2015

Peer Pressure and Moral Hazard in Teams: Experimental Evidence

Brice Corgnet
Chapman University, corgnet@chapman.edu

Roberto Hernán-González
Nottingham University

Stephen Rassenti
Chapman University, rassenti@chapman.edu

Recommended Citation

Corgnet, B., Hernán-González, R., and Rassenti, S. (2015). "Peer Pressure and Moral Hazard in Teams: Experimental Evidence," *Review of Behavioral Economics*: Vol. 2: No. 4, pp 379-403.

Peer Pressure and Moral Hazard in Teams: Experimental Evidence

Review of Behavioral Economics, 2015, 2: 379–403

Peer Pressure and Moral Hazard in Teams: Experimental Evidence

Brice Corgnet[1], Roberto Hernán-González[2] and Stephen Rassenti[3]

[1]*Argyros School of Business and Economics and Economic Science Institute, Chapman University, USA*
[2]*Nottingham University, UK*
[3]*Economic Science Institute, Chapman University, CA, USA*

ABSTRACT

Team incentives have been found to be particularly effective both in the lab and in the field despite the moral hazard in teams problem identified by Holmström (1982). In a newly developed virtual workplace, we show that, in line with Holmström, moral hazard in teams is indeed pervasive. Subsequently, we find strong evidence for the conjecture of Kandel and Lazear (1992) that peer pressure may resolve the moral hazard in teams problem. Organizations equipped with a very weak form of peer monitoring (anonymous and without physical proximity, verbal threats or face-to-face interactions) perform as well as those using individual incentives.

Keywords: Incentives, Free-riding, Monitoring, Peer pressure, Organization theory

JEL Codes: C92, D23, M52

> "An organization can secure the efforts necessary to its existence, then, either by the objective inducements it provides or by changing states of mind.... We shall call the process of offering objective incentives "the method of incentives"; and the processes of changing subjective attitudes "the method of persuasion."
>
> —Barnard (1938, p. 142)

Online Appendix available from:
http://dx.doi.org/10.1561/105.00000040_app
ISSN 2326-6198; DOI 10.1561/105.00000040

1 Introduction

1.1 Team Incentives in the Theory of Organizations

As a point of departure for the analysis of organizations and the development of an economic theory of the firm, theorists have put forward the pervasiveness of free-riding behaviors in teams in which it is difficult to observe and verify the contribution of each partner (Alchian and Demsetz, 1972; Holmström, 1982).

At the empirical level, the evidence of free riding behavior in teams has been limited (e.g., Leibowitz and Tollison, 1980; Encinosa *et al.*, 2007). Instead, team incentives have been found to be particularly effective both in laboratory experiments (Van Dijk *et al.*, 2001; Dohmen and Falk, 2011) and in field studies (Dumaine, 1990, 1994; Kruse, 1992; Manz and Sims, 1993; Ichniowski *et al.*, 1996; Hansen, 1997; Ichniowski *et al.*, 1997; Hamilton *et al.*, 2003; Babcock *et al.*, 2015). For example, Hamilton *et al.* (2003) show that equal sharing of production bonuses within teams stimulates cooperation, information sharing, monitoring and even mutual training, generating a productivity increase (relative to individual incentives) despite the expected free-rider problem. In a recent paper, Babcock *et al.* (2015) show that team incentives can outperform individual incentives in fostering students' attendance to the university gym Club.

Admittedly, shirking behaviors have been observed pervasively in the public good game literature (Ledyard, 1995). However, the present paper as well as the previously mentioned literature focus on the effect of incentives on shirking behaviors. The aim is to assess the difference, if any, in shirking behaviors when people are paid according to team versus individual incentives.

In this study, we are able to compare team and individual incentives while controlling for team-specific features that may interfere in the empirical assessment of team incentives. This is crucial because the empirical difficulty in identifying the free-riding behaviors in teams is likely due to the lack of control over crucial aspects of work teams. For example, peer monitoring, interpersonal relations or implicit incentives (e.g., firing threats) may have acted as confounding factors in the evaluation of team incentives.

Our experimental environment differs from previous experimental settings as it introduces a long real-effort *work task* as well as real-time access to leisure activities (Internet browsing). We find that production levels were on average 32.8% lower under team incentives than under individual incentives. This corresponds to a decrease in subjects' hourly production of \$1.65 or 20.7% of the minimum hourly wage where the experiments were conducted. This result was driven by extensive shirking behaviors in the team incentives treatment in which subjects spent on average 28.5% of their time browsing the Internet. These results are consistent with incentives theory (see Holmström (1979) and

Laffont and Martimort (2002) for a review) as they confirm the sound premise that performance is increased by the use of high-powered incentives schemes.

As a second step of our analysis, we introduced a real-time monitoring technology in our virtual organizations so as to assess whether the poor performance of team incentives could be mitigated by peer pressure.

1.2 Supervision and Peer-monitoring in the Theory of Organizations

Supervision is an important aspect of the theory of the firm that was mentioned by preeminent scholars as one of the raison d'être of organizations (Jensen and Meckling, 1976; Barzel, 1982; Chandler, 1992). Alchian and Demsetz (1972) put forward the need for centralized supervision in a context of asymmetric information between managers and their subordinates in a team context. They stress that peer monitoring is not an efficient mechanism because the agents would tend to shy away from monitoring activities. However, other theories view peer monitoring as a highly-effective mechanism (Kandel and Lazear, 1992; Mohnen *et al.*, 2008; Carpenter *et al.*, 2009; Gill and Stone, 2015). For example, Kandel and Lazear stress the role of shame arising when workers produce less than the group average as an important mechanism in understanding the effectiveness of peer pressure. Carpenter *et al.* (2009) emphasize the role of negative reciprocity as a behavioral mechanism leading contributors to voluntary incur private costs to punish free riders. Evidence of such behaviors has been found in public good experiments (Fehr and Gächter, 2000; Sefton *et al.*, 2007) as well as in contests between groups (Abbink *et al.*, 2010; Grosse *et al.*, 2011) stress the popularity of peer monitoring devices in a modified version of the public good game in which subjects could vote on whether to use a central monitor or rely on a decentralized monitoring system (peer monitoring).[1] The authors found that subjects mostly relied on peer monitoring as a disciplining device challenging the idea of Alchian and Demsetz that a central monitor is needed to avoid free riding behaviors in the provision of monitoring.

Peer effects have been reported in a series of field experiments. For example, Sacerdote (2001) and Zimmerman (2003) report peer effects on students' grades among college roommates. Falk and Ichino (2006) found that students who worked for fixed wages to stuff envelopes performed significantly better when working in pairs than when working alone. Mas and Moretti (2009) studied the case of supermarket cashiers and found positive peer effects on the number of items scanned by cashiers. The authors considered workers' visual contact and frequency of interactions as measures of peer pressure.

[1]Under peer monitoring, each subject decided how much to invest in the monitoring technology which precision determined the allocation of team profits. In particular, the proportion of the team profits which was allocated according to individual contributions increased in the precision of the monitoring technology.

Our approach differs from previous field works as it introduces anonymous real-time supervision in a controlled laboratory environment. Our aim is to measure peer pressure with as much precision as possible by recording the amount of time subjects spent watching others or the activities which were completed by the subjects who were being watched. Furthermore, our anonymous supervision mechanism allows us to isolate the effects of possible cofounds that may appear in a face to face interaction such as fear of retaliation. Our peer monitoring technology was such that each team member could monitor peers' activities at any point in time during the experiment. As a result, subjects could shape their monitoring strategy by deciding upon which subjects to monitor and when to do so.[2] Monitors were informed in real-time about the activities undertaken by supervisees and could therefore identify whether they were browsing the Internet or producing for the organization. In the peer pressure monitoring treatment, subjects were notified on their screen whenever they were being watched by others. This feature induced social pressure which is defined by Mas and Moretti (2009) as a case in which workers experience disutility when they are observed behaving selfishly by their peers.

Our environment offers a unique opportunity to provide a detailed analysis of peer monitoring activities. In the peer pressure treatment, a large proportion of subjects (88.3%) decided to monitor others. However, subjects dedicated only a small proportion of their time to monitoring activities (4.4%), compared with the proportion of their time subjects spent working (82.5%) or browsing the Internet (13.1%). Yet, all subjects were being watched for an average of 22.4% of their time. Team members shared the monitoring burden and maintained peer pressure during the whole duration of the experiment.

Our main contribution to the peer pressure literature is to show that team incentives combined with peer monitoring led to levels of performance and shirking (Internet usage) that were remarkably similar to individual incentives in the absence of punishments devices, communication technologies or physical proximity among subjects. In the peer pressure treatment, hourly production was \$1.55 higher (47.1% higher) and Internet usage was 54.1% lower than under team incentives alone. These findings confirm the conjecture of Kandel and Lazear (1992) that peer pressure may be an effective solution to the moral hazard in teams problem identified by Holmström (1982).

In contrast to public good games with monetary punishments (Carpenter, 2007a, 2007b; Fehr and Gächter, 2000), both effort and efficiency were increased by the introduction of peer monitoring. This was the case because subjects spent little time watching others as they shared the monitoring burden to limit the cost of monitoring.

[2]This endogenous aspect of our monitoring technology can be linked to search experiments in which subjects decide whether to observe or not their relative performance (Falk *et al.*, 2006; Burks *et al.*, 2013).

In addition, we were able to answer the question of "How is peer pressure generated? (Kandel and Lazear, 1992, p. 805)". We started by analyzing the dynamics of peer monitoring and showed that subjects were less likely to switch from the work task to the Internet if they had been watched in the previous five minutes. We shed some light on this effect by conducting experiments in which organizational members could watch each other's activities without being noticed by their peers. We show that, in contrast to visible peer monitoring, the invisible monitoring technology did not reduce shirking. These results indicate that effective peer monitoring crucially hinges on social pressure.

2 Experimental Design and Hypotheses

2.1 Virtual Organization

We develop a framework in which subjects could undertake a real-effort organizational task, have access to Internet, and monitor other subjects' behavior in real-time.[3]

2.1.1 The Work Task

We use a long, repetitive and effortful task to ensure that individual performance is mostly driven by effort considerations. Subjects were asked to sum up tables of 36 numbers for one hour and 40 minutes. Each table completed correctly generated a 40-cent profit while a penalty of 20 cents was subtracted from individual production for each incorrect answer. After each subject completed a table, the accumulated individual production was updated so that subjects knew whether their answer was correct or not. At the end of each period, and only then, participants were informed about the total amount of money generated by all 10 participants' *work task* during the period.[4]

2.1.2 Internet Browsing

At any point during the experiment, participants could switch from the *work task* to the leisure activity that consisted of browsing the Internet. Each activity was undertaken separately, in a different screen but the Internet browser was embedded in the software so that the experimenter could keep a record of the switching times between activities as well as the exact amount of

[3] A video presentation of the software is available at http://sites.google.com/site/vopeerpressure/home/videos.

[4] Similar real-effort tasks are used by Eriksson *et al.* (2009), Niederle and Vesterlund (2007), Bartling *et al.* (2009), and Dohmen and Falk (2011).

time subjects spent on each activity. To switch from one activity to another subjects simply had to click on the corresponding option of the drop-down menu at the bottom-right of their screens. Internet browsing introduced temptation in the spirit of recent self-control experiments providing on-the-job distraction activities such as watching a humorous video (Bucciol *et al.*, 2013).[5]

The consideration of leisure-related issues in the experimental literature was first introduced in the analysis of labor supply by Dickinson (1999). Participants had to undertake a two-hour typing task on four different days. In one of the two treatments (the combined experiment), subjects could leave the laboratory whenever they had achieved a certain output level. This aimed at capturing off-the-job leisure activities. In the intensity experiment, the author assesses on-the-job leisure by studying the pace at which workers completed the task. In our experimental design we measure on-the-job leisure directly by recording the exact amount of time each subject spent on the Internet.[6]

2.1.3 The Click Pay

In addition to the previously mentioned activities, each subject could click on a yellow box moving slowly from left to right at the bottom of their screen. A new yellow box appeared every 25 seconds whether the subject was currently *working* or *browsing the Internet*. Each time subjects clicked on the box they earned *5 cents*. Given that the experiment consisted of 5 periods of 20 minutes each, subjects could receive a total of $12.00 just by clicking on all the 240 yellow boxes that appeared on the screen during the experiment. This aimed at representing the pay that workers obtain just for being present at their workstation regardless of their commitment to the *work task*. This was implemented to mimic real work environments in which only a portion of employees' compensation is pay-for-performance. We also wanted to allow participants to earn money even when they decided to shirk. This is in the spirit of Mohnen *et al.* (2008) who paid subjects (0.10€) to take a 25-second time-out during which they were not able to work on the incentivized counting task. The idea is to construct laboratory environments in which on-the-job shirking is commonly observed so as to be able to uncover the incentive effects (Corgnet *et al.*, 2011; Corgnet *et al.*, 2015).

2.1.4 Real-time Monitoring

In the monitoring treatments, subjects were able to monitor others' activities in real time. We allowed subjects to monitor their peers' activities at any time

[5]Note that usage of cell phones was not allowed in the lab so that Internet usage, if any, was exclusively embedded in the experimental platform.

[6]Other on-the-job leisure activities, such as giving subjects access to magazines, are used in related studies (Eriksson *et al.*, 2009; Charness *et al.*, 2010).

during the experiment by selecting the *Watch* option. In that respect, our monitoring technology offers a unique opportunity to assess the effect of peer pressure over time and examine the conjecture that peer effects may fade away as time passes (Falk and Ichino, 2006).

Monitoring activities were undertaken in a separate screen so that subjects could not participate in the *work task* or the leisure activity while monitoring others. In the monitoring screen, subjects could decide whether to monitor only a subset or all the other subjects at the same time. The information was displayed in a table, where each column showed information regarding the activities completed by a given subject. Monitors were informed about the activities undertaken by each subject (*Internet*, *Work Task*, or *Watch*), their current production as well as their contribution to the *work task* (in % terms).

In the peer pressure treatment, subjects were notified with a message stating the experiment *ID* of the watcher jointly with an eye picture whenever they were being watched. We also conducted a treatment in which subjects were not notified when they were being watched by others (invisible monitoring) so as to isolate the role of social pressure in peer monitoring. Note that social pressure, though minimal, is not totally eliminated in the invisible monitoring treatment since workers may still feel that they are watched by their peers even if they are not notified about it.

The monitoring technology used in the present paper allows for precise control over the supervision activities which is difficult to obtain in the field (Bandiera *et al.*, 2005; Falk and Ichino, 2006; Mas and Moretti, 2009). For example, we can measure the exact amount of time subjects were watched by others as well as the amount of time they spent watching others. It is also possible to identify the watchers as well as the subjects who were being watched.

Another distinctive feature of our monitoring technology is that subjects could freely decide upon their monitoring strategy. Subjects could choose who to monitor and when to do so. This feature of the supervision technology allows us to study subjects' monitoring behavior.

Note that our virtual monitoring technology does not allow workers to hide from the watcher as would be the case in traditional work environments. However, we believe that our technology may be a good representation of modern and highly-computerized workplaces that use virtual monitoring to track employees' activities (e.g., *Virtual Monitoring*™).

2.2 Treatments

We ran four treatments. In the baseline, subjects were rewarded on the *work task* according to their individual production (treatment I). In the second treatment (treatment T), the total production of the 10 subjects participating in the experiment was equally distributed among them. Our third experiment

was the peer pressure treatment (treatment *TP*) which was equivalent to treatment *T* except that all ten subjects could monitor others using the peer monitoring technology. Treatment *TPN* was similar to treatment *TP* except that organizational members were not notified on their screen when they were being watched by another subject.

In all treatments subjects could individually obtain the full rewards ($2.40 per period) for clicking on yellow boxes. The instructions for each treatment are available online.[7]

2.3 Conceptual Framework

We build our conceptual framework on the moral-hazard in teams problem introduced by Holmström (1982) and on its extension to the presence of peer monitoring which was proposed by Kandel and Lazear (1992) (see online appendix for details and proofs). We summarize our findings in the following hypotheses.

Hypothesis 1 (**Individual incentives versus team incentives**). *Production is expected to be greater and Internet usage is expected to be lower under individual incentives than under team incentives.*

Regarding the comparison of the team incentives and the peer pressure treatments, we use the work of Kandel and Lazear (1992) and provide an illustration of their peer pressure function by referring to the work of Andreoni and Bernheim (2009) on social image (see online appendix for details and proofs). In particular, we consider that workers care about their social image and feel pride if they produce more than a given benchmark level of effort and feel shame if they fall short of the benchmark. In order to assess the interaction between audience effects and peer monitoring, we introduce a distinction between visible (treatment TP) and invisible audiences (treatment TPN). We assume that a person is more affected by social image concerns when the audience is visible than when it is not.

In the presence of concerns for social image, we expect to find higher levels of production in the peer monitoring treatments than in the team incentives treatment without peer monitoring (Hypothesis 2).

Hypothesis 2 (**Peer monitoring**). *Production is expected to be greater and Internet usage is expected to be lower in the peer monitoring treatments, than in the team incentives treatment without peer monitoring as long as workers are sufficiently concerned with their social image. In that case, workers are expected to dedicate part of their time to peer monitoring activities so as to foster the effort of the other workers.*

[7] http://sites.google.com/site/vopeerpressure/home/instructions. Instructions for treatment *TPN* were the same as for treatment *TP* except for slide 36 which was removed.

We should also recognize that one might expect peer monitoring activities to backfire generating distrust among workers. Recent research has emphasized this negative aspect of monitoring and put forward that trusting employees can lead to higher levels of effort than intensive supervision (Frey, 1993; Falk and Kosfeld, 2006; Fehr *et al.*, 2007a, 2007b; Dickinson and Villeval, 2008). We do not consider crowding-out of effort as our primary hypothesis because the disciplining effect of supervision has been found to be dominant in the absence of interpersonal relationships among workers as is the case in our experimental design (Frey, 1993; Dickinson and Villeval, 2008). In addition, crowding-out effects are likely to be stronger in a principal-agent relationship or in any situation in which the monitor has some authority on the supervisee's work. In our design, we consider a multi-agent monitoring structure in which there is no principal and no hierarchy among subjects.

Finally, we conjecture that the impact of social image concerns is diminished under invisible monitoring compared with the peer pressure treatment. As a result, we expect invisible monitoring to have a lower impact on production and internet usage than visible monitoring (peer pressure treatment).

Hypothesis 3 (**Peer pressure and invisible monitoring**). *Production is expected to be lower and Internet usage is expected to be greater in the invisible monitoring treatment than in the peer pressure treatment.*

Interestingly, the invisible monitoring treatment will also help us assess whether any effect of the peer monitoring technology can be accounted for by the access to continuous feedback on others' production levels. In particular, if the effect of peer monitoring on workers' production levels is driven by the access to feedback rather than to social pressure we should observe invisible monitoring to perform as well as the peer pressure treatment (see Nikiforakis (2010) for the study of feedback in public good games with monetary punishments).

2.4 Procedures

Our subject pool consisted of students from a major American university with a diverse population. Participants were recruited by emails from a pool of more than 2,000 students who had signed up to participate in experiments. The experiments took place in December 2010 and February 2011. In total, 246 subjects participated in the experiment, divided in 25 sessions. We ran seven sessions for treatment I, and six sessions for each of treatments T, TP, and TPN. Ten students participated in each session, except for two sessions of 8 students in treatment I.[8]

[8]We conducted two to four sessions (maximum capacity of the lab) at a time rendering the identification of session partners difficult. Also, subjects were called sequentially for payments and left the facility right away.

The instructions were displayed on subjects' computer screens. Subjects had exactly 20 minutes to read the instructions. Three minutes before the end of the instructions period, a monitor entered the room announcing the time remaining and handing out a printed copy of the summary of the instructions. None of the participants asked for extra time to read the instructions.

At the end of the experiment, subjects were paid their earnings in cash, rounded up to the nearest quarter. Participants in treatments I, T, TP, and TPN earned on average \$27.25, \$24.45, \$27.10, and \$24.95 respectively. This includes a \$7.00 show-up fee. Experimental sessions lasted on average two hours and fifteen minutes.

3 Results

3.1 Team Incentives versus Individual Incentives

3.1.1 Individual Production

We define production as the monetary amount generated by a subject's answers on the *work task* divided by the reward for each correct answer (40 cents). It can be interpreted as the total number of correct tables completed by a given subject discounted by the number of incorrect answers. In both the treatments, period production steadily increased except for the third period as is
as evidence of a learning effect which is commonly observed in long arithmetic tasks (Charness and Campbell, 1988). The increase in production overtime should not be seen as evidence of limited shirking under team incentives. For example, the positive trend is less pronounced under team incentives than under individual incentives (see Table II.1 in the online appendix for regression analyses). Also, subjects who produced more than the group average increased their production overtime under individual incentives whereas they failed to do so under team incentives (see Table II.2 in the online appendix).

Average individual production per period was equal to 4.21 tables under individual incentives compared with 2.83 tables under team incentives. This corresponds to a 48.8% production gap (\$1.65 in hourly production) between individual and team incentives (see Table A1 in the appendix for statistical analyses).[9] This finding also holds when comparing individual production across treatments for each of the five periods separately.

[9]We use clustered t-tests and clustered rank-sum tests. The clustered version of the Wilcoxon rank-sum test was performed using Datta and Satten test (2005) while the clustered version of the t-test followed (Donner *et al.*, 1981). We aim at controlling for the fact that individual production in a given session may be affected by group production. This correction is especially relevant for the treatment with team incentives in which case the contributions of other group members, displayed on a subject's screen at the end of each period, may affect an individual's motivation.

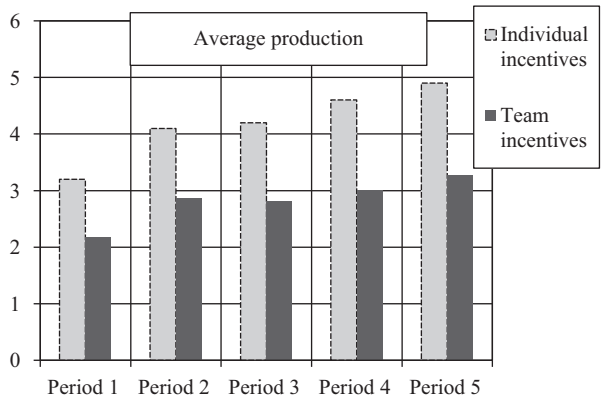

Figure 1: Average production per period across treatments.

Result 1 (*Work task* production: Individual versus team incentives)**.** *Total production and period production were significantly greater in the individual incentives treatment than in the team incentives treatment.*

Result 1 is not surprising in the light of incentive theory (Hypothesis 1) but constitutes an essential step in the empirical analysis of incentives given the limited evidence of free riding behaviors in teams in controlled environments using real-effort tasks. A related analysis was conducted by Nalbantian and Schotter (1997) in an abstract experimental setting in which the authors compared different types of group incentives programs ranging from revenue sharing (team incentives) to target-based and team-tournament incentives.[10]

3.1.2 Internet Usage

We report a positive trend in Internet usage in both treatments which is most pronounced under team incentives (see Table II.3 in the online appendix for regression analyses). This suggests that the treatment effect became stronger over time as subjects' fatigue and boredom set in. Under team incentives subjects spent on average 28.5% of their time browsing the Internet while this percentage was only equal to 11.9% under individual incentives. We reject the

[10]We do not study different types of group incentives schemes. Rather, we focus on team incentives (revenue sharing) schemes. Notice that Nalbantian and Schotter (1997) did not compare individual and team incentives. Indeed, individual incentives schemes in an abstract effort setting automatically lead subjects to choose the efficient level of effort e^*. Instead, the authors study the more interesting case of a wage-cum-supervision mechanism. In that case, agents are paid a wage W as long as they are not caught by the principal choosing an effort level below e* in which case they would receive a lower wage ($W^- < W$). Under this scheme, the principal can only verify the chosen level of effort with a given probability p.

hypothesis that Internet usage was identical for individual and team incentives whether it is for the whole experiment or for each of the five periods analyzed separately (see Table A1 in the appendix). These findings are in line with Hypothesis 1.

Result 2 (Internet usage: Individual versus team incentives)**.**

(i) Internet usage was significantly lower in the individual incentives treatment compared with team incentives. This result also holds when analyzing each period separately.

(ii) The increase in Internet usage over time was significantly more pronounced in the team incentives treatment than in the individual incentives treatment.

Related experiments have stressed the relevance of off-the-job leisure activities that were assessed by analyzing quitting behaviors (Dickinson, 1999; Falk and Huffman, 2007) but these studies have not assessed the impact of on-the-job leisure on performance.

This finding emphasizes that, in an environment with a long and real-effort task in which fatigue was likely to set in, high-powered incentives were very effective in bringing down Internet usage. Indeed, subjects spent almost three times as long on the Internet under team incentives than under individual incentives. The introduction of Internet as an alternative activity is a crucial feature of our environment that may have led subjects to consider leisure activities to be as salient as the *work task*. Yet, many subjects never consulted the Internet (40.9% and 11.7% under individual and team incentives, respectively) focusing exclusively on completing the *work task*. The incentives effects identified in the current study are unlikely to be identified in an experimental environment in which subjects do not have access to on-the-job leisure activities (Corgnet *et al.*, 2015).[11]

Note that in addition to the *work task* and *Internet browsing*, subjects could obtain earnings from clicking on boxes appearing at the bottom of their screen. Unsurprisingly, no significant differences were observed across treatments as subjects successfully clicked on the box in 98% (97%) of its appearances under individual (team) incentives. We report no differences across treatments (all $p\text{-}values > 0.5$).

[11] In this paper, we study the effect of not allowing the subjects to access the internet during the experiment for both individual and team incentives. We find that, under team incentives, production levels were significantly lower when Internet browsing was available than when it was not. Under individual incentives, however, no differences in production levels were observed between the treatment in which Internet was available and the treatment in which it was not.

3.2 Peer Monitoring

We start the analysis of peer monitoring by providing general statistics on watching activities for both monitoring treatments, with (treatment *TP*) and without notification (treatment *TPN*).

3.2.1 Watching Activities

Subjects were watched 22.4% (29.9%) of the time in treatment *TP* (*TPN*) while subjects' dedication to monitoring activities was limited to 4.4% (5.3%) of their available time. This occurred because most watchers, regardless of the monitoring treatment, decided to monitor all subjects at the same time. As a result, the amount of time subjects were being watched during the experiment was similar across subjects. In particular, subjects with different levels of performance were being watched for the same amount of time (see Table II.4 in the online appendix).[12]

On average, subjects monitored their peers 5.7 (6.9) times during the experiment for an average duration of 46 (45) seconds per watching episode in treatment *TP* (*TPN*). It is interesting to note that subjects were willing to dedicate a significant amount of their time to monitor others even in the case in which monitors could not exert peer pressure on other subjects (treatment *TPN*). This suggests that, besides exerting peer pressure, subjects monitored others to obtain feedback about their relative performance as well as to scrutinize others' behavior in the organization. Comparing monitoring treatments, we observe no statistically significant differences regarding the amount of time subjects spent watching others (see Table A2 in the appendix). However, we find that subjects were watched significantly more often under the invisible monitoring treatment than under the peer pressure treatment. This follows from the fact that in the invisible monitoring treatment subjects were significantly less likely to watch only a subset of the other nine organizational members (5.1% of the watching episodes) than under peer pressure (11.1% of the watching episodes). These findings are consistent with the fact that, in treatment *TPN*, monitoring was driven by the willingness to observe others' behaviors and compare oneself with the group while in treatment *TP*, monitoring could have been partly driven by concerns for exerting peer pressure. Even though we cannot disentangle the exact motive of watchers in the peer pressure treatment, we know that watching others induced peer pressure (whether it is intentional or not) because those who were being watched were explicitly notified.

[12]This result should be interpreted with care as it may simply be due to the fact that monitoring all subjects was quick and easy (the option to "monitor all" was used in 93.3% of the watching episodes) making the decision of whom to monitor less relevant. In our setting, monitoring all subjects could be done at no extra cost by clicking on the "monitor all" button.

Interestingly, monitoring did not fade away over time. The proportion of their time subjects spent watching others was equal to 4.2% (4.8%) in the first period compared to 5.6% (5.4%) in the last period in treatment *TP* (*TPN*). Considering the experiment as a whole, only 7 out of 60 (3 out of 60) of the subjects did not spend any time monitoring their peers in treatment *TP* (*TPN*). In our experiment, monitoring entailed an opportunity cost since subjects who watched others had to leave the *work task* screen affecting their production negatively. However, this monitoring cost was shared among team members because subjects were paid according to team incentives. As a result, any decline in production due to monitoring activities would affect all workers in the same magnitude. Our environment differs from the model presented by Alchian and Demsetz (1972) and tested by Grosse *et al.* (2011) in which subjects who are paid according to their individual contribution would incur an individual cost for monitoring.

Interestingly, subjects rarely watched the same person at the same time. This occurred only in 16.7% and 17.4% of the watching episodes in treatments *TP* and *TPN* (proportion test, p-value $= 0.924$). It is not surprising to report that all subjects were watched during the experiment for at least 12 minutes (16 minutes) in treatment *TP* (*TPN*). In treatment *TP*, we also observe that subjects were more likely to be watched right after watching others (see Table II.5 in the online appendix for the analysis).[13]

Result 3 (Watching activities)**.**

(i) Regardless of the monitoring treatment, watching activities were limited to a small percentage of subjects' available time. Nevertheless, all subjects were being watched during the experiment for an average of 22.4% and 29.9% of the duration of the experiment in treatments TP and TPN, respectively.

(ii) Regardless of the monitoring treatment, watching activities did not fade away across periods.

(iii) The magnitude of watching activities was similar across monitoring treatments. However, monitors were more likely to watch only a subset of subjects in treatment TP than in treatment TPN. As a result, subjects were more likely to be watched in treatment TPN than in treatment TP.

3.2.2 Comparison of Individual Production Across Treatments

Average production was 47.1% larger (hourly production was \$1.55 higher) in the peer pressure treatment than in the team incentives treatment. The

[13]This could be seen as evidence of retaliation. We thank an anonymous referee for suggesting this analysis.

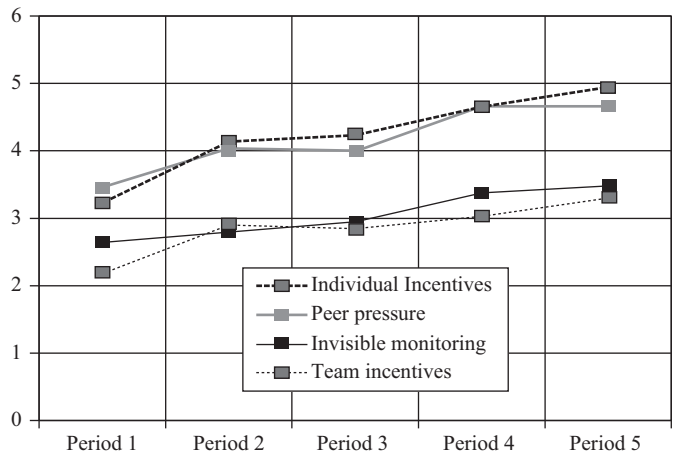

Figure 2: Average production per period for all treatments.

statistical significance of our results holds across periods (see Table A1 in the appendix). Average total production under peer pressure (20.6) was remarkably close to the case of individual incentives (21.0). Similarly to previous treatments, individual production in the monitoring treatments increased over time (see Figure 2). These findings are in line with Hypothesis 2.

In our design, peer monitoring was costing money to the firm by detracting employees from their work. In the peer pressure treatment, employees spent 4.4% of their time monitoring others but this opportunity cost was offset by the strong positive impact of peer monitoring on employee productivity.[14]

Invisible monitoring did not lead to any significant increase in either total or period production with respect to the team incentives treatment without monitoring (see Figure 2 and Table A1 in the appendix). At the same time, invisible monitoring led to average production levels which were 26.9% and 28.4% lower (hourly production was \$1.33 and \$1.43 lower) than in the peer pressure and individual incentives treatments. These findings are consistent with Hypothesis 3.

Result 4 (*Work task* production: Peer monitoring versus team and individual incentives)**.**

(i) Total production and period production were significantly greater in the peer pressure treatment than in the team incentives treatment.

(ii) Total production and period production were not significantly different between the peer pressure and the individual incentives treatments.

[14]The monitoring cost did not include a fixed monetary cost such as the one incurred, for example, for purchasing the virtual monitoring technology.

(iii) Total production and period production were not significantly different between the invisible monitoring treatment and the team incentives treatment without monitoring.

(iv) Total production and period production were significantly lower in the invisible monitoring treatment than in the peer pressure treatment.

The absence of any positive effect on production levels in the invisible monitoring treatment suggests that social pressure is a crucial element of the effectiveness of the monitoring technology. We conducted additional analyses and showed that being watched by others in a given time span of five minutes increased one's own production in the next five to ten minutes in the peer pressure treatment (see Table A3 in appendix). By contrast, watching others in a given time span of five minutes did not affect one's own production in the following minutes. Interestingly, the increase in production that follows from being watched was more pronounced when a subject was watched by more than one watcher (see Table II.6 in the online appendix).[15] This result confirms the intuition that the effect of peer pressure is magnified by the number of supervisors. This positive effect is also present in the Kandel and Lazear (1992) model of peer pressure.

Our results also show the robustness of peer effects encountered in the field (Falk and Ichino, 2006; Mas and Moretti, 2009) in a controlled laboratory experiment. We were able to obtain strong peer monitoring effects under anonymity and in the absence of monetary punishments. In field studies such as the one designed by Mas and Moretti (2009), workers were not anonymous and could potentially face retaliation for non-cooperative behaviors. In our design, the interaction between subjects was anonymous so as to prevent any form of retaliation after the experiment. In contrast to field studies (Falk and Ichino, 2006; Mas and Moretti, 2009) and public good games with threats (Masclet *et al.*, 2003), subjects were not allowed to communicate in our experiment. The effectiveness of our peer monitoring technology did not rely on physical proximity, verbal threats or face to face interactions. The fact that our monitoring technology was highly effective despite the absence of physical proximity and face to face communication is especially relevant given the growing interest for virtual monitoring devices within firms. A large number of programs such as *Spector Soft*, *Virtual Monitoring*™, *Employee Monitoring* or *Webwatcher* are already available to monitor employees' activities in real time. These platforms closely resemble our experimental implementation of virtual monitoring by allowing monitors to scrutinize employees' activities and track the time spent

[15]This is the case because in our regression analysis the interaction term between the amount of time a subject was watched and the number of watchers (in $t-2$) is positive and significant. This shows that the positive effect of peer monitoring is not solely driven by the amount of time a subject was watched but also by the number of subjects who were watching. We thank an anonymous referee for suggesting this analysis.

on various applications whether using their desktop computers, laptops or smart phones. Importantly, our findings complement previous research on peer effects by showing that peer monitoring can exactly offset the loss in production resulting from the use of team incentives schemes which are pervasive in firms.

In contrast to other supervision mechanisms, peer monitoring does not seem to induce crowding-out of effort which has been reported in recent experimental works (Frey, 1993; Falk and Kosfeld, 2006; Dickinson and Villeval, 2008). These authors stress that supervision may be perceived as a signal of distrust and, as a result, undermine workers' effort. Frey (1993) as well as Dickinson and Villeval (2008) put forward that the crowding-out effect that results from monitoring activities dominates its disciplining effect when there exist interpersonal relationships between managers and employees, whereas the opposite tends to be true in the absence of such relationships. In that respect, our findings are consistent with these previous works since our experimental design is characterized by the absence of interpersonal relationships among workers.

3.2.3 Comparison of Internet Usage Across Treatments

Similarly to previous treatments, we identify a positive trend in Internet usage for both monitoring treatments. The proportion of time subjects dedicated to Internet in treatment *TP* [*TPN*] in the first two periods was only 7.7% [9.7%] on average compared with 16.7% [26.4%] in the last three periods. Peer monitoring had a considerable impact on Internet usage (see Figure 3), however.[16] The average proportion of time subjects spent on Internet was significantly lower in the peer pressure treatment (13.1%) than in the team incentives treatment (28.5%). This difference in Internet usage was significant whether considering total Internet usage or Internet usage per period (see Table A1 in the appendix). Interestingly, we find slightly significant differences in Internet usage between the invisible monitoring treatment (19.8%) and the team incentives treatment (28.5%). This supports the conjecture that social pressure may not be fully eliminated in the invisible monitoring treatment. Subjects may refrain from using the Internet so as to avoid being caught by an invisible monitor. Nevertheless, Internet usage was significantly lower in the peer pressure treatment than under invisible monitoring.

The evolution of Internet usage was remarkably similar for the peer pressure and the individual incentives treatments (see Table A1 in the appendix). By

[16] The results reported in this section are similar if we analyze working time (time spent on the *work task*) rather than Internet usage (see Table A1). Using working time instead of Internet usage allows us to control for the fact that monitoring activities may have been used by subjects as an alternative leisure activity. One could argue that the low Internet usage in peer monitoring treatments is due to the substitutability between monitoring and Internet activities.

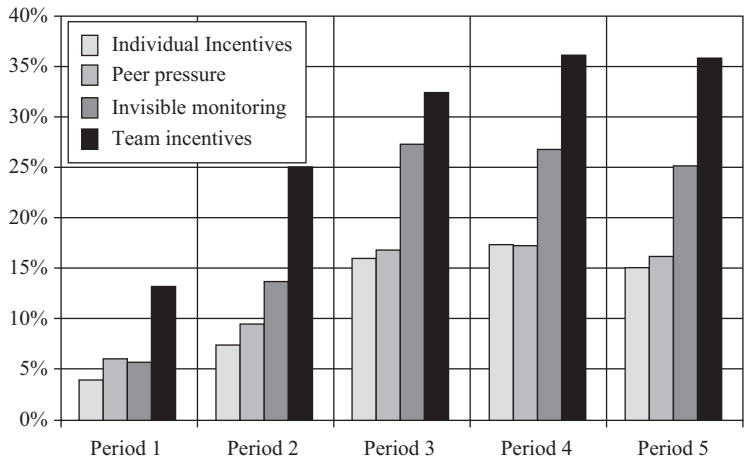

Figure 3: Average Internet usage (in %) for all treatments across periods.

contrast, internet usage was significantly higher under invisible monitoring (19.8%) than under individual incentives (11.9%).

In addition, we analyze whether knowing that they were being watched affected subjects' use of the Internet during the experiment. We show that in the peer pressure treatment, subjects were less likely to switch from the work task to the Internet if they had been watched by others in the previous five minutes (see Table II.7 in the online appendix). In sum, the introduction of peer monitoring in our experimental design brings down Internet usage. This is an important finding given the growing concern for cyber-slacking (Malachowski, 2005; Young, 2006).

Result 5 (Internet usage: Peer monitoring versus team and individual incentives)**.**

(i) Internet usage was significantly lower in the peer pressure treatment than in the team incentives treatment. Also, Internet usage was marginally lower in the invisible monitoring treatment than in the team incentives treatment.

(ii) Internet usage was not significantly different between the peer pressure and the individual incentives treatments. However, Internet usage was significantly higher in the invisible monitoring treatment than in the individual incentives and peer pressure treatments.

4 Conclusions

We studied peer monitoring as an example of mechanism that may allow organizations to recover the efficiency loss provoked by the use of weak incentives. We found that using team incentives in combination with peer monitoring allowed organizations to reach production levels that were as high as in the case of individual incentives. In contrast to public good games with punishments, both effort and efficiency were increased by the use of peer monitoring. To our knowledge, ours is the first controlled experiment showing that peer monitoring can offset the loss in efficiency resulting from the use of low-powered incentives schemes. Peer monitoring was particularly effective because subjects spent a limited amount of time watching others while sharing the monitoring burden so that all subjects were being watched at least once during the experiment. It is as if people possessed natural skills for peer monitoring and understood both its positive effect on productivity as well as the negative consequences of its intensive use.

Peer monitoring is traditionally seen as a decisive advantage of organizations where its effectiveness usually relies on face to face and repeated interactions among parties that are inherent to the organizational environment (Bandiera *et al.*, 2005; Falk and Ichino, 2006; Mas and Moretti, 2009). Interestingly, the implementation of virtual monitoring devices of the type used in our study may mitigate the comparative advantage of traditional organizations vis-à-vis virtual organizations and other decentralized organizational structures.

Appendix

		Period 1	Period 2	Period 3	Period 4	Period 5	Total
Treatment I vs. Treatment T	Prod.	*+48.2%*	*+40.3%*	*+50.1%*	*+51.6%*	*+52.0%*	*+48.8%*
		0.025	**0.026**	**0.025**	**0.006**	**0.004**	**0.002**
		(0.029)	**(0.028)**	**(0.043)**	**(0.025)**	**(0.020)**	**(0.009)**
	Internet usage[17]	*−70.3%*	*−70.9%*	*−50.7%*	*−51.7%*	*−58.1%*	*−58.2%*
		0.003	**0.002**	**0.024**	**0.006**	**0.001**	**<0.001**
		(0.021)	**(0.019)**	**(0.050)**	**(0.013)**	**(0.002)**	**(0.008)**
Treatment T vs. Treatment TP	Prod.	*−36.1%*	*−27.3%*	*−29.8%*	*−34.5%*	*−29.1%*	*−31.3%*
		0.012	**0.061**	**0.060**	**0.039**	**0.035**	**0.019**
		(0.028)	**(0.124)**	**(0.094)**	**(0.095)**	**(0.084)**	**(0.049)**

Table A1: *Percentage differences* and p-values for clustered t-tests (rank-sum tests) assessing differences in production, Internet usage and working time across treatments.

[17]Working time p-values are identical to Internet usage, as there are no other activities available in these treatments.

		Period 1	Period 2	Period 3	Period 4	Period 5	Total
	Internet	*+122.5%*	*+161.5%*	*+92.0%*	*+109.9%*	*+123.4%*	*+117.2%*
	usage	**0.034**	**0.011**	**0.044**	**0.008**	**0.001**	**0.003**
		(0.128)	**(0.032)**	**(0.083)**	**(0.021)**	**(0.007)**	**(0.010)**
	Work	*−3.5%*	*−13.9%*	*−14.2%*	*−18.6%*	*−18.1%*	*−13.4%*
	time	0.383	**0.049**	0.144	**0.045**	**0.025**	**0.035**
		(0.280)	(0.332)	(0.560)	(0.121)	**(0.047)**	**(0.094)**
Treatment *I* vs.	Prod.	*−5.3%*	*+1.9%*	*+5.4%*	*−0.8%*	*+7.8%*	*+2.1%*
Treatment *TP*		0.755	0.900	0.757	0.959	0.581	0.866
		(0.778)	(0.639)	(0.819)	(0.932)	(0.537)	(0.712)
	Internet	*−33.9%*	*−23.9%*	*−5.4%*	*+1.4%*	*−6.5%*	*−9.1%*
	usage	0.283	0.438	0.874	0.963	0.842	0.728
		(0.187)	(0.785)	(0.490)	(0.747)	(0.265)	(0.754)
	Work	*+6.9%*	*+6.4%*	*+6.7%*	*+5.2%*	*+8.5%*	*+6.7%*
	time	**0.003**	**0.063**	0.341	0.456	0.192	0.114
		(0.003)	**(0.058)**	**(0.058)**	**(0.097)**	**(0.011)**	**(0.058)**
Treatment *I* vs.	Prod.	*+23.6%*	*+47.9%*	*+43.2%*	*+37.6%*	*+43.9%*	*+39.6%*
Treatment *TPN*		0.198	**0.007**	**0.052**	**0.037**	**0.012**	**0.009**
		(0.185)	**(0.020)**	**(0.075)**	**(0.065)**	**(0.028)**	**(0.024)**
	Internet	*−31.8%*	*−47.0%*	*−41.7%*	*−34.9%*	*−40.2%*	*−39.6%*
	usage	0.405	**0.046**	**0.058**	0.121	**0.072**	**0.042**
		(0.230)	(0.104)	**(0.069)**	**(0.081)**	**(0.021)**	**(0.047)**
	Work	*+7.4%*	*+16.0%*	*+23.9%*	*+21.5%*	*+22.3%*	*+17.6%*
	time	**0.007**	**0.001**	**0.005**	**0.020**	**0.009**	**0.001**
		(0.017)	**(0.003)**	**(0.008)**	**(0.007)**	**(0.003)**	**(0.005)**
Treatment *T* vs.	Prod.	*−16.6%*	*+5.5%*	*−4.6%*	*−9.2%*	*−5.3%*	*−6.0%*
Treatment *TPN*		0.325	0.779	0.816	0.652	0.744	0.687
		0.325	0.779	0.816	0.652	0.744	0.687
	Internet	*+129.8%*	*+82.0%*	*+18.3%*	*+34.7%*	*+42.9%*	*+44.3%*
	usage	**0.040**	**0.073**	0.533	0.233	0.102	0.110
		(0.129)	(0.188)	(0.390)	(0.168)	**(0.022)**	(0.114)
	Work	*−3.0%*	*−6.2%*	*−0.3%*	*−5.9%*	*−7.7%*	*−4.6%*
	time	0.479	0.436	0.976	0.616	0.418	0.531
		(0.522)	(0.631)	(0.468)	(0.916)	(0.813)	(0.898)
Treatment *TP* vs.	Prod.	*+30.6%*	*+45.2%*	*+35.9%*	*+38.7%*	*+33.5%*	*+36.7%*
Treatment *TPN*		**0.096**	**0.043**	0.111	0.108	**0.069**	**0.054**
		(0.176)	**(0.087)**	(0.122)	(0.156)	**(0.099)**	**(0.099)**
	Internet	*+3.3%*	*−30.4%*	*−38.4%*	*−35.8%*	*−36.0%*	*−33.6%*
	usage	0.941	0.265	0.100	0.127	0.106	**0.097**
		(0.895)	(0.226)	(0.160)	(0.146)	(0.148)	**(0.067)**
	Work	*+0.5%*	*+9.0%*	*+16.1%*	*+15.5%*	*+12.7%*	*+10.1%*
	time	0.883	**0.076**	**0.072**	0.116	0.128	**0.070**
		(0.839)	**(0.085)**	(0.152)	**(0.092)**	(0.105)	**(0.060)**

Table A1: Continued

	Watching time	Length of watching episodes	Proportion of watching episodes for which only one (all) subject(s) is monitored	Amount of time being watched	Number of times a subject is watched
Regression type	Tobit with random effects	Tobit with random effects	Probit with random effects	Tobit with random effects	Poisson with random effects
P-value associated with the TPN treatment dummy	0.165	0.573	0.041 (0.075)	<0.001	<0.001

Table A2: *P*-values associated with the TPN treatment dummy capturing differences across monitoring treatments.

Note: All regressions are completed at the minute level and all include a trend. These results are robust to the cases of the 5-minute analysis as well as to the case of the analysis per period.

	Coefficients
Intercept	-0.803^{***}
Being watched in *t-1*	0.001
Being watched in *t-2*	0.002^{***}
Watching in *t-1*	-0.001
Watching in *t-2*	0.001
Trend	0.066^{***}
Number of observations	$n = 1080$
and Log likelihood	398 left-censored
	$Loglikelihood = -1324.595$, $\text{Prob} > \chi^2 = 0$

Table A3: Tobit regression with random effects for individual production in a 5-minute time span[18]

Note: $^{*}p$-value < 0.10, $^{**}p$-value < 0.05, and $^{***}p$-value < 0.01.

In Table A3, we use a 5-minute time frame to assess the impact of watching activities on real-time production. The independent variables related to watching activities are referred to as *Watching* and *Being watched*. These variables measure the amount of time (in seconds) that a subject spent watching others (*Watching*) and the amount of time (in seconds) a subject was watched (*Being watched*) by at least one subject in a given time span of five minutes. We introduce independent variables with lags so as to mitigate possible endo-

[18] An independent variable accounting for the number of watchers is not statistically significant when introduced in the specification of the regression. This may be due to the fact that the information on the number of watchers was not made particularly salient. In case a subject was watched by more than one person, the following indication was printed on the screen: "more than one subject is watching you."

geneity issues.[19] We include a trend as independent variable so as to control for the steady increase of production across periods.[20]

References

Abbink, K., J. Brandts, B. Herrmann, and H. Orzen. 2010. "Intergroup Conflict and Intra-group Punishment in an Experimental Contest Game". *American Economic Review*. 100(1): 420–447.

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

Andreoni, J. and B. D. Bernheim. 2009. "Social Image and the 50–50 Norm: A Theoretical and Experimental Analysis of Audience Effects". *Econometrica*. 77(5): 1607–1636.

Babcock, P., K. Bedard, G. Charness, J. Hartman, and R. H. 2015. "Letting Down the Team? Evidence of Social Effects of Team Incentives". *Journal of the European Economic Association*. 13(5): 841–870.

Bandiera, O., I. Barankay, and I. Rasul. 2005. "Social Preferences and the Response to Incentives: Evidence from Personnel Data". *Quarterly Journal of Economics*. 120(3): 917–962.

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

Barzel, Y. 1982. "Measurement Costs and the Organization of Markets". *Journal of Law and Economics*. 25(1): 27–48.

Bucciol, A., D. Houser, and M. Piovesan. 2013. "Temptation at Work". *PLoS ONE*. 8(1): e53713. DOI: 10.1371/journal.pone.0053713.

Burks, S. V., J. P. Carpenter, L. Goette, and A. Rustichini. 2013. "Overconfidence and social signalling". *The Review of Economic Studies*. 80: 949–983.

Carpenter, J. P. 2007a. "Punishing Free Riders: How Group Size Affects Mutual Monitoring and the Provision of Public Goods". *Games and Economic Behavior*. 60(1): 31–51.

[19] Endogeneity issues may arise if we introduce the current amount of time subjects spent watching others as well as the current amount of time they were being watched by others as independent variables. Indeed, one may expect that individual production could cause changes in watching behaviors. For example, subjects with low levels of production may feel ashamed (Kandel and Lazear, 1992) and decide to avoid consulting the performance of others.

[20] Similar results are obtained when controlling for beginning or end of period effects. For example, the nature of our results is unchanged when introducing in our regression analysis a dummy variable that takes value one if the five minute time span corresponds to the first (last) five minutes of the period.

Carpenter, J. P. 2007b. "The demand for punishment". *Journal of Economic Behavior and Organization*. 62: 522–542.

Carpenter, J. P., S. Bowles, H. Gintis, and S. H. Hwang. 2009. "Strong Reciprocity and Team Production: Theory and Evidence". *Journal of Economic Behavior and Organization*. 71(2): 221–232.

Chandler, A. D. 1992. "Organizational Capabilities and the Economic History of the Industrial Enterprise". *Journal of Economic Perspectives*. 6(3): 79–100.

Charness, G., D. Masclet, and M. C. Villeval. 2010. "Competitive Preferences and Status as an Incentive: Experimental Evidence". *IZA Discussion Paper 5034, Bonn, and GATE WP 10–16*. halshs-00497974.

Charness, N. and J. I. D. Campbell. 1988. "Acquiring Skill at Mental Calculation in Adulthood: A Task Decomposition". *Journal of Experimental Psychology: General*. 117: 115–129.

Corgnet, B., R. Hernán-Gonzalez, and S. Rassenti. 2011. "Real effort, real leisure and real-time supervision: Incentives and peer pressure in virtual organizations". Working paper, Chapman University, Economic Science Institute 11–05.

Corgnet, B., R. Hernán-Gonzalez, and E. Schniter. 2015. "Why Real Leisure Really Matters: Incentive Effects on Real Effort in the Laboratory". *Experimental Economics*. 18: 284–301.

Datta, S. and G. Satten. 2005. "Rank-sum Tests for Clustered Data". *Journal of the American Statistical Association*. 100: 908–915.

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

Donner, A., N. Birkett, and C. Buck. 1981. "Randomization by Cluster: Sample Size Requirements and Analysis". *American Journal of Epidemiology*. 114: 906–914.

Dumaine, B. 1990. "Who needs a boss?" *Fortune*. 121: 52–60.

Dumaine, B. 1994. "The Trouble with Teams". *Fortune*. 130(5): 86–92.

Encinosa, W., M. Gaynor, and J. Rebitzer. 2007. "The Sociology of Groups and the Economics of Incentives: Theory and Evidence on Compensation Systems". *Journal of Economic Behavior and Organization*. 62: 187–214.

Eriksson, T., A. Poulsen, and M. C. Villeval. 2009. "Feedback and Incentives: Experimental Evidence". *Labour Economics*. 16(6): 679–688.

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 U. Sunde. 2006. "Self-Confidence and Search". *IZA Discussion Paper*.

Falk, A. and A. Ichino. 2006. "Clean Evidence on Peer Effects". *Journal of Labor Economics*. 24: 39–58.

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

Fehr, E. and S. Gächter. 2000. "Cooperation and Punishment in Public Goods Experiments". *American Economic Review*. 90(4): 980–994.

Fehr, E., A. Klein, and K. M. Schmidt. 2007a. "Fairness and Contract Design". *Econometrica*. 75(1): 121–154.

Fehr, E., A. Klein, and K. M. Schmidt. 2007b. "Adding a Stick to the Carrot? The Interaction of Bonuses and Fines". *American Economic Review papers and proceedings*. 97(2): 177–181.

Frey, B. 1993. "Does Monitoring Increase Work Effort? The Rivalry with Trust and Loyalty". *Economic Inquiry*. 31: 663–670.

Gill, D. and R. Stone. 2015. "Desert and Inequity Aversion in Teams". *Journal of Public Economics*. 123: 42–54.

Grosse, S., L. Putterman, and B. Rockenbach. 2011. "Monitoring in Teams: A Model and Experiment on the Central Monitor Hypothesis". *Journal of the European Economic Association*. 9(4): 785–816.

Hamilton, B., J. Nickerson, and H. Owan. 2003. "Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation". *Journal of Political Economy*. 111(3): 465–497.

Hansen, D. G. 1997. "Worker Performance and Group Incentives: A Case Study". *Industrial and Labor Relations Review*. 51(1): 37–49.

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

Holmström, B. 1982. "Moral Hazard in Teams". *Bell Journal of Economics*. 13: 324–340.

Ichniowski, C., T. A. Kochan, D. Levine, C. Olson, and G. Strauss. 1996. "What Works at Work: Overview and Assessment". *Industrial Relations: A Journal of Economy and Society*. 35(3): 299–333.

Ichniowski, C., K. Shaw, and G. Prennushi. 1997. "The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines". *American Economic Review*. 87(3): 291–313.

Jensen, M. C. and W. H. Meckling. 1976. "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure". *Journal of Financial Economics*. 3(4): 305–360.

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

Kruse, D. L. 1992. "Profit Sharing and Productivity: Microeconomic Evidence from the United States". *Economic Journal*. 102(410): 24–36.

Laffont, J. J. and D. Martimort. 2002. *The Theory of Incentives: The Principal-Agent Model*. Princeton, N.J.: Oxford: Princeton University Press.

Ledyard, J. 1995. *Public Goods: A Survey of Experimental Research, in Handbook of experimental economics*. Ed. by J. H. Kagel and A. E. Roth. Princeton: Princeton University Press. 111–94.

Malachowski, D. 2005. "Wasted Time at Work Costing Companies Billions". *San Francisco Chronicle*. July. URL: http://www.sfgate.com/cgi-bin/article.cgi?f=/g/a/2005/07/11/wastingtime.TMP.

Mas, A. and E. Moretti. 2009. "Peers at Work". *American Economic Review*. 99(1): 112–145.

Masclet, D., C. Noussair, S. Tucker, and M. C. Villeval. 2003. "Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism". *American Economic Review*. 93(1): 366–380.

Mohnen, A., K. Pokorny, and D. Sliwka. 2008. "Transparency, Inequity Aversion, and the Dynamics of Peer Pressure in Teams: Theory and Evidence". *Journal of Labor Economics*. 26: 693–720.

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

Niederle, M. and L. Vesterlund. 2007. "Do Women Shy Away from Competition? Do Men Compete Too Much?" *The Quarterly Journal of Economics*. 3(8): 1067–1101.

Nikiforakis, N. 2010. "Feedback, Punishment and Cooperation in Public Good Experiments". *Games and Economic Behavior*. 68: 689–702.

Sacerdote, B. 2001. "Peer Effects with Random Assignment: Results for Dartmouth Roommates". *Quarterly Journal of Economics*. 116(2): 681–704.

Sefton, M., R. Shupp, and J. Walker. 2007. "The Effect of Rewards and Sanctions in Provision of Public Goods". *Economic Inquiry*. 45: 671–690.

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

Young, K. S. 2006. "The Internet and Workplace Transformation". *Advances in Management Information Systems*. 7: 193–204.

Zimmerman, D. J. 2003. "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment". *Review of Economics and Statistics*. 85(1): 9–23.