hired in that particular round. The explanatory variable of main interest is 'firing_{t-1}'. The dismissal of another worker appears to have a significantly positive effect on a worker's effort choice for both contract types.

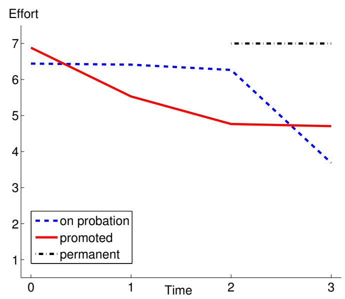
The interaction with the noise dummy shows no difference in this effect between Noise and Full information treatments. There is also no additional effect of firing multiple workers in a given round. Newly hired workers do not choose different effort levels than those already in the team. This suggests 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).

The final column in Table 5 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 does not have an impact. Under the Probation contract newly hired workers choose significantly lower effort levels than the non-permanent workers already in the team (note that newly hired and being permanent are mutually exclusive). This might be the case because of the overall decrease in effort levels, or because workers and unemployed subjects only observed output and not others' individual productivities (which precludes that newly hired can exactly mimic workers still on probation). So it is not the case that the selection of new members improves the team's production.

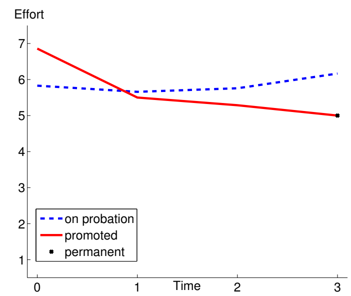
The regression results for the Probation contract suggest that there is a 'structural break' directly after a worker gets tenure. Indeed, workers who survive their probation phase immediately reduce their effort level on average by 1.1 units in FP and by 1.8 units in NP in the first round after being permanent compared to the last round of probation.[22] 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 5. 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 (solid red line) and of all the other, non-permanent workers (dashed blue line).[23] The two bottom panels consider the situation where there

[22]Under FP we have 30 permanent workers. After becoming permanent, 19 of these choose exactly the output of the previous round, 5 an effort below it, and 6 an effort level above it (3 out of these 6 decrease their effort level though). For treatment NP these numbers are 7, 3, and 17 (11), respectively. (Note that these numbers include all the permanent workers, not only those who received permanent status first in their group.)

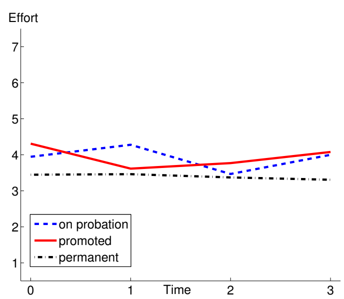
[23]Note that new permanent workers can also arrive to the group at time 2 or 3 if they get promoted just one or two rounds later than another worker in the treatment. This explains the third (dotted-dashed black) line in the upper left panel for treatment FP. As that panel depicts only the first worker getting permanent status in a group at time 1, the averages



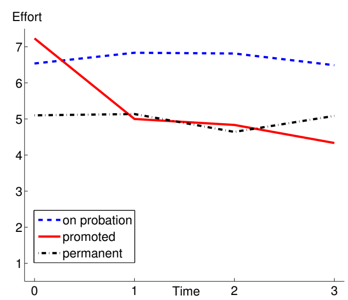
(a) First permanent workers - FP



(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 where 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 where 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 5: Effort decisions when getting promoted

are already other permanent workers around at time 0. These panels include a third (dotted-dashed black) 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 5, 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 or efficient effort during probation, i.e. those who appear to be exemplary workers (this is the case for 83% and 78% of the permanent workers in FP and NP, respectively). Remarkably, some of these workers strongly drop their effort level once they are assigned a permanent job. In FP, in one matching group three workers get a permanent job after round 5 in which they all three choose an

for 'permanent' in Figure 5a are based on a few observations (1 observation for time 2, 3 observations for time 3). These workers come from 2 groups, one that managed to maintain high effort levels, and another one that could not.

effort level of 7. One of these workers chooses effort 1 in round 6 and 7, and then tries a higher effort again, but at that point of time it is too late; another permanent worker also chooses 1, and this is the output till the end. In a second matching group, two workers get a permanent job in round 5 with effort levels equal to 6 and 7 respectively. The worker with effort level 6 then reduces effort to 1 in period 6, but then increases back to an effort of 5 in period 7. However, in that period the other permanent worker chooses 2. In the end this group manages to climb back to an effort of 4. In a third matching group, almost the same process as in the first group is observed. That is, 5 workers receive a permanent job after round 5, in which one chooses an effort level of 8 while all other choose an effort of 7. After all choose 7 in round 6, one worker chooses an effort of 1 in round 7, and tries to revert back in round 8. However, that is too late, another permanent worker chooses 1 in round 8, and all group members immediately jump to effort 1. Overall, the following picture emerges for these groups. The first workers to get permanent status are 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 and become the team's weakest link. This resembles Lazear's (2004) explanation for the "Peter principle". To paraphrase him by replacing his notion of ability with work motivation (Lazear, 2004, p. S146): "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."[24]

Possibly, these dynamics are affected by the scarce information provided to the workers. In all the treatments, workers are only informed of the minimum productivity in their group. The dynamics after some workers receive a permanent job look very similar to what is normally observed in a standard minimum effort game after subjects have played a couple of rounds. It is remarkable that after having been assigned a permanent job subjects proceed as if there is no shared history of playing the game cooperatively. The fear of exclusion is effective in getting people to cooperate, but it is not able to generate the trust that is needed for subjects to cooperate when the fear of exclusion has disappeared.

6 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 whether productivity was noisy or not.

Our design allowed us to address two main themes. First, we studied how important it is to maintain the fear of exclusion. Can managers afford (i.e.

[24]Among the other groups of FP, two groups manage to maintain high effort levels of 7 throughout the game: one group where all workers receive permanent jobs, and one group where nobody is assigned a permanent job. There is also one inefficient group that stays inefficient after the first workers get a permanent job (here 3 workers receive a permanent job in round 5 with effort levels 6, 6, 7 and output 2 – in the next round their efforts are 2, 4, 6, respectively.

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 tended to steadily deteriorate after some of its members have been assigned a permanent contract. Especially these members immediately reduced their effort upon getting a permanent status, an observation similar in spirit to the well-known Peter principle. Overall our results suggest that the fear of exclusion facilitates cooperation among people, but it might not be the right tool to teach people to trust each other. This is nicely illustrated in the probation treatment where trust collapses after the fear of exclusion is lifted, and groups cannot maintain high effort levels. We thus think that our experiment highlight an important drawback of probation contracts.

Second, we investigated the effectiveness of the fear of exclusion when performance is noisy. Noisy performance may dampen the role of the fear of exclusion, because cooperation is no longer supported in an equilibrium of the stage game and because it becomes harder for managers to distinguish the role of effort from luck. We find that workers were disciplined swiftly and teams performed efficiently when the manager had the discretion to fire workers in every single round. In stark contrast, when workers were completely protected from being fired, their performance gradually deteriorated and overall performance was substantially worse. These patterns emerge in the data independent of whether performance was noisy or not.

Our results offer substantial support for the incentive effect and only limited support for the selection effect of exclusion. The fear of exclusion is only effective as long as it is present. It is not the case that it helps select the right people who then continue cooperating after the fear is lifted.

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

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.

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 density of output if everybody chooses the same effort.[25] 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.[26] Using these two tables, we calculate expected losses from deviating to a lower effort level. Table 8 presents these losses.[27]

[25]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).

[26]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 $P_6(7|5) = q * (p+q)^5$.

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

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

From Table 8 we observe that the expected loss of deviating one unit downwards 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.[28]

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

$4 - P_6(7|5) \cdot 7 - P_6(6|5) \cdot 6 - P_6(5|5) \cdot 5 - P_6(4|5) \cdot 4 - P_6(3|5) \cdot 3) = 6.84.$

[28] Note that since workers never observe other workers' 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	-	-

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 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.[29]

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.[30]

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.

In this appendix we provided the analysis of the repeated game with an indefinite ending. In an experiment, it is impossible to literally implement a repeated game with indefinite length, because it is incredible that subjects will be kept in the laboratory after the building closes. To mimic a game with an indefinite length, we communicated to the subjects that the experiment would last at least 25 rounds and at most 40 rounds. We find it highly unlikely that subjects applied backward induction in our game. Formally, however, we cannot exclude that subjects apply backward induction from round 40. With backward induction, the same range of equilibria will survive when there is no noise (because the range of equilibria is already supported in the stage game).

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

[30] 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.

The analysis of the repeated game with noise would change though. Given that the stage game has a unique inefficient equilibrium, players should anticipate that this will be played in the last round. Applying backward induction, the result would then be that the inefficient equilibrium would also be played in all previous rounds.

A.3 Firing using a threshold strategy

We finally consider equilibria in which the manager plays an active role. Suppose that the manager chooses a threshold level P^* and fires workers who have a productivity lower than this threshold. 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 become unemployed forever after. Furthermore, we assume that all workers, including newly hired ones; use the same strategy, and that deviating from this strategy does not result in a change in behavior from the other parties. In particular, all workers choose effort level e^* independent of the history of the game, and the manager always fires workers with a productivity lower than P^* independent of the history of the game.

Under Full information, any threshold $1 \leq P^* \leq 9$ can be supported as equilibrium threshold. To see this, suppose the manager uses threshold P^*. From all symmetric effort profiles with $e \leq 7$ being an equilibrium of the stage game, it it immediately follows that all workers choosing the same effort level $e^* \leq 7$ weakly above P^* is an equilibrium. For effort levels 8 and 9 a sufficiently high discount factor is required.[31] With all workers choosing e^* using a threshold weakly below is a best response for the manager. If she would deviate to a strictly higher threshold, she would have to fire all workers which would reault in positive firing costs, but no future benefits. Note that in these type of equilibria, firing can not occur on the equilibrium path. Given that in all subsequent rounds all workers will choose e* again, firing serves no purpose and only brings about firing costs. Refraining from firing is then always strictly better for the manager, making firing a non-equilibrium response. To sum up, under Full information we constructed a simple type of equilibrium in which any threshold $2 \leq P^* \leq 9$ can be supported as equilibrium threshold level, and workers choose $e^* \geq P^*$.

Under Noise, we can construct similar equilibria. Now workers can make sure that they are never fired by choosing an effort level $e^* \geq P^* + 2$. In this case the only candidates for an equilibrium are of this form $e^* = P^* + 2$. To see this, notice that if $e^* > P^* + 2$, then a worker can deviate to $e^* - 1$ without risking being fired, resulting in a higher expected payoff (see Table 9). Since we assume that workers choose the same effort level irrespective of the history, this deviation is not punished, so in equilibrium we cannot have $e^* > P^* + 2$. Furthermore, the punishment of deviation is more severe in this case with the

[31] 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 by reducing effort cost. If $P^* = 8$, then $e^* = 8$ is supported in equilibrium if $\delta \geq \frac{1}{9}$. For $P^* = 9$ and $e^* = 9$ we need that workers' discount factor is at least $\frac{2}{9}$.

Table 10: Equilibrium strategies and workers' threshold discount factor under Noise

P^*	7	6	5	4	3	2
e^*	9	8	7	6	5	4
δ^*	0.844	0.472	0.632	0.668	0.708	0.752

manager than in the previous case of Appendix A.2 where the manager is absent and deviations are punished by the other workers using grim-trigger strategies. This is because now a deviator might become unemployed forever, resulting in a low payoff of 30 in every period. Table 10 lists the equilibrium effort levels, threshold levels and the threshold discount factor of the workers.[32] Note that these threshold discount factors are lower than the ones in Table 9 because of the more severe punishment. If workers have a higher discount factor than the one in Table 10, they do not have an incentive to deviate to a lower effort level. Deviating to a higher effort level is not beneficial as it only increases the expected output slightly, but has substantial additional effort costs. Turning to the manager, note that in this equilibrium she does not fire workers on the equilibrium path. Deviating to a lower threshold level does not change anything, as workers still perform better. Deviating to a higher threshold level does not change the output (by assumption workers stick to their effort level), but comes with a positive expected firing cost, so it is not a beneficial deviation. Thus, the equilibrium strategies described by Table 10 constitute an equilibrium for any discount factor of the manager.[33]

For the Medium contract, we assume that the manager uses a threshold strategy based on the minimum of the productivity levels in the last 3 periods. If this minimum is below the threshold, she fires the worker, otherwise she does not. Under Full information the same equilibria can be established as for the Spot contract. Under Noise, we have similar equilibria, though the equilibrium threshold discount factors might change, as the probability of detecting a deviation changes given the manager has more rounds to observe.

Finally, for the Probation contract under Full information essentially the same equilibria emerge as for FS. Only effort levels 8 and 9 cannot be sustained with these simple equilibria, as these effort levels require a potent firing threat (and thus no permanent workers). 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. In that case, workers would need to apply trigger strategies in order to maintain high effort levels. It is possible to construct equilibria where in a first phase all the workers get a permanent job (for example,

[32] The threshold discount factor was again calculated by looking at the expected payoff of adhering the strategy, and comparing it to the best deviation.

[33] It is possible to describe equilibria in which there is a positive probability of firing on the equilbrium path. Note however, that in order to enforce firing on the equilibrium path, workers have to punish a deviating manager, as firing has a strictly positive cost.

doing so can be a best response if the manager believes that all workers will switch to 1 if they are not offered permanent jobs when it is possible to do so) and then in the second phase an equilibrium of the Long term case is implemented similar to those in Appendix A.2. An alternative equilibrium in which firing does occur on the equilibrium path is where the manager uses as threshold strategy $P^* > 1$ but, on top of that, fires all workers every fifth round to avoid them becoming permanent (and choosing $e = 1$ onwards). Workers choose effort level e^* when they are on probation, and $e = 1$ when they are permanent. For e^* and δ sufficiently high the gain in future expected output then exceeds the immediate firing costs of 120.

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.

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:

Payoff worker = 50 + 20 * output - 10 * 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 Efficiency

In Section 5.1 we have observed that teams managed to perform better in terms of effort levels in the Spot treatments (and to a lesser extent in the Medium treatments) than in the Probation or Longterm treatments. In this Appendix we investigate whether this result extends to efficiency. Do managers use their power efficiently under Spot and Medium 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.

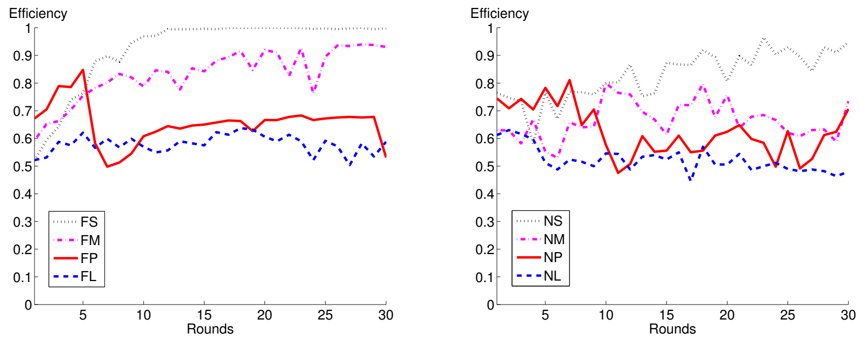


Figure 6: Average efficiency level over time for the Full information (left panel) and the Noise case (right panel)

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. Notice that in this analysis we exclude the earning of the unemployed. In this sense this appendix relates to the team's efficiency in the light of different firing regimes.

Figure 6 presents these relative efficiency levels over time per treatment. The figure shows that relative efficiency is highest in the Spot treatments, and in fact, it is almost always 1 for FS after round 12. Efficiency levels of the Medium contract are in between those of Spot and Longterm under both production processes, always closer to Spot than to Longterm (and under Full information further to Probation as well). Furthermore, efficiency is increasing in the first half of the experiment for the Spot and (to a lesser extent) for the Medium treatments, but is rather 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 11 presents the average efficiency levels (both including and excluding the manager's payoffs) 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 in the second half of the experiment. In contrast, no differences are found between the Spot and the Medium contract for Full information. Under Noise Spot marginally outperforms Medium. Furthermore, the

Table 11: Average relative efficiency

Panel A - average efficiency levels			
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**)
FM	0.77 (0.76)	0.90 (0.89)	0.05** (0.05**)
FP	0.65 (0.66)	0.66 (0.68)	0.92 (0.60)
FL	0.57 (0.57)	0.59 (0.60)	0.25 (0.12)
NS	0.76 (0.75)	0.90 (0.88)	0.03** (0.04**)
NM	0.66 (0.66)	0.68 (0.67)	0.46 (0.46)
NP	0.66 (0.66)	0.59 (0.60)	0.17 (0.12)
NL	0.54 (0.54)	0.50 (0.52)	0.35 (0.92)

Panel B - comparison across treatments									
	FS	FM	FP	FL		NS	NM	NP	NL
FS		0.63 (0.52)	0.26 (0.34)	0.05* (0.05*)	NS		0.08* (0.08*)	0.20 (0.23)	0.02** (0.01**)
FM	0.14 (0.09*)		0.42 (0.63)	0.11 (0.11)	NM	0.05* (0.04**)		0.75 (0.63)	0.04** (0.04**)
FP	0.03** (0.02**)	0.11 (0.11)		0.63 (0.52)	NP	0.01*** (0.01***)	0.52 (0.52)		0.05** (0.04**)
FL	0.03** (0.02**)	0.04** (0.04**)	0.75 (0.75)		NL	0.00*** (0.00***)	0.05* (0.05*)	0.42 (0.34)	

Notes: ***: significant at 1% level, **: significant at 5%, *: 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 ($n = 6$). In Panel B the above diagonal depicts differences in the first half, and below diagonal the differences in the second half.

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 and FM than in NS and NM in the second half of the experiment.[34] 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 is 0.997, in treatment NM is 0.73 in the second half of the experiment. Neither of these values are significantly different from the actual efficiency in FS and FM, respectively ($p > 0.19$).

If we exclude the manager's surplus from the efficiency measure, the same picture emerges (see Table 11). Keeping the production process fixed, workers gain a lot from Spot contracts. The possibility to always fire thus not only

[34] p-values are the following: $p^2_{FS-NS} = 0.00$, $p^2_{FM-NM} = 0.02$, all other p-values are at least 0.15.

benefits managers, but also the workers themselves.

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.

Bernard, M., Fanning, J., and Yüksel, S. (2016). Finding cooperators: sorting through repeated interaction. *Working paper.*

Berninghaus, S., Bleich, S., and Güth, W. (2013). The consequences of (de) regulation on employment duration and efficiency: An experimental study. *Economic Inquiry*, 51(1):1050–1065.

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

Brown, M., Falk, A., and Fehr, E. (2004). Relational Contracts and the Nature of Market Interactions. *Econometrica*, 72(3):747–780.

Brown, M., Falk, A., and Fehr, E. (2012). Competition and relational contracts: the role of unemployment as a disciplinary device. *Journal of the European Economic Association*, 10(4):887–907.

Charness, G. and Yang, C.-L. (2014). Starting small toward voluntary formation of efficient large groups in public goods provision. *Journal of Economic Behavior & Organization*, 102:119–32.

Chaudhuri, A., Schotter, A., and Sopher, B. (2009). Talking Ourselves to Efficiency: Coordination in Inter-Generational Minimum Games with Private, Almost Common and Common Knowledge of Advice. *The Economic Journal*, 119:91–122.

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.

Corgnet, B., Hernán-González, R., and Rassenti, S. (2015). Firing threats: Incentive effects and impression management. *Games and Economic Behavior*, 91:97–113.

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.

Dugar, S. (2010). Nonmonetary sanctions and rewards in an experimental coordination game. *Journal of Economic Behavior & Organization*, 73(3):377–386.

Falk, A., Huffman, D., and Macleod, W. B. (2015). Institutions and contract enforcement. *Journal of Labor Economics*, 33(3 Part 1):571–590.

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.

Loewenstein, G. (1999). Experimental economics from the vantage-point of behavioural economics. *The Economic Journal*, 109(453):25–34.

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.

Masclet, D., Noussair, C., Tucker, S., and Villeval, M.-C. (2003). Monetary and nonmonetary punishment in the voluntary contributions mechanism. *The American Economic Review*, 93(1):366–380.

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. (2016). Efficient Coordination in Weakest-Link Games. *Review of Economic Studies*, 83(2):737–767.

Salmon, T. C. and Weber, R. A. (2017). Maintaining efficiency while integrating entrants from lower performing groups: An experimental study. *The Economic Journal*, 127(600):417–444.

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.

Solda, A. and Villeval, M. C. (2017). Exclusion and reintegration in a social dilemma. *GATE Working Paper*, (1720).

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.