Harnessing the Power of Social Incentives to Curb Shirking in Teams

Brice Corgnet, Brian Gunia and Roberto Hernán González[1]

In-Press at Journal of Economics & Management Strategy

[1] Brice Corgnet: Univ Lyon, emlyon business school, GATE UMR 5824, F-69130 Ecully, France. Email: corgnet@em-lyon.com. Brian C. Gunia: Carey Business School, Johns Hopkins University, 100 International Drive Baltimore, MD 21202, USA +1 (410) 234-9423, brian.gunia@jhu.edu. Roberto, Hernán-González, CEREN EA 7477, Burgundy School of Business, Université Bourgogne Franche-Comté, Dijon, France. Email: Roberto.hernan-gonzalez@bsb-education.com. Brice Corgnet acknowledges that this research was performed within the framework of the LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon, within the program Investissements d'Avenir (ANR-11-IDEX-007) operated by the French National Research Agency. The authors acknowledge the continuous support of the Economic Science Institute at Chapman University. Roberto Hernán-González also acknowledges financial support from Burgundy School of Business, the Spanish Ministerio de Economía y Competitividad [2016/00122/001], Ministerio de Ciencia, Innovación y Universidades [PGC2018-093506-B-I00], and Proyectos de Excelencia de la Junta Andalucía. This research started when the first author was at Chapman Unviersity and the third author was visiting Chapman University.

Harnessing the Power of Social Incentives to Curb Shirking in Teams

Brice Corgnet, Brian Gunia and Roberto Hernán González

Abstract

We study several solutions to shirking in teams, each of which triggers social incentives by reshaping the workplace social context. Using an experimental design, we manipulate social pressure at work by varying the type of workplace monitoring and the extent to which employees are allowed to engage in social interaction. This design allows us to assess the effectiveness as well as the appeal of each solution. Despite similar effectiveness in boosting productivity, only organizational systems involving social interaction (via chat) were comparably appealing to a baseline treatment. This suggests that solutions involving social interaction are more likely to be effective in the long-run than solutions involving monitoring alone.

Keywords: Social Incentives; Social Pressure; Moral Hazard in Teams; Laboratory Experiments.

JEL codes: C92, D23, D91, M54.

1. Introduction

Shirking in teams is a core topic addressed by economic theories of incentives (Holmström, 1982). In the absence of accurate and verifiable information regarding individual contributions, managers must typically rely on team incentives. However, such compensation contracts may provide insufficient incentives because they do not fully reward individual effort. Because team incentives are used when individual contributions cannot be contracted (see Holmström, 2017)[2], solutions to shirking in teams are thus of a non-contractual nature.

Numerous solutions to shirking issues rely on social incentives (Bandiera, Barankay and Rasul, 2010; Ashraf and Bandiera, 2018), which refer to the effect of the social context on an individual's motivation to complete work (see Tamir and Hugues, 2018; Corgnet, Hernan-Gonzalez, and Mateo, 2019). The social context is especially relevant in the case of teamwork because team members often interact frequently, virtually or in-person (see Miller and Schuster, 1987; Ledford, Lawler and Mohrman, 1995; Hamilton, Nickerson and Owan, 2003; Lazear and Shaw, 2007; Nyberg et al., 2018).

The current paper aims to compare the effectiveness and acceptability of various social incentive schemes intended to curb shirking and foster team performance. In other words, we not only study the impact of these schemes on work effort but also measure workers' willingness to embrace them, thus integrating disparate streams of research on team shirking and employee satisfaction. A practical goal is to help practitioners identify obstacles in the implementation of the various systems. We suggest that effective shirking solutions are those that promote work effort and appeal to workers at the same time.

1.1. Free Riding in Teams and Social Incentives

Social incentives typically rely on either peer pressure or social preferences. Peer pressure can be seen as the mechanism by which observing others or being observed by others affects one's own behavior (e.g., Falk and Ichino, 2006; Mas and Moretti, 2009; Guryan, Kroft, and Notowidigdo, 2009; Corgnet, Hernan-Gonzalez, and Rassenti, 2015a; see Herbst and Mas, 2015 for a review), whereas social preferences are defined as a person's inclination to care about others' payoffs in addition to their own (e.g., Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002; Fehr and Fischbacher, 2002). Either type of social incentive can influence individual behavior.

[2] Holmström comments on the difficulty of obtaining reliable information about production in his first job at Ahlström (a Finnish company) after graduation: "The integrity of the data therefore seemed questionable for technical as well as strategic reasons." (p. 414).

Specifically, social incentives have been shown to help mitigate shirking in teams. Rotemberg (1994) and Dur and Sol (2010) suggested that the presence of altruistic motives tend to reduce shirking because altruistic workers refrain from behaviors that would hurt others' welfare. This suggests that triggering prosocial motives might represent an effective solution to shirking in teams. Growing evidence suggests that prosocial concerns indeed foster cooperation (e.g., Fehr and Fischbacher, 2002; Carpenter and Seki, 2011; Chaudhuri, 2011). The challenge is to promote such prosocial concerns in organizations in which shirking is pervasive.

One mechanism that often induces prosocial concerns toward team members is the activation of group identity (Akerlof and Kranton, 2000, 2005). A series of experimental papers have shown that inducing group identity triggers prosocial and cooperative behavior (e.g., Goette, Huffman, and Meier, 2006; Charness, Rigotti, and Rustichini, 2007; Charness, Cobo-Reyes, and Jiménez, 2014). These findings corroborate the results of previous research in social psychology showing that inducing a "minimal group identity" (e.g., grouping people according to self-reported preferences on paintings; e.g., Tajfel et al., 1971; Tajfel et al., 1979; Rabbie, Schot, and Visser, 1989; Mummendey et al., 1992; Yamagishi, Jin, and Kiyonary, 1999) could promote group cooperation. Dugar and Shahriar (2012) have also shown that group identity fostered cooperation, whether it was induced using a "minimal group" paradigm or via real, existing group identities. We thus expect organizational policies and practices that increase or activate group identity to alleviate shirking in teams. As examples, such practices might include team-building exercises (e.g., Charness, Cobo-Reyes, and Jiménez, 2014) or enhanced opportunities for communication and social interaction (e.g., Chen and Li, 2009; Gioia, 2017). Because social interactions can foster group identity, they might foster altruism among group members and thus facilitate cooperation (Dur and Sol, 2010).

Continuous social interaction can be seen as a distinctive feature of organizations versus markets, and these social interactions can foster the internalization of work ethics that proscribe shirking[3] (Ramalingam and Rauh, 2014). In particular, we focus on the impact of granting workers' access to a peer chat platform as a mechanism fostering social interaction (Dawes, 1991; Chen and Li, 2009) and thus boosting team production. Communication has been found to sharply increase cooperation in social dilemmas, especially in larger groups (see Sally, 1995; Balliet, 2010 for reviews). Similarly, a vast literature on public goods games (see Ledyard, 1995; Zelmer, 2003 for an overview) has shown that the introduction of communication can increase cooperation (see e.g., Ostrom and Walker, 1991; Palfrey

[3] This argument also relates to the study of norms in Kandel and Lazear (1992).

and Rosenthal, 1991; Ostrom, Walker, and Gardner, 1992; Davis and Holt, 1993; Gardner, Ostrom, and Walker, 1994; Sally, 1995; Bohnet and Frey, 1999; Bochet, Putterman, and Page, 2006; Bochet and Putterman, 2009). Additionally, communication fosters group identity and commitment (Kerr and Kaufman-Gilliland, 1994) as well as the development of social norms (Bicchieri, 2002). Our first hypothesis is stated as follows and is formally derived in Appendix A following the models of Rotemberg (1994) and Dur and Sol (2010).

Hypothesis 1 (Chat). *Teams endowed with the ability to engage in peer chat will exhibit higher production levels and less shirking than teams not endowed with peer chat*.

Another mechanism that can foster prosocial behavior in teams is peer pressure.[4] Models of peer pressure often incorporate feelings of guilt or shame that emerge when an individual exerts less effort than team members (e.g., Kandel and Lazear, 1992; Barron and Gjerde, 1997). These feelings make shirking psychologically costly. Indeed, accumulating evidence shows that being watched by one's coworkers effectively deters shirking in teams by increasing the amount of shame an individual experiences. For example, Mas and Moretti (2009) reported positive peer effects on the number of items scanned by supermarket cashiers. These positive effects emerged when cashiers were observed by highly productive workers, but not comparably productive workers, suggesting that the feelings of shame that emerge when others deem an individual a lower producer are especially relevant in understanding peer effects. Mas and Moretti (2009) refer to mechanisms based on shameful feelings to explain social pressure. They emphasize that the effectiveness of social pressure in reducing shirking hinges upon people's desire to be seen as prosocial, and thus their susceptibility to shame. This mechanism has been modeled by Kandel and Lazear (1992) as well as Bénabou and Tirole (2006) and further validated by the experimental tests in Corgnet, Hernan-Gonzalez, and Rassenti (2015a) and Corgnet, Hernan-Gonzalez, and Mateo (2019).

In addition, experimental papers have reported a positive relationship between being watched and prosocial behavior (e.g., Hoffman, McCabe, and Smith, 1996; Burnham and Hare, 2007; Andreoni and Bernheim, 2009). Since this effect is stronger when more people are watching (Diener, 1980; Reyniers and Bhalla, 2013), teams in which more workers can monitor each other would be expected to

[4] We abstract away from the possibility of monetary punishments toward free riders (e.g., Fehr and Gächter, 2000; Carpenter 2007a, 2007b; Nikiforakis, 2008).

outperform those in which only a few workers can monitor. Our second hypothesis is stated as follows and is formally derived in Appendix A building on the model of Kandel and Lazear (1992).

> **Hypothesis 2 (Monitoring).** *Teams endowed with the ability to engage in peer monitoring will exhibit higher production levels and less shirking than teams not endowed with peer monitoring*.

In our model in Appendix A, we assume that the effects of peer chat and peer monitoring on work effort are additive. Our model posits that organizations that use both mechanisms will outperform those that only use one. This means we do not consider cases, for example, in which the effect of peer monitoring is either magnified or weakened by the presence of peer chat.

1.2. Incentives and Work Satisfaction

Our model implies that both peer monitoring and peer chat should reduce shirking in teams. The availability of multiple solutions to the shirking problem testifies to the richness of the theory, but it also puts the practitioner in the delicate situation of choosing among and combining solutions that appear comparably effective. How might a practitioner make this choice? Our aim is to show that organizational systems that produce similar incentive effects might, however, generate strikingly different levels of work satisfaction. Dissatisfied workers, in turn, might well be less productive or even less likely to remain in the organization, immediately but particularly over the longer-run. These insights may help practitioners choose an appropriate solution.

In contract theory, the distinction between incentive effects and work satisfaction is formalized by the incentive compatibility and participation constraints (see e.g., Laffont and Martimort, 2002; Bolton and Dewatripont, 2005). Incentive compatibility constraints measure the extent to which a compensation contract fosters work effort, whereas participation constraints assess a worker’s satisfaction (measured in utility terms) compared to available alternatives.

Despite these two features of any given work arrangement, practitioners may be tempted to focus on incentive effects and downplay workers’ satisfaction under the assumption that workers will find the costs of leaving their job prohibitive in the short-term. Even when workers cannot credibly leave the company immediately, however, it is crucial for managers to take into account workers’ well-being and satisfaction on the job (Danna & Griffin, 1999). One reason is that disgruntled workers will likely resist organizational changes like the implementation of a new shirking solution, perhaps by engaging in counterproductive organizational behaviors (Niehoff and Moorman, 1993). Additionally, dissatisfied workers will likely exhibit low levels of motivation, thus reducing their inclination to exert effort without explicit incentives (see Frey, 1997; Fehr and Falk, 2002; Gneezy, Meier, and Rey-Biel, 2011). As the managers interviewed in Bewley's study (1995, p. 252) made clear: "Workers have so many opportunities to take advantage of employers that it is not wise to depend on coercion and financial incentives alone as motivators." Because employment contracts are inherently incomplete, it is impossible to provide explicit incentives for all dimensions of a job (Holmström and Milgrom, 1991, 1994; Itoh, 1991; Maskin and Tirole, 1999). Ultimately, managers have to rely on employees' intrinsic motivation (Deci, 1971; Frey, 1997; Deci and Ryan, 2000; Fehr and Falk, 2002), or their inherent enjoyment of and satisfaction with the job and task at hand (Kinicki et al., 2002; Ryan, 1982; Spector, 1985) to sustain their performance.

Self-determination theory, which has been formalized in Economics by Bénabou and Tirole (2002, 2003), suggests that the three main drivers of intrinsic motivation and closely allied construct of work satisfaction are *competence*, *autonomy* and *relatedness* (Deci and Ryan, 1985, 2000; Ryan and Deci, 2000; Deci et al., 2001; Gagne and Deci, 2005). In other words, workers will report high levels of satisfaction on the job when they feel good at what they are doing (competence), feel they are doing it out of their own volition (autonomy), and feel socially connected to others (relatedness).

In line with self-determination theory, workplace surveillance mechanisms tend to reduce the perceived autonomy of workers, thus lowering job satisfaction (e.g., Frey, 1997; Ambrose and Alder, 2000; Stanton, 2000a,b; Ariss, 2002; Alder, Noel and Ambrose, 2006; Falk and Kosfeld, 2006). Commentators report that employees "feel degraded, stressed, and dehumanized" by a surveillance system (Ariss, 2002: 555), which "has a detrimental effect on employee morale, increases worker stress, and engenders negative job attitudes" (Alder, Noel, and Ambrose, 2006, p. 895). Despite its strong incentive effects (see Mas and Moretti, 2009; Corgnet, Hernan-Gonzalez, and Rassenti, 2015a; Herbst and Mas, 2015), then, monitoring might reduce workers' autonomy and temper their satisfaction at work.

By contrast, peer chat is unlikely to threaten autonomy, as workers will be free to initiate or stop any conversation. In addition, peer chat should increase perceptions of relatedness to other workers, thus fostering work satisfaction. This leads us to the following hypothesis.

Hypothesis 3 (Work satisfaction).

i) The effects of an organizational shirking solution on workers' job satisfaction will diverge from the solution's incentive effects.

ii) Organizational shirking solutions involving chat will lead to greater worker satisfaction than solutions without chat.

iii) The net effect of monitoring will be mixed because this solution will increase workers' incentives to exert effort, thus boosting productivity, but it will also make the task less enjoyable, and thus decrease work satisfaction.

As we show in Appendix A, the effect of peer chat and peer monitoring on work satisfaction should not affect the magnitude of the incentive effects captured in Hypotheses 1 and 2. This implies that the magnitude of incentive effects and workers' satisfaction are not necessarily aligned.

Work satisfaction should, however, foster work motivation in the longer-run (Westover et al., 2010) and affect work behavior on job dimensions that are not contractually incentivized. As suggested above, for example, a large literature has shown that satisfied workers are less likely to engage in counterproductive work behaviors that are inappropriate and harmful to the firm and their coworkers (Dalal, 2005). A satisfied workforce is also more likely to engage in organizational citizenship behavior (Niehoff and Moorman, 1993), thus going beyond the contractual definition of their job to help their coworkers and add value to the company. In addition, satisfied employees will be less likely to leave the company, thus minimizing turnover costs. Because workers who want to leave as a result of a change in the organizational setup might not be able to do so immediately, the negative impact of some of these organizational changes might only be seen in the longer-run. This is why managers might be tempted to favor organizational changes that produce strong incentive effects in the short run at the risk of generating long-term costs. Our results suggest they might wish to reconsider carefully the pros and cons of this approach.

1.3. Experimental Tests and Findings

To test our hypotheses, we used a laboratory workplace in which workers undertook a real-effort task but also had access to the Internet for leisure purposes (see Corgnet, Hernan-Gonzalez, and Schniter, 2015). We included six main treatments in a 2×3 factorial between-subject design, which served to manipulate social interactions and monitoring among workers. Social interaction was manipulated at one of two levels including treatments in which workers did or did not have access to a chat platform to communicate with other team members. The monitoring dimension was manipulated at three levels: All workers could monitor each other's activities, only one of the workers could monitor everyone else, or no workers could monitor.

In our baseline treatment in which neither chat nor monitoring was present, we observed substantial shirking. Workers spent about 30% of their time on the internet instead of working on the task. In line with Hypotheses 1 and 2, shirking was substantially reduced by any of the treatments in which we introduced peer chat, monitoring, or both. In any of these treatments, workers spent about 10% more time on the work task and produced about 40% more than in the baseline.

To test Hypothesis 3, we designed an additional experiment that assessed participants' willingness to work in a given organizational system. Unlike our first study, in which the organizational system was set exogenously by the experimenter, participants in this study could state their preferences for each of the six systems previously studied. The system that received the highest average rating across team members was then implemented. Alternatively, we could have used a survey to elicit participants' work satisfaction (e.g., Spector, 1985; Deci and Ryan, 2000) in each of the six systems. However, we wanted to employ a research design in which participants would have an incentive to truthfully reveal their preference for each organizational system. Our design encouraged truth-telling because workers who do not reveal their true preferences could end up working in an organizational system they dislike.

In line with Hypothesis 3, we found that organizations involving peer chat but no monitoring tended to be more popular than those involving monitoring but no chat. In addition, organizational systems involving only monitoring rated significantly lower than the baseline whereas those involving only peer chat rated directionally (but not significantly) higher than the baseline. Unexpectedly, organizational systems involving monitoring and peer chat together were as popular as those systems involving peer chat without monitoring. This implies that workers' negative reaction toward peer monitoring was fully offset by the presence of peer chat. This interaction effect was not part of Hypothesis 3 or our model (see Appendix A), but it suggests the interesting possibility that the negative effect of monitoring systems in terms of work satisfaction might be alleviated by fostering social interaction between workers. In other words, workers might be less reluctant to be monitored by others if they can communicate with them—possibly because they can then voice their concerns regarding what could be perceived as abusive monitoring.

2. Design

The current research included two studies that both used an interactive, virtual environment to test our hypotheses in a tightly-controlled fashion. Interdependent individuals performed an analytical task that also allowed them to check the internet, replicating many features of a real-world work environment.

Study 1: Solutions to Shirking in Teams (Hypotheses 1 & 2)

Design. To investigate the first two hypotheses, we used a 2×3 between-subject factorial design in which the chat dimension was either present or absent and monitoring was absent, given to one team member, or given to all team members (see Table 1). For each of the six treatments, 60 participants were recruited on the basis of a power calculation; at the beginning of the experiment, participants were randomly placed into 10-person groups and remained with their group for the whole experiment.

TABLE 1—2×3 FACTORIAL DESIGN

Organizational systems		Chat availability	
		Absent	Present
Monitoring availability	Absent	No Chat-No Monitor (Baseline)	Chat-No Monitor
	One worker	No Chat-One Monitor	Chat-One Monitor
	All workers	No Chat-All Monitor	Chat-All Monitor

The work task. The instructions indicated that participants could choose among several activities, including the work task. Adapted from previous research using summation tasks (e.g., Eriksson, Poulsen, and Villeval, 2009), the work task was a particularly long and laborious task intended to resemble the monotony that can accompany organizational life and prompt shirking at work. The task required participants to sum up tables of 36 numbers without using a pen, scratch paper, or a calculator (see Figure 1). Each table had six rows and six columns of randomly-generated integers between zero and ten. Before providing the grand total in the bottom-right cell, participants had to provide a separate subtotal for each of the 12 rows and columns. Calculating these subtotals did not directly generate earnings but could help participants compute the grand total, which generated a 40-cent profit if correct and a 20-cent penalty if incorrect. After completing a table, participants learned whether their answers were correct and how much money they earned. At the end of each period, participants learned the total amount of money generated by all ten participants' efforts on the work task.

	Column1	Column2	Column3	Column4	Column5	Column6	Sum Row:
	3.00	6.00	3.00	0.00	6.00	0.00	
	10.00	5.00	1.00	5.00	2.00	3.00	
	8.00	3.00	5.00	4.00	8.00	7.00	
	1.00	6.00	0.00	9.00	8.00	0.00	
	3.00	7.00	0.00	8.00	10.00	4.00	
	3.00	10.00	10.00	6.00	10.00	0.00	
Sum Column:							

Figure 1. Work Task

At any point during the experiment, all participants could switch from the work task to internet browsing. Depending on their experimental treatment, they might also have the ability to monitor or communicate with their peers. Participants could spend as much or as little time as they wanted on the various activities, each of which was undertaken on a separate screen. To switch activities, participants simply chose the corresponding option from a drop-down menu at the bottom-right of their screens.

Internet. If participants chose the internet, the work task window was replaced by an internet window (embedded in the software; see Figure 2). Within the bounds of university policy, participants could use the internet however they liked, including email. Their confidentiality was assured and maintained, but the software tracked the exact amount of time spent on each activity. Although participants could not complete the work task while browsing the internet, switching was quick and easy.

Figure 2. Embedded internet Screen

In the 'No Monitor' treatments, participants cannot monitor or be monitored. In the 'All Monitor' treatments, all participants could choose to watch the activities of their peers. In the 'One Monitor' treatments, only one of the ten participants was given the ability to watch everyone else, and everyone else was aware of this ability. If participants selected the monitoring option from the drop-down menu, they were directed to a separate window where they could choose whom to monitor (anywhere from one to all other participants) and to actually perform the monitoring. For each selected participant, a column in a table listed their activities (e.g., switched to the internet, provided a subtotal), their current earnings, and their percentage contribution to the team total. As the current research is concerned with reactions to monitoring, we did not focus on the choice to monitor but rather the experience of being monitored. Participants who were being monitored saw a box indicating that "[Experiment ID of the monitor] is watching you" (see Figure 3).

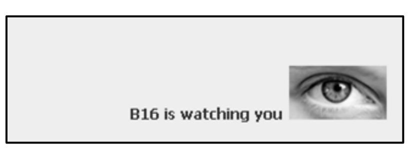

Figure 3: Being Monitored

Peer chat. In the ‘Chat’ treatments, participants could choose to exchange instant messages with their teammates. This virtual form of communication was chosen to maintain anonymity, and because it represents a simple form of communication, bereft of potential social confounds (Gunia et al., 2012). Participants who chose to communicate by selecting that option from the drop-down menu entered a chat room in which they could send a message to one or more people. Participants with whom others wanted to communicate saw a pop-up window displaying the sender’s experiment ID and message content (see Figure 4).

Figure 4. Peer Chat

Variables.

Production. This variable is calculated as the value of the number of sums solved correctly (40 cents each) minus the number of mistakes (20 cents each) in a period.

Usage of time. We calculated the total amount of time workers spent on each activity: work task, internet, chat or monitoring. We report these values either in seconds or as a percentage of the total time.

Task ability. We created a dummy variable, ‘First sum is correct’, as a measure of ability (see Corgnet et al., 2015 for further discussions on the use of this ability measure). This variable takes a value of one if the first table completed by a subject was correct, and value zero otherwise.

Procedures. The experiment was conducted using the proprietary Virtual Organizations software developed by CYDeveloper LLC for the authors. The software facilitated a multi-party team task, controlled centrally by an experimenter. Upon arrival at the lab, participants were directed to private computer terminals and prompted to read a detailed set of computerized instructions.[5] Participants had exactly 20 minutes to read the instructions, with a timer displayed on a large screen at the front of the lab. The instructions indicated that they were one of ten members of a virtual team, which undertook a

[5] The full set of instruction is available here: **https://tinyurl.com/utryu2v**.

1-hour and 40-minute task, broken up into 20-minute periods. Three minutes before the end of the instruction period, the experimenter announced the time remaining and hands out a printed summary of the instructions. At the end of the instruction round, the experimenter closed the instructions and launched the experiment from the server. Subjects were paid according to team incentives; that is, they received one tenth of the group production-based pay in each period. Participants' total earnings were calculated as the sum of their earnings in each of the five periods.

Participants were 360 undergraduates (48.95% male; average age 20.12) enrolled in a subject pool at a Western U.S. university. Specifically, we conducted six sessions of ten participants for each of the six treatments. Based on previous findings using the same real-effort task (see Corgnet, Hernan-Gonzalez, and Rassenti, 2015a,b; Corgnet, Hernan-Gonzalez, and Schniter, 2015), we calculated that recruiting 60 participants for each treatment ensured 80% power to detect a 20% increase in workers' production with respect to the baseline.

Participants responded to an email offering $7 plus an unspecified amount of bonus money for participation in an experiment lasting 2.5 hours. On average, participants earned a total of $26.55, and the experiment lasted for 2 hours.

Study 2: Workers' Satisfaction (Hypothesis 3)

Study 2 used the same task as Study 1 to investigate the same six organizational systems. It substantially extended Study 1, however, by focusing on participants' subjective reactions to these systems (following Zweig and Webster, 2002) and by allowing them to actually experience the system that elicited the most favorable reactions. By allowing participants to choose and experience a system three times, participants were able to fine-tune their reactions if necessary.[6]

Design. The design differed from Study 1 because participants rated each of the six organizational systems (e.g., 'Chat-No Monitor') before each period, and the system that received the highest average rating across team members was announced and implemented. Participants were presented with a summary of the six organizational systems, and they answered the following, general question: "How much do you want to work in each of the following organizations?" (1 = not at all, 7 = very much so; see Figure 5). In the one instance in which two systems tied for the highest rating, a system was randomly selected.

[6] The full set of instruction is available here: **https://tinyurl.com/s3xpvjc**.

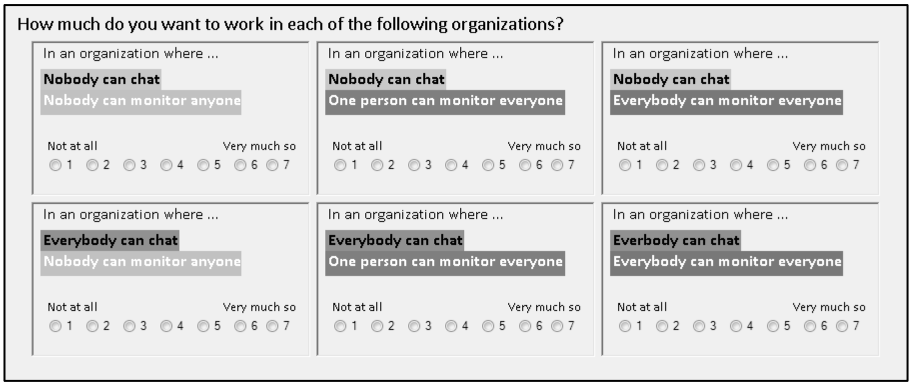

Figure 5. Organization Ratings Screen

Our experimental design thus provides an incentivized elicitation of workers' preferences for the various organizational systems. In our setup, increasing one's own rating for an organizational system increased the chances that this system would be implemented. Workers who did not reveal their true preferences regarding a given system could end up working in a system they disliked in the next period (see Smith, 1982). Our approach thus differs from standard survey techniques used to elicit work satisfaction (e.g., Spector, 1985) or work motivation (Deci and Ryan, 2000).

We do not focus on the production and shirking data from Study 2 because this study primarily aimed to measure work satisfaction. Additionally, incentive effects in this study might be biased by selection effects, in that workers who rated organizational systems differently were likely to differ in terms of individual characteristics such as ability on the task. Since we were not focused on incentive effects, we used three periods instead of six and shortened the length of each period to 10 minutes.

Variables.

Ratings. In this study, workers rated each organizational system on a 1 to 7 Likert-type scale on three occasions.

Procedures. Fifty undergraduate students (48% male; average age 19.71, SD = 1.69) from the same participant pool as Study 1, but who had not participated in Study 1, participated in Study 2. They responded to an email offering $7 plus an unspecified amount of bonus money for participation in an experiment lasting 1 hour. Five separate sessions of ten workers were conducted; on average, participants earned a total of $16.25.

3. Results[7]

3.1. Study 1: Shirking in Teams

We report the descriptive statistics of our main variables in Table B1 in Appendix B including our proxy for workers' ability levels ('First sum is correct'). However, this ability measure should be interpreted with caution because it is endogenous to the treatment. For example, a treatment that induces high levels of production might lead workers to be more focused when completing the first task, thus leading to higher values for the 'First sum is correct' variable. However, we do not find significant differences across treatments regarding our ability proxy at the 5% significance level, except for the differences between the Chat-No Monitor treatment (65% of workers classified as high ability) and the No-Chat-One Monitor treatment (40%) and between the Chat-No Monitor treatment and the baseline (47%) (see Table B2 in Appendix B).[8] Therefore, in the main text, we present our results without controlling for workers' ability levels. In Table B3 in Appendix B, we present additional regression analyses showing that our results continue to hold when controlling for our proxy of workers' ability levels.

In line with Hypotheses 1 and 2, all of the organizational systems involving chat, monitoring, or both achieved a higher level of production than the baseline organizational system in which neither chat nor monitoring was present (see left panel of Figure 6 and Table B1 in Appendix B). On average, a worker involved in any of the organizational systems endowed with chat, monitoring, or both produced 35.27% more ($7.67, SD = $5.18) than the baseline organizational system (Cohen's $d = 0.40$). Another measure of workers' effort is the amount of time they spent online, as browsing the internet does not have any positive effect on workers' productivity for this task (see Corgnet, Hernan-Gonzalez, and Schniter, 2015). Instead, browsing the internet simply distracts the worker, thus reducing his or her productivity—a set of activities often called cyberloafing (Henle and Blanchard, 2008), which occurs when an employee uses the internet during the work period for non-work purposes (Lim, 2002; Wagner et al., 2012).[9]

---

[7] The data that support the findings of this study are available from the corresponding author upon request.
[8] These two comparisons are not significant when applying Bonferroni-Holm corrections for multiple comparisons (Holm, 1979) (see Table B2 in Appendix B).
[9] Thus, using the internet over lunch, using it for work purposes, or using an offline application would not qualify as cyberloafing. What would qualify is any personal activity, conducted during the work period (e.g., Web browsing, email, social media; Kallman, 1993). Obvious in theory, these distinctions can blur in practice, as employees may, for example, encounter irrelevant websites during legitimate searches or open personal emails to retrieve work-related information. Occasionally, they may also "abuse" the internet to cope with stress or to stimulate their creativity (Henle and Blanchard, 2008). While recognizing the inevitable "grey area" between use and abuse, the current research makes a rigid distinction by

The time participants spent online (see right panel of Figure 6 and Table B1 in Appendix B) corresponded to 12.35% (SD = 23.29%) of the total available time in any of the organizational systems involving chat, monitoring, or both, versus 28.52% (SD = 34.80%) in the baseline (Cohen's *d* = 0.81). The comparison of internet usage across organizational systems should, however, take into account the fact that the six systems differed in the number of activities available to workers. It follows that monitoring or chatting activities could potentially be used as substitutes for internet usage, thus mechanically lowering the time spent online in any of the organizational systems endowed with chat, monitoring, or both. To alleviate this concern, we also used the time spent on the work task screen as a measure of workers' effort. We find that, despite having more options available, workers dedicated more time to the work task (82.38%) in the organizational systems involving chat, monitoring, or both, as compared to the baseline (71.48%) (Cohen's *d* = 0.52) (see Figure 6, right panel). We do not observe differences in the time spent chatting across the three treatments involving chat (*p*-values for all three pairwise comparisons are greater than 0.1 using *t*-tests or Rank-Sum tests). In the same vein, we do not see statistical differences in the time spent monitoring between the two 'All Monitor' treatments and between the two 'One Monitor' treatments (*p*-values for the two pairwise comparisons are greater than 0.1 using T-tests or Rank-Sum tests).[10]

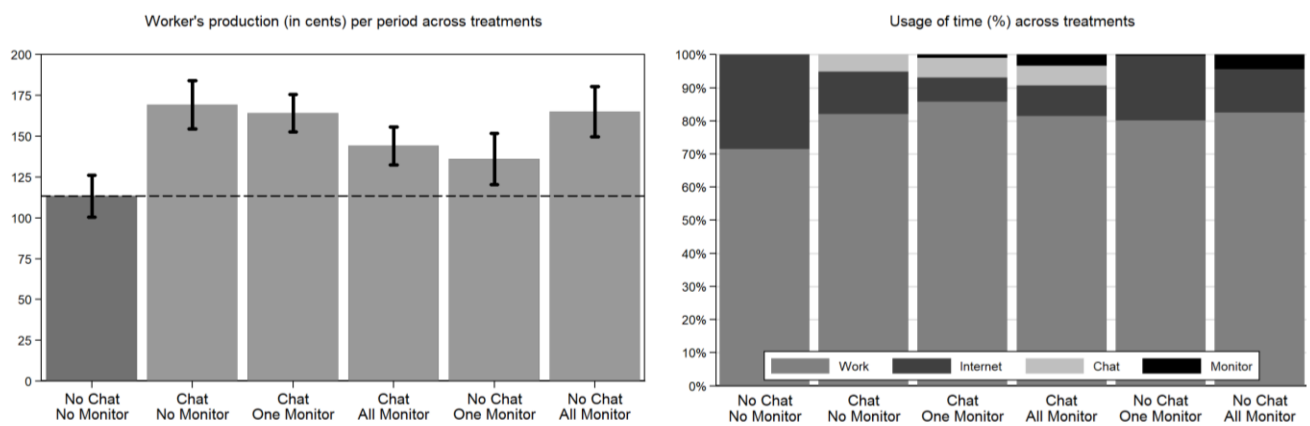

Figure 6. Worker's Production (in cents, including 95% confidence intervals) (left panel) and Usage of Time (%) (right panel) Across Organizational Systems

focusing on clear cases of abuse, which interrupt work (Jett and George, 2003) and are thus counterproductive (Henle and Blanchard, 2008).

[10] This means the one monitor spent as much time monitoring in 'One Monitor' treatments as the average amount of time the multiple monitors spent monitoring in 'All monitor' treatments.

Table 2 shows the statistical significance of differences in production, internet usage, and time on the task between the baseline and the other organizational systems involving chat, monitoring, or both.[11]

TABLE 2—LINEAR PANEL REGRESSION WITH RANDOM EFFECTS FOR WORKERS' PRODUCTION (IN CENTS), INTERNET USAGE AND TIME ON THE TASK (IN SECOND)

Dependent variable	Production (in cents)		Internet Usage[12] (in seconds)		Time on the Task (in seconds)	
	[1]	[2]	[3]	[4]	[5]	[6]
Intercept	59.733***	85.667***	238.892***	139.738***	940.906***	1,060.261***
	(10.655)	(6.586)	(52.020)	(49.645)	(52.715)	(49.645)
Chat-No Monitor Dummy	55.933**	18.433	-188.746***	-87.588	126.095*	-27.615
	(25.606)	(14.222)	(60.877)	(57.876)	(68.821)	(70.026)
Chat-One Monitor Dummy	38.667**	-6.733	-255.424***	-92.100*	171.469***	-25.626
	(18.565)	(21.208)	(52.429)	(54.677)	(51.779)	(55.417)
Chat-All Monitor Dummy	30.800**	-13.400	-232.191***	-64.314	119.660*	-93.787*
	(14.342)	(16.188)	(55.780)	(51.454)	(61.849)	(53.315)
No Chat-One Monitor Dummy	22.733	2.933	-108.941	-48.786	104.183	40.417
	(19.117)	(20.513)	(69.360)	(51.448)	(69.370)	(51.646)
No Chat-All Monitor Dummy	51.600**	42.900***	-184.701***	-82.293	132.349 **	44.243
	(20.456)	(15.167)	(58.092)	(52.290)	(58.730)	(54.044)
Period	17.911***	9.267***	34.447***	67.498***	-27.714***	-67.499***
	(1.713)	(3.520)	(4.999)	(8.030)	(5.752)	(8.030)
Chat-No Monitor Dum.× Period	-	12.500**		-33.719***		51.237***
		(6.174)		(11.782)		(12.475)
Chat-One Monitor Dum.× Period	-	15.133***		-54.441***		65.698***
		(5.098)		(10.612)		(10.549)
Chat-All Monitor Dum.× Period	-	14.733***		-55.959***		71.149***
		(4.523)		(11.533)		(11.571)
No Chat-One Monitor Dum.× Period	-	6.600		-20.051		21.255
		(4.286)		(14.825)		(15.048)
No Chat-All Monitor Dum.× Period	-	2.900		-34.135***		29.369**
		(4.125)		(12.996)		(12.888)
P-values (coefficient comparisons)						
Chat-No Monitor vs Chat-One Monitor [× period]	0.532	[0.675]	0.056	[0.061]	0.335	[0.218]
Chat-No Monitor vs Chat-All Monitor [× period]	0.315	[0.701]	0.280	[0.063]	0.912	[0.116]
Chat-No Monitor vs No Chat-One Monitor [× period]	0.236	[0.295]	0.166	[0.367]	0.739	[0.056]
Chat-No Monitor vs No Chat-All Monitor [× period]	0.881	[0.081]	0.926	[0.975]	0.909	[0.115]
Chat-One Monitor vs Chat-All Monitor [× period]	0.657	[0.932]	0.365	[0.888]	0.151	[0.613]
Chat-One Monitor vs No Chat-One Monitor [× period]	0.464	[0.054]	0.003	[0.016]	0.159	[0.002]
Chat-One Monitor vs No Chat-All Monitor [× period]	0.573	[0.004]	0.020	[0.100]	0.199	[0.003]
Chat-All Monitor vs No Chat-One Monitor [× period]	0.659	[0.030]	0.018	[0.016]	0.791	[0.001]
Chat-All Monitor vs No Chat-All Monitor [× period]	0.291	[0.001]	0.185	[0.097]	0.780	[0.001]
No Chat-One Monitor vs No Chat-All Monitor [×period]	0.217	[0.256]	0.165	[0.382]	0.610	[0.617]
Observations (organizations)	1800	1800	1800	1800	1800	1800
Prob > χ^2	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
R^2	0.063	0.068	0.098	0.105	0.042	0.054

Notes: Estimation output using robust standard errors clustered at the organization level (in parentheses). Similar results are obtained using bootstrapping techniques for standard errors. The p-values reported in columns [1], [3], and [5] correspond to the comparison of the treatment dummies, e.g. 'Chat-No monitor' vs. 'Chat-One Monitor',

[11] Similar results are obtained when controlling for ability on the task, as measured by whether the first table completed by each participant was solved correctly or not (see Table B3 in Appendix B). Similar results are also obtained when controlling for the average production of a group in the previous period (Avg Group Production (t-1)) (see Table B4 in Appendix B).

[12] An alternative definition of internet usage which would facilitate comparisons between treatments with a different subset of available activities is the ratio of time spent on the internet divided by the time spent on the task or on the internet (thus excluding any time dedicated to monitoring or chatting). Doing so, we obtain qualitatively similar results as those reported in Table 2 (regressions [3] and [4]).

in the absence of any interaction term with the period variable. In columns [2], [4], and [6], we report instead the comparison of the interaction effect between treatment dummies and period, e.g. 'Chat-No Monitor Dum.× Period' vs. 'Chat-One Monitor Dum.× Period'.
*** Significant at the 1 percent level, ** Significant at the 5 percent level, * Significant at the 10 percent level.

In Table 2, the coefficients associated with each organizational system dummy in regressions [1] and [5] are positive and significant except for 'No Chat-One Monitor Dummy', which is positive yet not significant, *p*-values = 0.234 and 0.133). The coefficients associated with each organizational system dummy in regression [3] are negative and significant except for 'No Chat-One Monitor Dummy', which is negative yet not significant, *p*-value = 0.116. This is consistent with our model (see Appendix A) and the work of Kandel and Lazear (1992), according to which a lower number of monitors would tend to reduce the amount of peer pressure, thus reducing the corresponding positive effect on workers' effort.

In Table 2 (lower panel), the pairwise comparisons of coefficients in regressions [1] and [5] indicate that the organizational systems involving chat, monitoring, or both do not significantly differ in terms of production and time dedicated to the task. Differences in coefficients across treatments regarding internet usage (regression [3]) might thus be due to the number of different activities available across treatments.

The absence of significant differences in workers' production levels across organizational systems endowed with chat, monitoring, or both are not inconsistent with Hypotheses 1 and 2, which only specify a significant effect with respect to the baseline. However, our model (Appendix A) suggests that the effect of chat and monitoring should be additive such that the 'Chat-All Monitor' ('Chat-One Monitor') treatment should outperform 'No Chat-All Monitor', 'Chat-No Monitor' or 'Chat-One Monitor' ('No Chat-One Monitor' or 'Chat-No Monitor'). As is shown in the lower panel of Table 2, none of these comparisons are statistically significant with the exception of the decrease in internet usage in 'Chat-One Monitor' compared to 'No Chat-One Monitor' (*p*-value = 0.003) and to 'Chat-No Monitor' (*p*-value = 0.056). A possible explanation for this lack of statistical differences could be a *ceiling effect*, by which the level of performance achieved using only chat or monitoring is close to the maximum level of performance of a team. To assess the validity of this claim, we use the data on workers' performance on the same task under individual incentives and in the absence of either chat or monitoring (see Corgnet, Hernan-Gonzalez, and Rassenti, 2015a). In line with the *ceiling effect* argument, when comparing the performance of workers under individual incentives with the five treatments of the current study involving monitoring, chat, or both, we obtain *p*-values that are greater than 0.1 in all cases (*p*-values = 0.12, 0.90, 0.95, 0.42 and 0.12) using panel regressions similar to the ones in Table 1.[13] The absence of significant differences between each of these organizational systems and the case of individual incentives

[13] The comparison of individual incentives with the baseline treatment yields a p-value < 0.001.

suggests that workers' performance is already at a high (possibly maximum) level when chat or monitoring alone is present. An alternative explanation for the lack of differences in production between non-baseline treatments could relate to differences in workers' ability levels. However, Table B3 in Appendix B indicates that our findings are robust when controlling for a proxy of workers' ability.

Even though Table 2 reports the results of panel regression analyses at the period level, similar findings are obtained using standard parametric and non-parametric tests that compare total workers' production, internet usage, and time on the task across treatments (see Table B5 in Appendix B).[14] In Table B5, we show that all treatments led to significantly higher levels of production than the baseline except for 'No Chat-One Monitor'. Although 'Chat-No Monitor' led to the highest level of production (see Figure 6), it only significantly outperformed 'No Chat-One Monitor' which was the non-baseline treatment achieving the lowest level of production (p-value = 0.055 for the Rank-Sum test). Looking at the time spent on the task, all treatments significantly outperformed the baseline with the notable exception of 'Chat-All Monitor' (p-value = 0.443 for the Rank-Sum test). One possible reason is that this was the only treatment in which all workers had access to as many as three alternative activities: chatting, monitoring and browsing the web. In line with this argument, workers spent more time on the work task in the 'Chat-All Monitor' than in the baseline in all periods but the first one, during which they might have tried all possible alternatives. Regarding internet usage, all treatments outperform the baseline and no significant differences are observed across treatments except for a lesser browsing time in 'Chat-One Monitor' than in 'No Chat-One Monitor'.

In regressions [2], [4], and [6] of Table 2, we also assess the dynamics of production, internet usage, and time on the task across treatments (see also Figure 7). For all treatments involving chat, the interaction coefficients between organizational dummies and the number of periods (variable 'period') are positive and significant for production and time on the task (regressions [2] and [6]), while being negative and significant for internet usage (regression [4]).[15] Thus, the positive impact of chat (and chat with monitoring) on workers' effort tends to increase over time. This might be the case because workers need time to get familiar with the chat feature. Alternatively, and as we argue in the hypotheses section, building the necessary team identity to trigger workers' prosocial concerns may require time.

[14] Similar results are also obtained when using a non-parametric test with clusters at the session level (Somers' d, Somers, 1962).

[15] We find similar results for internet usage if we exclude the time participants spent on chat and monitoring activities.

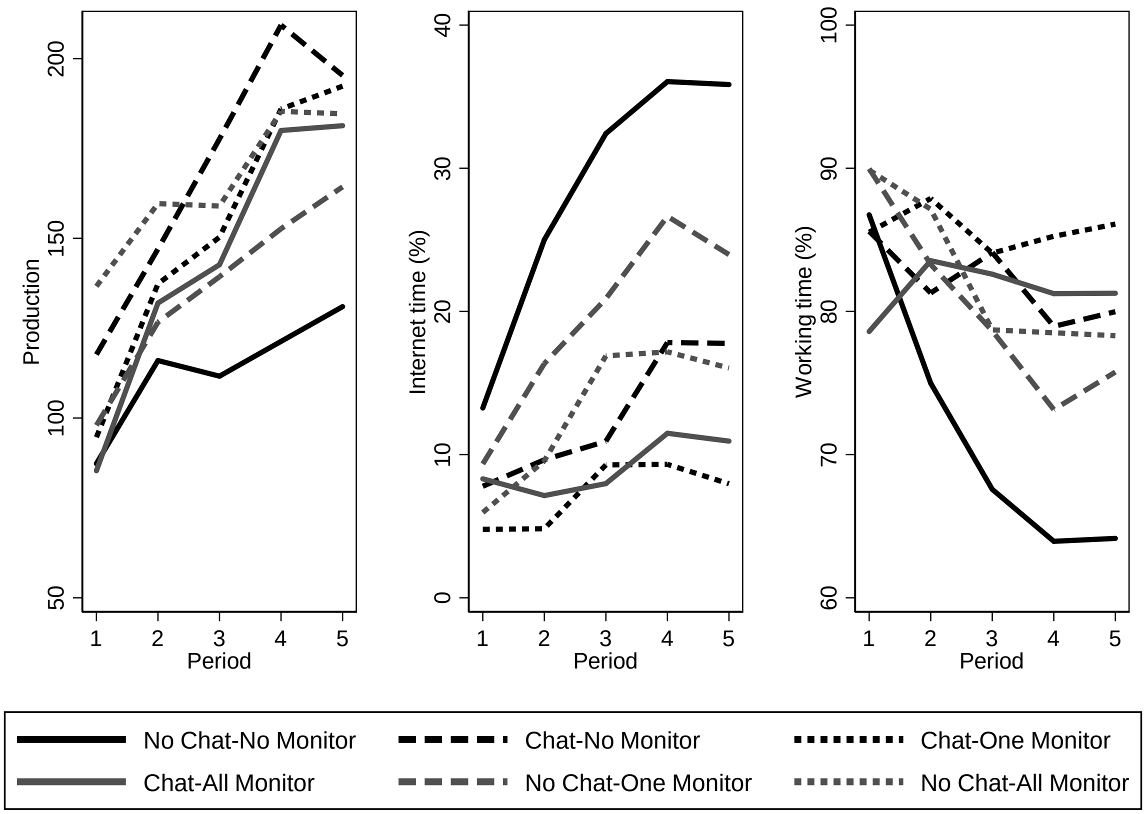

Figure 7. Average workers production (in cents), time spent on internet, and time spent on the work Task (in seconds) across organizational systems and periods

By contrast with chat, the positive effect of monitoring on workers' production (regression [2]) does not increase over time (the coefficients for 'No Chat-One Monitor Dum.×Period' and 'No Chat-All Monitor Dum.×Period' are not significant). In addition, the difference between the coefficients 'Chat-All Monitor×Period' and 'No Chat-All Monitor×Period' is significant (p-value = 0.001), as is the difference between the coefficients 'Chat-One Monitor×Period' and 'No Chat-One Monitor×Period' (p-value = 0.054). This means production is more likely to increase over time when chat is present than when it is absent, given a particular level of monitoring ('One Monitor' or 'All Monitor'). This might occur because the effect of chat relies partly on building team identity or fostering cooperative norms, both of which may require time.

Regarding the dynamics of internet usage or time on the task, we observe an effect that increases over time for the treatment 'No Chat-All Monitor Dummy', which could be due to workers' learning how to use the monitoring features over the course of the experiment.[16] This effect is not significant for the treatment ‘No Chat-One Monitor Dummy’, however.

In Table B6 (see Appendix B), we focus on the content of the conversations in the organizational systems involving chat. Two of the authors independently read and inductively extracted categories, resolving disagreements through discussion. Two graduate student coders who were unaware of the hypotheses then independently assigned each of 354 messages to one of the 13 final categories. The coders agreed on the categorization for 69.50% of the messages, for an acceptable Cohen’s Kappa of 0.65. A sizable proportion of messages (17.50%) were social in nature (e.g., introductions or jokes), thus possibly triggering team identity, as in standard greeting procedures used in the literature (e.g., Chen and Li, 2009; Gioia, 2017). In addition, a large proportion of messages (61.60%) contained content that could be considered normative (e.g., asking or informing about performance, encouraging performance). This suggests that chat might also have induced norms of cooperation (as in Kandel and Lazear, 1992), thus fostering workers’ effort. These norms of cooperation might be especially salient when workers have had the chance to get to know each other via chat. In the end, chat would tend to promote prosocial concerns either by enhancing altruism toward coworkers who share a common team identity (as in Dur and Sol, 2010) or by promoting norms of cooperation across workers who do not necessarily feel altruistic toward each other (Kandel and Lazear, 1992). These two possible mechanisms are likely both present at the same time, and our setup does not seek to isolate them.

In line with Hypotheses 1 and 2, this study shows that both monitoring- and chat-based organizational solutions to shirking in teams are effective compared to a baseline. Although no significant differences are observed across the solutions in terms of workers’ performance and effort levels, Hypothesis 3 suggests that organizational systems involving chat (versus those involving monitoring) will tend to be more appealing to workers. We test these claims in Study 2.

3.2. Study 2: Organizational Systems Ratings

Figure 8 displays average ratings across the ten organizational members for each period. The ordering of organizational systems is the same whether the first or last rating is used, and no statistically significant differences are observed between the two ratings except for ‘Chat-One Monitor’ and ‘Chat-All Monitor’, which became less popular over time (see Table B7 in Appendix B). Even though the

[16] Note that when conducting a linear panel regression (as in Table 2) with the time spent monitoring by workers as a function of the number of periods in the ‘All Monitor’ treatments, we do not observe a significant increase of the amount of time spent watching over time. So, the dynamics of the monitoring activity would not seem to explain this pattern.

popularity of ‘Chat-All Monitor’ went down, it was still selected in three out of the five teams in the last period. All teams tried this organizational system in the first period, but two decided to switch to either the baseline organizational system or ‘Chat-One Monitor’. One explanation for this reduced popularity over time is that chatting requires time to effectively boost workers’ production and thus increase workers’ revenues, as is shown in our dynamic analysis of production in regression [2] of Table 2 in Study 1.[17]

Regardless of the dynamics of ratings, we find that, in line with Hypothesis 3, adding chat to a given organizational system tends to increase its popularity.[18] This effect is statistically significant when we consider the first rating, which was not influenced by workers’ experience with a given organizational system (see the statistical analyses in Table B8 in Appendix B). That is, the treatments ‘Chat-No Monitor’ (‘Chat-One Monitor’) [‘Chat-All Monitor’] led to significantly higher first ratings than ‘No Chat-No Monitor’ (‘No Chat-One Monitor’) [‘No Chat-All Monitor’]. These findings also hold when considering the last rating and average ratings, except that the difference between ‘Chat-No Monitor’ and ‘No Chat-No Monitor’ is not statistically significant in that case (see Tables B9 and B10). This follows from the fact that workers reduced their ratings for the ‘Chat-No Monitor’ system over time.

By contrast, adding monitoring to an organizational system without monitoring does not increase workers’ ratings. Actually, the effect is systematically negative and, in most cases, statistically significant. That is, the treatment ‘No Chat-No Monitor’ (‘Chat-No Monitor’) led to higher ratings than ‘No Chat-One Monitor’ and ‘No Chat-All Monitor’ (‘Chat-One Monitor’ and ‘Chat-All Monitor’). These differences are always significant except for the comparison between ‘Chat-All Monitor’ and ‘Chat-No Monitor’ for the first and average ratings, and the comparison between ‘No Chat-All Monitor’ and ‘No Chat-No Monitor’ for the first ratings. Thus, the negative effect of monitoring on workers’ enjoyment of the task may more than offset the positive effect of monitoring on production levels and workers’ revenues. All in all, workers are less willing to join a team when monitoring is present.

[17] In Study 2, the experiment was substantially shorter than in Study 1 because there were two periods less and each period was half-shorter.

[18] The popularity of chat was not due to the possibility of using the platform to discuss ratings of organizations and possibly attempting to find a consensus between workers. Indeed, we could not identify any instance in which participants attempted to coordinate their ratings in the chat platform. In line with this observation, we do not observe greater ratings consensus in later periods. Actually, the dispersion of group ratings (as measured with, for example, standard deviation) for the selected organization increased rather than decreased over time. The overall standard deviation of all groups was equal to 2.07, 2.34 and 2.53 in periods 1, 2 and 3.

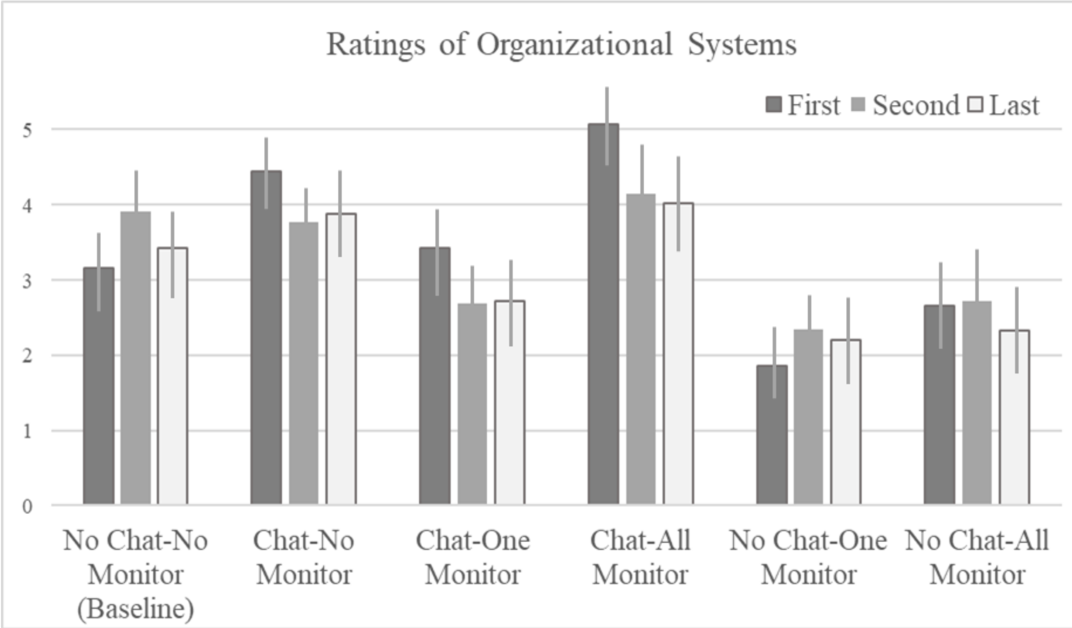

Figure 8. Average Ratings (with 95% confidence intervals bars) Across Organizational Systems for First, Second and Last Periods

Form a practitioner standpoint, it thus follows that, among the organizational solutions to shirking under investigation, promoting chat among peers may be preferred. Indeed, none of the organizational solutions in which chat was absent generated higher ratings than the baseline. This means that organizational solutions relying only on monitoring reduce workers' satisfaction despite leading to higher organizational performance and higher workers' revenues. In Figure 9, we show that workers' satisfaction ratings are not aligned with organizational performance (measured as production in Study 1).[19] In particular, the organizational system that received the highest ratings ('Chat-All Monitor') was ranked fourth out of the five solutions in terms of organizational performance (using production data from Study 1). By contrast, the organizational system involving peer monitoring but no chat ('No Chat-All Monitor') received the second-lowest ratings while leading to the second-highest organizational performance. The solutions to shirking that should probably be favored are in the top right corner (shaded area) in Figure 9. These are organizational systems that produce ratings at least as high as the baseline (Study 2), while increasing workers' performance substantially (Study 1).

[19] We use production in Study 1 to avoid endogeneity effects that might occur as the organizational systems were determined by participants votes in Study 2.

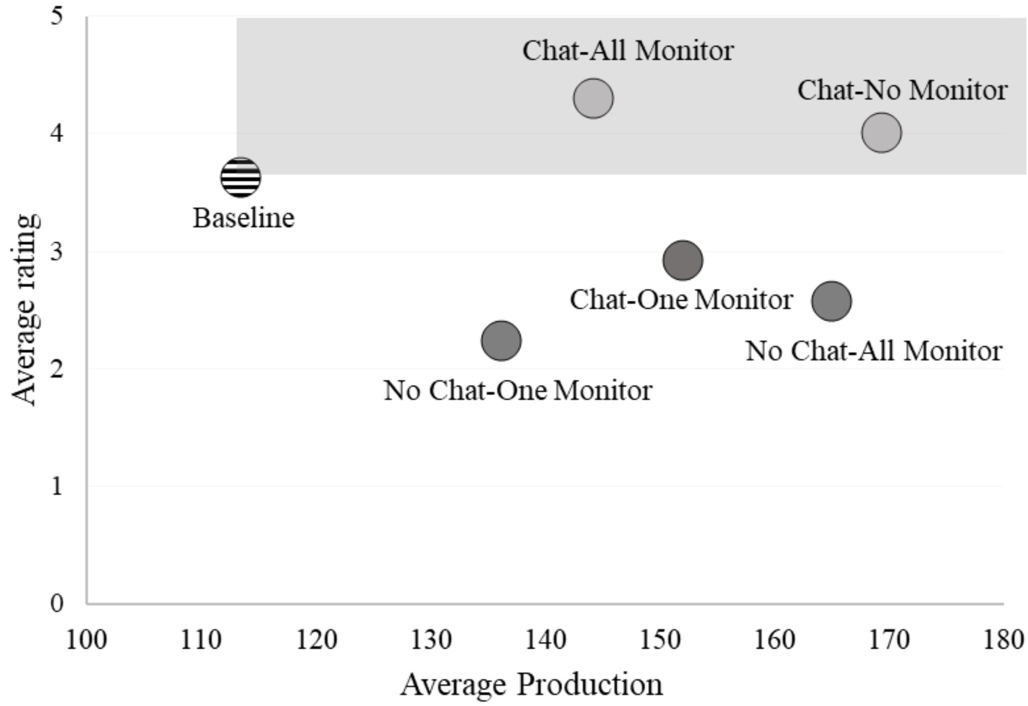

Figure 9. Average Production (Study 1) and Average Ratings (Study 2) Across Organizational Systems

It is also noteworthy that organizational systems involving both monitoring and chat produced ratings similar to those involving chat only. This suggests the negative effect of monitoring is offset by the presence of chat. Although this result was not predicted by our model, it might be understood a posteriori as a positive interaction effect between chat and monitoring on workers' intrinsic motivation (Deci, 1971; Frey, 1997; Deci and Ryan, 2000; Fehr and Falk, 2002). Granting workers the possibility to voice their concerns about intrusive monitoring or otherwise build a positive social relationship might offset the excessive control (lack of autonomy) associated with monitoring (see e.g., Wagner et al., 1997; Corgnet, Hernan-Gonzalez, and McCarter, 2015). In line with this argument, we observe that 'Chat-All Monitor' is the only treatment in which workers exchanged messages regarding the monitoring strategy. Actually, 22% of the messages sent by workers in this treatment dealt with issues of monitoring, most of which (about two-thirds) indicated that monitoring should be used with parsimony and that any abusive monitoring strategy would reflect a lack of trust and an insufficient dedication to the work task. The following message of a participant illustrates this point: "Don't spend so much time watching… Speed up production."

Practitioners who are already using monitoring solutions that would be costly to dismantle might thus foster communication between peers about the monitoring system, as a means of restoring workers' motivation. Concretely, organizations might foster peer communication by promoting user-friendly chat platforms at work and encouraging employees to use them, particularly when communicating about issues of monitoring and performance. Indeed, the time for such solutions seems rife, as a multitude of professional instant messaging platforms, conference technologies, chat rooms, blogs, and billboards have become available, many of which can dramatically increase the ease and lower the costs of communication. In addition, enterprise social networks, which are internal private social networks (e.g., Socialcast) that facilitate communication among employees, have boomed in recent years (e.g., Mishra, Walker and Mishra, 2014). Indeed, technologies like these may take on particular importance in the COVID-19 era. Of course, organizations not subject to social distancing requirements could also encourage peer communication in "old-fashioned" ways like task meetings or open-door policies.

4. Discussion

Shirking in teams is a major incentive-related issue in economics for which many solutions, often based on monitoring technologies, have been proposed. For example, monitoring technologies are a popular solution to curb cyberloafing, which is a modern manifestation of the shirking problem in teams (Blanchard and Henle, 2008). Indeed, a large majority of organizations have implemented systems to monitor their employees' internet use (Alge, 2001), creating an internet monitoring industry now valued at more than $300 million (Alder, Noel and Ambrose, 2006). With the advent of widespread virtual work in the face COVID-19, the trend toward monitoring seems likely to increase further (Cutter, Chen, & Krouse, 2020; Hernandez, 2020). Despite a sustained uptick in monitoring systems, however, the efficacy of these systems remains unclear (Niehoff and Moorman, 1993; Stanton and Weiss, 2000).

This paper highlights both the positive effect of monitoring systems on workers' performance and their negative impact on workers' satisfaction. In addition, we show that other organizational solutions fostering team identity and promoting prosocial concerns can achieve the same level of worker performance without putting work satisfaction at risk. In particular, we show that organizational systems promoting peer communication are more popular yet comparably effective solutions to cyberloafing, and possibly to other forms of shirking.

Several organizations have already recognized the potential side effects of monitoring their employees excessively and the need for alternative solutions. A General Motors executive, for example, said: "The company's philosophy is that the workplace is an environment of mutual trust and respect. This precludes a policy of accessing employee email" (Agarwal and Rodhain, 2002, p. 3). Our research supports this position. Yet, most organizations continue to monitor workers extensively (Alder, Noel and Ambrose, 2006)—and perhaps especially in the context of virtual work (e.g., Cutter et al., 2020)—thereby downplaying the long-term consequences of a dissatisfied and unmotivated workforce. Our work might motivate employers to reconsider these policies, or at least consider them carefully.

5. References

The data that support the findings of this study are available from the corresponding author upon request.

Agarwal, R., & Rodhain, F. (2002). *Mine or ours: Email privacy expectations, employee attitudes, and perceived work environment characteristics.* Paper presented at the 35th Hawaii International Conference on Systems Sciences.

Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, *115*(3), 715-753.

Akerlof, G. A., & Kranton, R. E. (2005). Identity and the Economics of Organizations. *Journal of Economic perspectives*, *19*(1), 9-32.

Alder, G. S., Noel, T. W., & Ambrose, M. L. (2006). Clarifying the effects of Internet monitoring on job attitudes: The mediating role of employee trust. *Information & Management, 43*(7), 894-903.

Alge, B. J. (2001). Effects of computer surveillance on perceptions of privacy and procedural justice. *Journal of Applied Psychology*, *86*(4), 797-804.

Allport, F. H. (1924). *Social psychology*. Boston: Houghton Mifflin.

Ambrose, M. L., & Alder, G. S. (2000). Designing, implementing, and utilizing computerized performance monitoring: Enhancing organizational justice. *Research in Personnel and Human Resource Management, 18*(187-219).

Andreoni, J., & Bernheim, B. D. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, *77*(5), 1607-1636.

Ariss, S. S. (2002). Computer monitoring: Benefits and pitfalls facing management. *Information & Management, 39*(7), 553-558.

Ashraf, N., & Bandiera, O. (2018). Social incentives in organizations. *Annual Review of Economics*, *10*, 439-463.

Balliet, D. (2010). Communication and cooperation in social dilemmas: A meta-analytic review. *Journal of Conflict Resolution*, *54* (1), 39-57.

Bandiera, O., Barankay, I., & Rasul, I. (2010). Social incentives in the workplace. *The Review of Economic Studies*, *77*(2), 417-458.

Barron, J. M., & Gjerde, K. P. (1997). Peer pressure in an agency relationship. *Journal of Labor Economics*, *15*(2), 234-254.

Bartling, B., & Von Siemens, F. A. (2010). The intensity of incentives in firms and markets: Moral hazard with envious agents. *Labour Economics*, *17*(3), 598-607.

Bewley, T. F. (1995). A depressed labor market as explained by participants. *The American Economic Review*, *85*(2), 250-254.

Bénabou, R., & Tirole, J. (2002). Self-confidence and personal motivation. *The Quarterly Journal of Economics*, *117*(3), 871-915.

Bénabou, R., & Tirole, J. (2003). Intrinsic and extrinsic motivation. *The Review of Economic Studies*, *70*(3), 489-520.

Bénabou, R., & Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, *96*(5), 1652-1678.

Bicchieri, C. (2002). Covenants without swords: Group identity, norms, and communication in social dilemmas. *Rationality and Society*, *14*(2), 192-228.

Blanchard, A. L., & Henle, C. A. (2008). Correlates of different forms of cyberloafing: The role of norms and external locus of control. *Computers in Human Behavior, 24*(3), 1067-1084.
Bochet, O., Page, T., & Putterman, L. (2006). Communication and punishment in voluntary contribution experiments. *Journal of Economic Behavior & Organization*, *60*(1), 11-26.
Bochet, O., & Putterman, L. (2009). Not just babble: Opening the black box of communication in a voluntary contribution experiment. *European Economic Review*, *53*(3), 309-326.
Bohnet, I., & Frey, B. S. (1999). Social distance and other-regarding behavior in dictator games: Comment. *American Economic Review*, *89*(1), 335-339.
Bolton, P., & Dewatripont, M. (2005). *Contract Theory*. MIT press.
Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, *90*(1), 166-193.
Burnham, T. C., & Hare, B. (2007). Engineering human cooperation. *Human Nature*, 18(2), 88-108.
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.
Carpenter, J. P. (2007b). The demand for punishment. *Journal of Economic Behavior & Organization*, *62*(4), 522-542.
Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, *117*(3), 817-869.
Charness, G., Cobo-Reyes, R., & Jiménez, N. (2014). Identities, selection, and contributions in a public-goods game. *Games and Economic Behavior*, *87*, 322-338.
Charness, G., Rigotti, L., & Rustichini, A. (2007). Individual behavior and group membership. *American Economic Review*, *97*(4), 1340-1352.
Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics*, *14*(1), 47-83.
Chen, Y., & Li, S. X. (2009). Group identity and social preferences. *American Economic Review*, 99(1), 431-457.Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. *Advances in Experimental Social Psychology*, *24*, 201-234.
Corgnet, B., Hernán-Gonzalez, R., & Mateo, R. (2019). Rac(g)e Against the Machine?: Social Incentives When Humans Meet Robots. *Social Incentives When Humans Meet Robots. GATE WP*.
Corgnet, B., Hernan-Gonzalez and McCarter, M. (2015). The role of the decision-making regime on cooperation in a workgroup social dilemma: an examination of cyberloafing" *Games*, *6*(4), 588-603.
Corgnet, B., Hernan-Gonzalez, R., & Rassenti, S. (2015a). Peer pressure and moral hazard in teams: experimental evidence. *Review of Behavioral Economics*, *2*(4), 375-403.
Corgnet, B., Hernan-Gonzalez, R., & Rassenti, S. (2015b). Firing threats and tenure in virtual organizations: incentives effects and impression management. *Games and Economic Behavior*, *91*, 97-113.
Corgnet, B., Hernan-Gonzalez, R., & Schniter, E. (2015). Why real leisure really matters: incentive effects on real effort in the laboratory. *Experimental Economics 18*(2), 284-301.
Corgnet, B., Gómez-Miñambres, J., & Hernán-González, R. (2015). Goal Setting and Monetary Incentives: When Large Stakes Are Not Enough. *Management Science*, 61(12), 2926-2944.
Cutter, C., Chen, T. & Krouse, S. (2020). You're working from home, but your company is still watching You. May, 11th, *Wall Street Journal*.
Dalal, R. S. (2005). A meta-analysis of the relationship between organizational citizenship behavior and counterproductive work behavior. *Journal of Applied Psychology*, 90(6), 1241–1255.
Danna, K., & Griffin, R. W. (1999). Health and well-being in the workplace: A review and synthesis of the literature. *Journal of Management*, *25*(3), 357-384.

Davis, D. D., & Holt, C. A. (1993). *Experimental Economics*. Princeton university press.
Dawes, R. M. (1991). Social dilemmas, economic self-interest, and evolutionary theory. In *Frontiers of Mathematical Psychology* (pp. 53-79). Springer, New York, NY.
Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology, 18*(1), 105.
Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.
Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*(4), 227-268.
Deci, E. L., Ryan, R. M., Gagne, M., Leone, D. R., Usunov, J., & Kornazheva, B. P. (2001). Need satisfaction, motivation, and well-being in the work organizations of a former eastern bloc country: A cross-cultural study of self-determination. *Personality and Social Psychology Bulletin, 27*(8), 930-942.
Diener, E. (1980). Deindividuation: The absence of self-awareness and self-regulation in group members. *The Psychology of Group Influence*, *209242*.
Dugar, S., & Shahriar, Q. (2012). Group identity and the moral hazard problem: Experimental evidence. *Journal of Economics & Management Strategy*, *21*(4), 1061-1081.
Dur, R., & Sol, J. (2010). Social interaction, co-worker altruism, and incentives. *Games and Economic Behavior*, *69*(2), 293-301.
Englmaier, F., & Wambach, A. (2010). Optimal incentive contracts under inequity aversion. *Games and Economic Behavior*, *69*(2), 312-328.
Eriksson, T., Poulsen, A., & Villeval, M. C. (2009). Feedback and incentives: Experimental evidence. *Labour Economics, 16*(6), 679-688.
Falk, A., & Ichino, A. (2006). Clean evidence on peer effects. *Journal of Labor Economics, 24*(1), 39-57.
Falk, A., & Kosfeld, M. (2006). The hidden costs of control. *American Economic Review, 96*(5), 1611-1630.
Fehr, E., & Falk, A. (2002). Psychological foundations of incentives. *European Economic Review, 46*(4-5), 687-724.
Fehr, E., & Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives*, *14*(3), 159-181.
Frey, B. S. (1997). Not just for the money.
Fehr, E., & Fischbacher, U. (2002). Why social preferences matter–the impact of non-selfish motives on competition, cooperation and incentives. *The Economic Journal*, *112*(478), C1-C33.
Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics, 114*(3), 817-868.
Frey, B. S. (1993). Does monitoring increase work effort: The rivalry with trust and loyalty. *Economic Inquiry, 31*(4), 663-670.
Gagne, M., & Deci, E. L. (2005). Self-determination theory and work motivation. *Journal of Organizational Behavior, 26*(4), 331-362.
Gioia, F. (2017). Peer effects on risk behaviour: the importance of group identity. *Experimental Economics*, *20*(1), 100-129.
Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives*, *25*(4), 191-210.
Goette, L., Huffman, D., & Meier, S. (2006). The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups. *American Economic Review*, *96*(2), 212-216.

Gunia, B. C., Wang, L., Huang, L., Wang, J. W., & Murnighan, J. K. (2012). Contemplation and conversation: Subtle influences on moral decision making. *Academy of Management Journal, 55*(1), 13-33.
Guryan, J., Kroft, K., & Notowidigdo, M. J. (2009). Peer effects in the workplace: Evidence from random groupings in professional golf tournaments. *American Economic Journal: Applied Economics, 1*(4), 34-68.
Hamilton, B. H., Nickerson, J. A., & Owan, H. (2012). Diversity and productivity in production teams. In *Advances in the Economic Analysis of participatory and Labor-managed Firms* (pp. 99-138). Emerald Group Publishing Limited.
Henle, C. A., & Blanchard, A. L. (2008). The interaction of work stressors and organizational sanctions on cyberloafing. *Journal of Managerial Issues*, *20*(3), 383-400.
Herbst, D., & Mas, A. (2015). Peer effects on worker output in the laboratory generalize to the field. *Science*, *350*(6260), 545-549.
Hernandez, K. (2020). Even if you're working from home, your employer is still keeping track of your productivity—here's what you need to know. March, 20th, *CNBC*.
Hoffman, E., McCabe, K., & Smith, V. L. (1996). Social distance and other-regarding behavior in dictator games. *The American Economic Review*, *86*(3), 653-660.
Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics, 6(2), 65-70.
Holmström, B. (1982). Moral hazard in teams. *The Bell Journal of Economics,* 324-340.
Holmstrom, B., & Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law Economics & Organization*, *7*, 24.
Holmstrom, B., & Milgrom, P. (1994). The firm as an incentive system. *The American Economic Review*, 972-991.
Holmström, B. (2017). Pay for performance and beyond. *American Economic Review*, *107*(7), 1753-77.
Itoh, H. (1991). Incentives to help in multi-agent situations. *Econometrica*, *59*(3), 611-636.
Jett, Q. R., & George, J. M. (2003). Work interrupted: A closer look at the role of interruptions in organizational life. *Academy of management Review*, 28(3), 494-507.
Kallman, E. (1993). Electronic monitoring of employees: Issues and guidelines. *Journal of Systems Management, 44*(6), 17-21.
Kandel, E., & Lazear, E. P. (1992). Peer pressure and partnerships. *Journal of Political Economy, 100*(4), 801-817.
Kerr, N. L., & Kaufman-Gilliland, C. M. (1994). Communication, commitment, and cooperation in social dilemma. *Journal of Personality and Social Psychology*, *66*(3), 513.
Kinicki, A. J., McKee-Ryan, F. M., Schriesheim, C. A., & Carson, K. P. (2002). Assessing the construct validity of the job descriptive index: A review and meta-analysis. *Journal of Applied Psychology*, 87(1), 14-32.
Laffont, J. J., & Martimort, D. (2002). The Principal Agent Model. *Princeton University*.
Lazear, E. P., & Shaw, K. L. (2007). Personnel economics: The economist's view of human resources. *Journal of Economic Perspectives*, *21*(4), 91-114.
Ledford Jr, G. E., Lawler III, E. E., & Mohrman, S. A. (1995). Reward innovations in Fortune 1000 companies. *Compensation & Benefits Review*, *27*(4), 76-80.
Ledyard, J. O., Kagel, J. H., & Roth, A. E. (1995). Handbook of experimental economics. *Public Goods: A Survey of Experimental Research*, 111-194.
Lim, V. K. G. (2002). The IT way of loafing on the job: Cyberloafing, neutralizing, and organizational justice. *Journal of Organizational Behavior, 23*, 675-694.

Niehoff, B. P., & Moorman, R. H. (1993). Justice as a mediator of the relationship between methods of monitoring and organizational citizenship behavior. *Academy of Management Journal, 36*(3), 527-556.

Mas, A., & Moretti, E. (2009). Peers at work. *American Economic Review*, *99*(1), 112-45.

Maskin, E., & Tirole, J. (1999). Unforeseen contingencies and incomplete contracts. *The Review of Economic Studies*, *66*(1), 83-114.

Miller, C. S., & Schuster, M. H. (1987). Gainsharing plans: A comparative analysis. *Organizational Dynamics*.

Mishra, K., Walker, K., & Mishra, A. (2014). Using Social Media in the Workplace: How “Lenovo Central” Brings Employees Together. In *Integrating Social Media into Business Practice, Applications, Management, and Models* (pp. 146-161). IGI Global.

Niehoff, B. P., & Moorman, R. H. (1993). Justice as a mediator of the relationship between methods of monitoring and organizational citizenship behavior. *Academy of Management Journal*, *36*(3), 527-556.

Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves?. *Journal of Public Economics*, *92*(1-2), 91-112.

Mummendey, A., Simon, B., Dietze, C., Grünert, M., Haeger, G., Kessler, S., ... & Schäferhoff, S. (1992). Categorization is not enough: Intergroup discrimination in negative outcome allocation. *Journal of Experimental Social Psychology*, *28*(2), 125-144.

Nyberg, A. J., Maltarich, M. A., Abdulsalam, D. D., Essman, S. M., & Cragun, O. (2018). Collective pay for performance: a cross-disciplinary review and meta-analysis. *Journal of Management, 44*(6), 2433-2472.

Ostrom, E., & Walker, J. (1991). Communication in a commons: cooperation without external enforcement. *Laboratory Research in Political Economy*, 287-322.

Ostrom, E., Walker, J., & Gardner, R. (1992). Covenants with and without a sword: Self-governance is possible. *American Political Science Review*, *86*(2), 404-417.

Ostrom, E., Gardner, R., & Walker, J. (1994). *Rules, games, and common-pool resources*. University of Michigan Press.

Palfrey, T. R., & Rosenthal, H. (1991). Testing for effects of cheap talk in a public goods game with private information. *Games and Economic Behavior*, *3*(2), 183-220.

Rabbie, J. M., Schot, J. C., & Visser, L. (1989). Social identity theory: A conceptual and empirical critique from the perspective of a behavioural interaction model. *European Journal of Social Psychology*, *19*(3), 171-202.

Ramalingam, A., & Rauh, M. T. (2010). The firm as a socialization device. *Management Science*, *56*(12), 2191-2206.

Rey-Biel, P. (2008). Inequity aversion and team incentives. *Scandinavian Journal of Economics*, *110*(2), 297-320.

Reyniers, D., & Bhalla, R. (2013). Reluctant altruism and peer pressure in charitable giving. *Judgment and Decision Making*, *8*(1), 7-15.

Rotemberg, J. J. (1994). Human relations in the workplace. *Journal of Political Economy*, *102*(4), 684-717.

Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, 43(3), 450-461.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68-78.

Sally, D. (1995). Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992. *Rationality and Society*, *7*(1), 58-92.

Sherif, M. (1936). A study of some social factors in perception. *Archives of Psychology*, *27*, 187.
Smith, V. L. (1982). Microeconomic systems as an experimental science. *The American Economic Review*, *72*(5), 923-955.
Smith, M. J., & Amick, B. C. (1989). Electronic monitoring at the workplace: Implications for employee control and job stress. In S. L. Sauter, J. J. Hurrell & C. L. Cooper (Eds.), *Job Control and Worker Health* (pp. 275-289). New York: John Wiley & Sons, Ltd.
Somers, R. H. (1962). A new asymmetric measure of association for ordinal variables. *American Sociological Review, 27*(6), 799-811.
Spector P. (1985). Measurement of human service staff satisfaction: Development of the Job Satisfaction Survey. *American Journal of Community Psychology*. *13*, 693-713.
Stanton, J. M. (2000a). Reactions to employee performance monitoring: Framework, review, and research directions. *Human Performance, 13*(1), 85-113.
Stanton, J. M. (2000b). Traditional and electronic monitoring from an organizational justice perspective. *Journal of Business and Psychology, 15*(1), 129-147.
Stanton, J. M., & Weiss, E. M. (2000). Electronic monitoring in their own words: An exploratory study of employees' experiences with new types of surveillance. *Computers in Human Behavior, 16*(4), 423-440.
Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*(2), 149-178.
Tajfel, H., Turner, J. C., Austin, W. G., & Worchel, S. (1979). An integrative theory of intergroup conflict. *Organizational Identity: A Reader*, 56-65.
Tamir, D. I., & Hughes, B. L. (2018). Social Rewards: From basic social building blocks to complex social behavior. *Perspectives on Psychological Science*, *13*(6), 700-717.
Wagner, D. T., Barnes, C. M., Lim, V. K. G., & Ferris, D. L. (2012). Lost sleep and cyberloafing: Evidence from the laboratory and a Daylight Saving Time quasi-experiment. *Journal of Applied Psychology, 97*(5), 1068-1076.
Westover, J. H., Westover, A. R., & Westover, L. A. (2010). Enhancing long-term worker productivity and performance: The connection of key work domains to job satisfaction and organizational commitment. *International Journal of Productivity and Performance Management*, *59*(4), 372-387.
Yamagishi, T., Jin, N., & Kiyonari, T. (1999). Bounded generalized reciprocity: Ingroup boasting and ingroup favoritism. *Advances in Group Processes*, *16*(1), 161-197.
Zelmer, J. (2003). Linear public goods experiments: A meta-analysis. *Experimental Economics*, *6*(3), 299-310.
Zweig, D., & Webster, J. (2002). Where is the line between benign and invasive? An examination of psychological barriers to the acceptance of awareness monitoring systems. *Journal of Organizational Behavior, 23*(5), 605-633.

Appendix A. Theoretical model

We rely on previous social preferences and social pressure models to study the effect of monitoring and peer chat on effort provision (see Kandel and Lazear, 1992; Rotemberg 1994; Fehr and Schmidt, 1999; Rey-Biel, 2008; Bartling and von Siemens, 2010; Dur and Sol, 2010; Englmaier and Wambach, 2010). We derive our hypotheses using the moral-hazard in teams' model introduced by Holmström (1982). We consider n workers producing a total output $\mathrm{f} \coloneqq \mathrm{f}(\mathrm{e}_1, \mathrm{e}_2, \dots, \mathrm{e}_\mathrm{n})$ which depends on each worker's effort $e_i \geq 0$ where $i \in \{1, \dots, n\}$. We assume that $f(\cdot)$ is linear and separable in workers' efforts, $f \coloneqq \sum_{i=1}^{n} a_i e_i$, where $a_i > 0$ is the marginal product of effort of worker *i*. By assuming separability in workers' effort, our production function allows us to identify each worker's individual contribution. This is the type of production function we use in our experimental design. The cost of effort is represented by $C(e_i)$ where $C'(e_i) \geq 0$ and $C''(e_i) \leq 0$. Each worker in the team is paid according to team incentives thus collecting a share $\frac{1}{n}$ of total production. The utility function of worker *i* is thus:

$$v_i \coloneqq \frac{f}{n} - C(e_i) \qquad [1]$$

Hypothesis 1. (Chat)

We extend the utility function of worker *i* in [1] to account for the effect of the presence of chat and monitoring. Following Dur and Sol (2010), we assume that chat will foster social interaction between workers thus promoting altruistic motives. We capture worker *i*'s altruism with a parameter $\xi_i \geq 0$. An altruistic person ($\xi_i > 0$) values other workers' pay positively so that under peer chat a worker's utility function becomes:

$$u_i \coloneqq v_i + \frac{(n-1)f}{n} \xi_i \qquad [2]$$

We derive our first hypothesis by relying upon the fact that peer chat will induce stronger altruistic motives among team partners, in line with the model of Dur and Sol (2010). Our first hypothesis abstracts away from participation constraint so that we are going to assume that workers have already accepted to work under certain organizational conditions, in this case peer chat. We thus focus on the incentive effect associated to peer chat looking into the incentive compatibility constraint of workers, which follows directly from workers' utility maximization. In the case of an altruistic worker, the first order condition (see [3]) is such that an increase in altruistic concerns (ξ_i) which follows from peer chat will lead to an increase in the level of effort exerted by workers given the assumptions on the cost of effort function:

$$C'(e_i) = \frac{1+(n-1)\xi_i}{n} a_i \qquad [3]$$

This completes our proof of Hypothesis 1. In addition to altruism, concerns regarding inequality aversion have been shown to be prevalent (see Charness and Rabin, 1993; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). However, in the context of team incentives, all workers are paid the same so that any consideration regarding inequality in payoffs induced by peer chat would not have any effect.

Hypothesis 2. (Monitoring)

Following Kandel and Lazear (1992) modeling of peer pressure in team production setups, we consider that being observed by other team workers will affect their incentive to exert effort. At the empirical level, Mas and Moretti (2009) and Corgnet, Hernan-Gonzalez, and Rassenti (2015a) have shown that workers exert higher effort when observed by other team members. We can think of a variety of reasons why workers would produce more when being observed by others. The first possibility relates to audience effects à la Andreoni and Bernheim (2009) in which case people wants to be seen as fair. In particular, the authors put forward that people are inclined to split outcomes equally when seen by others as they want to be perceived as egalitarian. In the context of team incentives, all workers are paid the same so that fairness concerns regarding strict pay equality do not apply to our setup. However, workers might still be motivated to work hard because they want to be seen as complying with a social norm of effort and production (Kandel and Lazear, 1992). The work of Corgnet, Hernan-Gonzalez, and Rassenti (2015a) shows that peer pressure in teams is effective for both low and high producers suggesting that being observed by others does not foster a common production norm although it can certainly induce a norm of high effort. One way to model the emergence of a high-effort norm under peer monitoring is to consider that team members feel shame whenever they slack off because this directly hurts others' payoffs by reducing team production. We thus model social pressure as workers' willingness not to hurt the payoffs of their team members. We capture this effect with the parameter $\chi_{i,j} \geq 0$ which measures the peer pressure worker *i* suffers from worker *j* so that the utility function of a worker who is subject to peer monitoring can be written as:

$$w_i \coloneqq v_i + \frac{f}{n}\sum_j \chi_{i,j} \qquad [4]$$

where $j \in M$ and M stands for the set of workers who observe worker i's performance.

Our peer pressure model is such that a worker who is observed will feel shame and thus value the payoff of other workers positively. Peer pressure thus triggers shame leading team members to behave as if they were altruistic. But, the difference between altruism and social pressure is that altruistic workers [see 3] will exert higher effort when working in a team regardless of whether they are observed or not by their team members. As for Hypothesis 1, we derived our hypothesis regarding peer monitoring using the first order conditions of the worker's utility function. The first order condition below shows that peer monitoring, by enlarging the set of monitors *M*, will make the term $\sum_j \chi_{i,j}$ larger thus boosting workers' effort.

$$C'(e_i) = \frac{1+\sum_j \chi_{i,j}}{n} a_i \quad [5]$$

This completes our proof of Hypothesis 2.

Hypothesis 3. (Work satisfaction)

So far, we have assumed that the participation constraint was satisfied so that only incentive effects were studied. However, as we argue in our hypotheses section, peer chat and peer monitoring induce different participation constraints. Peer chat increases work motivation by provide a social context to workers, which has been shown to be a crucial element of well-being at work. By contrast, peer monitoring by inducing further control and restricting autonomy will have the opposite effect. We can thus write the participation constraint of worker *i* as follows:

$$\frac{f}{n} + \frac{f}{n}\{(n-1)\xi_i + \sum_j \chi_{i,j}\} - C(e_i) + pc_i - pm_i \geq v_0 \quad [6]$$

where v_0 is the utility level obtained by a worker in the next-best alternative, and pc_i represents the utility gain of worker *i* from being in a team which can engage in peer chat and $-pm_i$ represents the utility loss of worker *i* from being in a team in which peer monitoring is present. From [6], it directly follows that peer chat will induce greater work satisfaction (left-hand side) than a baseline treatment in which there is neither chat nor monitoring and in which the participation constraint would be such that: $\frac{f}{n} - C(e_i) \geq v_0$. This follows from the fact that peer chat is positively valued by workers ($pc_i > 0$) and it fosters altruism ($\xi_i > 0$). The effect of peer monitoring on the participation constraint is mixed because it increases workers' revenues ($\frac{f}{n}\sum_j \chi_{i,j} > 0\}$ while being negatively value by workers ($-pm_i < 0$). This leads to Hypothesis 3.

It is important to note that the effect of peer chat and peer monitoring on the participation constraint do not affect the magnitude of incentive effect. However, we believe they are crucial because they might affect work behavior and in particular promote counter productive work behavior such as theft or absenteeism. It is also the case that lower work satisfaction, by decreasing the left-hand side of the participation constraint, will push workers to leave the firm thus creating additional turnover costs. Workers who want to leave because the participation constraint is not met as a result of a change in organizational design (such as the introduction of peer monitoring) might not be able to do so immediately in which case the negative impact of a poorly accepted organizational change will be seen only in the longer run. This is why managers may sometimes bypass the participation constraint and focus on incentive effect which will produce positive effect in the short run.

Appendix B. Additional tables

Table B1. Study 1 Descriptive Statistics

Mean (Standard deviation)	Worker's production per period (in cents)	Percentage of time spent on each activity:				First Table Correct
		Work Task	Internet	Chat	Monitoring	
No Chat-No Monitor (Baseline)	113.44 (92.48)	71.48% (27.07%)	28.52% (27.07%)	-	-	46.67%
Chat-No Monitor	169.44 (105.68)	81.99% (18.54%)	12.79% (16.89%)	5.22% (5.44%)	-	65.00%
Chat-All Monitor	144.24 (75.12)	81.45% (17.58%)	9.17% (13.69%)	5.96% (9.39%)	3.41% (3.57%)	55.00%
Chat-One Monitor	152.133 (117.006)	87.32% (13.58%)	6.62% (11.60%)	5.19% (5.08%)	0.87% (2.88%)	56.67%
No Chat-One Monitor	136.24 (123.28)	80.16% (23.36%)	19.44% (23.61%)	-	0.40% (1.39%)	40.00%
No Chat-All Monitor	165.04 (116.5)	82.51% (21.65%)	13.13% (20.39%)	-	4.36% (3.92%)	56.67%

Table B2. Study 1 Pairwise Comparisons Between Treatments

P-values for PROPORTION TESTS FOR FIRST TABLE CORRECT (BONFERRONI-HOLM CORRECTION)

Organizational system	No Chat-No Monitor (Baseline)	Chat-No Monitor	Chat-All Monitor	Chat-One Monitor	No Chat-All Monitor
Chat-No Monitor	0.043 (0.605)				
Chat-All Monitor	0.361 (>0.999)	0.264 (>0.999)			
Chat-One Monitor	0.273 (>0.999)	0.350 (>0.999)	0.854 (>0.999)		
No Chat-All Monitor	0.273 (>0.999)	0.350 (>0.999)	0.854 (>0.999)	0.999 (>0.999)	
No Chat-One Monitor	0.461 (>0.999)	0.006 (0.092)	0.099 (>0.999)	0.068 (0.8801)	0.068 (0.8801)

Table B3. Linear panel regression with random effects for workers' production (in cents), internet usage and time on the task (in seconds), including ability measure (First sum is correct).

Dependent variable	Production (in cents)		Internet Usage (in seconds)		Time on the Task (in seconds)	
	[1]	[2]	[3]	[4]	[5]	[6]
Intercept	17.921 (11.501)	43.854*** (6.219)	259.169*** (54.332)	160.015*** (53.009)	916.600*** (54.767)	1035.954*** (53.392)
Chat-No Monitor Dummy	39.507* (22.040)	2.007 (9.922)	-180.781*** (61.392)	-79.622 (58.322)	116.546* (69.471)	-37.164 (70.244)
Chat-One Monitor Dummy	29.707* (16.263)	-15.693 (16.074)	-251.080*** (53.563)	-87.755 (56.434)	166.261*** (52.844)	-30.835 (57.223)
Chat-All Monitor Dummy	23.333 (14.664)	-20.867 (13.832)	-228.571*** (56.861)	-60.694 (52.969)	115.320* (63.086)	-98.128* (54.516)
No Chat-One Monitor Dummy	28.707* (16.429)	8.907 (17.655)	-111.838 (69.872)	-51.683 (52.743)	107.655 (69.892)	43.889 (53.142)
No Chat-All Monitor Dummy	42.640** (19.145)	33.940*** (12.266)	-180.356*** (58.864)	-77.948 (53.193)	127.141** (59.646)	39.035 (54.965)
Period	17.911*** (1.713)	9.267*** (3.521)	34.447*** (5.001)	67.499*** (8.033)	-27.714*** (5.754)	-67.499*** (8.033)
First sum is correct	89.598*** (9.226)	89.598*** (9.238)	-43.449* (23.537)	-43.449* (23.570)	52.086** (24.854)	52.086** (24.889)
Chat-No Monitor Dum.× Period		12.500** (6.175)		-33.719*** (11.786)		51.237*** (12.479)
Chat-One Monitor Dum.× Period		15.133*** (5.100)		-54.441*** (10.615)		65.699*** (10.553)
Chat-All Monitor Dum.× Period		14.733*** (4.527)		-55.959*** (11.537)		71.149*** (11.575)
No Chat-One Monitor Dum.× Period		6.600 (4.287)		-20.052 (14.830)		21.255 (15.053)
No Chat-All Monitor Dum.× Period		2.900 (4.126)		-34.136*** (13.000)		29.369** (12.892)
Observations (organizations)	1800	1800	1800	1800	1800	1800
Prob > χ^2	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
R^2	0.188	0.192	0.102	0.11	0.0484	0.0607

Table B4. Linear panel regression with random effects for workers' production (in cents), internet usage and time on the task (in seconds), including group production in the previous period (Avg Group Production (t-1)).

Dependent variable	Production (in cents)		Internet Usage (in seconds)		Time on the Task (in seconds)	
	[1]	[2]	[3]	[4]	[5]	[6]
Intercept	60.298*** (11.435)	80.144*** (11.555)	305.499*** (61.881)	269.258*** (89.757)	871.190*** (62.839)	930.783*** (89.871)
Chat-No Monitor Dummy	45.270** (20.028)	24.881 (18.359)	-194.004*** (70.772)	-207.453** (102.131)	129.221* (77.265)	55.756 (117.464)
Chat-One Monitor Dummy	36.040** (15.115)	4.604 (27.252)	-278.203*** (61.050)	-196.492** (95.543)	198.452*** (59.636)	98.395 (110.477)
Chat-All Monitor Dummy	30.785*** (11.795)	5.173 (15.904)	-263.116*** (62.925)	-205.231** (94.031)	158.694** (68.023)	75.592 (94.507)
No Chat-One Monitor Dummy	19.384 (13.950)	2.760 (25.710)	-114.932 (78.988)	-96.067 (96.916)	108.679 (78.316)	89.815 (96.711)
No Chat-All Monitor Dummy	35.975** (16.471)	26.316 (18.602)	-184.752*** (67.677)	-124.841 (96.013)	125.819* (67.122)	101.282 (97.498)
Period	7.179*** (2.474)	2.765 (2.129)	38.368*** (6.905)	47.698*** (11.347)	-35.373*** (6.691)	-47.693*** (11.352)
Avg Group Production (t-1)	0.317*** (0.084)	0.277*** (0.086)	-0.474* (0.244)	-0.442* (0.229)	0.592** (0.245)	0.441* (0.233)
Chat-No Monitor Dum.× Period	-	6.445 (4.310)	-	3.337 (16.145)	-	23.313 (16.161)
Chat-One Monitor Dum.× Period	-	9.362* (5.553)	-	-23.656* (13.603)	-	30.012 (20.121)
Chat-All Monitor Dum.× Period	-	7.616** (3.532)	-	-16.782 (15.268)	-	24.861* (13.804)
No Chat-One Monitor Dum.× Period	-	4.981 (5.447)	-	-5.579 (16.505)	-	6.256 (16.576)
No Chat-All Monitor Dum.× Period	-	3.348 (3.709)	-	-17.597 (13.256)	-	9.215 (13.916)
Observations (organizations)	1440	1440	1440	1440	1440	1440
Prob > χ^2	142.99***	196.54***	124.59***	311.8***	80.828***	214.31***
R^2	0.0948	.0899	0.106	0.107	0.061	0.0596

Table B5. Study 1 Pairwise Comparisons Between Treatments

P-values for RANK-SUM TESTS AND T-TESTS FOR TOTAL PRODUCTION |INTERNET USAGE| (TIME ON TASK) PER WORKER

<table>
<tr><th>Organizational system</th><th></th><th>No Chat-No Monitor (Baseline)</th><th>Chat-No Monitor</th><th>Chat-One Monitor</th><th>Chat-All Monitor</th><th>No Chat-One Monitor</th></tr>
<tr><td rowspan="2">Chat-No Monitor</td><td>Rank-Sum test</td><td>0.005
|<0.001|
(0.129)</td><td></td><td></td><td></td><td></td></tr>
<tr><td>T-test</td><td>0.003
|<0.001|
(0.014)</td><td></td><td></td><td></td><td></td></tr>
<tr><td rowspan="2">Chat-One Monitor</td><td>Rank-Sum test</td><td>0.039
|<0.001|
(0.043)</td><td>0.416
|0.273|
(0.453)</td><td></td><td></td><td></td></tr>
<tr><td>T-test</td><td>0.023
|<0.001|
(<0.001)</td><td>0.343
|0.042|
(0.215)</td><td></td><td></td><td></td></tr>
<tr><td rowspan="2">Chat-All Monitor</td><td>Rank-Sum test</td><td>0.069
|<0.001|
(0.443)</td><td>0.159
|0.648|
(0.350)</td><td>0.783
|0.314|
(0.047)</td><td></td><td></td></tr>
<tr><td>T-test</td><td>0.047
|<0.001|
(0.018)</td><td>0.136
|0.200|
(0.871)</td><td>0.610
|0.418|
(0.144)</td><td></td><td></td></tr>
<tr><td rowspan="2">No Chat-One Monitor</td><td>Rank-Sum test</td><td>0.476
|0.026|
(0.053)</td><td>0.055
|0.182|
(0.466)</td><td>0.255
|0.013|
(0.813)</td><td>0.379
|0.069|
(0.125)</td><td></td></tr>
<tr><td>T-test</td><td>0.255
|0.053|
(0.062)</td><td>0.116
|0.079|
(0.636)</td><td>0.425
|<0.001|
(0.116)</td><td>0.666
|0.004|
(0.733)</td><td></td></tr>
<tr><td rowspan="2">No Chat-All Monitor</td><td>Rank-Sum test</td><td>0.015
|<0.001|
(0.043)</td><td>0.811
|0.492|
(0.353)</td><td>0.616
|0.733|
(0.498)</td><td>0.300
|0.801|
(0.017)</td><td>0.114
|0.073|
(0.816)</td></tr>
<tr><td>T-test</td><td>0.008
|<0.001|
(0.015)</td><td>0.831
|0.922|
(0.888)</td><td>0.502
|0.058|
(0.334)</td><td>0.248
|0.214|
(0.769)</td><td>0.190
|0.120|
(0.569)</td></tr>
</table>

*p -value<.10, ** p-value<.05, and *** p-value<.01.

Table B6. Study 1 - Communication Categories.

Master categories	**Category Number**	**% of messages**	**Category description**
Social interaction	1	2.80%	Greetings (Hello/Goodbye)
	2	3.40%	Distracting others (jokes, stories)
	3	11.30%	Personal chat (talking about likes and dislikes)
	All	**17.50%**	
Encouragement and help	4	13.30%	Encouraging others to produce
	5	3.40%	Thanking other for their cooperative behavior
	6	26.80%	Help others complete the task
	7	11.00%	Ask others for help and hints to complete the task
	All	**54.50%**	
Performance assessment and comparisons	10	1.40%	Ask others' performance on the task
	11	5.70%	State your own performance
	All	**7.10%**	
Discouragement	8	0.50%	Discouraging others to produce
	9	2.00%	Asking others what is the point of producing anything
	All	**2.50%**	
Non-strategic comments on the experiment	12	14.7%	General comments about the experiment and its goals
	13	3.7%	Other specific comments on the experiment
	All	**18.40%**	

Table B7. Study 2 Pairwise Comparisons between first and last ratings for each organizational system
P-values for t-tests (Sign-Rank tests)

Organizational system	No Chat-No Monitor (Baseline)	Chat- No Monitor	Chat- One Monitor	Chat- All Monitor	No Chat- One Monitor	No Chat- All Monitor
P-values	0.414 (0.352)	0.127 (0.136)	0.040 (0.012)	<0.001 (0.001)	0.229 (0.663)	0.175 (0.091)

Table B8. Study 2 Pairwise Comparisons between treatments for first ratings
P-values for t-tests (Sign-Rank tests)[20]

Organizational system	No Chat- No Monitor (Baseline)	Chat-No Monitor	Chat-One Monitor	Chat-All Monitor	No Chat-One Monitor
Chat-No Monitor	**0.003** **(0.003)**				
Chat-One Monitor	0.620 (0.591)	*0.025* *(0.037)*			
Chat-All Monitor	<0.001 (<0.001)	*0.101* *(0.131)*	<0.001 (<0.001)		
No Chat-One Monitor	*<0.001* *(0.002)*	<0.001 (<0.001)	**<0.001** **(<0.001)**	<0.001 (<0.001)	
No Chat-All Monitor	*0.191* *(0.183)*	<0.001 (<0.001)	0.109 (0.107)	**<0.001** **(<0.001)**	0.008 (0.012)

*p -value<.10, ** p-value<.05, and *** p-value<.01.

[20] Similar results are obtained when using a non-parametric test with clusters at the session level (Somers' *d*, Somers, 1962).

Table B9. Study 2 Pairwise Comparisons between treatments for last ratings

P-values for t-tests (sign rank tests)[21]

Organizational system	No Chat-No Monitor (Baseline)	Chat-No Monitor	Chat-One Monitor	Chat-All Monitor	No Chat-One Monitor
Chat-No Monitor	**0.401 (0.398)**				
Chat-One Monitor	0.182 (0.207)	*0.003 (0.007)*			
Chat-All Monitor	0.755 (0.338)	*0.006 (0.936)*	0.182 (0.007)		
No Chat-One Monitor	*0.002 (0.014)*	0.001 (0.001)	**0.164 (0.008)**	0.348 (0.002)	
No Chat-All Monitor	*0.016 (0.012)*	0.003 (0.003)	0.394 (0.365)	**<0.001 (<0.001)**	0.754 (0.563)

*p -value<.10, ** p-value<.05, and *** p-value<.01.

[21] Similar results are obtained when using a non-parametric test with clusters at the session level (Somers' *d*, Somers, 1962).

Table B10. Study 2 Pairwise Comparisons between treatments for average ratings
P-values for t-tests (Sign rank tests)[22]

Organizational system	No Chat-No Monitor (Baseline)	Chat-No Monitor	Chat-One Monitor	Chat-All Monitor	No Chat-One Monitor
Chat-No Monitor	**0.195 (0.104)**				
Chat-One Monitor	0.216 (0.238)	*0.003 (0.012)*			
Chat-All Monitor	0.098 (0.090)	*0.279 (0.448)*	<0.001 (<0.001)		
No Chat-One Monitor	*<0.001 (<0.001)*	<0.001 (<0.001)	**0.004 (<0.001)**	<0.001 (<0.001)	
No Chat-All Monitor	*0.018 (0.026)*	<0.001 (<0.001)	0.359 (0.278)	**<0.001 (<0.001)**	0.154 (0.098)

*p -value<.10, ** p-value<.05, and *** p-value<.01.

[22] Similar results are obtained when using a non-parametric test with clusters at the session level (Somers' *d*, Somers, 1962).