Watching or Not Watching? Access to Information and the Incentive Effects of Firing Threats

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

This version: June 2021
(February 2018)

Barcelona GSE Working Paper Series

Working Paper nº 1023

Watching or Not Watching?

Access to Information 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 (Zayed University)
Carles Solà (City University of New York, College of Staten Island)

June 2021

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 employees. 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, firing threats based on observing organizational performance alone increase employees' 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, such as bonuses, individual and team incentives, that rely on a similar amount of information about employees.

JEL classification: C92, D23, D82

Keywords: Firing threats, incentives, monitoring, laboratory experiments

Acknowledgements: The authors gratefully acknowledge financial support from the Spanish Ministry of Economics and Competitiveness through Grant: ECO2017-88130 and through the Severo Ochoa Program for Centers of Excellence in R&D (CEX2019-000915-S), the *Generalitat de Catalunya* (Grant: 2017 SGR 1136).

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 employees' 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 illustrated by empirical works showing low levels of effort of salaried employees (De Paola et al. 2014; Corgnet, Hernán-González and Rassenti, 2015a, CHR henceforth). Thus, it is 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 employees 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 good 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; Bonroy et al. 2019; Dannenberg, Haita-Falah and Zitzelsberger, 2020; Solda and Villeval, 2020). Croson et al. 2015; report a positive effect of excludability of the lowest performer in a broad range of settings covering a public good game, a weakest-link game and a best-shot game. 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 is practically relevant because the use of firing threats is often motivated by the lack of information available to supervisors regarding employees' effort (e.g. Shapiro and Stiglitz, 1984). 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 were sufficient to incentivize employees 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 employees 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 employees were only paid a fixed salary that 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 employees under various informational conditions. In what we call the firing treatments, the boss could fire one (out of nine) employees at the end of each of five periods (except for the first period). Firing treatments differed in the amount of information that could be collected by the boss during the monitoring of employees.

In the *complete information* treatment, bosses had access to real-time information about each employee's production as well as about their current activity (either working, chatting or browsing the Internet). In the *partial information* treatment, bosses could not observe employees' production levels but could see the current activity they were undertaking. In the *minimal information* treatment, bosses could neither observe employees' 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 collective firing threats according to which the boss could fire employees at random. In the *peer information* treatment, all employees were endowed with real-time monitoring of their co-employees (as in the *complete information* treatment) but the boss could not monitor. The only way a boss could access individual information about employees was by chatting with them. In the *baseline treatment*, the boss had complete information but could not fire anyone.

The information conditions in which bosses cannot directly collect any individual information about employees (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) that provides a ranking of the firing treatments in terms of organizational performance. In particular, the *minimal* treatment is 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 that remains a crucial empirical question.

In line with our *hypothesis*, we find that employees' production and task dedication were significantly higher in all the firing treatments than in the baseline in which firing employees 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 employees. The relatively large size effect 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 employees. 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 $d \approx 1$). These findings show that firing threats are remarkably effective even when the supervision of employees is minimal.

Our results echo previous research in public good games showing that a collective punishment mechanism that is triggered in case of low group performance increases cooperation rates (Fatas, Jimenez-Jimenez and Morales, 2010; Fatas, Morales and Ubeda, 2010; Bonroy et al., 2019). This collective-punishment mechanism resembles firing under *minimal information* because it only relies on group performance information. However, our setup differs from the one used in these studies in various important dimensions. Beyond methodological differences related to the use of real-effort and real-leisure tasks (Corgnet, Hernán-González and Schniter, 2015, henceforth, CHS), our design captures a principal-agent relationship (e.g. Laffont and Martimort, 2002; Bolton and Dewatripont, 2005) instead of a team production task à la Holmström (1982). In our setup, employees are paid a fixed wage and are under the supervision of a boss who is responsible for firing decisions. We made this choice purposefully so as to mimic the context in which firing threats have been studied in the literature on incentives (e.g. Shapiro and Stiglitz, 1984).

We also compare the results of our firing treatments 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 employees as in the *complete information* treatment would retain 10% of the production of all employees while assigning the remaining 90% to the nine employees at the end of each period. We also compared firing threats to team and individual incentives treatments using data from CHR and CHS. 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 employees' effort (e.g. Shapiro and Stiglitz, 1984). By assessing the effectiveness of firing threats across different levels of information and comparing them to other popular incentive schemes, 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, access the Internet for leisure purposes or chat with other employees or the boss. 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 now describe the main tasks.

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

[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.

[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 tables correctly.

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

net number of correct answers, discounted by the equivalent (in monetary terms) number of mistakes.

2.1.2. *Alternatives to the work task*

The use of a real-effort task is motivated by the fact that our *hypothesis* posits treatment differences but no point predictions, thus not requiring us to induce a specific monetary cost of effort function. Thus, we would expect our *hypothesis* to hold for a broad range of cost of effort functions.

An issue with the use of a real-effort task is that participants may simply work as hard as possible, regardless of the incentives and of other aspects of the environment (see van Dijk, Sonnemans and van Winden, 2001). For this reason, we gave participants access to the Internet. Internet browsing is attractive and distracts employees from completing the work task, thus serving as a measure of on-the-job leisure (see CHS; Goerg, Kube and Radbruch, 2019).

Participants could access an Internet browser at any point during the experiment. Because we did not allow participants to use their mobile phones, Internet use was an appropriate measure of on-the-job-leisure. Subjects were explicitly told in the instructions that their usage of the Internet was strictly confidential.[5]

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 earn $2.40 just by clicking on the boxes.[6] This feature was used to mimic the pay obtained for being present at the workstation regardless of one's 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.[7]

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

[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.

[7] Although participants do engage in the alternative activities to the work task (see Section 4), these activities may not be as attractive as on-the-job leisure in natural settings. If that is true, then observed treatment differences should be seen as lower bounds of differences in natural settings.

2.1.3 *Chatting*

The boss and all employees 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. The chat worked in the same way in all the treatments.

2.2. *Main treatments*

Table 1 summarizes the main features of our treatments together with the number of subjects. In our baseline treatment, employees were rewarded a fixed wage of 200¢ each period; they were not incentivized based on their performance on the *work task*.[8] Bosses received the output produced by all subjects (including themselves) on the *work task*, but were not paid a fixed wage. The boss could monitor the nine employees during the experiment. They had access to a separate monitoring screen with real-time information about each of the employees they selected to monitor. The monitoring screen showed, for each selected employee, the activity they were undertaking (Internet, Chat or Work Task), their updated accumulated production as well as their contribution to the work task (in % terms). At the end of each period, the boss had access to a monitoring summary which included information regarding employees' activities during the period, their production levels as well as their contribution to total production.

In all other treatments (the 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 the employees could be fired in an 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*

[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).

employees and the boss.[10] They were not able to chat with the members of the organization, and they could not be rehired. [11]

Table 1.
Summary of the treatments.

Treatment	Description	Number of sessions (subjects)
No Firing Complete information *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. The boss and the employees had access to the Internet and could chat.	6 (60)
Firing *Complete information*	As in the baseline, the boss could monitor employees' activities and individual production. In addition, they could fire one employee at the end of periods 2, 3 and 4. Payment to employees was the same as in in the baseline but the boss also kept the fixed wage of dismissed employees.	6 (60)
Firing *Partial information*	Same as *complete information* except that bosses could *only* monitor employees' activities not accessing any information regarding their individual production.	6 (60)
Firing *Minimal information*	Same as *complete information* 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)
Firing *Peer information*	Same as *complete information* except that bosses could *not* monitor employees while employees could monitor each other. Bosses only had access to the total production of the organization when deciding upon firing employees. But, they could collect additional information by chatting with employees.	6 (60)

The firing treatments differed in the monitoring activity and the information bosses could obtain from it. In the *complete information* treatment, bosses had the same monitoring information as in the *baseline*. Thus, the baseline and the *complete information* treatments aimed at representing the

[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.
[11] This contrasts with Solda and Villeval (2020) study the costs and benefits of reintegration of previous excluded players from a social dilemma game.

case in which bosses have access to exhaustive information about employees' performance and work dedication. These treatments replicate CHS.

We then considered treatments in which bosses had access to less information to make their firing decisions. In the *partial information* treatment, the monitoring screen only showed the activity undertaken by selected workers without showing any individual information about accumulated production and contribution to the work task. Therefore, employees might be on the task screen pretending to be working whereas they put no effort on the task (as in CHR). 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 employees' productivity levels. However, in our *partial information* 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 information* treatment is likely to be much less costly to collect than in the *complete information* treatment.

In the remaining firing treatments (*minimal information* and *peer information*) bosses did not have access to the monitoring technology. As a result, the only piece of information available to the boss is the total output of the organization, which is released at the end of the period. However, in the *peer info* treatment and unlike the other firing treatments, employees could monitor each other with the same supervision technology as in the *complete information* treatment, whereas the boss could not. 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). In the *peer information* treatment, we changed two features of the design compared to the *complete information* treatment, by turning off monitoring for the boss while turning it on for the employees. Therefore, we consider this treatment as an additional robustness check of the effect of firing threats under limited information and do not focus on the comparison between the *peer information* treatment and other firing treatments.The monitoring summary available to the monitor(s) at the end of each period presented only the information available for each treatment, as explained above.

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 their 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. We used six groups of ten people per treatment following previous studies using the same task with groups of ten people (e.g. Corgnet, Hernán-González, and Rassenti, 2015b; CHR; CHS). Using these datasets, we determined the number of groups we would need to detect differences in production across treatments of 20% with a power of 80% at a 5% significance level. A priori, this ensured we could detect an effect size of 0.3 (Cohen's *d*) or 0.03 (η^2), which is relatively small.[12]

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 survey. Subjects playing the role of an employee (boss) in the *complete*, *partial*, *peer*, *minimal information,* and baseline treatments earned €29.36 (€97.47), €28.06 (€95.58), €28.24 (€83.81) €29.21 (€76.54), and €29.09 (€54.26) on average, respectively.[13] 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 employees who exert low levels of effort. This threat is especially effective when the boss can observe with precision the individual production of each employee 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 employees'

[12] Ex-post, our findings show that we are able to detect differences in production between the baseline and the firing treatments which were all greater than 69.2%. In line with our ex-ante power calculations we are able to detect a 25.9% (28.2%) difference in production between the minimal and the partial (complete) information treatments. Consistently with our ex-ante power calculations, we are not able to detect a 9.7% (1.8%) difference in production between the minimal and the peer information (complete and partial) treatments.

[13] Earnings related to production and salaries were, on average, €9.52 (€72.37), €9.52 (€67.57), €9.55 (€60.96), €9.74 (€53.10), and €10 (€34.40) for employees (bosses) in the *complete*, *partial*, *peer*, *minimal info*, and baseline treatments, respectively. Remaining earnings correspond to a €5 show-up fee, survey payments, and the gains obtained from clicking on yellow boxes.

output and effort levels. In this treatment, we expect firing threats to be especially powerful in incentivizing employees because shirkers would very likely be identified in which case they would be dismissed and would suffer a monetary loss.

The effectiveness of 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 employee 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 employee's performance via chatting. However, it is unlikely that employees 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 employees by dismissing one of them at random when the level of performance of the organization is excessively low. In that context, employees might decide to avoid shirking to prevent the boss from firing one of the employees which would lead to their dismissal in 1 out of 9 cases.[14] 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 an employee at random because they might inadvertently dismiss high producers. We thus expect firing threats to be weaker in the *minimal information* treatments compared to the *partial* and *complete information* treatments in which the boss can implement threats at the individual level.[15]

Below, we state our hypothesis regarding the performance ranking of the baseline and the *complete*, *partial* and *minimal information* treatments.

[14] This the case in which no workers had yet been fired. This probability increases as the size of the organization shrinks over time.
[15] 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.

Hypothesis (Performance ranking)

i. *Employees' output will be the highest in the complete information treatment followed by the partial information treatment.*
ii. *Employees' output will be lower in the minimal information treatment than in the partial information treatment.*
iii. *Employees' output will be the lowest in the baseline.*

In addition to testing the performance ranking hypothesis, we will also report on how the *peer information* treatment ranks with respect to the other treatments introduced above. As already mentioned this analysis is exploratory because the *peer information* treatment is not directly comparable to the other treatments. Indeed, in the *peer information* treatment, we changed two features of the design compared to the *complete information* treatment, by turning off monitoring for the boss while turning it on for the employees.

In the *peer information* treatment, the boss does not directly observe any individual information about employees, as in the *minimal information* treatment. However, the boss could instruct an employee to collect information about other employees. Therefore, the boss might have access to more information than in the *minimal information* treatment. In addition, peer pressure might be more intense in the *peer information* treatment compared to the other treatments because, unlike other treatments, employees are potentially monitored by other employees (see Corgnet, Hernan-Gonzalez and Rassenti, 2015b). The previous arguments suggest that the *peer information* treatment will outperform the *minimal information* treatment. However, there is no reason to believe that employees will transmit reliable information to the boss as this is cheap talk. Employees, who are directly competing for keeping their employment, might systematically downplay the performance of other employees thus impeding the boss to gather any reliable information. This behavior is especially likely because any information about employees' individual performance cannot be verified a posteriori by the boss. In sum, our conjecture regarding the ranking of the *peer information* treatment is not clear-cut. Conservatively, we predict that employees' output under the *peer information* treatment will be at least as high as in the *minimal information* treatment.

4. Results

In Section 4.1 we analyze production, work dedication and profits in the first four of the five periods of our experiment which correspond to the periods in which firing threats had monetary consequences. In Section 4.2, we focus on firing decisions. 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.

4.1. *Employees' production, work dedication and firm profits*

We present some descriptive statistics of individual production and work dedication in Tables A.1 and A.2 in Appendix A. To analyze employees' production, we only include subjects who belong to the organization in a given period, thus excluding fired subjects. This means that in the three treatments where firing was possible, subjects who had already been fired were excluded from the analysis.

Figure 1 shows the average production of employees in the first four periods separately and jointly (see '1 to 4' in the figure). The overall pattern observed is qualitatively consistent with the ranking of treatments stated in our h*ypothesis*. One can see that, for periods 1 to 4, average output is the highest in the *complete information* treatment (152.95) followed by the *partial information* treatment (150.24) and the *minimal information* treatment (119.25). In addition, we observe that all firing treatments lead to higher levels of production than the baseline (70.46). We also observe that the production level for *peer information* (130.68) is in between the *partial information* and *minimal information* treatments.

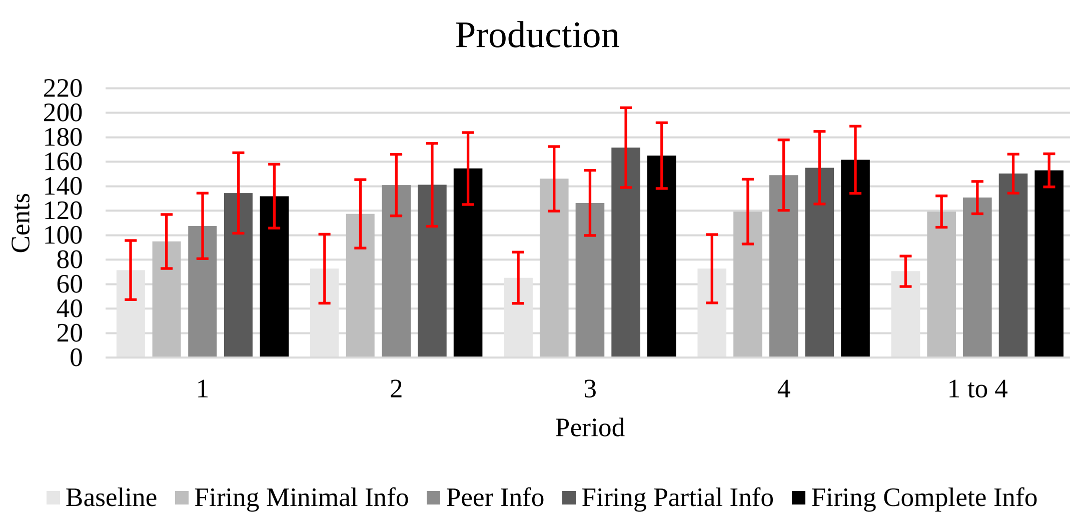

Figure 1. Employees' average production across treatments for periods 1 to 4. Subjects who have been fired before a current period are excluded. The bars show 95% confidence interval.

To test our *hypothesis*, we use a GLS random-effects panel model with robust standard errors clustered at the session level for employees' production.[16] Using the Breusch-Pagan Lagrange Multiplier test, we cannot reject the random effects specification. The statistical analysis is reported in Table 2. Robustness checks are conducted in Appendix B. Tables B.1 and B.2 use pairwise comparison between treatments in separate regressions, and Tables B.3 to B.6 introduce additional controls and consider all periods instead of the first four. Tables B.7 to B.9 use non-parametric tests instead of panel regressions. We show that our findings continue to hold regardless of the robustness check.

The regression results in Table 2 show that all the firing treatments yield higher employees' production than the baseline (*Hypothesis iii*) because all the firing dummies are positive and significant (all p-values < 0.01). The results of the F-tests comparing treatment coefficients (see lower panel of Table 2) show that *complete information* and *partial information* treatments yield higher production than *minimal information* (*Hypothesis ii*) (p-values = 0.018 and 0.020, respectively). However, in contrast with *Hypothesis i*, we observe no statistically significant difference between *complete information* and *partial information* treatments (p-values = 0.802). In addition, we report that production levels under *peer information* are not significantly different from *minimal information, partial information,* and *complete information* treatments (p-values = 0.316, 0.140 and 0.170, respectively). None of the results are qualitatively affected by including the fired subjects in the analysis and considering their production to be zero (see Table B.10 in Appendix B in which we replicated Table 2 using firm profits, defined as total production of employees minus paid wages, as the dependent variable).

In addition to production levels, we analyze 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.

[16] 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 that are shown to be statistically significant using robust standard errors continue to be significant when using the wild bootstrap procedure.

Table 2

GLS regressions (Baseline treatment as baseline condition) with random effects for individual production and working time (periods 1 to 4). Robust standard errors. Excluding fired workers.

	Individual Production	Working Time
Constant	-41.37*	781.35***
	(18.07)	(62.29)
Complete info[+]	74.51***	364.29***
	(13.92)	(47.98)
Partial info[++]	78.40***	387.36***
	(15.40)	(49.43)
Minimal info[+++]	43.34***	346.08***
	(13.26)	(49.55)
Peer info[++++]	56.13***	279.95***
	(13.46)	(53.23)
Ability	6.06***	- 4.38
	(.99)	(2.71)
Gender	16.83*	19.09
	(9.70)	(27.69)
Observations	1024	1024
R^2	0.2064	0.2891

[+]Dummy that takes the value 1 for *complete information* treatment and 0 otherwise.
[++] Dummy that takes the value 1 for *partial information* treatment and 0 otherwise.
[+++] Dummy that takes the value 1 for *minimal information* treatment and 0 otherwise.
[++++] Dummy that takes the value 1 for *peer information treatment* and 0 otherwise.
*p-value<0.1, **p-value<.0.05, and ***p-value<.0.01. (Standard deviation in parentheses)

Results of F-tests of equality between treatment coefficients:

	Production		**Working time**	
Treatment comparison	**chi2(1)**	**Prob > chi2**	**chi2(1)**	**Prob > chi2**
Complete info vs partial info	0.06	0.8017	1.53	0.2166
Complete info vs minimal info	5.59	0.0180	0.78	0.3768
Complete info vs peer info	1.88	0.1702	8.95	0.0028
Partial info vs minimal info	5.43	0.0198	3.32	0.0686
Partial info vs peer info	2.18	0.1403	14.01	0.0002
Minimal info vs peer info	1.00	0.3161	4.54	0.0330

In Figure 2 and in line with our analysis of production, we observe that employees spent considerably less time on the work task in the baseline (62.13%) than in any of the other three firing treatments (91.17%) (all p-values < 0.01 in Table 2). In line with production results, the lower panel of Table 2 shows that employees spent as much time working in the *partial information* treatment (95.38%) as in the *complete information* (92.51%) (p-value = 0.217). In addition, employees spent more time working in the *partial information* treatment than in the *minimal information* treatment (91.25%) although the difference is only marginally significant (p-value = 0.069). Under *peer information*, employees spent significantly less time working than in any other firing treatments (see lower panel in Table 2). This finding is not surprising because employees had access to an additional alternative activity (monitoring) to the work task in the *peer information* treatments which occupied 2.57% of their time on average.

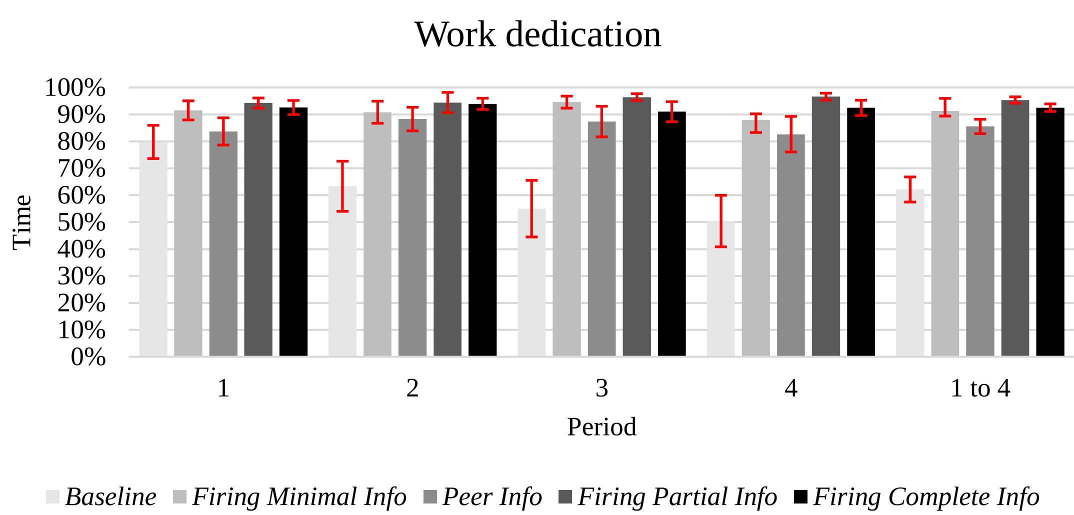

Figure 2. Employees' average working time (%) across treatments for periods 1 to 4. Subjects who have been fired before an actual period are excluded. 95% confidence interval bars.

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 the cases in treatments *complete information*, *partial information*, *peer information*, *minimal information* 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 information* 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 information* and *peer information* firing treatments (Cohen's $d \approx 0.63$) that do not rely on any individual information about employees. Our results put forth that firing threats are 'robust' incentive schemes (Hart and Holmström, 1986) which, regardless of the information available to the boss, effectively increase employees' production.

4.2. *Firing decisions*

In this section, we analyze bosses' firing decisions in the firing treatments to see whether firing patterns are consistent with the *informativeness* principle that underlies our *hypothesis*. In Table 3, we show the average production of fired and non-fired employees per period and per treatment. Focusing on the last column which shows total production, one can see that in the *complete information*, *partial information* and *peer information* treatments fired worked were producing significantly less than the rest of employees prior to being fired.

By contrast, in the *minimal information* treatment there was no significant difference in the average production levels between fired employees 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 employees could be accessed, whereas they had to fire at random when no such individual information was available.[17]

The fact that bosses in the *peer information* 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 information* 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 information* 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 bosses, however, as only few messages were dedicated to reveal information about other employees (1.4) or themselves (5.3).[18] Thus, only a limited amount

[17] In the last column of Table 3, 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¢).

[18] For the other firing treatments these numbers of messages were, on average, slightly lower: 0.2 and 4.42, respectively. This implies that almost no information was shared by employees about the performance of their coworkers.

of reliable information could have been transmitted to the boss via chat. This might explain why we do not observe significant differences in production levels between the *peer* and *minimal information* treatments.

Table 3

Firing decisions per period across treatments.

Treatment		Period 2	Period 3	Period 4	Total
Complete Information	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**)
Partial Information	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**)
Peer Information	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**)
Minimal Information	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. The number above the p-value is the value of the test statistic.

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 employees.

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 three other relevant incentive schemes: discretionary bonuses (Nalbantian and

Schotter, 1997; Gerhart, Rynes and Fulmer, 2009; Bryson et al. 2012), individual incentives, and team incentives. To make these comparisons, we use data from CHR and CHS that employed the same real-effort task, thus allowing for a direct comparison with our findings. 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 employees was equally shared among them as well as an individual incentives treatment in which all ten employees were paid exactly the amount they individually produced. In CHR, an individual incentives treatment was also conducted along with two of the treatments in the current study (baseline and *complete information*).[19] Because these prior studies did not consider the incentive effect of discretionary bonuses, we conducted an additional treatment. In the discretionary bonus treatment, the boss kept 10% of the value of all output produced by the nine employees in the organization. Bosses could monitor employees' activities and individual production (as in the baseline and *complete information* treatments) and distributed the remaining 90% of the total production of the organization to employees at the end of each period. In addition, bosses were paid the value of their own production.

We start by showing that there were not statistical differences in production levels between our data and CHR for the baseline and *complete information* treatments (see Table G.1 and Figure G.1 in 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 information* treatments) do not significantly underperform any of the incentive schemes which also make use of individual information about employees such as individual incentives and discretionary bonuses (see Tables F.1 and F.2a,b,c in Appendix F, first two columns).[20] This result is especially notable for the *partial information* treatment which, despite using less information than individual incentives and discretionary bonuses, achieves a similar level of performance. In addition, both *partial* and *complete information* treatments outperform team incentives (see Table F.3, first two columns). Firing threats which cannot be based on reliable individual information about employees

[19] The individual incentives treatment in CHR differs from CHS because employees were paid a 200¢ fixed wage in addition to the piece rate.
[20] In Appendix F, we show graphical comparisons of production levels in the baseline and the four firing treatments with the discretionary bonus treatment (Figure F.1), the treatments from CHR and CHS (Figure F.2) and the team incentives treatment (Figure F.3).

(*minimal* and *peer information* treatments) significantly underperform individual incentives and discretionary bonuses (see Tables 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 perform at least as well as 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 employees' 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 employees. For example, the *complete information* 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 information* treatment does not significantly outperform the *partial information* 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 information* treatments (see Table 3) suggests that monitoring employees may have been as intensive in the *partial information* treatment as in the *complete information* treatment. Indeed, bosses spent about the same time monitoring in the *complete information* treatment (14.58%) as in the treatment with *partial information* (10.40%) (see Table A.3 in Appendix A). This difference was not statistically significant (see Table A.4). In addition, this reasonable amount of monitoring does not seem to correspond to a case of high monitoring intensity as 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 employees 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 employees' production. In Appendix C, we present additional analyses that show that an employee 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 information* (p-value = 0.366 for the dummy variable that takes the value 1 if an employee has been watched in the previous 5 minutes) and *complete information* (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 information* and *partial information* treatments might simply be due to ceiling effects. This interpretation is consistent with the fact that the organizational output in the *partial information* 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*

The presence of social motives is a potential explanation for why performance is high in the *minimal information* and *peer information* compared to the baseline. One notable difference across our firing treatments is that firing decisions, and thus employees' compensation, are based on different type of information about performance. In the *minimal information* and *peer information* treatments, firing decisions are likely to be based on total production, which is the most relevant piece of information available to bosses. By contrast, firing decisions will likely depend on individual information in the *partial information* and *complete information* treatments. It follows that an employee's pay in the *minimal information* and *peer information* treatments depends 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 information* and *complete information* treatments as pay should only depend on one's own performance. It follows that in the *minimal information* and *peer information* 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.

The explanation of our findings based on altruistic motives implies that the presence of altruistic employees will have a more positive effect on organizational performance in the *minimal information* and *peer information* treatments than in the *partial information* and *complete information* treatments. However, our data are not consistent with this interpretation of our findings. In Table D.1 in Appendix D, we show that the interaction term (Altruism × SP, where SP is a dummy that takes value 1 for the *minimal* and *peer information*) is actually negative and marginally significant.

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 information* and *peer information* treatments (10.15% of available time is spent chatting) compared to *partial information* and *complete information* (4.35%).[21] In addition, the proportion of time employees spent encouraging and helping their peers (categories 4, 5 and 7) was not higher in the *minimal information* and *peer information* treatments compared to the *partial information* and *complete information* treatments (see Appendix E).

In addition to altruistic motives, it could also be that treatments differ in reciprocal cooperation (see e.g. Fehr, Fischbacher and Gächter, 2002; Fehr and Fischbacher, 2003; Burlando and Guala, 2005; Fischbacher and Gachter, 2010). Employees might be more likely to increase their production when their peers do so in the *minimal information* and *peer information* treatments compared to the *partial information* and *complete information* treatments. To test this conjecture, we study the relationship between individual production and the production of the other employees in the previous period (see Table D.2 in Appendix D). We find that individual production does not depend on peers' production, regardless of the treatment. This finding is not consistent with an interpretation of our findings based on reciprocal cooperation.

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 their incentive effects across various information levels. 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.

[21] This result continue to hold if we only consider chatting activities between peers.

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 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 employees.

Our second major finding relates to the comparison of incentive schemes that make use of individual information about employees. We show that firing threats based on partial information of employees' individual contributions are particularly appealing to employers. This is the case because they perform similarly to incentive schemes based on detailed information about employees' individual contributions including individual incentives and discretionary bonuses whereas their implementation costs are likely to be small. These findings are consistent with the widespread use of firing threats based on limited but uncontroversial and easily measured information about employees' 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 schemes in combination with other organizational policies (Milgrom and Roberts, 1995; Ichniowski and Shaw, 2003; Roberts, 2007; Brynjolfsson and Milgrom, 2013). Future research could also extend the current study to assess the long-term effects of firing threats, especially in cases in which limited information is used. It could be that employees' motivation weakens over time when the firing policy is perceived as procedurally unfair. This concern is likely to apply to the case of the minimal information treatment. Likewise, further research should study whether employees will be willing to apply to a firm that uses firing threats on a regular basis, especially when it is based on minimal information.

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.

Bonroy, O., A. Garapin, S. Hamilton and D. Souza Monteiro (2019), “Free-Riding on Product Quality in Cooperatives: Lessons from an Experiment,” *American Journal of Agricultural Economics*, 101(1), 89-108.

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.

Burlando, R. M., and F. Guala (2005): “Heterogeneous Agents in Public Goods Experiments,” *Experimental Economics*, 8(1), 35-54.

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 (2015a, CHR): "Firing Threats: Incentive Effects and Impression Management," *Games and Economic Behavior,* 91, 97-113.

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.

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

Croson, R., E. Fatas, T. Neugebauer and A. Morales (2015), *Journal of Economic Behavior and Organization,* 114, 13-26.

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.

Fatas, E., F. Jimenez-Jimenez and A. Morales (2010): “Blind Fines in Cooperatives,” *Applied Economic Perspectives and Policy*, 32(4), 564-587.

Fatas, E., A. Morales and P. Ubeda (2010): “Blind Justice: An Experimental Analysis of Random Punishment,” *Journal of Economic Psychology*, 31, 358-373.

Fehr, E. and U. Fischbacher (2003): "The Nature of Human Altruism," *Nature*, 425(6960), 785-791.

Fehr, E., Fischbacher, U. and S. Gächter (2002): "Strong Reciprocity, Human Cooperation, and the Enforcement of Social Norms," *Human Nature*, 13(1), 1-25.

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

Fischbacher, U. and S. Gachter (2010): "Social Preferences, Beliefs, and the Dynamics of Free Riding In Public Goods Experiments," *American Economic Review*, 100(1), 541-56.

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.

Goerg, S. J., Kube, S., & Radbruch, J. (2019): "The Effectiveness of Incentive Schemes in the Presence of Implicit Effort Costs," *Management Science*, 65(9), 4063-4078.

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.

Holmström, B. (1982): "Moral Hazard in Teams," *The Bell Journal of Economics*, 324-340.

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.

Solda, A. and M.C. Villeval (2020): "Exclusion and Reintegration in a Social Dilemma," *Economic Inquiry* 58(1), 120- 149.

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.

Van Dijk, F., J. Sonnemans, and F. van Winden (2001): "Incentive Systems in a Real Effort Experiment," *European Economic Review*, 45, 187-214.

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

Table A.3
Period evolution of monitoring activities (% of total time).

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

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

	Complete Information vs. Partial Information	Complete Information vs. Baseline	Complete Information vs. Peer Information
Constant	117.94***	135.29***	11.19
	(46.02)	(46.78)	(9.79)
Complete info+	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

+Complete info is a dummy variable that takes value 1 for the Complete Information treatment, and 0 otherwise.
*p-value<.10, **p-value <.05, and ***p-value <.01. (Standard deviation in parentheses)

Appendix B. Robustness analyses

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

	Complete Info vs. Partial Info	Complete Info vs. Peer Info	Complete Info vs. Minimal Info	Complete Info vs. Baseline	Partial Info vs. Peer Info	Partial Info vs. Minimal Info	Partial Info vs. Baseline	Peer Info vs. Minimal Info	Peer Info vs. Baseline	Minimal Info vs. Baseline
Constant	- .86	10.94	- 4.72	- 22.56	- 10.26	- 26.10	- 41.06	11.58	- 7.80	- 12.10
	(29.50)	(24.78)	(24.27)	(25.48)	(28.88)	(28.87)	(29.34)	(21.11)	(23.48)	(22.22)
Treatment^{+}	- 6.88	18.06	30.74**	76.13***	23.58	36.37**	81.41***	12.50	57.24***	44.70***
	(15.45)	(13.48)	(13.22)	(13.97)	(15.34)	(15.42)	(15.43)	(12.67)	(13.27)	(13.11)
Ability[22]	8.47***	5.58***	5.45***	5.62***	7.67***	7.43***	7.29***	4.08***	4.62***	4.49***
	(1.66)	(1.44)	(1.29)	(1.34)	(1.91)	(1.69)	(1.69)	(1.27)	(1.38)	(1.23)
Gender^{++}	19.61	23.95	27.03*	8.19	17.75	21.63	4.30	29.71*	7.96	12.30
	(16.20)	(15.34)	(14.60)	(14.84)	(16.86)	(15.54)	(15.79)	(14.97)	(14.92)	(14.39)
Observations	419	386	423	426	385	422	425	389	392	429
R^2	0.1980	0.1553	0.1591	0.2566	0.1774	0.1756	0.2545	0.1102	0.1795	0.1424

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

Table B.2

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

	Complete Info vs. Partial Info	Complete Info vs. Peer Info	Complete Info vs. Minimal Info	Complete Info vs. Baseline	Partial Info vs. Peer Info	Partial Info vs. Minimal Info	Partial Info vs. Baseline	Peer Info vs. Minimal Info	Peer Info vs. Baseline	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 B.3[23] (Additional controls)
GLS regression with random effects for individual production (periods 1–4) across treatments. Robust standard errors. Excluding fired employees.

	Complete Info vs. Partial Info	Complete Info vs. Peer Info	Complete Info vs. Minimal Info	Complete Info vs. Baseline	Partial Info vs. Peer Info	Partial Info vs. Minimal Info	Partial Info vs. Baseline	Peer Info vs. Minimal Info	Peer Info vs. Baseline	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.
Aheadness and behindness aversion are defined as in Bartling et al. (2009).
*p-value<0.1, **p-value<.0.05, and ***p-value<.0.01. (Standard deviation in parentheses)

[23] 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 (Additional controls)[24]

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

	Complete Info vs. Partial Info	Complete Info vs. Peer Info	Complete Info vs. Minimal Info	Complete Info vs. Baseline	Partial Info vs. Peer Info	Partial Info vs. Minimal Info	Partial Info vs. Baseline	Peer Info vs. Minimal Info	Peer Info vs. Baseline	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.
Aheadness and behindness aversion are defined as in Bartling et al. (2009).
*p-value<0.1, **p-value<.0.05, and ***p-value<.0.01. (Standard deviation in parentheses)

[24] 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.5 (Additional controls)[25]

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

	Complete Info vs. Partial Info	Complete Info vs. Peer Info	Complete Info vs. Minimal Info	Complete Info vs. Baseline	Partial Info vs. Peer Info	Partial Info vs. Minimal Info	Partial Info vs. Baseline	Peer Info vs. Minimal Info	Peer Info vs. Baseline	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.
Aheadness and behindness aversion are defined as in Bartling et al. (2009).
*p-value<0.1, **p-value<.0.05, and ***p-value<.0.01. (Standard deviation in parentheses)

[25] 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.6 (Additional controls)[26]

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

	Complete Info vs. Partial Info	Complete Info vs. Peer Info	Complete Info vs. Minimal Info	Complete Info vs. Baseline	Partial Info vs. Peer Info	Partial Info vs. Minimal Info	Partial Info vs. Baseline	Peer Info vs. Minimal Info	Peer Info vs. Baseline	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)
Obsrvations	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.
Aheadness and behindness aversion are defined as in Bartling et al. (2009).
*p-value<0.1, **p-value<.0.05, and ***p-value<.0.01. (Standard deviation in parentheses)

[26] 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.7
Wilcoxon Rank Sum tests for production between treatments periods 1 to 4. Fired workers excluded.

Treatment comparison	z-value	p-value
Complete information vs partial information	1.031	0.3024
Complete information vs minimal information	3.610	0.0003
Complete information vs peer information	2..165	0.0304
Complete information vs baseline	9.148	0.0000
Partial information vs minimal information	2.496	0.0126
Partial information vs peer information	1.004	0.3155
Partial information vs baseline	8.284	0.0000
Minimal information vs peer information	-1.570	0.1164
Minimal information vs baseline	6.395	0.0000
Peer information vs baseline	7.407	0.0000

Table B.8
Wilcoxon Rank Sum tests for working time between treatments periods 1 to 4 fired workers excluded.

Treatment comparison	z-value	p-value
Complete information vs partial information	-4.279	0.0000
Complete information vs minimal information	-0.821	0.4115
Complete information vs peer information	4.999	0.0000
Complete information vs baseline	9.457	0.0000
Partial information vs minimal information	3.091	0.0020
Partial information vs peer information	8.385	0.0000
Partial information vs baseline	11.543	0.0000
Minimal information vs peer information	5.207	0.0000
Minimal information vs baseline	9.575	0.0000
Peer information vs baseline	5.990	0.0000

Table B.9
Wilcoxon Rank Sum tests for firm's profits between treatments periods 1 to 4.

Treatment comparison	z-value	p-value
Complete information vs partial information	0.186	0.8527
Complete information vs minimal information	2.704	0.0068
Complete information vs peer information	1.781	0.0750
Complete information vs baseline	5.507	0.0000
Partial information vs minimal information	2.352	0.0187
Partial information vs peer information	1.450	0.1470
Partial information vs baseline	5.156	0.0000
Minimal information vs peer information	-0.790	0.4294
Minimal information vs baseline	4.518	0.0000
Peer information vs baseline	4.751	0.0000

Table B.10 (Firm profits)
GLS regressions (Baseline treatment as baseline condition) with random effects for firm profit (periods 1 to 4). Robust standard errors.

	Firm Profit
Constant	634.17***
	(92.03)
Complete info[+]	754.17***
	(145.78)
Partial info[++]	732.5***
	(159.24)
Minimal info[+++]	449.17***
	(116.76)
Peer info[++++]	555.83***
	(138.61)
Observations	116
R^2	0.4096

[+] Complete info takes the value 1 for Complete Information treatment and 0 otherwise.
[++] Partial info takes the value 1 for Partial Information treatment and 0 otherwise.
[+++] Minimal info takes the value 1 for Minimal Information treatment and 0 otherwise.
[++++] Peer info takes the value 1 for Peer Information treatment and 0 otherwise.
*p-value<0.1, **p-value<.0.05, and ***p-value<.0.01. (Standard deviation in parentheses)

Results of F-tests of equality between treatment variables:

Treatment comparison	**chi2(1)**	**Prob > chi2**
Complete information vs partial information	0.02	0.8999
Complete information vs minimal information	5.18	0.0228
Complete information vs peer information	1.67	0.1960
Partial information vs minimal information	3.64	0.0564
Partial information vs peer information	1.13	0.2879
Minimal information vs peer information	0.72	0.3977

Appendix C. 5-minute analysis.

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

	Complete Information	Partial Information
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 an employee 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 employees' production (periods 1–4) across treatments where firing was allowed. Robust standard errors clustered by session. Excluding fired employees.

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 information* and *peer information* treatments and 0 for the *complete* and *partial information* treatments. Excluding fired subjects.
*p-value<0.1, **p-value<.0.05, and ***p-value<.0.01. (Standard deviation in parentheses)

Table D.2
GLS regressions with random effects for individual production by treatment (periods 2 to 5). Robust standard errors.

	Complete Information	Partial Information	Minimal Information	Peer Information	Baseline
Constant	173.74***	109.37***	116.67***	113.99***	59.64**
	(30.06)	(30.88)	(27.06)	(22.20)	(24.27)
Group production previous period+	- .02	.03	- .01	.01	.01
	(.02)	(.02)	(.02)	(.02)	(.04)
Observations	203	203	209	170	216
R^2	0.0049	0.0039	0.0015	0.0123	0.0001

+Total group production minus worker individual production in the previous period.

*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 E.3). Then, we computed the Cohen's Kappa coefficient for each category to assess inter-rater agreement (see Table E.1).[27] 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 E.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 E.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 *Complete Information* (19.87), *Partial Information* (16.25%), *Peer Information* (19.99%), and *Minimal Information* (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 *Complete Information* treatment (11.04% for category 4, and 11.69% for category 13) compare to the baseline, *Partial Information*, *Peer Information*, and *Minimal Information* 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

[27] 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.

the baseline treatment (31.54%) than in *Complete Information* (4.85%), *Partial Information* (3.85%), *Peer Information* (10.21%), and *Minimal Information* (10.12%) (see Table E.4).

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

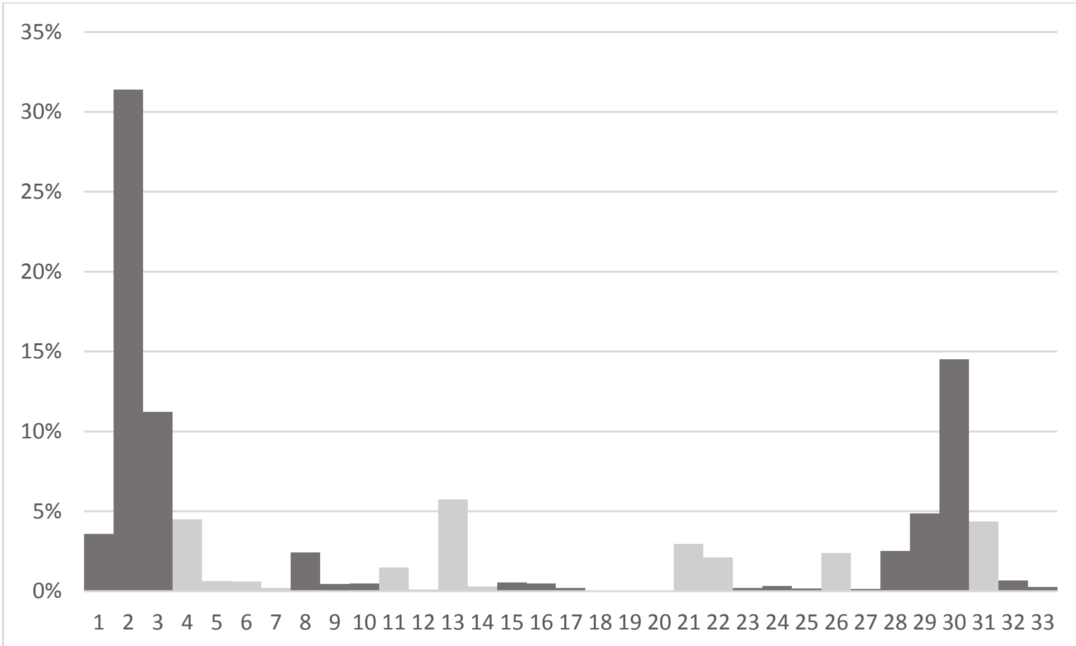

Figure E.1. Histogram of categorization of messages for all treatments.

Table E.2 Percentage of categories by treatment.

Category	Baseline	Complete Information	Partial Information	Peer Information	Minimal Information
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 *Bs* they will all be paid the same if they achieve a certain level of total production
	20	*C* telling all *Bs* 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 employees.

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 information* and *partial information* treatments and 0 for the *complete information* and *partial information* 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 employees.

	Complete Info vs. Disc. Bonus Complete Info	Partial Info vs. Disc. Bonus Complete Info	Peer Info vs. Disc. Bonus Complete Info	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)

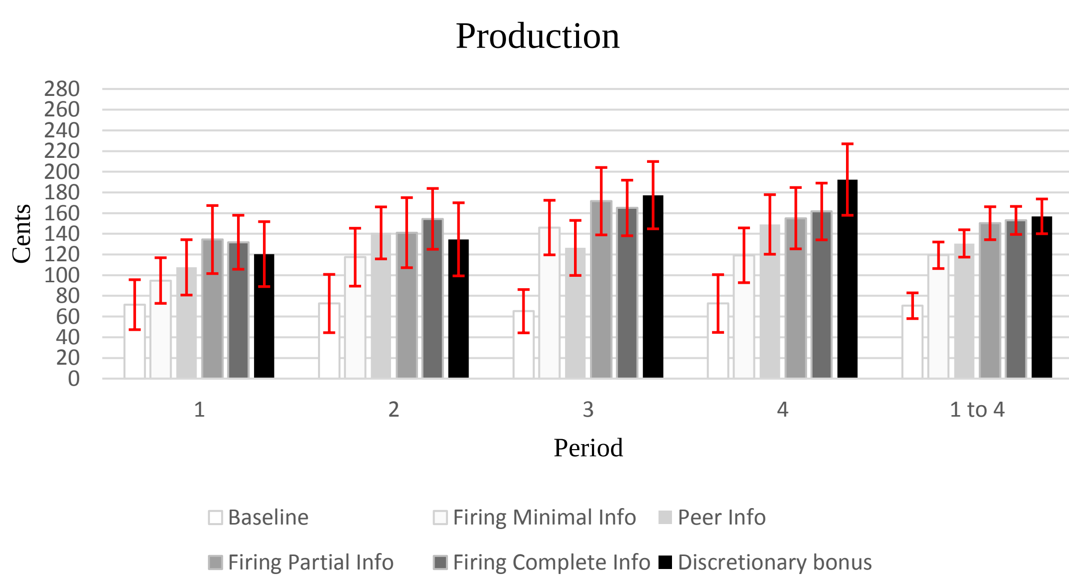

Figure F.1. Employees' average production across treatments for periods 1 to 4. Subjects who have been fired before a current period are excluded. The bars show 95% confidence interval.

Table F.2a

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

	Complete Info vs. Individual Incentives (CHR)	Partial Info vs. Individual Incentives (CHR)	Peer Info vs. Individual Incentives (CHR)	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 employees.

	Complete Info vs. Individual Incentives (CHS)	Partial Info vs. Individual Incentives (CHS)	Peer Info vs. Individual Incentives (CHS)	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 employees.

	Complete Info vs. Individual Incentives (CHR & CHS)	Partial Info vs. Individual Incentives (CHR & CHS)	Peer Info vs. Individual Incentives (CHR & CHS)	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)

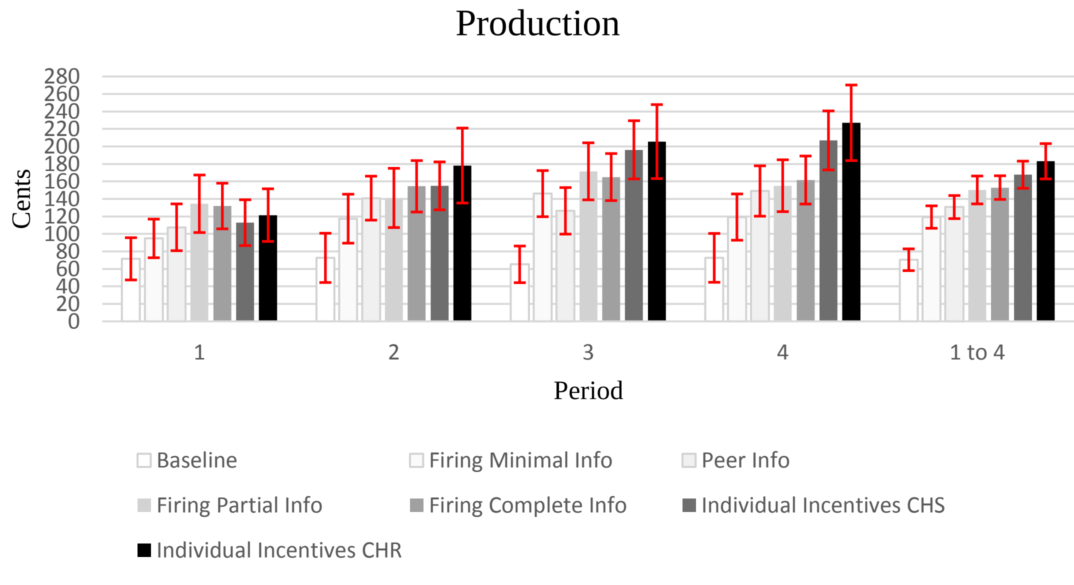

Figure F.2. Employees' average production across treatments for periods 1 to 4. Subjects who have been fired before a current period are excluded. The bars show 95% confidence interval.

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

	Complete Info vs. Team Incentives (CHS)	Partial Info vs. Team Incentives (CHS)	Peer Info vs. Team Incentives (CHS)	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)

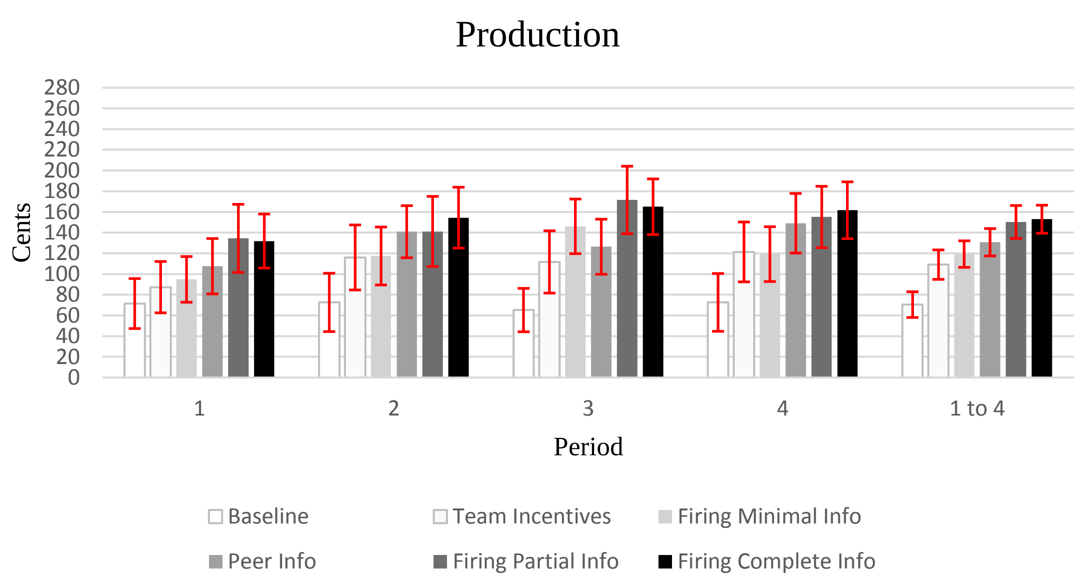

Figure F.3. Employees' average production across treatments for periods 1 to 4. Subjects who have been fired before a current period are excluded. The bars show 95% confidence interval.

Appendix G. Replication

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

	Complete Info vs. 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)

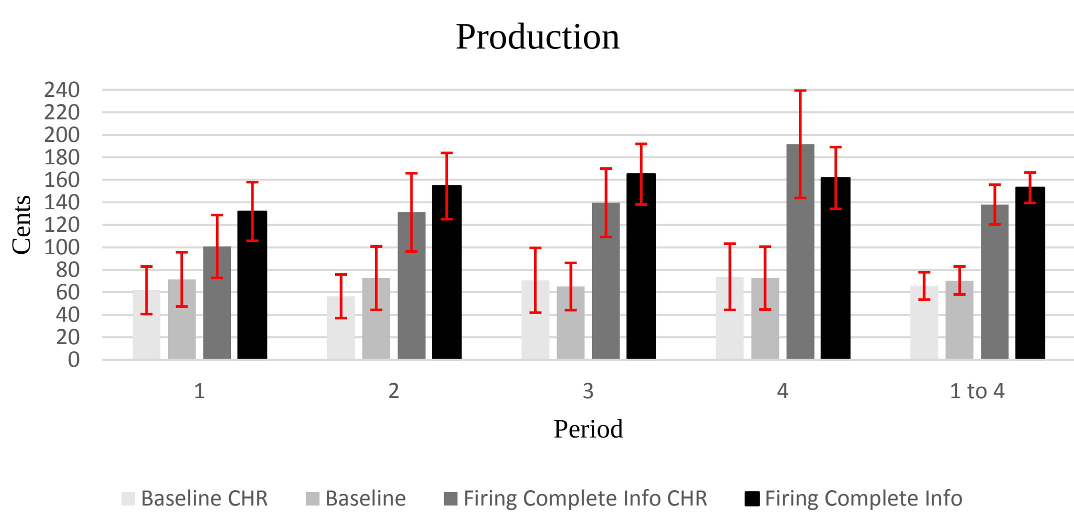

Figure G.1. Employees' average production across treatments for periods 1 to 4. Subjects who have been fired before a current period are excluded. The bars show 95% confidence interval.

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.

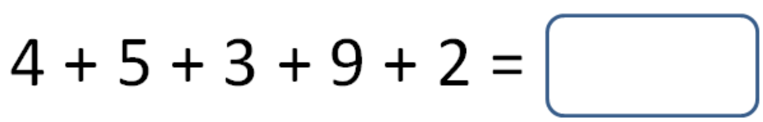

Figure O1.1. Example of Adding Task question.

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