Watching or Not Watching? Supervision Technology and the Incentive Effects of Firing Threats

Jordi Brandts
Brice Corgnet
Roberto Hernán-González
José M. Ortiz
Carles Solà

This version: August 2020
(February 2018)

Barcelona GSE Working Paper Series

Working Paper nº 1023

Watching or Not Watching?

Supervision Technology and the Incentive Effects of Firing Threats

Jordi Brandts (IAE(CSIC) and Barcelona GSE)
Brice Corgnet (EM Lyon, GATE)
Roberto Hernán-González (Burgundy School of Business)
José Mª Ortiz (Middlesex University)
Carles Solà (City University of New York, College of Staten Island)

August 2020

Abstract

A common rationale for the use of salary contracts is that they can produce substantial incentive effects when coupled with firing threats. However, enforcing firing threats may require close supervision of employees, thus possibly offsetting the very reasons salaries are commonly used, such as lowering monitoring costs and granting autonomy to workers. We design a series of experiments to study the effectiveness of firing threats when only limited information is available to supervisors.. We show that light and unobtrusive supervision can produce large incentive effects. Compared to salary contracts alone, firing threats based on observing organizational performance alone increase workers' output by 70% whereas only observing how long an employee works doubles output. These findings show that salaries can produce large incentive effects even in the absence of intensive supervision. Finally, we show that salary contracts with firing threats perform at least as well as other popular incentive schemes (bonuses, individual and team incentives) that rely on a similar amount of information about workers.

JEL classification: C92, D23, D82

Keywords: Firing threats, incentives, monitoring, laboratory experiments

Acknowledgements: We thank the Spanish Ministry of Economics and Competitiveness through Grant: ECO2017-88130 and through the Severo Ochoa Program for Centers of Excellence in R&D (SEV2015-0563) and the *Generalitat de Catalunya* (Grant: 2017 SGR 1136) for financial support.

1. Introduction

Salary contracts are commonly used by organizations despite their supposedly weak incentive effects (Baker, Jensen and Murphy, 1988; Chiappori and Salanié, 2000; Laffont and Martimort, 2002; Bolton and Dewatripont, 2005). Many reasons have been advanced to explain the popularity of such compensation contracts which are not in any way tied to performance. One obvious argument for not rewarding performance directly is that it economizes on the *monetary* costs associated with employees' close supervision (Alchian and Demsetz, 1972; Demougin and Fluet, 2001). In addition, the use of salaries might avoid the *psychological* costs of monitoring linked to the negative impact of excessive control on employees' motivation (see Alge, 2001; Bénabou and Tirole, 2003; Falk and Kosfeld, 2006; Belot and Schröder, 2016). Easing supervision will increase workers' perception of autonomy thus enhancing their intrinsic motivation to complete the task (Deci and Ryan, 2000; Corgnet, Gunia and Hernán-González, 2020) and their overall job satisfaction (e.g. Fried and Ferris, 1987; Sousa-Poza and Sousa-Poza, 2000; Humphrey, Nahrgang and Morgeson, 2007). An illustration of the *psychological* costs of excessive supervision is the negative reaction of the public to the recent surge in the use of invasive technologies to monitor employees working from home (Lazzarotti, 2020).

Although important, autonomy is often not sufficient to guarantee high work output as is illustrated by empirical works showing low levels of effort of salaried workers (De Paola et al. 2014; Corgnet, Hernán-González and Rassenti, 2015, CHR henceforth). It is thus not surprising that the payment of fixed salaries is often coupled with implicit incentives, such as firing threats (Becker and Stigler, 1974; Klein and Leffler, 1981; Shapiro and Stiglitz, 1984; MacLeod and Malcomson, 1989). Firing threats have been found to be effective in boosting the work performance of salaried workers in both archival (Riphahn, 2004; Ichino and Riphahn, 2005) and laboratory studies (CHR; Falk, Huffman and MacLeod, 2015; Charness et al. 2017; Kopányi-Peuker, Offerman and Sloof, 2018; Dannenberg, Haita-Falah and Zitzelsberger, 2020). Related works using public goods games have also reported a positive effect of ostracizing group members on cooperation levels (Masclet, 2003; Cinyabuguma, Page and Putterman, 2005; Maier-Rigaud, Martinsson and Staffiero, 2010; Feinberg, Willer, and Schultz, 2014; Dannenberg, Haita-Falah and Zitzelsberger, 2020).

However, little is known about the conditions under which firing threats are effective incentive schemes. In particular, practitioners would want to know how much information supervisors need

to have to ensure that firing threats produce large incentive effects. This question is what motivated our study. If very detailed information was required, then combining salary contracts with firing threats might not be a viable option for companies because of the large monitoring costs. By contrast, if rather imprecise or limited information was sufficient to incentivize workers with firing threats, then such incentive schemes would be more appealing.

We used laboratory experiments to enable us to precisely vary the quantity of information available across the various firing treatments. The experimental methodology allowed us to assess the causal link between the quantity of information available to supervisors and the effectiveness of firing threats. We deployed an organizational context in which a boss was in charge of monitoring and firing workers who could choose between three activities in a number of consecutive production periods: doing a real-effort task, spending their time chatting or browsing the web. In order to isolate the incentive effect of firing threats, we introduced dismissals as the only incentive mechanism available to bosses so that workers were only paid a fixed salary which could not be made contingent on the information available to the boss.

To study firing threats we conducted four treatments in which the boss could fire workers under various informational conditions. In the firing treatments, the boss could fire one (out of nine) workers at the end of each of five periods (except for the first period). Firing treatments differed in the amount of information which could be collected by the boss during the monitoring of workers.

In the *complete information* treatment, bosses had access to real-time information about each worker's production as well as about the current activity workers were undertaking (either working, chatting or browsing the Internet). In the *partial information* treatment, bosses could not observe workers' production levels but could see the current activity each of them was undertaking. In the *minimal information* treatment, bosses could neither observe workers' production nor could they observe their current activity. They were only informed about the total production of the organization which was the only piece of information available to guide bosses' firing decisions. Thus, the only type of threats in this treatment were random *collective firing threats* according to which the boss could fire workers at random if group production was below a certain level. In the *peer information* treatment, all workers were endowed with real-time monitoring of their co-workers (as in the *complete information* treatment) but the boss could not

monitor. The only way a boss could access individual information about workers was by chatting with workers.

In the *baseline treatment*, the boss had complete information but could not fire anyone. We also conducted a treatment involving the payment of discretionary bonuses by the boss (the *discretionary bonuses* treatment) which was used to gauge the magnitude of incentive effects of firing threats with respect to an incentive scheme in which, unlike in the fixed salary baseline, performance could be rewarded.

The information conditions in which bosses cannot directly collect any individual information about workers (i.e. *minimal information* and *peer information* treatments) would be particularly challenging to study in the field because of their potential ineffectiveness. Our experimental paradigm thus offers a unique testbed for these unusual supervision and firing practices.

In our analysis we focus on the comparative statics of increasing the information available to bosses on organizational performance in line with the approach discussed in Schotter (2015). In our case our comparative statics Hypothesis follows directly from the *informativeness* principle (Holmström, 1979) which provides a ranking of the various treatments in terms of organizational performance. In particular, *minimal* and *peer information* treatments are expected to outperform the no-firing baseline while underperforming the *partial information* treatment, and the *complete information* treatment is expected to achieve the highest level of output. Importantly, the *informativeness* principle does not tell us about the size of these effects which remain a crucial empirical question.

In line with our Hypothesis, we find that workers' production and task dedication were significantly higher in all four firing treatments than in the baseline in which firing workers was not possible. In particular, organizational production was 70% higher in the *minimal information* treatment compared to the baseline. The *minimal information* treatment reached production levels similar to the *peer information* treatment which is the other supervision protocol in which bosses could not obtain any type of individual information about workers. The relatively large effect size of the *minimal* and *peer information* treatments (Cohen's $d \approx 0.63$; see Cohen, 1988) is especially striking because they do not rely on any individual information about workers. To our knowledge, these are the first results showing substantial incentive effects of *collective firing threats*. Organizational production in the *minimal (peer) information* treatment was, however, 27% (19%) lower than in the other two firing treatments (*partial* and *complete information*) which led to

similar levels of output. The incentive effect of supervision based on *partial information*, as measured by a comparison of organizational performance with the no-firing baseline, was very large (Cohen's $d \approx 1$). In addition to providing support for the *informativeness principle*, these findings show that firing threats are remarkably effective even when the supervision of workers is minimal.

Finally, we compare our results to three other commonly-used incentive schemes: bonuses, team and individual incentives (Gerhart, Rynes and Fulmer, 2009; Bryson et al. 2012). In the *discretionary bonus* treatment, a boss who had the same access to information about workers as in the *complete information* treatment would retain 10% of the production of all workers while assigning the remaining 90% to the nine workers at the end of each period. We also compared firing threats to team and individual incentives treatments using data from Corgnet, Hernán-González and Schniter (2015) (henceforth, CHS) and CHR. We show that even when firing threats rely on *partial information* they yield similar levels of performance as bonuses, team and individual incentives.

Our results suggest that firing threats are popular within organizations because they provide substantial incentive effects while economizing on supervision costs. This result is practically insightful because firing threats are likely to be prevalent in organizations in which fixed salaries are used and supervisors rely on limited information about workers' effort (e.g. Shapiro and Stiglitz, 1984). By assessing the effectiveness of firing threats across different levels of information, our work also answers the early call of Hart and Holmström (1986) to study the 'robustness' of incentive effects.

Finally, we show that our findings cannot be explained by non-informational mechanisms such as the crowding out of motivation due to excessive monitoring and social preferences.

2. Design

2.1. *Lab workplace*

Our computerized experimental environment consists of a virtual organization with one boss (referred to as C) and nine employees (referred to as Bs) and multiple periods following CHR.[1] In our setting, employees can, at any point in time during the experiment, complete a real-effort task,

[1] We chose to have ten people in each organization so as to represent a small company, which both in the EU and the US is typically defined as comprising at least 10 people.

access the Internet for leisure purposes or chat with other employees. Only one activity could be completed at the same time allowing the experimenter a precise measurement of the time spent on each activity by each subject.

A session consisted of five production periods of 20 minutes each. The length of the experiment was chosen so as to be able to observe fatigue and uncover incentive effects (e.g. CHS). The software allows for the boss to monitor employees' activities in real time and to track their experimental IDs across periods. We describe the main tasks briefly.

2.1.1. *The work task*

We use a *work task* that required a significant level of effort. All subjects, the employees and the boss, had to add up numbers from tables with 36 numbers for one hour and 40 minutes.[2] The reason for having bosses work on the same task as the employees was to allow them to assess its difficulty and, thus, to make firing decisions knowingly.[3] Each table completed correctly generated a 40-cent profit while a penalty of 20 cents was subtracted from individual production for each incorrect answer.[4] At the end of each period, the total amount of money generated by all ten subjects during the period was displayed in the history panel located at the bottom of their screens. We define *individual production* as the monetary amount generated by a given subject on the work task divided by the reward for a correct answer (40 cents). This measure can be interpreted as the net number of correct answers, discounted by the equivalent (in monetary terms) number of mistakes.

2.1.2. *Internet browsing*

At any point in time, subjects could access an Internet browser. Because of privacy concerns, subjects were explicitly told in the instructions that their usage of the Internet was strictly confidential.[5] Because Internet does not help workers complete the *work task* (CHS), it was a distraction and served as a measure of on-the-job leisure. This measure of on-the-job leisure was

[2] Different variations of this task have been used by Bartling et al. (2009), Dohmen and Falk (2010), and Abeler et al. (2011). A counting task that consisted of summing up the number of zeros in a table randomly filled with ones and zeros was also used in Falk and Huffman (2007). A long typing task was used in Dickinson's (1999) experiment for which subjects had to come during four days for a two-hour experiment. Falk and Ichino (2006) used a four-hour mailing task in their field experiment on peer effects. In another field experiment by Gneezy and List (2006), subjects were asked to enter data into a computer database for six hours.

[3] In the work task, subjects were not allowed to use a pen, scratch paper or calculator. This rule amplified the level of effort subjects had to exert in order to add up the matrices correctly.

[4] Penalties did not apply when individual accumulated production was equal to zero so that individual production could not be negative.

[5] Subjects were expected to follow the norms set by the university regarding the use of Internet on campus.

accurate in our setting because subjects could not access their personal phone during the experiment.

2.1.3. *Chatting activities*

The boss and all workers could also enter a chat room to communicate with one another during the experiment. All incoming messages were identified by the experiment *ID* of the sender. Subjects could direct their messages to all subjects or a subset of them.

2.1.4. *Monitoring activities*

In the *partial* and *complete information* treatments as well as in the baseline and *discretionary bonus* treatment, the boss could monitor the nine workers during the experiment. Depending on the treatment, the boss received information in real time about the activities undertaken by the selected subject, their current total production, as well as their contribution to the *work task* (in % terms). This platform, which allows monitoring in real time, was designed to capture the main features of the supervision tools used in firms (e.g. Camarinha-Matos, Afsarmanesh and Ollus, 2008; Lazzarotti, 2020). At the end of each period, the boss had access to a monitoring summary, depending on the treatment, which included information regarding employees' activities during the period, their production levels as well as their contribution to total production.

In addition to working, browsing the web or chatting, subjects could collect a steady inflow of cash by clicking on a box moving slowly at the bottom of their screen. Each period, subjects could thus earn \$2.40 just by clicking on the boxes.[6] This feature was used to mimic the pay workers obtain for being present at their workstation regardless of their commitment to the *work task*. Actually, subjects could click on the boxes even when browsing the Internet. We added this task to alleviate the common issue of *active participation* in laboratory experiments, which was raised by Lei, Noussair and Plott (2001) in the context of asset markets. In the context of our lab workplace, CHS (page 285) stress that "*subjects may engage actively in a focal work task because of expectations, rewards, and lack of desirable alternatives*". When desirable alternatives are present, active participation in effortful work may be traded off to some degree, revealing subtle incentive effects.

[6] The box appeared at the bottom of a subject's screen every 25 seconds independently of whether the subject was currently working on the work task, chatting, or browsing the Internet.

2.2. *Treatments*

Table 1 summarizes the main features of our treatments together with the number of subjects in each cell.

Table 1.
Summary of the treatments.

Treatment	Description	Number of sessions (subjects)
No Firing Complete info (*Baseline*)	Employees were paid a fixed wage of 200¢ per period. The boss kept the value of all output produced by all employees in the organization. In addition, bosses were paid the value of their own production. Bosses could monitor employees' activities and individual production but had no possible recourse.	6 (60)
Firing *Complete info*	The boss could monitor employees' activities and individual production, and could fire one employee at the end of periods 2, 3 and 4. Payment to workers was the same as in in the baseline but the boss also kept the fixed wage of dismissed employees.	6 (60)
Firing *Partial info*	Same as Firing *complete info* except that bosses could *only* monitor employees' activities not accessing any information regarding their individual production.	6 (60)
Firing *Peer info*	Same as Firing *complete info* except that bosses could *not* monitor employees and thus only had access to the total production of the organization when deciding upon firing employees. In addition, workers could monitor each other.	5 (50)[7]
Firing *Minimal info*	Same as Firing *complete info* except that bosses could *not* monitor employees and thus only had access to the total production of the organization when deciding upon firing employees.	6 (60)
Discretionary bonus Complete info	The boss kept 10% of the value of all output produced by all employees in the organization. In addition, bosses were paid the value of their own production. Bosses could monitor employees' activities and individual production and distributed the remaining 90% of the total production of the organization to the nine employees at the end of each period.	6 (60)

[7] In one of the six sessions, the server crashed and the information could not be completely retrieved from the log file. We thus cannot include a thorough analysis of this session in the paper. We thus decided to drop all the data for this session.

In all treatments, employees were rewarded a fixed wage of 200¢ each period; they were not incentivized based on their performance on the *work task*.[8] In all treatments except the *discretionary bonus* treatment the bosses received the output produced by all subjects (including themselves) on the *work task*, but was not paid a fixed wage. In all firing treatments, the boss could fire one employee at the end of each of periods 2, 3 and 4.[9] The boss kept the fixed wage of dismissed employees in the following periods. Our aim was to conduct a conservative test for the effects of firing threats on employees' production by considering that only one of them could be fired each period. Hence, up to one third (three out of nine) of employees could be fired in a given experiment.

Dismissed employees could only browse the Internet. They were rewarded solely for their earnings on the clicking task which were reduced to 1¢ per box instead of 5¢ per box for the *active* employees and the boss.[10] They were not able to chat with the members of the organization, and they could not be rehired. The *complete info* treatment aimed at representing the case in which bosses have access to exhaustive information about workers' performance and work dedication. This treatment replicates CHS. In practice, employers have access to less information and our aim is to assess the size of incentive effects in those circumstances. We start by considering the *partial info* treatment which corresponds to the more realistic case in which bosses can only assess how much time workers dedicate to the task without knowing the precise contribution of each individual worker. This treatment relates to the case envisioned by Kopányi-Peuker, Offerman and Sloof (2018) in which a team member could be permanently excluded of a weak-link game by a boss who observed noisy information about workers' productivity levels. However, in our *partial info* treatment, unlike in Kopányi-Peuker, Offerman and Sloof (2018), the boss does not observe any direct measure of productivity. The information collected by the boss in the *partial info* treatment is likely to be much less costly to collect than in the *complete info* treatment.

In the two remaining supervision treatments (*Firing Peer info* and *Firing Minimal info*) bosses do not have access to any individual information about workers, which is the case in which

[8] The choice of 200¢ was made so that, at least some *employees* would not be able to produce that value thus inducing the boss to fire workers. This value was calibrated using previous related experiments (e.g. CHR).

[9] We do not allow for firing in period 1 because of the large learning effects observed in the great majority of real-effort experiments that makes the first period substantially different from the rest of the experiment (e.g. see Charness and Campbell, 1988).

[10] As a result, the maximum period earnings of dismissed subjects on the clicking task were equal to 48¢ instead of 240¢ for *active* employees.

supervision is inexistent and thus entail no costs. The only piece of information available to the boss (total output of the organization) is observed at the end of the period without the need for monitoring workers. In the same vein, the boss cannot monitor workers in the *peer info* treatment. However, workers can monitor each other with the same supervision technology as in the *complete info* treatment. This treatment aims at capturing a common situation in which monitoring is performed at the level of the team instead of being centralized in the hands of the boss (see e.g. Grosse, Putterman and Rockenbach, 2011).

Finally, we consider a *discretionary bonus* treatment so as to compare the performance of firing treatment to an incentive scheme in which, unlike the baseline, employees could be rewarded on their individual performance. In this treatment, the boss kept 10% of organizational production and distributed the rest to the nine employees at the end of each period. Bosses could monitor employees as in the *complete info* treatment.

2.3. *Survey data*

For each session, we collected survey data for a number of items (see Appendix O.1). This information was used to provide controls for our statistical analysis.

Adding skills. Subjects were asked to sum as many sets of five one-digit numbers as they could during two minutes in the spirit of Dohmen and Falk (2011). Each correct answer was rewarded with 10 cents. The number of correct answers is what we refer to as "ability". To ensure that this measure was not affected by fatigue and treatment differences, it was collected upon arrival at the lab and before receiving instructions for the corresponding treatment.

At the end of the experiment, subjects were asked to fill out a 10-minute survey including questions regarding demographics, cognitive skills and social preferences. We collected these measures at the end of the experiment because they are less central to our study than adding skills and were not planned to be used as main controls in our analysis.

Demographics. We asked subjects about their name, age and gender. We also asked them how many hours a week they usually worked for pay or volunteer. We also collected data regarding which degree they were currently studying.

Cognitive skills. We measured cognitive reflection using the CRT developed by Frederick (2005). Our CRT measure sums the number of correct answers on the test.

Social preferences. We elicited social preferences following Bartling et al. (2009) and Corgnet, Espin and Hernan-Gonzalez (2015). We asked subjects to make six choices between two possible

allocations of money between themselves and another anonymous and randomly assigned subject in the experiment. In each experimental session, two subjects and one of the six decisions were selected at random for payment. The choice of the first subject in the selected decision was used to allocate payoffs between the two subjects. All decisions were anonymous. The allocation decisions are described in Table O1 in the online Appendix O1.

2.4. *Procedures*

Our subject pool consisted of students from two major Spanish Universities. The experiments took place between December 2014 and June 2016. In total, 240 subjects participated in the experiments, divided into 24 groups of 10 subjects each, that is six groups for each treatment. All of the interaction was anonymous. Subjects had 20 minutes to read the instructions (see Appendix O2) on their screens. Three minutes before the end of the instructions period, a monitor announced the time remaining and handed out a printed copy of the summary of the instructions. None of the subjects asked for extra time to read the instructions. The interaction between the experimenter and the subjects was negligible.

At the end of the experiment, subjects were paid their earnings in cash, rounded up to the nearest quarter. Individual earnings at the end of the experiment were computed as the sum of all earnings in the *5* periods plus the earnings from the adding and social preferences tasks included in the questionnaire. Subjects playing the role of an employee (boss) in the *complete*, *partial*, *peer*, *minimal info, discretionary bonus*, and *baseline* treatments earned €29.36 (€97.47), €28.06 (€95.58), €28.24 (€83.81) €29.21 (€76.54), €35.51 (€26.64), and €29.09 (€54.26) on average, respectively. This includes a five euro show-up fee. Experimental sessions lasted on average two hours and thirty minutes.

3. Hypothesis

In what follows we lay out our main hypothesis regarding the comparative statics of the performance of the firing treatments. The ranking of the treatments is based on applying the *informativeness* principle (Holmström, 1979) to our context. The *informativeness* principle puts forth that every piece of information can be used to tie pay and performance more closely thus strengthening incentives. In our case, information helps the boss dismiss those workers who exert low levels of effort. This threat is especially effective when the boss can observe with precision the individual production of each worker as well as their dedication to the task and the amount of

time spent on leisure activities. This case is captured in the *complete information* treatment in which the boss has access to extensive individual information about workers' output and effort levels. In this treatment, we expect firing threats to be especially powerful in incentivizing workers because shirkers would very likely be identified in which case they would be dismissed and suffer a monetary loss.

The effectiveness of the firing threats is expected to be lower when the boss can only observe some of the information. In the *partial information* treatment, the boss cannot observe the exact level of output of each worker although they can observe the time spent on the work task. However, this measure is only a proxy for work effort as people can spend time at their work station while not effectively exerting effort. We thus expect firing threats to be less effective in enhancing work output under *partial information* than under *complete information*.

In the *minimal information* treatment, firing threats are weak because the boss cannot collect any individual-level information. Instead, the boss observes the aggregate production of the organization. The boss could still attempt to obtain information about each worker's performance via chatting. However, it is unlikely that workers will truthfully reveal their performance in a context in which they can be fired at the end of the period for underperforming. It follows that any threat of dismissal cannot be targeted at the least productive individual. A boss can still threaten workers by dismissing one of them at random when the level of performance of the organization is excessively low. In that context, workers might decide to avoid shirking to prevent the boss from firing one of the workers which would lead to their dismissal in 1 out of 9 cases.[11] One additional caveat with respect to the effectiveness of the random threats is that they might not be credible. Indeed, it is not necessarily beneficial for the boss to fire a worker at random because they might inadvertently dismiss high producers. The threat can, however, become credible if bosses reciprocate negatively to any shirking behavior of workers. In that context the boss would punish workers by firing one of them at random even at the risk of a cost to the company (Fehr and Gächter, 1998, 2000).

Even though this type of *collective firing threats* could, in theory, be effective, we are not aware of any empirical evidence. In any case, firing threats will be weaker in the *minimal*

[11] This the case in which no workers had yet been fired. This probability increases as the size of the organization shrinks over time.

information treatments compared to the *partial* and *complete information* treatments in which the boss can implement threats at the individual level.[12]

Finally, we expect the *peer information treatment* to achieve a level of performance similar to the *minimal information* treatment. This is the case because the boss will not observe directly any individual of information about workers, exactly as in the *minimal information* treatment. A crucial difference between the two treatments is that, under *peer information*, the boss could instruct a worker to collect information about other workers. However, since workers are directly competing for keeping their employment, they may downplay the performance of other workers thus impeding the boss to gather any reliable information. This behavior is especially likely because any information about workers' individual performance cannot be verified a posteriori by the boss. We thus predict that the *peer* and *minimal information* treatments will lead to similar performance.

We summarize our expected performance ranking of the different treatments in the following Hypothesis.

Hypothesis:
Workers' output will be the highest in the complete information treatment followed by the partial information treatment. The minimal and peer information treatments will lead to lower output than the complete and partial information treatments while outperforming the baseline.

In addition to testing the performance ranking Hypothesis, we will conduct an exploratory analysis by comparing the various firing incentive schemes to popular incentive schemes such as bonuses, individual and team incentives.

4. Results

In sections 4.1 and 4.2 we analyze production, work dedication and firing decisions in the first four of the five periods of our experiment which correspond to the periods in which firing threats had monetary consequences. In Appendix A, we report results when pooling data for all five periods (see Tables A.3 and A.4). In section 4.3 we compare our results with those of other popular incentive schemes. In section 4.4, we assess whether non-informational mechanisms could explain our findings.

[12] The *partial* and *complete information* treatments can also allow bosses to implement collective firing threats because they will observe, as in the *minimal information* treatment, the total output of the organization.

4.1. *Employees' production and work dedication*

At every step of the analysis, we only include subjects who belong to the organization in each period excluding fired subjects. This means that in the three treatments where firing was possible, subjects who had been fired before an actual period were excluded from the analysis. However, none of the results are qualitatively affected by including the fired subjects in the analysis and considering their production to be zero.

Figure 1 shows the average production of employees in the first four periods separately and jointly (see '1 to 4' in the figure). One can see that for periods 1 to 4, the rank of treatment in terms of production levels is consistent with our *Hypothesis*. In particular, all the firing treatments outperform the baseline and the *partial* and *complete info* treatments perform better than the *minimal* and *peer info* treatments. We also observe that the production level for *peer info* is higher than *minimal info*. In addition, production levels are higher in *complete info* than in *partial info*. The ranking of production is, hence, consistent with our *Hypothesis*.[13]

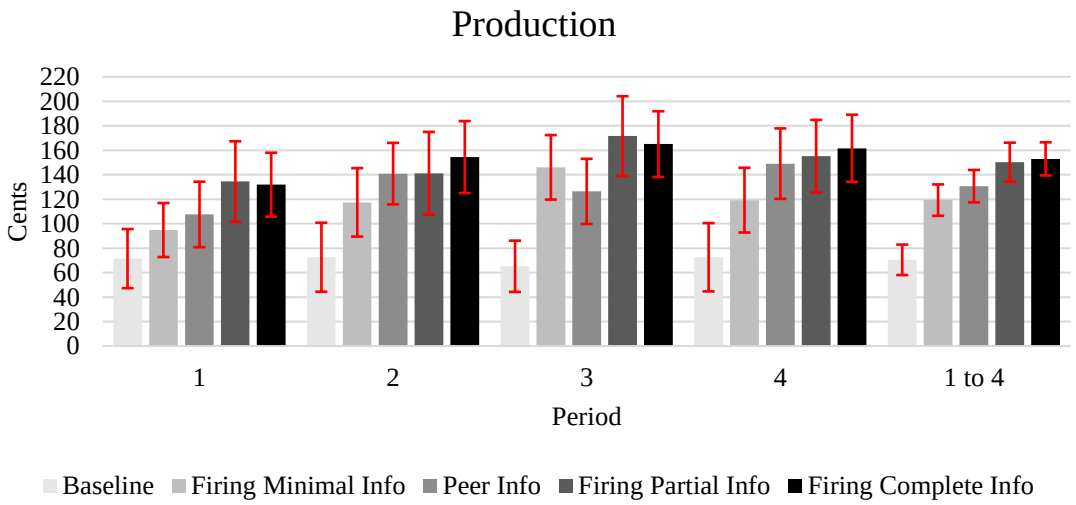

Figure 1. Employees' average production across treatments for periods 1 to 4. Subjects who have been fired before a current period are excluded. The bars show 95% confidence interval.

We use a GLS random-effects model (see Table A.1 in Appendix A)[14] to check for any statistically significant differences in employees' average production across treatments. Given the nature of our data, we use panel regressions and robust standard errors clustered at the session

[13] A similar pattern can be observed for each individual period except for some exceptions in period 3.
[14] Appendix B contains several tables with a robustness analysis of the regressions reported in the main text.

level.[15] We use random effects in all the regressions reported in this paper. Using the Breusch-Pagan Lagrange Multiplier test, we cannot reject the random effects specification.

The results show that individual production for employees is significantly higher in treatments where firing is available compared with the *baseline* without firing threats. Production is 70% higher under minimal information than in the baseline. We also find that *complete* and *partial* information yield higher production than *minimal* information. Average production of employees is about 30% higher in the firing treatments in which the boss has access to individual information about employees (*complete info* and *partial info* treatments) compared to the *minimal info* treatment. This difference is statistically significant (see the last column in Table A.1). The *peer info* treatment ranks in between the *partial* and *minimal info* treatments in terms of performance although it is not significantly different from any of those (see Table A.1). Likewise, average production in the *complete info* (152.8) and *partial info* treatments (150.4) do not differ significantly. In sum, our finding are largely consistent with our Hypothesis. As predicted, *complete* and *partial info* yield higher output than *minimal info*. *Minimal* and *peer info* treatments do no significantly differ, and all firing treatments outperform the baseline. Most surprisingly, the *partial info* treatment does not significantly underperform the *complete info* treatment, thus suggesting imprecise information about workers produce substantial incentive effects.[16]

In addition to production levels, we also report working time for the first four periods as these two measures can lead to different results if employees exert non-productive effort (being present at the workstation without completing the task). Figure 2 shows working time defined as the percentage of time employees spent on the task screen instead of browsing the web or chatting with other subjects. Working time is thus a measure of work dedication, which negatively correlates with on-the-job leisure which can be calculated as the time spent browsing and chatting. In all the firing treatments but *peer info*, chatting and browsing are the only alternative activities to the work task for employees.

[15] Following Cameron and Miller (2011), we also estimated standard errors using the wild bootstrap procedure. Using this procedure, we obtained very similar p-values to the ones reported in the results section. In particular, the effects which are shown to be statistically significant using robust standard errors continue to be significant when using the wild bootstrap procedure.

[16] The magnitude of the firing threats vary across periods. Being fired in the second period entails larger costs in terms of foregone wages than being fired at the end of the fourth period. However, as is shown in Table A.3, workers' production is almost identical in periods 2 and 4 across all treatments. It follows that the positive effect of firing threats is as pronounced in period 2 as in period 4, regardless of the firing treatment.

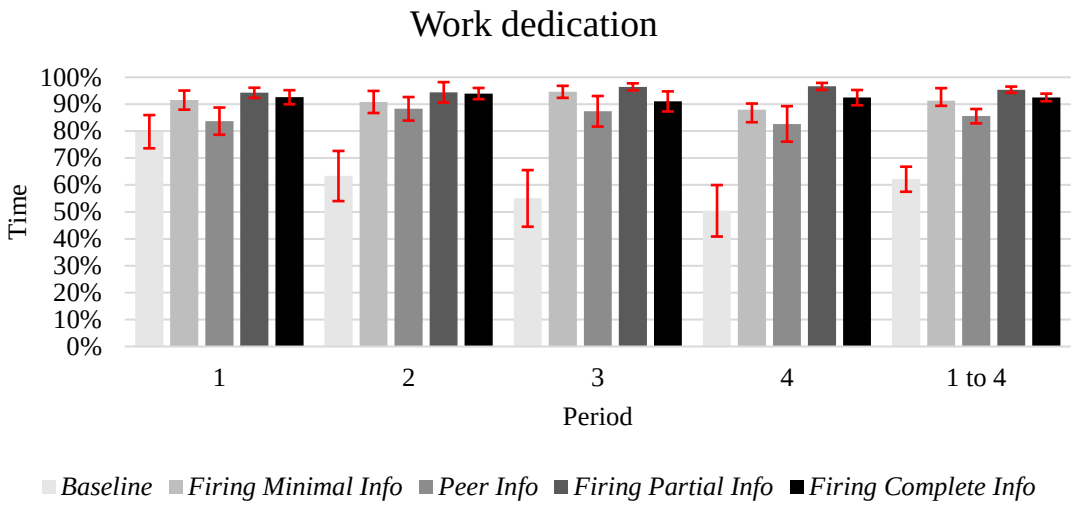

Figure 2. Employees' average working time (%) across treatments for periods 1 to 4. Subjects who have been fired before an actual period are excluded. 95% confidence interval bars.

In Figure 2 we observe that employees spend significantly less time on the work task in the *baseline* (62.13%) than in any of the four firing treatments (91.17%) (see Table A.2 in Appendix A for the statistical analyses of overall treatment differences and Table A.7 for changes of work dedication over time in the different treatments). In line with the production results, we do not find any significant differences in the time spent on the task between the firing treatments with *complete info* (92.51%) and *partial info* (95.38%) (see Table A.2 in Appendix A).[17] In line with the production findings, the time spent on the work task was higher in the *complete info* and *partial info* treatments compared to the *minimal info* and *peer info* treatment (91.25% and 85.55%, respectively). The differences in working time across firing treatments are, however, of limited magnitude and only significant when comparing *partial info* and *minimal info* treatments (see Table A.2) and when comparing *peer info* with the rest of the firing treatments. This last finding is not surprising because employees had access to an additional alternative activity (monitoring) to the work task in the *peer info* treatments which occupied 2.57% of their time on average.[18] Employees also spent more time in the chat when we compare *minimal + peer info* treatments and *complete + partial info* treatments (see Table F.4)

[17] We do not present treatment comparisons for bosses in the main text because of their limited number per treatment. However, we show these results in Tables A.3 and A.4 in Appendix A.

[18] When we add work time and monitoring time together, there are statistical differences between *peer info* and *complete* and *partial info* treatments, p-value =0.051 and p-value = 0.010, respectively. However, the difference disappears when we compare *peer info* with *minimal info*, p-value = 0.230 (The p-values were calculated using GLS regressions as in Table A.2 with work time plus monitoring time as a dependent variable).

Finally, subjects could also obtain earnings from clicking on yellow boxes appearing every 25 seconds at the bottom of their screen. No significant differences were observed across treatments regarding the clicking task. Subjects successfully clicked on the box in 96.46%, 93.77%, 93.27% 92.59% and 94.25% of cases in treatments *complete info, partial info, peer info, minimal info* and *baseline*, respectively.

Overall, our findings show that bosses who had access to individual employee information when making firing decisions achieve higher levels of performance than those who could only observe organizational performance. In addition, the size of the effects appears to be substantial. In particular, the incentive effect of firing threats based on either *partial* or *complete* information is very large (Cohen's $d \approx 1$). This extends the findings of Kopányi-Peuker, Offerman and Sloof (2018) regarding the positive effect of exclusion in a chosen-effort weak-link game in which the boss only observed a noisy measure of employees' decisions. In particular, our *partial info* treatment is strikingly effective despite the fact that the boss does not have access to any individual productivity data about employees.

Even more surprising is the relatively large incentive effect of the *minimal* and *peer information* firing treatments (Cohen's $d \approx 0.63$) which do not rely on any individual information about workers. To our knowledge, this constitutes the first evidence of the incentive effects of *collective firing threats*. Our results also put forth that firing threats are 'robust' incentive schemes (Hart and Holmström, 1986) which, regardless of the information available to the boss, effectively increase workers' production.

4.2. *Firing decisions*

In this section, we analyze bosses' firing decisions in the four firing treatments to see whether firing patterns are consistent with the *informativeness* principle which underlies our *Hypothesis*. In Table 2, we show the average production of fired and non-fired workers per period and per treatment. Focusing on the last column which shows total production, one can see that in the *complete, partial* and *peer info* treatments fired worked were producing significantly less than the rest of workers prior to being fired.

By contrast, in the *minimal info* treatment there is no significant difference in the average production levels across fired workers and those who were not fired. These findings are consistent with the *informativeness* principle. That is, bosses fired the least productive individuals in all

treatments in which individual information about workers could be accessed, whereas they had to fire at random when no such individual information was available.

In the *peer info* treatment, the fact that fired employees had lower productivity levels than those who were kept in the organization suggests that some reliable information was effectively transmitted between monitoring peers and the boss via chat.[19] This might explain why firing threats were about 10% more effective in boosting workers' production in the *peer info* treatment compared to *minimal info* although this difference is not statistically significant (p-value = 0.32, Wilcoxon rank-sum test). The fact that bosses in the *peer info* treatment were able to fire low-productivity employees, despite not being able to observe them directly, suggests chatting activities might have permitted the communication of some relevant productivity information in that treatment. Although the total number of messages per session that were sent in the *peer info* treatment (144.6) was about a third of the baseline (399.7), it was more than twice higher than in the other firing treatments (61.7). In addition, bosses were actively using chat to obtain information from employees. In particular, 48.21% of all boss messages in the *peer info* treatment were sent to a single recipient compared to 10.89% in the baseline and 19.95% in the other firing treatments (see Table E.2 in Appendix E for detailed information on chatting across treatments). Employees were not fully responsive to the call of managers, however, as only few messages were dedicated to reveal information about other employees (1.4) or themselves (5.3).[20] Thus, only a limited amount of reliable information could have been transmitted to the boss via chat. This explains why we do not observe significant differences in production levels between the *peer* and *minimal info* treatments.

It is also worth noting that only a few employees were actually fired (less than one third of the number of employees that could have been fired). This is in line with our results showing that firing threats induce high organizational performance compared to the baseline which, in turn, implies that the boss does not need to fire workers.

[19] In Table 1, one can see that production differences (as well as significance levels) between fired and non-fired workers were higher for the *complete* (496.65¢) and *partial info* (591.66¢) treatments compared to *peer info* (428.21¢).
[20] For the other firing treatments these numbers of messages were slightly lower: 0.2 and 4.42, respectively. This implies that almost no information was shared by employees about the performance of their coworkers.

Table 2.
Firing decisions per period across treatments.

		Period 2	Period 3	Period 4	Total
	Total [maximum possible] number of fired subjects	2 [6]	2 [6]	3 [6]	7 [18]
	Average production of employees before being fired	20	40	66.67	271.43
	Average production of other employees	159.61	170	167.66	768.08
	p-value†	2.112 (**0.0347**)	1.789 (**0.0735**)	1.765 (**0.0776**)	3.220 (**0.0013**)
	Total [maximum possible] number of fired subjects	2 [6]	3 [6]	1 [6]	6 [18]
	Average production of employees before being fired	0	100	0	156.67
	Average production of other employees	146.54	175.92	158.33	748.33
	p-value	2.209 (**0.0272**)	1.043 (0.2968)	1.595 (0.1107)	3.469 (**0.0005**)
	Total [maximum possible] number of fired subjects	1 [5]	2 [5]	3 [5]	6 [15]
	Average production of employees before being fired	120	0	26.67	213.33
	Average production of other employees	141.36	132.38	151.28	641.54
	p-value	0.232 (0.8163)	2.098 **(0.0359)**	2.311 **(0.0208)**	3.040 (**0.0024**)
	Total [maximum possible] number of fired subjects	1 [6]	1 [6]	2 [6]	4 [18]
	Average production of employees before being fired	80	120	170	211.90
	Average production of other employees	118.11	146.54	117.2	530
	p-value	0.258 (0.7962)	0.197 (0.8441)	-1.099 (0.2719)	1.041 (0.2979)

†This p-value refers to the Wilcoxon rank-sum test that assesses whether average production is the same for subjects who were fired and for those who were not fired.

4.3. On the comparative performance of firing threats and other common incentive schemes

The large incentive effects of firing threats pose the question of their relative effectiveness compared to other popular incentive schemes such as individual and team incentives as well as discretionary bonuses (Nalbantian and Schotter, 1997; Gerhart, Rynes and Fulmer, 2009; Bryson et al. 2012). For the sake of these comparisons, we conducted an additional treatment with discretionary bonuses (see Table 1). In addition, we use the data from the studies of CHS and CHR that employed the same real-effort task setup, thus allowing for a direct comparison with our findings above. The experimental setup in these studies was exactly the same as the current design except for the incentive schemes that were used. In CHS, a team incentives treatment was conducted in which the production of the ten workers was equally shared among them as well as an individual incentive treatment in which all ten workers were paid exactly the amount they individually produced. In CHR, an individual incentive treatment was also conducted along with two of the treatments in the current study (*baseline* and *complete info*).[21]

We start by showing that there were not statistical differences in production levels between our data and CHR for the *baseline* and *complete info* treatments (Appendix G). In addition to providing a valuable replication of CHR findings, this reassures us that the comparison of incentive schemes across datasets is meaningful.

We find that firing threats with individual information (*complete* and *partial info* treatments) do not significantly underperform any of the incentive schemes which also make use of individual information about workers such as individual incentives and discretionary bonuses (see Tables F.1 and F.2a,b,c in Appendix F, first two columns).[22] This result is especially notable for the *partial info* treatment which, despite using less information than individual incentives and discretionary bonuses, achieves a similar level of performance. In addition, both *partial* and *complete info* treatments outperform team incentives (see Table F.3, first two columns). Firing threats which cannot be based on reliable individual information about workers (*minimal* and *peer info* treatments) significantly underperform individual incentives and discretionary bonuses (see Tables

[21] In both treatments, workers were also given a fixed wage of 200¢ as in the current study. That is, the only difference between the individual incentive treatments in CHS and CHR is that in the latter workers were paid a fixed wage in addition to the piece rate.

[22] See also figures F.1, F.2 and F.3 for graphical comparisons comparing production levels in the first five treatments separately with the discretionary bonus treatment (F.1), the treatments from CHR and CHS (F.2) and the team incentives treatment (F.3).

F.1 and F.2a,b,c in Appendix F, last two columns). However, they lead to higher production levels, although not significantly so, than team incentives (see Table F.3, last two columns).

Overall, firing threats thus never underperformed incentive schemes that relied on the same type of information.

4.4. *Non-informational mechanisms*

In this section, we explore whether alternative mechanisms to the *informativeness* principle could account for our findings. In particular, we focus on the crowding out effect of intensive monitoring and on the positive effect of social motives on workers' output.

4.4.1. *Monitoring intensity*

Because firing treatments vary in the type of supervision they rely on, they might also differ in the intensity of monitoring of workers are subject to. For example, the *complete info* treatment may have generated an excessive amount of monitoring that was detrimental to organizational production as employees could have perceived this intense supervision as distrust (e.g. Dickinson and Villeval, 2008). This might then explain why the *complete info* treatment does not significantly outperform the *partial info* treatment. However, we have several pieces of evidence that do not support the monitoring-intensity explanation of our findings. First, the fact that bosses fired employees according to their relative performance levels in both the *complete* and *partial info* treatments (see Table 2) suggests that monitoring employees may have been as intensive in the *partial info* treatment as in the *complete info* treatment. Indeed, bosses spent about the same time monitoring in the *complete info* firing treatment (14.58%) as in the treatment with *partial info* (10.40%) (see Appendix A, Table A.5. This difference was not statistically significant (see Table A.6). In addition, this reasonable amount of monitoring does not seem to correspond to a case of high monitoring intensity as is described by Dickinson and Villeval (2008) and which can entail distrust and crowding out of motivation (Alge, 2001). Besides, the negative effect of monitoring identified by Dickson and Villeval (2008) only appeared when workers had friendship ties, which is not the case in our experiments.

Finally, we show that the monitoring activities of the boss did not have a negative effect on subsequent workers' production. In Appendix C, we present additional analyses that show that a worker who was being watched by the boss in a given time span of five minutes did not significantly modify his or her own production in the next time span of five minutes in both the *partial info* (p-value = 0.366 for the dummy variable that takes the value 1 if a worker has been

watched in the previous 5 minutes) and *complete info* (p-value = 0.934) treatments (see Table C.1 in Appendix C). In sum, our findings do not seem to be consistent with an explanation based on excessive monitoring and control. Instead, we posit that the lack of statistically significant differences between the *complete* and *partial info* treatment might simply be due to ceiling effects. This interpretation is consistent with the fact that the organizational output in the *partial info* treatment was already similar to the case of high-powered incentive schemes such as individual incentives and discretionary bonuses (see Section 4.3).

4.4.2. *Social motives: Team identity and fairness concerns*

The presence of social motives is a potential explanation for why performance is high in the *minimal* and *peer info* compared to the baseline. One notable difference across our firing treatments is that firing decisions and thus workers' compensation are expected to be driven by total production in the *minimal* and *peer info* treatments whereas they are expected to only depend on individual information in the *partial* and *complete info* treatment.[23] It follows that an employee's pay in the *minimal* and *peer info* treatments depend on others' effort which might then trigger social motives (e.g. Kandel and Lazear, 1992; Bandiera, Barankay and Rasul, 2010; Corgnet, Hernan-Gonzalez and Rassenti, 2015b). By contrast, social motives are less likely to play a role in the *partial* and *complete info* treatment as pay should only depend on one's own performance. It follows that in the *minimal* and *peer info* treatments, employees may exert high effort because they care about other organizational members (e.g. Rotemberg, 1994; Dur and Sol, 2010). Rotemberg (1994) and Dur and Sol (2010) stress that altruism can reduce shirking in organizations.

Our data, however, are not consistent with an interpretation of our findings based on social motives. First, we show that the interaction effect between altruism and firing treatments is only marginally in explaining workers' production (see Table D.1 in Appendix D). In addition, our analysis of chat activities shows that the magnitude of chatting activities, which are key to trigger social motives (see Dur and Sol, 2010), was only slightly higher in the *minimal* and *peer info* treatments (10.15% of available time is spent chatting) compared to *partial* and *complete info*

[23] Because our data reports only a limited number of fired employees we cannot reliably estimate the claim that group production influence firing decisions more significantly in the *partial* and *minimal info* treatments compared to *complete info*.

(4.35%).[24] In addition, the proportion of time employees spent encouraging and helping their peers (categories 4, 5 and 7) was not higher in the *minimal* and *peer info* treatments compared to the *partial* and *complete info* treatments (see Appendix E).

5. Conclusions

Because the use of firing threats is often justified in contexts in which information is scarce (e.g. Shapiro and Stiglitz, 1984), we assessed its incentive effects across various information levels.[25] By studying the performance of firing schemes under different informational conditions, our work also responds to the call of Hart and Holmström (1986) to study the robustness of commonly-used incentive schemes.

Our main finding is that firing threats exhibit large incentive effects even when no reliable individual information is available. This suggests *collective firing threats* that are triggered when the production of the group is insufficient are particularly powerful. In addition, we show that *collective firing threats* achieve a level of performance which that is at least as high as other popular incentive schemes (bonuses, individual and team incentives) that rely on a similar amount of information about workers.

Our second major finding relates to the comparison of incentive schemes that make use of individual information about workers. We show that firing threats based on partial information of workers' individual contributions are particularly appealing to employers. This is the case because they perform similarly to incentive schemes based on detailed information about workers' individual contributions including individual incentives and discretionary bonuses whereas their potential costs can be relatively small. These findings are consistent with the widespread use of firing threats based on limited but uncontroversial and easily measured information about workers' dedication to their job such as absenteeism (Banerjee and Duflo, 2006; Duflo, Hanna and Ryan, 2012).

Because ours is a study of boundary conditions, we promptly recognize that further robustness checks should also be applied to validate our findings in different contexts. This might include assessing the effectiveness of firing threats in more complex organizational settings in which employers could, for example, use various combinations of incentive schemes or use these

[24] This result continue to holds if we only consider chatting activities between peers.

[25] A weaker assumption would be to consider that even though precise individual information is observable by employers it is not verifiable (MacLeod and Malcomson 1989).

schemes in combination with other organizational policies (Milgrom and Roberts, 1995; Ichniowski and Shaw, 2003; Roberts, 2007; Brynjolfsson and Milgrom, 2013).

References

Abeler, J., A. Falk, L. Goette, and D. Huffman (2011): "Reference points and effort provision," *American Economic Review*, 101(2), 470-492.

Alchian, A. and H. Demsetz (1972): "Production, Information Costs, and Economic Organization," *American Economic Review* 62(5), 777-795.

Alge, B. J. (2001): "Effects of Computer Surveillance on Perceptions of Privacy and Procedural Justice," *Journal of Applied Psychology*, 86(4), 797-804.

Baker, G., M. Jensen and K. Murphy (1988): "Compensation and Incentives: Practice vs. Theory," *Journal of Finance* 43(3), 593-616.

Bandiera, O., I. Barankay, and I. Rasul (2010): "Social Incentives in the Workplace," Review of Economic Studies 77(2), 417-458.

Banerjee A. and E. Duflo (2006): "Addressing absence," *Journal of Economic Perspectives*, 20, 117-132.

Bartling, B., E. Fehr, M. A. Maréchal, and D. Schunk (2009): "Egalitarianism and Competitiveness," *American Economic Review: Papers & Proceedings*, 99(2), 93–98.

Becker, G. S., and G. J. Stigler (1974): "Law Enforcement, Malfeasance, and Compensation of Enforcers," *The Journal of Legal Studies*, 3(1), 1-18.

Belot, M., and M. Schröder (2016): "The Spillover Effects of Monitoring: A Field Experiment," *Management Science*, 62(1), 37-45.

Bénabou, R., and J. Tirole (2003): "Intrinsic and Extrinsic Motivation," *Review of Economics Studies*, 70, 489-520.

Bolton, P. and M. Dewatripont (2005): "Contract Theory," *MIT Press*, Cambridge, MA.

Brynjolfsson, E. and P. Milgrom (2013): "Complementarity in Organizations," *Handbook of Organizational Economics*, 11-55.

Bryson, A., Freeman, R., Lucifora, C., Pellizzari, M., and V. Perotin (2012) "Paying for Performance: Incentive Pay Schemes and Employees' Financial Participation," CEP Discussion Paper, 1112.

Camarinha-Matos, L. M., Afsarmanesh, H. and M. Ollus (2008): "Methods and Tools for Collaborative Networked Organizations," *Springer Science*.

Cameron, A. C and D. Miller (2011): "Robust Inference with Clustered Data," *Handbook of Empirical Economics and Finance*, ed. by A. Ullah and D. E. A. Giles. Boca Raton, CRC Press, 1-28.

Charness, N. and J. Campbell (1988): "Acquiring skill at mental calculation in adulthood: a task decomposition," *Journal of Experimental Psychology: General* 117, 115-129.

Charness, G., R. Cobo-Reyes, N. Jimenez, J. A. Lacomba, and F. Lagos (2017): "Job security and long-term investment: An experimental analysis," *European Economic Review*, 95, 195-214.

Chiappori P. and B. Salanié (2000): "Testing Contract Theory: A Survey of Some Recent Work," in: M. Dewatripont, L.P. Hansen, S. Turnovski, Advances in Economic Theory, Eighth World Congress of the Econometric Society, Cambridge University Press.

Cinyabuguma, M., T. Page, and L. Putterman (2005): "Cooperation under the Threat of Expulsion in a Public Goods Experiment," *Journal of Public Economics*, 89(8), 1421-1435.

Cohen J. (1988): *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge Academic.

Corgnet, B., A. Espin and R. Hernán-González (2015), "The cognitive basis of social behavior: cognitive reflection overrides antisocial but not always prosocial motives," Frontiers of Behavioral Neuroscience., 9, 00287.

Corgnet B., Gunia B., and R. Hernán-González (2020): "Harnessing the Power of Social Incentives to Curb Shirking in Teams," Working paper, GATE 2020-06.

Corgnet, B., R. Hernán-González, and S. Rassenti (2015, CHR): "Firing Threats: Incentive Effects and Impression Management," *Games and Economic Behavior*, 91, 97-113.

Corgnet, B., R. Hernán-González, and E. Schniter, (2015): "Why real leisure really matters: incentive effects on real effort in the laboratory," *Experimental Economics*, 18(2), 284-301.

Corgnet, B., R. Hernán-González, and S. Rassenti (2015b): "Peer Pressure and Moral Hazard in Teams: Experimental Evidence," *Review of Behavioral Economics*, 2(4), 379-403.

Dannenberg, A., C. Haita-Falah, and S. Zitzelsberger (2020): "Voting on the Threat of Exclusion in a Public Goods Experiment," *Experimental Economics*, 23(1), 84-109.

Deci, E. L., and R. Ryan (2000): "The "What" and "Why" of Goal Pursuits: Human Needs and the Self-determination of Behavior," *Psychological Inquiry*, 11(4), 227-268.

Demougin, D. and C. Fluet (2001): "Monitoring Versus Incentives'" *European Economic Review*, 45(9), 1741-1764.

De Paola, M., Scoppa, V. and V. Pupo (2014): "Absenteeism in the Italian Public Sector: The Effects of Changes in Sick Leave Policy," *Journal of Labor Economics*, 32(2), 337-360.

Dickinson, D.(1999): "An experimental examination of labor supply and work intensities," *Journal of Labor Economics*, 17, 638–670.

Dickinson, D., and M. C. Villeval (2008): "Does Monitoring Decrease Work Effort? The Complementarity Between Agency and Crowding-Out Theories," *Games and Economic Behavior*, 63, 56-76.

Dohmen, T., and A. Falk (2010): "You get what you pay for: incentives and selection in the education system," *Economic Journal*, 120 (546), 256–271.

Dohmen, T., and A. Falk (2011): "Performance Pay and Multi-Dimensional Sorting: Productivity, Preferences and Gender," *American Economic Review*, 101 (2), 556-590.

Duflo E., R. Hanna, and S. Ryan (2012): "Incentives work: getting teachers to come to school," American Economic Review 102, 1241–1278.

Dur, R. and J. Sol (2010): "Social Interaction, Co-Worker Altruism, and Incentives," *Games and Economic Behavior* 69(2), 293-301.

Falk, A., and A. Ichino (2006): "Clean evidence on peer effects," *Journal of Labor Economics*. 24 (1), 39–58.

Falk, A., and D. Huffman (2007): "Studying labor market institutions in the lab: minimum wages, employment protection, and workfare," *Journal of Institutional and Theoretical Economics*, 163 (1), 30–45.

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

Falk, A. and M. Kosfeld (2006): "The Hidden Costs of Control," *American Economic Review*, 96(5), 1611-1630.

Fehr, E. and S. Gächter (1998): "Reciprocity and Economics: The Economic Implications of Homo Reciprocans," *European Economic Review*, 42(3-5), 845-859.

Fehr, E. and S. Gächter (2000): "Fairness and Retaliation: The Economics of Reciprocity," *Journal of Economic Perspectives*, 14(3), 159-181.

Feinberg, M., Willer, R. and M. Schultz (2014): "Gossip and Ostracism Promote Cooperation in Groups," *Psychological Science* 25(3), 656-664.

Frederick, S. (2005): "Cognitive reflection and decision making," *Journal of Economic perspectives*, 19(4), 25-42.

Fried, Y. and G. R Ferris (1987): "The Validity of the Job Characteristics Model: A Review and Meta-Analysis," *Personnel Psychology*, 40, 287-322.

Gerhart, B., Rynes, S. L. and I. S. Fulmer (2009): "Pay and Performance: Individuals, Groups, and Executives," *The Academy of Management Annals*, 3(1), 251-315.

Gneezy, U., and J.A. List (2006): "Putting behavioral economics to work: testing for gift exchange in labor markets using field experiments," *Econometrica*, 74 (5), 1365–1384.

Grosse, S., Putterman, L. and B. Rockenbach (2011): "Monitoring in Teams: Using Laboratory Experiments to Study a Theory of the Firm," *Journal of the European Economic Association*, 9(4), 785-816.

Hart, O.D. and B. Holmstrm (1986): "The Theory of Contracts," Working Paper, MIT 418.

Holmström, B. (1979): "Moral Hazard and Observability," *Bell Journal of Economics*, 10(1), 74-91.

Humphrey, S. E., Nahrgang, J. D. and F. P. Morgeson (2007): "Integrating Motivational, Social, and Contextual Work Design Features: A Meta-Analytic Summary and Theoretical Extension of the Work Design Literature," *Journal of Applied Psychology*, 92(5), 1332.

Ichino, A. and R. Riphahn (2005): "The Effect of Employment Protection on Worker Effort: Absenteeism During and After Probation," *Journal of the European Economic Association*, 3(1), 120-143.

Ichniowski, C. and K. Shaw (2003) "Beyond Incentive Pay: Insiders' Estimates of the Value of Complementary Human Resource Management Practices," *Journal of Economic Perspectives* 17(1), 155-180.

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

Klein, B., and K. B. Leffler (1981): "The Role of Market Forces in Assuring Contractual Performance," *Journal of Political Economy*, University of Chicago Press, 89(4), 615-41.

Kopányi-Peuker, A., Offerman, T., and R. Sloof (2018): "Team Production Benefits From a Permanent Fear of Exclusion," *European Economic Review* 103, 125-149.

Laffont, J. J. and D. Martimort (2002): The theory of incentives: The principal-agent model. Princeton University Press: Princeton, N.J.; Oxford.

Lazzarotti, J. (2020): "Out of Sight is Not Out of Mind – Monitoring Workers Working From Home," *The National Law Review*, April 7th.

Lei, V., Noussair, C.N., and C.R. Plott (2001): "Nonspeculative bubbles in experimental asset markets: Lack of common knowledge of rationality vs. actual irrationality," *Econometrica* 69(4), 831-859.

MacLeod, W. B., and J. Malcomson (1989): "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment," *Econometrica*, 57(2), 447-80.

Maier-Rigaud, F. P., Martinsson, P. and G. Staffiero (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.

Milgrom, P. and J. Roberts (1995) "Complementarities and Fit Strategy, Structure, and Organizational Change in Manufacturing," *Journal of Accounting and Economics* 19(2-3), 179-208.

Nalbantian, H. and A. Schotter (1997): "Productivity Under Group Incentives: An Experimental Study," *American Economic Review*, 87(3), 314-341.

Riphahn, R. (2004): "Employment Protection and Effort Among German Employees," *Economics Letters*, 85(3), 353-357.

Roberts, J. (2007): The Modern Firm: Organizational Design for Performance and Growth. Oxford University Press.

Rotemberg, J. (1994). "Humans relations in the workplace," *Journal of Political Economy*, 102: 684-717.

Schotter, A., (2015), "On the Relationship Between Economic Theory and Experiments," in Handbook of Experimental Economic Methodology, Guillaume R. Fréchette and Andrew Schotter (eds.), Oxford University Press, 360-390.

Shapiro, C., and J. Stiglitz (1984): "Equilibrium Unemployment as a Worker Discipline Device," *American Economic Review* 74(3), 433-44.

Sousa-Poza, A., and A. A. Sousa-Poza (2000): "Well-Being at Work: A Cross-National Analysis of the Levels and Determinants of Job Satisfaction," *Journal of Socio-Economics* 29(6), 517-538.

APPENDICES

Appendix A. Additional tables and regression analyses.

Appendix B. Robustness analyses.

Appendix C. 5-minute analysis.

Appendix D. Social motives analysis.

Appendix E. Chat analysis.

Appendix F. Discretionary bonuses, individual and team incentives.

Appendix G. Replication.

Appendix O1. Tests.

Appendix O2. Instructions.

Appendix A. Additional tables and regression analyses

Table A.1

GLS regression with random effects for individual production (periods 1–4) across treatments. Robust standard errors. Excluding fired workers.

	Firing Complete Info vs. Partial Info	Firing Complete Info vs. Peer Info	Firing Complete Info vs. Minimal Info	Firing Complete Info vs. Baseline	Firing Partial Info vs. Peer Info	Firing Partial Info vs. Minimal Info	Firing Partial Info vs. Baseline	Firing Peer Info vs. Minimal Info	Firing Peer Info vs. Baseline	Firing Minimal Info vs. Baseline	Firing Complete+ Partial vs. Peer + Minimal
Constant	- .86	10.94	- 4.72	- 22.56	- 10.26	- 26.10	- 41.06	11.58	- 7.80	- 12.10	-6.99
	(29.50)	(24.78)	(24.27)	(25.48)	(28.88)	(28.87)	(29.34)	(21.11)	(23.48)	(22.22)	(19.08)
Treatment[+]	- 6.88	18.06	30.74**	76.13***	23.58	36.37**	81.41***	12.50	57.24***	44.70***	27.04***
	(15.45)	(13.48)	(13.22)	(13.97)	(15.34)	(15.42)	(15.43)	(12.67)	(13.27)	(13.11)	(10.10)
Ability[26]	8.47***	5.58***	5.45***	5.62***	7.67***	7.43***	7.29***	4.08***	4.62***	4.49***	6.425***
	(1.66)	(1.44)	(1.29)	(1.34)	(1.91)	(1.69)	(1.69)	(1.27)	(1.38)	(1.23)	(1.130)
Gender[++]	19.61	23.95	27.03*	8.19	17.75	21.63	4.30	29.71*	7.96	12.30	22.88**
	(16.20)	(15.34)	(14.60)	(14.84)	(16.86)	(15.54)	(15.79)	(14.97)	(14.92)	(14.39)	(11.04)
Observations	419	386	423	426	385	422	425	389	392	429	808
R^2	0.1980	0.1553	0.1591	0.2566	0.1774	0.1756	0.2545	0.1102	0.1795	0.1424	0.1634

[+]Treatment is a dummy variable that takes value 1 for the first treatment (treatments in the last column) in the comparison and 0 otherwise. Excluding fired subjects.

[+]Gender is a dummy variable that takes value 1 for males.

*p-value<0.1, **p-value<.0.05, and ***p-value<.0.01. (Standard deviation in parentheses)

[26] We compute ability as the number of correct answers in the mathematical task subjects do before the experiment.

Table A.2

GLS regression with random effects for working time (in seconds) (periods 1–4) across treatments. Robust standard errors. Excluding fired workers.

	Firing Complete Info vs. Partial Info	Firing Complete Info vs. Peer Info	Firing Complete Info vs. Minimal Info	Firing Complete Info vs. Baseline	Firing Partial Info vs. Peer Info	Firing Partial Info vs. Minimal Info	Firing Partial Info vs. Baseline	Firing Peer Info vs. Minimal Info	Firing Peer Info vs. Baseline	Firing Minimal Info vs. Baseline
Constant	1108.56***	1050.38***	1123.65***	765.94***	1055.03***	1124.36***	764.85***	1177.44***	812.54***	813.45***
	(38.19)	(34.90)	(39.01)	(97.51)	(38.19)	(39.22)	(99.34)	(48.91)	(102.56)	(98.67)
Treatment[+]	- 25.24	82.84**	17.47	365.85***	100.40***	41.03*	393.61***	-67.55**	281.26***	349.62***
	(18.67)	(27.99)	(20.39)	(48.54)	(28.86)	(23.25)	(51.31)	(30.81)	(53.80)	(49.77)
Ability	- 2.44	- 7.25	- 4.21*	- 1.26	- 7.78	- 4.42*	- 1.07	- 9.79*	- 5.28	- 2.62
	(1.92)	(4.52)	(2.22)	(3.85)	(5.27)	(2.67)	(4.23)	(5.33)	(6.08)	(4.41)
Gender	39.64*	53.53	20.30	- 1.44	55.58	21.36	- 2.65	39.05	6.46	- 20.79
	(23.88)	(38.25)	(22.61)	(49.80)	(43.95)	(29.04)	(52.99)	(43.68)	(64.27)	(51.66)
Observations	419	386	423	426	385	422	425	389	392	429
R^2	0.0416	0.1080	0.0213	0.2591	0.1587	0.0435	0.2998	0.0872	0.1504	0.2360

[+]Treatment is a dummy variable that takes value 1 for the first treatment in the comparison and 0 otherwise. Excluding fired subjects.

*p-value<0.1, **p-value<.0.05, and ***p-value<.0.01. (Standard deviation in parentheses)

Table A.3
Average (median) [standard deviation] individual production across treatments.

	Treatment	Period 1	Period 2	Period 3	Period 4	Subtotal Periods 1-4	Period 5	Total
B subjects only (Workers)	Firing Complete Info (including fired subjects)	3.30 (3) [2.39]	3.86 (3.25) [2.69]	3.97 (4.25) [2.49]	3.74 (3.5) [2.55]	3.72 (3.5) [2.53]	2.72 (2) [2.81]	3.51 (3) [2.61]
	Excluding fired subjects	-	-	4.12 (4.5) [2.41]	4.04 (3.75) [2.41]	3.82 (3.5) [2.49]	3.13 (2) [2.79]	3.70 (3.5) [2.55]
	Firing Partial Info (including fired subjects)	3.36 (3) [3.01]	3.53 (2.5) [3.10]	4.13 (3.5) [2.98]	3.52 (3.5) [2.71]	3.63 (3) [2.95]	2.53 (2) [2.61]	3.41 (3) [2.92]
	Excluding fired subjects	-	-	4.29 (3.5) [2.93]	3.88 (4) [2.58]	3.76 (3) [2.92]	2.84 (2.5) [2.60]	3.59 (3) [2.88]
	Firing Peer Info (including fired subjects)	2.69 (2) [2.23]	3.52 (3.5) [2.09]	3.09 (3) [2.21]	3.48 (3) [2.42]	3.19 (3) [2.25]	1.83 (0) [2.53]	2.92 (3) [2.36]
	Excluding fired subjects	-	-	3.16 (3) [2.19]	3.73 (3.25) [2.31]	3.27 (3) [2.22]	2.11 (1) [2.61]	3.06 (3) [2.33]
	Firing Minimal Info (including fired subjects)	2.37 (1.75) [2.02]	2.93 (2.5) [2.56]	3.58 (3.25) [2.42]	2.87 (3) [2.40]	2.94 (2.5) [2.38]	1.18 (0) [1.93]	2.59 (2) [2.40]
	Excluding fired subjects	-	-	3.65 (3.5) [2.39]	2.98 (3) [2.38]	2.98 (2.5) [2.37]	1.28 (0) [1.98]	2.66 (2) [2.39]
	Baseline	1.79 (1) [2.21]	1.81 (1) [2.58]	1.63 (1) [1.92]	1.81 (1) [2.56]	1.76 (1) [2.32]	1.39 (0.25) [2.12]	1.69 (1) [2.28]
C subjects only (Bosses)	Firing Complete Info	4.17 (4.5) [2.79]	4.92 (3.5) [4.13]	5.17 (4.75) [5.32]	4.08 (3.5) [1.98]	4.58 (4) [3.55]	4.25 (4.25) [2.58]	4.52 (4) [3.34]
	Firing Partial Info	2.58 (2.5) [2.40]	3.33 (2) [2.54]	2.67 (2.75) [1.33]	4.33 (3.25) [3.14]	3.23 (2.5) [2.39]	2.42 (.75) [3.18]	3.07 (2.25) [2.52]
	Firing Peer Info	3.8 (4) [1.40]	3.7 (4) [1.40]	4.3 (3.5) [2.28]	5 (4) [1.37]	4.2 (4) [1.72]	4.1 (4.5) [1.56]	4.18 (4) [1.66]
	Firing Minimal Info	3.33 (3.25) [1.33]	3.42 (3.5) [1.46]	3.17 (2.25) [2.21]	3.08 (3.25) [1.46]	3.25 (3.25) [1.55]	3.25 (3.75) [2.16]	3.25 (3.5) [1.64]
	Baseline	1.67 (1.5) [1.86]	1.67 (1.5) [1.72]	3.17 (2.25) [3.33]	1.25 (1.25) [1.04]	1.94 (1.5) [2.14]	2.33 (2.25) [1.78]	2.02 (1.5) [2.05]

Table A.4
Average (median) [standard deviation] percentage of time subjects spent working across treatments.

	Treatment	Period 1	Period 2	Period 3	Period 4	Subtotal Periods 1-4	Period 5	Total
B subjects only (Workers)	Firing Complete Info (including fired subjects)	92.58 (96.62) [9.60]	93.95 (97.19) [7.64]	87.66 (95.73 [21.77]	85.59 (96.17) [26.25]	89.94 (96.67) [18.32]	65.79 (85.60 [37.64]	85.11 (96.05) [25.32]
	Excluding fired subjects	-	-	91.03 (95.96) [13.41]	92.44 (96.81 [9.98]	92.51 (96.84) [10.32]	75.59 (89.26) [29.59]	89.42 (96.41) [16.95]
	Firing Partial Info (including fired subjects)	94.22 (96.83) [6.94]	94.41 (97.99) [13.88]	92.88 (98.21) [18.94]	87.66 (97.92) [28.59]	92.29 (97.68) [18.89]	68.58 (89.82) [38.23]	87.55 (97.49) [25.76]
	Excluding fired subjects	-	-	96.45 (98.25) [4.63]	96.61 (98.06) [4.52]	95.38 (97.82) [8.52]	77.15 (92.35) [31.19]	91.98 (97.61) [16.98]
	Firing Peer Info (including fired subjects)	83.70 (91.62) [16.83]	88.29 (94.13) [14.56]	85.42 (93.84) [22.58]	77.16 (87.99) [29.19]	83.64 (92.92) [21.76]	52.64 (76.42) [41.75]	77.44 (90.61) [29.58]
	Excluding fired subjects	-	-	87.36 (93.88) [18.66]	82.68 (89.87) [21.15]	85.55 (93.22) [17.90]	60.74 (77.94) [38.90]	81.05 (91.76) [24.95]
	Firing Minimal Info (including fired subjects)	91.52 (96.36) [13]	90.83 (97.40) [15.03]	92.83 (98.07) [15.17]	84.75 (96.17) [23.60]	89.98 (97.03) [17.35]	49.29 (47.36) [37.79]	81.84 (96.11) [28.07]
	Excluding fired subjects	-	-	94.58 (98.08) [8.11]	88.01 (96.63) [16.91]	91.25 (97.18) [13.75]	53.23 (49.86) [36.48]	84.02 (96.35) [25.01]
	Baseline	79.78 (88.13) [22.59]	63.33 (72.72) [34.11]	55.01 (61.93) [38.51]	50.41 (54.41) [35.03]	62.13 (72.42) [34.74]	42.26 (32.27) [34.88]	58.16 (64.32) [35.60]
C subjects only (Bosses)	Firing Complete Info	79.70 (82.44) [15.53]	72.80 (70.38) [18.82]	67.87 (61.13) [21.24]	70.83 (72.88) [24.69]	72.80 (72.88) [19.49]	71.53 (69.62) [20.54]	72.55 (72.88) [19.35]
	Firing Partial Info	76.04 (67.40) [15.06]	77.18 (79.62) [16.22]	78.55 (79.75) [16.25]	88.18 (88.23) [6.04]	79.99 (84.80) [14.00]	82.83 (87.91) [10.74]	80.56 (85.14) [13.29]
	Firing Peer Info	91.74 (97.04) [8.90]	89.64 (90.99) [8.74]	83.83 (94.44) [22.84]	85.15 (85.60) [10.07]	87.59 (90.31) [13.22]	90.34 (91.99) [8.06]	88.14 (90.99) [12.27]
	Firing Minimal Info	96.40 (98.04) [4.28]	96.50 (98.46) [3.34]	95.83 (97.91) [5.47]	89.33 (95.66) [13.09]	94.52 (97.99) [7.72]	77.13 (93.39) [38.31]	91.04 (97.72) [18.72]
	Baseline	67.47 (73.27) [25.59]	63.84 (73.33) [34.34]	69.44 (75.33) [26.40]	69.99 (73.26) [24.32]	67.69 (73.82) [26.17]	72.02 (80.44) [26.38]	68.55 (74.96) [25.81]

Table A.5

Period evolution of monitoring activities (% of total time).

Treatment	Proportion of total time (in %) employees spent monitoring	Period 1	Period 2	Period 3	Period 4
Firing Complete Info	14.58%	14.24%	15.68%	15.14%	13.25%
Firing Partial Info	10.40%	13.71%	9.93%	13.07%	4.88%
Firing Peer Info (Employees)	2.77%	3.59%	2.82%	2.53%	2.09%
Baseline	12.93%	18.14%	12.47%	13.05%	8.06%

Table A.6

Tobit regression with random effects for monitoring time –in seconds- for periods 1 to 4.

	Firing Complete Info vs. Partial Info	Firing Complete Info vs. Baseline	Firing Complete Info vs. Peer Info
Constant	117.94***	135.29***	11.19
	(46.02)	(46.78)	(9.79)
Treatment[+]	53.91	35.93	162.25***
	(65.00)	(65.73)	(27.33)
Observations	*n* = 48 (3 left censored)	*n* = 48 (7 left censored)	*n* = 200 (68 left censored)
Log likelihood (L)	L = -284.011	L = -270.659	L = -800.39
	[Prob>χ2]=0.4069	[Prob>χ2]=0.5846	[Prob>χ2]=0.0000

[+]Treatment F is a dummy variable that takes value 1 for Treatment Firing Complete Info and 0 otherwise.
*p-value<.10, **p-value <.05, and ***p-value <.01. (Standard deviation in parentheses)

Table A.7

Tobit regression with random effects for working time –in seconds- per period for workers.

	Firing Complete Info	Firing Partial Info	Firing Peer Info	Firing Minimal Info	Baseline
Constant	1110.97***	1130.70***	1004.43***	1098.22***	957.42***
	(25.40)	(25.24)	(41.12)	(32.60)	(56.36)
Period 2	16.38	1.62	55.05	-8.29	-197.46***
	(32.21)	(33.82)	(42.49)	(40.28)	(46.28)
Period 3	-18.03	24.02	43.10	32.73	-308.06***
	(32.57)	(34.21)	(42.80)	(40.52)	(46.55)
Period 4	-2.04	25.81	-12.44	-45.82	-366.95***
	(32.95)	(34.78)	(43.44)	(40.75)	(46.64)
Period 5	-205.76***	-207.57***	-276.30***	-461.72***	-469.43***
	(33.55)	(34.98)	(44.57)	(41.21)	(46.79)
Observations	*n* = 257 (0 left censored)	*n* = 257 (0 left censored)	*n* = 215 (2 left censored)	*n* = 263 (0 left censored)	*n* = 270 (11 left censored)
Log likelihood (L)	L = -1701.259	L = -1700.363	L = -1471.650	L = -1803.411	L = -1858.559
	[Prob>χ2]<0.0001	[Prob>χ2]<0.0001	[Prob>χ2]<0.0001	[Prob>χ2]<0.0001	[Prob>χ2]<0.0001

*p-value<.10, **p-value <.05, and ***p-value <.01. (Standard deviation in parentheses)

Appendix B. Robustness analyses

Table B.1[27]

GLS regression with random effects for individual production (periods 1–4) across treatments. Robust standard errors. Excluding fired workers.

	Firing Complete Info vs. Partial Info	Firing Complete Info vs. Peer Info	Firing Complete Info vs. Minimal Info	Firing Complete Info vs. Baseline	Firing Partial Info vs. Peer Info	Firing Partial Info vs. Minimal Info	Firing Partial Info vs. Baseline	Firing Peer Info vs. Minimal Info	Firing Peer Info vs. Baseline	Firing Minimal Info vs. Baseline
Constant	.63	.84	1.34**	-.16	.11	-.42	-.79	.45	.11	.10
	(.82)	(.72)	(.64)	(.63)	(.78)	(.74)	(.73)	(.58)	(.61)	(.57)
Treatment[+]	-.24	.44	.84***	1.77***	.67*	1.04***	1.99***	.35	1.31***	1.02***
	(.38)	(.33)	(.32)	(.36)	(.38)	(.38)	(.37)	(.32)	(.34)	(.35)
Ability	.21***	.13***	.14***	.13***	.19***	.19***	.17***	.10***	.11**	.10***
	(.04)	(.03)	(.03)	(.03)	(.04)	(.04)	(.04)	(.03)	(.03)	(.03)
CRT	-.00	.03	-.00	.19	.04	-.00	.26	.04	.25	.22
	(.11)	(.10)	(.10)	(.15)	(.13)	(.12)	(.18)	(.10)	(.17)	(.16)
Gender	.33	.42	-.55	.05	.30	-.48	-.16	.68*	-.12	-.10
	(.40)	(.38)	(.37)	(.38)	(.41)	(.41)	(.42)	(.38)	(.40)	(.41)
Aheadness aversion	.25	.05	-.28	.25	.43	.07	.62	-.13	.51	.22
	(.49)	(.43)	(.47)	(.47)	(.45)	(.49)	(.49)	(.36)	(.41)	(.43)
Behindness aversion	-.86**	-.65*	-.46	-.73**	-.73*	-.65	-.67*	-.29	-.56*	-.46
	(.43)	(.39)	(.42)	(.36)	(.41)	(.43)	(.37)	(.36)	(.33)	(.35)
Observations	419	386	423	426	385	422	425	389	392	429
R^2	0.2170	0.1716	0.1711	0.2893	0.1956	0.1867	0.2925	0.1172	0.2245	0.1680

[+]Treatment is a dummy variable that takes value 1 for the first treatment in the comparison and 0 otherwise. Excluding fired subjects.

*p-value<0.1, **p-value<.0.05, and ***p-value<.0.01. (Standard deviation in parentheses)

[27] Similar results are obtained if we add as controls all the demographic questions used in the experimental survey: work and volunteering experience, age and academic degree.

Table B.2[28]

GLS regression with random effects for working time (in seconds) (periods 1–4) across treatments. Robust standard errors. Excluding fired workers.

	Firing Complete Info vs. Partial Info	Firing Complete Info vs. Peer Info	Firing Complete Info vs. Minimal Info	Firing Complete Info vs. Baseline	Firing Partial Info vs. Peer Info	Firing Partial Info vs. Minimal Info	Firing Partial Info vs. Baseline	Firing Peer Info vs. Minimal Info	Firing Peer Info vs. Baseline	Firing Minimal Info vs. Baseline
Constant	1130.78***	1095.50***	1147.48***	824.70***	1074.02***	1132.59***	809.31***	1211.85***	889.32***	824.45***
	(32.70)	(37.31)	(36.78)	(88.85)	(34.90)	(36.51)	(94.31)	(50.31)	(107.14)	(97.59)
Treatment[+]	-29.01*	81.50**	22.86	374.59***	108.34***	46.36**	398.54***	-68.43**	281.04***	341.63***
	(17.12)	(26.54)	(21.52)	(47.51)	(27.61)	(22.89)	(50.55)	(30.58)	(53.62)	(51.19)
Ability	-2.75	-8.27*	-4.28*	-1.77	-9.09*	-5.23*	-1.28	-11.43**	-6.34	-2.35
	(1.87)	(4.57)	(2.19)	(3.78)	(5.03)	(2.76)	(4.13)	(5.10)	(5.95)	(4.34)
CRT	2.62	9.94	-2.97	-10.15	20.41**	4.70	-4.25	17.17	-.53	-12.75
	(6.16)	(8.08)	(7.91)	(16.57)	(8.39)	(8.57)	(19.40)	(11.16)	(19.39)	(19.82)
Gender	32.26	29.81	-12.04	9.43	31.43	16.63	-1.80	10.19	-3.03	21.26
	(20.91)	(31.21)	(21.26)	(50.53)	(36.15)	(27.27)	(54.63)	(35.12)	(66.06)	(54.12)
Aheadness aversion	19.49	41,28	36.00	-40.17	61.66**	59.07**	-28.19	101.74**	-9.13	-12.41
	(19.85)	(27.80)	(24.78)	(59.30)	(27.75)	(29.02)	(58.30)	(37.54)	(67.74)	(67.87)
Behindness aversion	-33.73*	-62.90**	-44.60	-75.60	-53.48*	-40.34	-66.89	-74.16**	-101.24*	-74.01
	(19.84)	(30.41)	(29.50)	(53.12)	(29.08)	(30.26)	(50.30)	(37.66)	(60.64)	(59.36)
Observations	419	386	423	426	385	422	425	389	392	429
R^2	0.0584	0.1499	0.0361	0.2740	0.2149	0.0632	0.3093	0.1461	0.1687	0.2453

[+]Treatment is a dummy variable that takes value 1 for the first treatment in the comparison and 0 otherwise. Excluding fired subjects.
*p-value<0.1, **p-value<.0.05, and ***p-value<.0.01. (Standard deviation in parentheses)

[28] Similar results are obtained if we add as controls all the demographic questions used in the experimental survey: work and volunteering experience, age and academic degree.

Table B.3[29]

GLS regression with random effects for individual production (all periods) across treatments. Robust standard errors. Excluding fired workers.

	Firing Complete Info vs. Partial Info	Firing Complete Info vs. Peer Info	Firing Complete Info vs. Minimal Info	Firing Complete Info vs. Baseline	Firing Partial Info vs. Peer Info	Firing Partial Info vs. Minimal Info	Firing Partial Info vs. Baseline	Firing Peer Info vs. Minimal Info	Firing Peer Info vs. Baseline	Firing Minimal Info vs. Baseline
Constant	.82	.87	1.18*	-.27	.35	-.07	-.63	.64	.10	.03
	(.76)	(.68)	(.61)	(.61)	(.74)	(.70)	(.68)	(.59)	(.61)	(.53)
Treatment[+]	-.17	.55*	1.02***	1.70***	.70**	1.14***	1.87***	.45	1.16***	.79**
	(.36)	(.32)	(.30)	(.34)	(.36)	(.35)	(.34)	(.30)	(.32)	(.32)
Ability	.19***	.13***	.12***	.13***	.17***	.16***	.16***	.09**	.11**	.10***
	(.04)	(.03)	(.03)	(.03)	(.04)	(.04)	(.03)	(.03)	(.03)	(.03)
CRT	-.01	.00	-.01	.18	.03	-.00	.27	.02	.25	.22
	(.12)	(.10)	(.10)	(.15)	(.13)	(.11)	(.03)	(.09)	(.16)	(.16)
Gender	.25	.32	-.41	.05	.18	-.29	-.23	.47	-.19	-.03
	(.38)	(.35)	(.34)	(.37)	(.39)	(.38)	(.18)	(.36)	(.39)	(.39)
Aheadness aversion	.225	.20	-.07	.39	.36	.06	.54	.02	.56	.32
	(.46)	(.41)	(.45)	(.46)	(.41)	(.44)	(.47)	(.34)	(.41)	(.42)
Behindness aversion	-.78*	.73**	-.53	-.62*	-.74*	-.62	-.56	-.48	-.56*	-.45
	(.40)	(.37)	(.38)	(.34)	(.38)	(.38)	(.34)	(.33)	(.32)	(.33)
Observations	514	472	520	527	472	520	527	478	485	533
R^2	0.1718	0.1479	0.1435	0.2679	0.1616	0.1490	0.2667	0.0908	0.2055	0.1349

[+]Treatment is a dummy variable that takes value 1 for the first treatment in the comparison and 0 otherwise. Excluding fired subjects.
*p-value<0.1, **p-value<.0.05, and ***p-value<.0.01. (Standard deviation in parentheses)

[29] Similar results are obtained if we add as controls all the demographic questions used in the experimental survey: work and volunteering experience, age and academic degree.

Table B.4[30]

GLS regression with random effects for working time (in seconds) (all periods) across treatments. Robust standard errors. Excluding fired workers.

	Firing Complete Info vs. Partial Info	Firing Complete Info vs. Peer Info	Firing Complete Info vs. Minimal Info	Firing Complete Info vs. Baseline	Firing Partial Info vs. Peer Info	Firing Partial Info vs. Minimal Info	Firing Partial Info vs. Baseline	Firing Peer Info vs. Minimal Info	Firing Peer Info vs. Baseline	Firing Minimal Info vs. Baseline
Constant	1171.74***	1115.85***	1089.99***	748.69***	1095.46***	1124.60***	788.04***	1188.78***	856.88***	745.33***
	(36.22)	(52.08)	(45.58)	(87.63)	(47.50)	(43.20)	(94.04)	(61.92)	(109.99)	(93.64)
Treatment[+]	-22.87	104.67**	74.54***	390.49***	119.83***	92.67***	407.71***	-33.09	279.56***	304.85***
	(19.21)	(32.84)	(25.60)	(48.47)	(32.56)	(23.71)	(50.07)	(35.88)	(56.01)	(51.39)
Ability	-4.83**	-8.13*	-5.09*	-1.03	-11.18**	-7.41**	-2.29	-11.68**	-5.51	-1.91
	(2.12)	(4.91)	(2.76)	(3.83)	(4.93)	(2.90)	(4.10)	(5.38)	(5.98)	(4.37)
CRT	-3.82	-5.43	-8.05	-13.44	11.34	7.39	.87	10.06	-3.79	-7.20
	(8.45)	(10.49)	(9.75)	(18.05)	(10.07)	(9.53)	(20.64)	(12.69)	(20.82)	(20.70)
Gender	8.31	2.27	21.56	30.64	24.42	-12.25	-15.13	-14.50	-16.97	47.00
	(21.85)	(37.00)	(25.62)	(52.02)	(39.08)	(25.76)	(54.42)	(40.95)	(68.64)	(55.07)
Aheadness aversion	10.70	57.12	60.28*	-21.15	36.08	44.42	-49.26	112.80**	-8.20	2.13
	(24.78)	(36.02)	(30.97)	(64.94)	(35.11)	(31.53)	(62.22)	(46.05)	(73.68)	(72.47)
Behindness aversion	-25.16	-78.45**	-51.97	-58.81	-67.58**	-40.17	-59.93	-104.23**	-105.32*	-77.29
	(23.85)	(38.01)	(25.60)	(55.75)	(33.13)	(28.19)	(51.37)	(41.32)	(63.14)	(51.39)
Observations	514	472	520	527	472	520	527	478	485	533
R^2	0.0234	0.0868	0.0418	0.2467	0.1194	0.0578	0.2777	0.0660	0.1375	0.1619

[+]Treatment is a dummy variable that takes value 1 for the first treatment in the comparison and 0 otherwise. Excluding fired subjects.
*p-value<0.1, **p-value<.0.05, and ***p-value<.0.01. (Standard deviation in parentheses)

[30] Similar results are obtained if we add as controls all the demographic questions used in the experimental survey: work and volunteering experience, age and academic degree.

Appendix C. 5-minute analysis.

Table C.1. GLS regression with random effects for workers' production (all periods) across treatments where firing is allowed. Robust standard errors. Excluding fired workers.

	Firing Complete Info	Firing Partial Info
Constant	30.02***	24.50***
	(5.93)	(4.00)
Minute[+]	2.77	3.68**
	(2.12)	(1.53)
Watch[++]	.60	4.63
	(7.28)	(5.12)
Minute×Watch[+++]	- .60	- 1.09
	(2.83)	(2.23)
Observations	1028	1028
R^2	0.0032	0.0060

[+]Minute takes value 1 for the first 5 minutes of a period, 2 for next 5 minutes and so on until value 4.
[++]Watch is a dummy variable that takes the value 1 if a worker was observed in the previous 5-minute moment.
[+++]Minute×Watch id the interaction term between the previous variables.
*p-value<0.1, **p-value<.0.05, and ***p-value<.0.01. (Standard deviation in parentheses)

Appendix D. Social motives analysis.

Table D.1

GLS regression with random effects for workers' production (periods 1–4) across treatments where firing was allowed. Robust standard errors clustered by session. Excluding fired workers.

Constant	19.14
	(22.83)
SP[+]	8.92
	(23.78)
Altruism	10.79*
	(6.24)
SP × Altruism	-14.65*
	(8.35)
Ability	7.14***
	(1.06)
Observations	808
R^2	0.1634

[+]SP is a dummy variable that takes value 1 for the *minimal* and *peer info* treatments and 0 for the complete and partial info treatments. Excluding fired subjects.
*p-value<0.1, **p-value<.0.05, and ***p-value<.0.01. (Standard deviation in parentheses)

Appendix E. Chat analysis.

Each chat message was assigned to one of thirty-three categories by two graduate students coding messages independently (see Table F.3). Then, we computed the Cohen's Kappa coefficient for each category to assess inter-rater agreement (see Table F.1).[31] We dropped categories 18 and 19 from the analysis because they were empty and another seven categories (categories 7, 12, 17, 20, 23, 27, and 33) because the corresponding Cohen Kappa test was not significant at a 5% significance level. These categories represented only 1.17% of the messages (see Figure F.1). The most represented category (31.40%) corresponds to distracting messages (e.g. jokes and stories). General and nonstrategic messages constituted the great majority (68.26%) of chat messages. We consider as general and nonstrategic messages the ones that were assigned to categories related to either presentation (category 1), distraction (categories 2 and 3) or general observations about the experiment (categories 27, 28, 29 and 30). Most of the strategic messages consisted in subjects stating their own performance (category 13, 5.73% of all messages) and encouraging others to produce (category 4, 4.48% of all messages).

We present disaggregate data at treatment level of the percentage of messages of each category (see Table F.2). We can observe that 44.37% of messages in the baseline treatment are related to category 2 (jokes and stories). This percentage is relatively high compared to Firing Complete Info (19.87), Firing Partial Info (16.25%), Firing Peer Info (19.99%), and Firing Minimal Info (21.07%). In relation to strategic messages the highest differences we find are related to categories 4 (Encouraging others to produce) and 13 (State your own performance). We observe that the percentage of messages in these categories is much higher in the Firing Complete Info treatment (11.04% for category 4, and 11.69% for category 13) compare to the baseline, Firing Partial Info, Firing Peer Info, and Firing Minimal Info treatments (2.31%, 4.37%, 3.39% and 3.41% respectively for category 4, and 5.23%, 1.46%, 3.66% and 6.40% respectively for category 13).

In summary, chatting activities were mostly leisure activities. Indeed, similarly to Internet browsing, the average amount of time *B* subjects dedicated to chatting was significantly greater in the baseline treatment (31.54%) than in Firing Complete Info (4.85%), Firing Partial Info (3.85%), Firing Peer Info (10.21%), and Firing Minimal Info (10.12%).

[31] According to Landis and Koch (1977), Cohen Kappa coefficients between 0.4 and 0.6 correspond to a moderate agreement level and coefficients greater than 0.6 correspond to full agreement.

TABLE E.1 Inter-rater analysis of chat messages categorization.

Category	Agreement	Expected Agreement	Kappa	Standard Error	Z	Prob> Z
1	98.72%	93.42%	0.81	0.015	52.43	0
2	77.26%	54.92%	0.49	0.015	32.85	0
3	85.37%	80.08%	0.27	0.015	17.63	0
4	98.23%	93.18%	0.74	0.015	48.14	0
5	99.50%	98.71%	0.62	0.015	40.10	0
6	99.76%	98.87%	0.79	0.015	51.82	0
7	99.65%	99.65%	0.0004	0.0077	-0.06	0.5230
8	97.83%	94.89%	0.57	0.015	37.53	0
9	99.39%	99.01%	0.38	0.014	27.27	0
10	99.74%	99.18%	0.68	0.015	46.08	0
11	99.22%	97.41%	0.70	0.015	46.49	0
12	99.88%	99.88%	0.0006	0.015	-0.04	0.5150
13	97.00%	89.65%	0.71	0.015	46.50	0
14	99.76%	99.34%	0.64	0.015	42.17	0
15	99.20%	98.92%	0.26	0.014	18.62	0
16	99.76%	98.92%	0.78	0.015	50.82	0
17	-	-	-	-	-	-
18	-	-	-	-	-	-
19	-	-	-	-	-	-
20	99.98%	99.98%	0	-	-	-
21	97.64%	93.81%	0.62	0.015	40.82	0
22	97.75%	96.85%	0.29	0.013	22.13	0
23	99.55%	99.55%	0.002	0.015	-0.14	0.5562
24	99.83%	99.79%	0.22	0.013	17.34	0
25	99.72%	99.62%	0.25	0.015	16.30	0
26	98.56%	94.64%	0.73	0.015	47.99	0
27	99.93%	99.93%	0.0000	0.0000	0.00	0.5
28	96.74%	95.05%	0.34	0.014	23.58	0
29	96.05%	90.75%	0.57	0.014	39.56	0
30	83.45%	74.50%	0.35	0.014	25.22	0
31	95.74%	92.39%	0.44	0.014	32.51	0
32	98.75%	98.66%	0.07	0.010	6.71	0
33	99.53%	99.53%	0.0009	0.009	-0.09	0.5368

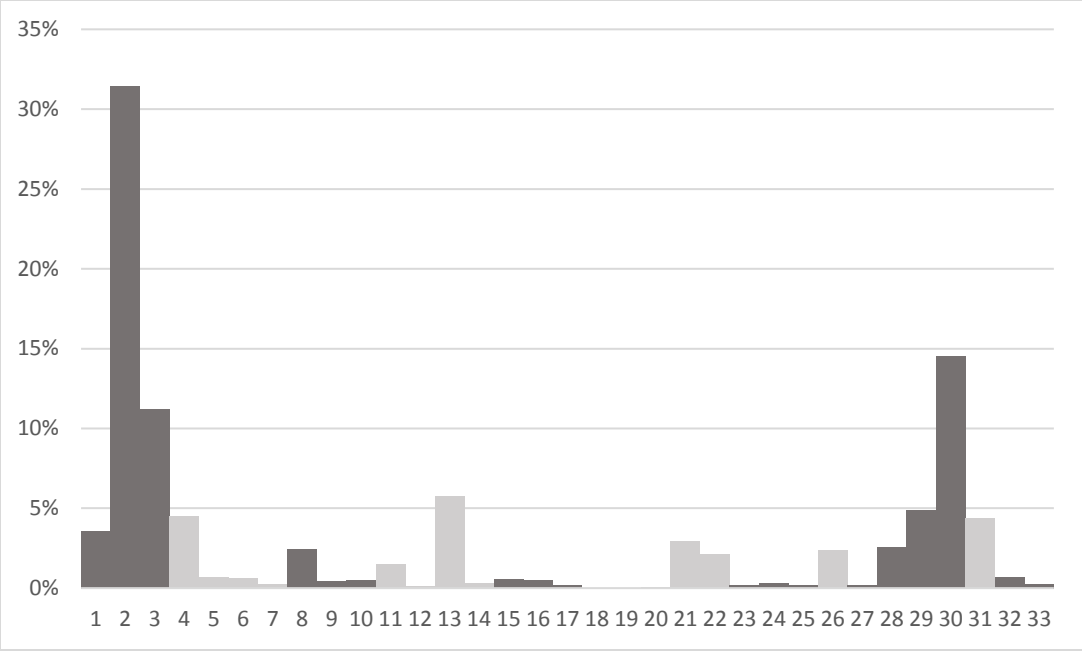

Figure E.1. Histogram of categorization of messages for all treatments.

Table E.2 Percentage of categories by treatment.

Category	Baseline	Complete Info	Partial Info	Peer Info	Minimal Info
1	3.48	2.34	4.37	3.11	3.82
2	44.37	19.87	16.25	19.99	21.07
3	8.46	3.64	9.79	22.68	14.05
4	2.31	11.04	4.37	3.39	3.41
5	0.60	0.52	1.46	0.55	0.72
6	0.21	3.90	0.83	0.28	0.00
7	0.06	0.91	0.21	0.07	0.31
8	1.50	3.77	6.04	2.56	5.68
9	0.42	1.04	0.83	0.07	0.93
10	0.15	2.60	0.83	0	0.41
11	0.96	1.04	0.21	1.38	3.72
12	0.04	0.13	0.00	0.07	0.10
13	5.23	11.69	1.46	3.66	6.40
14	0.17	0.00	0.00	0.97	0.62
15	0.56	0.13	0.83	0.35	0.93
16	0.00	2.86	2.08	0.21	1.14
17	0.00	0.00	0.00	0	0.00
18	0.00	0.00	0.00	0	0.00
19	0.00	0.00	0.00	0	0.00
20	0.00	0.00	0.00	0	0.10
21	3.94	2.86	1.87	2.56	1.34
22	1.06	1.69	2.50	1.94	3.20
23	0.02	0.00	1.46	0.21	0.83
24	0.08	0.13	0.00	0.21	0.10
25	0.00	0.00	0.00	0.14	1.45
26	2.33	3.51	7.29	2.77	1.96
27	0.06	0.00	0.00	0	0.00
28	1.94	3.12	2.92	4.56	1.76
29	4.32	3.90	9.79	5.67	4.44
30	13.45	15.58	14.79	18.12	14.05
31	3.31	3.38	7.29	3.73	5.99
32	0.83	0.39	1.04	0.28	0.52
33	0.10	0.00	1.46	0.28	0.41
Total messages	2398	385	240	723	484

Table E.3 Categories for chat messages.

Group Category	Category Number	Category
Social interaction	1	Greetings (Hello/Goodbye)
	2	Distracting others (jokes, stories)
	3	Personal chat (talking about likes and dislikes)
Positive feedback and help	4	Encouraging others to produce
	5	Thanking other for their cooperative behavior
	6	*C* give positive feedback about *B* contributions
	7	Help others complete the task
Discouragements	8	Discouraging others to produce
	9	Asking others what is the point of producing anything
	10	*C* give negative feedback about *B* contributions
Performance evaluation and comparison	11	Ask others' performance on the task
	12	*B* asks *C* about his/her own relative performance on the task
	13	State your own performance
	14	B talks to C about other B subjects' performance
Pay /firing threats	15	*B* threatening *C* not to produce anything
	16	*C* threatening others to fire them if they do not produce enough
	17	*C* telling *B* they will be paid based on their relative production
	18	*C* telling *B* they will be paid based on how much time they spent working instead of being online
	19	*C* telling all *B*s they will all be paid the same if they achieve a certain level of total production
	20	*C* telling all *B*s they will all be paid the same regardless of performance
Complaints about firing/supervision strategy/pay	21	Complaints about the supervision of the *C* subject
	22	Complaints about the firing/pay strategy of the *C* subject
Comments on firing/supervision/pay strategy	23	Suggesting/stating Firing strategy
	24	Suggesting/stating Supervising strategy
	25	Comments on effectiveness of firing policy
Envy	26	*B* envying the *C* subject
Non-strategic comments on the experiment	27	Ask others for help and hints to complete the task
	28	General comments about the experiment and its goals
	29	Specific comments on how earnings are calculated
	30	Other specific comments on the experiment
Influence and manipulation	31	Influencing *C* subject
Fairness	32	Negative comments on fairness of firing / pay policy
	33	Positive comments on fairness of firing / pay policy

Table E.4. GLS regression with random effects for chat time (in seconds) (periods 1–4) across treatments where firing was allowed. Robust standard errors clustered by session. Excluding fired workers.

Constant	2.64***
	(.37)
SP[+]	2.47**
	(.85)
Observations	808
R^2	0.0265

[+]SP is a dummy variable that takes value 1 for the minimal and partial info treatments and 0 for the complete and partial info treatments. Excluding fired subjects.
*p-value<0.1, **p-value<.0.05, and ***p-value<.0.01. (Standard deviation in parentheses)

Appendix F. Discretionary bonuses, individual and team incentives

Table F.1
GLS regression with random effects for individual production (periods 1–4) across treatments. Robust standard errors. Excluding fired workers.

	Firing Complete Info vs. Disc. Bonus Complete Info	Firing Partial Info vs. Disc. Bonus Complete Info	Firing Peer Info vs. Disc. Bonus Complete Info	Firing Minimal Info vs. Disc. Bonus Complete Info
Constant	9.103	-20.06	23.09	20.87
	(31.06)	(31.77)	(29.12)	(28.03)
Treatment[+]	-30.66*	-24.80	-45.87***	-58.68***
	(17.21)	(17.02)	(16.79)	(16.94)
Gender	19.64	13.09	17.68	21.94
	(17.24)	(17.11)	(18.44)	(16.65)
Ability	9.387***	12.44***	8.530***	8.184***
	(1.975)	(2.011)	(2.260)	(1.892)
Observations	415	414	381	418
R^2	0.1886	0.2370	0.1696	0.1725

+Treatment is a dummy variable that takes value 1 for the first treatment in the comparison and 0 otherwise. Excluding fired subjects.
*p-value<0.1, **p-value<.0.05, and ***p-value<.0.01. (Standard deviation in parentheses)

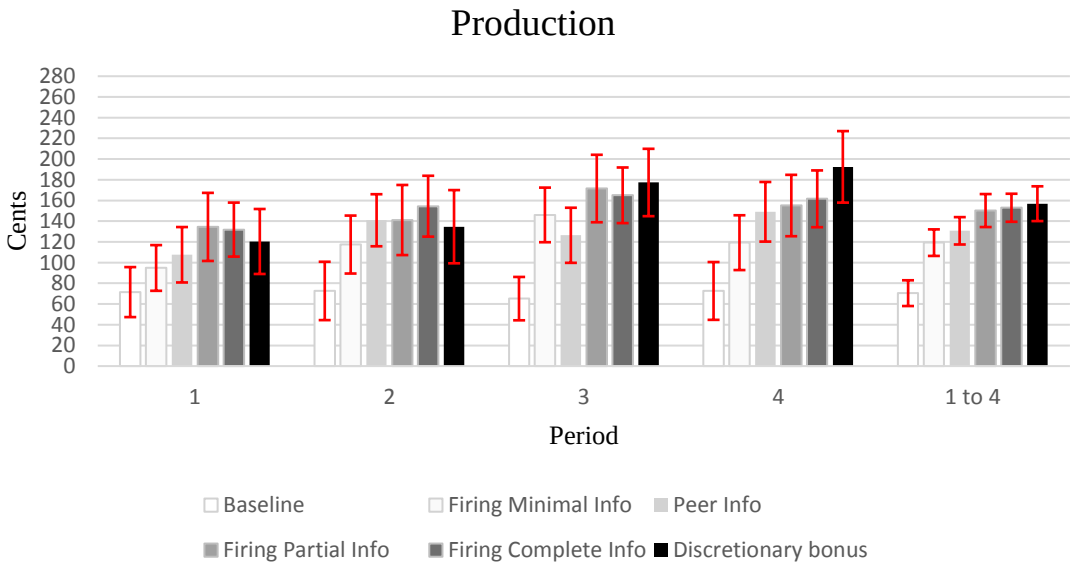

Figure F.1. Employees' average production across treatments for periods 1 to 4. Subjects who have been fired before a current period are excluded. The bars show 95% confidence interval.

Table F.2a

GLS regression with random effects for individual production (periods 1–4) across treatments. Robust standard errors. Excluding fired workers.

	Firing Complete Info vs. Individual Incentives (CHR)	Firing Partial Info vs. Individual Incentives (CHR)	Firing Peer Info vs. Individual Incentives (CHR)	Firing Minimal Info vs. Individual Incentives (CHR)
Constant	183.1***	183.1***	183.1***	183.1***
	(17.92)	(17.92)	(17.93)	(17.92)
Treatment^{+}	-32.02	-35.74	-53.90***	-64.25***
	(21.14)	(22.56)	(20.61)	(20.35)
Observations	426	425	392	429
R^2	0.0137	0.0146	0.0407	0.0603

+Treatment is a dummy variable that takes value 1 for the first treatment in the comparison and 0 otherwise. Excluding fired subjects.

*p-value<0.1, **p-value<.0.05, and ***p-value<.0.01. (Standard deviation in parentheses)

Table F.2b

GLS regression with random effects for individual production (periods 1–4) across treatments. Robust standard errors. Excluding fired workers.

	Firing Complete Info vs. Individual Incentives (CHS)	Firing Partial Info vs. Individual Incentives (CHS)	Firing Peer Info vs. Individual Incentives (CHS)	Firing Minimal Info vs. Individual Incentives (CHS)
Constant	167.7***	167.7***	167.7***	167.7***
	(13.02)	(13.02)	(13.02)	(13.02)
Treatment^{+}	-16.52	-20.21	-38.42**	-48.84***
	(17.17)	(18.89)	(16.51)	(16.21)
Observations	478	477	444	481
R^2	0.0039	0.0049	0243	0.0419

+Treatment is a dummy variable that takes value 1 for the first treatment in the comparison and 0 otherwise. Excluding fired subjects.

*p-Value<0.1, **p-value<.0.05, and ***p-value<.0.01. (Standard deviation in parentheses)

Table F.2c
GLS regression with random effects for individual production (periods 1–4) across treatments. Robust standard errors. Excluding fired workers.

	Firing Complete Info vs. Individual Incentives (CHR & CHS)	Firing Partial Info vs. Individual Incentives (CHR & CHS)	Firing Peer Info vs. Individual Incentives (CHR & CHS)	Firing Minimal Info vs. Individual Incentives (CHR & CHS)
Constant	174.5***	174.5***	174.5***	174.5***
	(10.77)	(10.77)	(10.77)	(10.77)
Treatment[+]	-23.46	-27.15	-45.36***	-55.73***
	(15.53)	(17.40)	(14.79)	(14.45)
Observations	694	693	660	697
R^2	0.0059	0.0070	0.0226	0.0385

+Treatment is a dummy variable that takes value 1 for the first treatment in the comparison and 0 otherwise. Excluding fired subjects.
*p-value<0.1, **p-value<.0.05, and ***p-value<.0.01. (Standard deviation in parentheses)

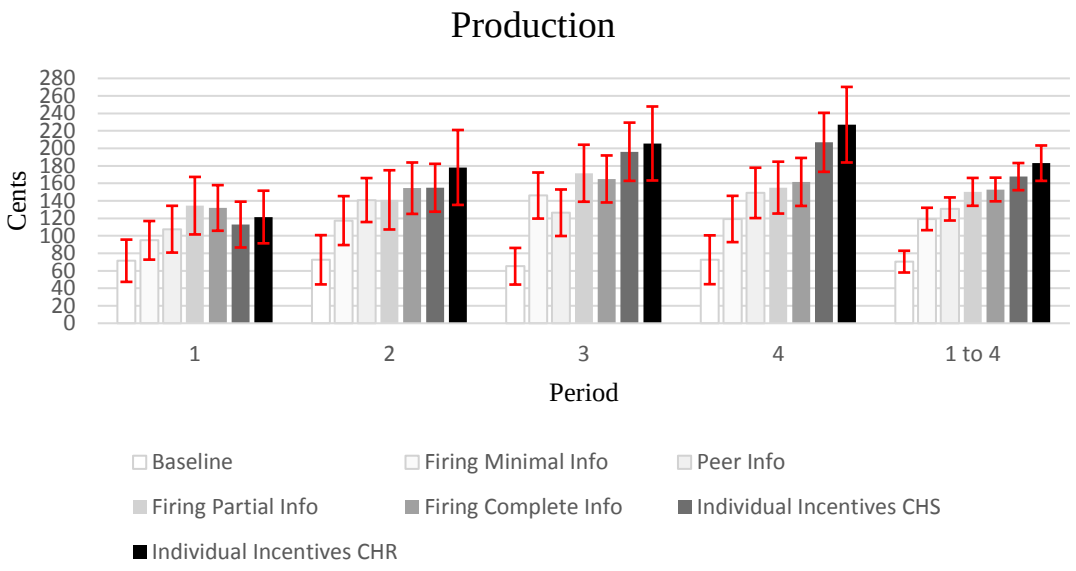

Figure F.2. Employees' average production across treatments for periods 1 to 4. Subjects who have been fired before a current period are excluded. The bars show 95% confidence interval.

Table F.3
GLS regression with random effects for individual production (periods 1–4) across treatments. Robust standard errors. Excluding fired workers.

	Firing Complete Info vs. Team Incentives (CHS)	Firing Partial Info vs. Team Incentives (CHS)	Firing Peer Info vs. Team Incentives (CHS)	Firing Minimal Info vs. Team Incentives (CHS)
Constant	109.1***	109.1***	109.1***	109.1***
	(11.93)	(11.93)	(11.94)	(11.93)
Treatment[+]	42.08**	38.37**	20.20	9.765
	(16.37)	(18.16)	(15.67)	(15.35)
Observations	450	449	416	453
R^2	0.0408	0.0314	0.0107	0.0024

+Treatment is a dummy variable that takes value 1 for the first treatment in the comparison and 0 otherwise. Excluding fired subjects.
*p-value<0.1, **p-value<.0.05, and ***p-value<.0.01. (Standard deviation in parentheses)

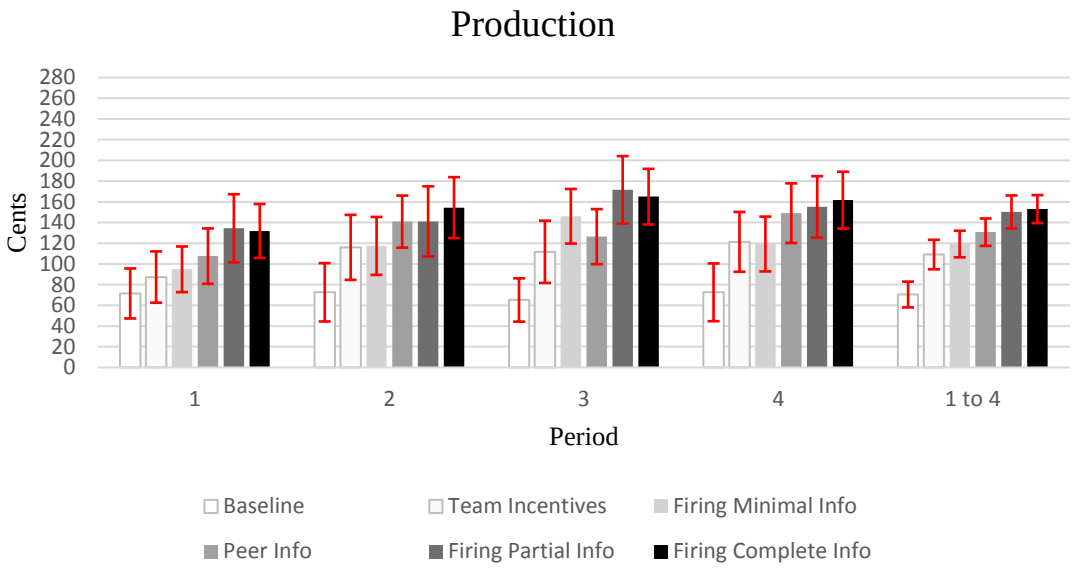

Figure F.3. Employees' average production across treatments for periods 1 to 4. Subjects who have been fired before a current period are excluded. The bars show 95% confidence interval.

Appendix G. Replication

Table G.1
GLS regression with random effects for individual production (periods 1–4) across treatments. Robust standard errors. Excluding fired workers.

	Firing Complete Info vs. Firing Complete Info (CHR)	Baseline vs. Baseline (CHR)
Constant	130.5***	65.67***
	(15.53)	(9.677)
Treatment[+]	20.51	4.796
	(19.16)	(14.18)
Observations	410	396
R^2	0.0044	0.0007

+Treatment is a dummy variable that takes value 1 for the first treatment in the comparison and 0 otherwise. Excluding fired subjects.
*p-value<0.1, **p-value<.0.05, and ***p-value<.0.01. (Standard deviation in parentheses)

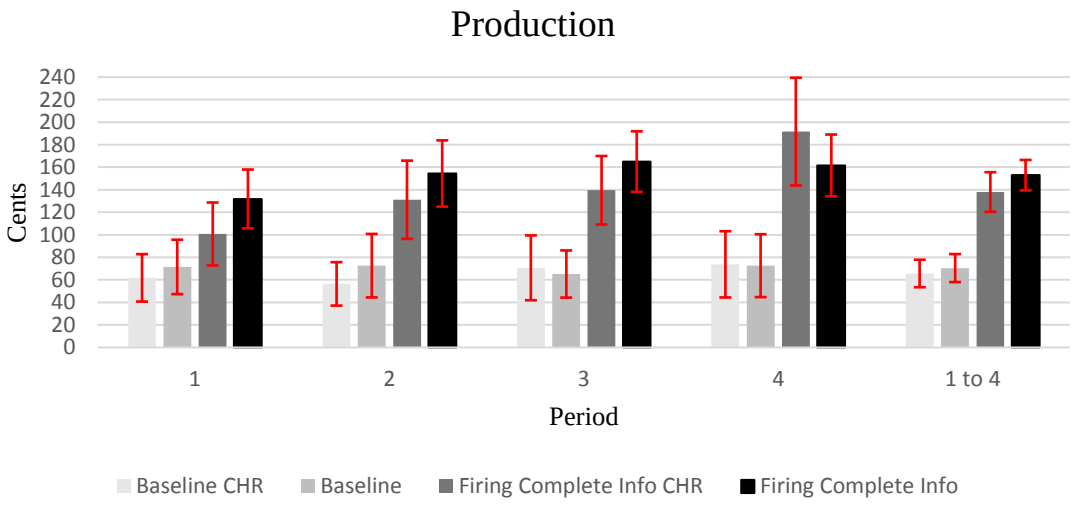

Figure G.1. Employees' average production across treatments for periods 1 to 4. Subjects who have been fired before a current period are excluded. The bars show 95% confidence interval.

Online Appendix.

Appendix O1. Tests

We detail below the tests which were completed by subjects as part of the one-hour survey conducted at the lab in which the experiment was performed. This survey was conducted at the beginning of the year, about six months before completion of Study 1.

Summation skills

The instructions for this task were as follows. Instructions:

This task consists in adding five one-digit numbers. During a period of 2 minutes you can solve as many problems as you want to. An example of the sum problem is displayed below. Next to the display, there is an input box and an O.K. button. You will have to enter the result into the box (only integer numbers are allowed) and then click on the O.K. button. For each sum problem that you solve correctly, you will receive 10 cents. If you enter a wrong result and click O.K., a message 'Last answer was not correct.' will be displayed. You will be informed about the number of problems you have solved correctly (on the right hand side of the screen). The time remaining in seconds will be displayed in the upper left corner of the screen.

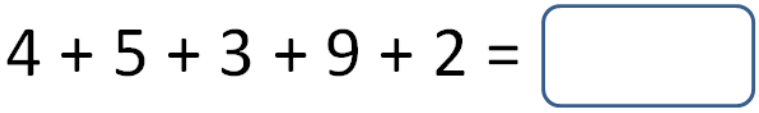

Figure O1.1. Example of Adding Task question.

Intrinsic motivation

To measure intrinsic motivation, we assess the extent to which people performed on the previous adding task in the absence of any monetary incentives. We then computed the intrinsic motivation score as the ratio between one's performance on the task without incentives and one's performance on the task in the presence of monetary incentives. The incentive version of the task was presented first, and the non-incentivized version of the task was presented at the end of the survey.

Social motives

Subjects made six choices between two possible allocations of money between themselves and another anonymous subject with whom they were randomly matched. In each experimental session (typically composed of 12 subjects), two subjects and one of the six decisions were selected at random for payment. The choice of the first subject in the selected decision was used to allocate payoffs between the two subjects. All decisions were anonymous. The first four decisions used the exact same payoffs as in Bartling et al. (2009). Decisions 5 and 6 were added by Corgnet, Espin and Hernán-González (2015).

All the allocation decisions are described in Table C.3. Option A always yielded an even distribution of money ($10 to both the self and the other subject) whereas option B yielded uneven payoffs. For each decision, we show in parentheses the envy/compassion parameter associated to choosing the egalitarian and non-egalitarian options (i.e. options A and B) and in square brackets the proportion of subjects who chose each option. Note that the model parameters associated to Decisions 1-4 are the same as in Study 1, except for the fact that in Decision 4 the threshold for the envy parameter is now 0.125 instead of 0.5.

Table O1.1. Decisions in the social preferences task (Study 1). For each option, we display the payoff for the decision-maker and the recipient, the associated model parameters (in parentheses) and the % of subjects choosing it (in square brackets).

Decision #	Option A self, other		Option B self, other	
1	$10,$10	[80%]	$10,$6	[20%]
2	$10,$10	[33%]	$16,$4	[67%]
3	$10,$10	[49%]	$10,$18	[51%]
4	$10,$10	[34%]	$11,$19	[66%]
5	$10,$10	[48%]	$12,$4	[52%]
6	$10,$10	[89%]	$8,$16	[11%]

The *altruism* index is calculated as the number of times one chooses Option A for decisions 1, 2 and 5 and Option B for decisions 3, 4 and 6. The higher the index the more likely a person values the other person's payoff positively.

Extended cognitive reflection test (CRT):

Taken from Frederick (2005):

(1) A bat and a ball cost $1.10 in total. The bat costs a dollar more than the ball. How much does the ball cost? ____ cents
[Correct answer: 5 cents; intuitive answer: 10 cents]

(2) If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? ____ minutes
[Correct answer: 5 minutes; intuitive answer: 100 minutes]

(3) In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? ____ days
[Correct answer: 47 days; intuitive answer: 24 days]

Taken from Toplack et al. (2014):

(4) If John can drink one barrel of water in 6 days, and Mary can drink one barrel of water in 12 days, how long would it take them to drink one barrel of water together? _____ days
[correct answer: 4 days; intuitive answer: 9]

(5) Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are in the class? _____ students
[correct answer: 29 students; intuitive answer: 30]

(6) A man buys a pig for $60, sells it for $70, buys it back for $80, and sells it finally for $90. How much has he made? _____ dollars
[correct answer: $20; intuitive answer: $10]

(7) Simon decided to invest $8,000 in the stock market one day early in 2008. Six months after he invested, on July 17, the stocks he had purchased were down 50%. Fortunately for Simon, from July 17 to October 17, the stocks he had purchased went up 75%. At this point, Simon has: a. broken even in the stock market, b. is ahead of where he began, c. has lost money
[correct answer: c; intuitive response: b]

Appendix O2. Instructions

Instructions for all treatments are available through this link:
https://drive.google.com/file/d/1emoL1h8x92B79Y0y0enZblNga-EQBfRf/view?usp=sharing