

# Ask Your Workers to Report Frequently, But Not Too Often

Kuangli Xie\*

November 11, 2018

**Abstract:** Firms have long had employees report progress on tasks to supervisors. Asking employees to report their progress frequently is generally thought to increase productivity. If this occurs it is unclear if it is due to individuals possessing some sort of intrinsic motivation to work which is intensified by the reporting, or if individuals work more diligently when they believe their actions are observed or if they engage in more effort solely to avoid negative feedback. It is also unclear how frequently employees should report. We use controlled experiments to examine how reporting frequency may affect workers' effort decisions, including how much effort they spend on work versus leisure, and how they allocate effort across different tasks available within the firm. We also conduct treatments to identify what aspect of the reporting process is the driver of any performance increases. We find that increasing the frequency of reporting leads workers to spend more time on firm tasks, as opposed to leisure. However, when the frequency of reporting is set too high, workers shift to performing tasks that are more likely to generate positive returns in the short-term but that lead to lower returns overall. When we try to uncover what aspect of the reporting mechanism is responsible for the performance increase, we find weaker effects than predicted by prior literature.

**Keywords:**

**JEL Classification Numbers:** C92, D53, G12

---

\*e-mail: [kxie@smu.edu](mailto:kxie@smu.edu). Southern Methodist University

# 1 Introduction

It is common practice in many firms to have employees report to their supervisors on the status of current projects at specified intervals. For example, firms, including Adobe Systems, Accenture PLC, Deloitte, and General Electric, encourage their managers to check in with their employees every week or every other week.<sup>1,2</sup> Companies such as Goldman Sachs and J.P. Morgan Chase are rolling out a new report and review system where managers can request their workers to report as often as the manager desires and can send ongoing feedback to workers.<sup>3</sup> Many high-tech firms such as Microsoft, Uber, and Walmartlabs encourage their teams to have daily stand-up meetings. The exact content and point of some of these types of meetings will vary from firm to firm but a common theme running through the design of most of them is for the employees to report on what they have accomplished since the prior meeting. In prior decades, high-frequency reporting as we see these days would have been very costly and in some cases impossible but firms, like General Electric, IBM and Amazon, have developed computer and mobile applications that enable frequent conversations between managers and workers to make daily reporting possible and maybe manageable.<sup>4</sup> With the advent of these sorts of capabilities, there is a reason to investigate them to determine what impacts this degree of monitoring might have on employees and what element of the monitoring system is responsible for any improvement in employee performance.

There are any number of possible uses for these meetings such as a manager passing along instructions to his or her workers about priorities and goals, workers seeking feedback on how best to go about a project and so on (Chhokar and Wallin, 1984; Lurie and Swaminathan, 2007; Balcazar et al., 1985) but one of the primary uses for monitoring employees like this is to verify that the employees are working on appropriate tasks and putting in effort towards company goals. That is because these monitoring schemes are of course an attempt to solve

---

<sup>1</sup><https://www.wsj.com/articles/the-never-ending-performance-review-1494322200>

<sup>2</sup><https://www.wsj.com/articles/ge-re-engineers-performance-reviews-pay-practices-1465358463>

<sup>3</sup><https://www.wsj.com/articles/goldman-goes-beyond-annual-review-with-real-time-employee-ratings-1492786653>

<sup>4</sup><https://hbr.org/2016/10/the-performance-management-revolution>

the standard principal-agent problem in which the firm wishes to extract high effort from employees who would otherwise prefer shirking, i.e., not working on firm-related projects.

In many of these monitoring schemes used in practice, financial incentives may be only weakly involved. That is, while eventual promotion and raise decisions may arise out of a year worth of weekly or daily meetings, it is rare for each meeting to involve explicit financial consequences. In fact, much of the point of these meetings is that they are thought to activate non-financial drivers of effort. This leads to an important question of determining if it is possible that frequent monitoring could potentially affect employee effort even absent any financial consequences. This will be the issue examined in this paper. It's not due to a contention that financial incentives do not matter, but rather an interest in determining if monitoring alone can have a significant impact on behavior and then if so, what is the channel through which this effect occurs.

There is a great deal of prior literature that one can look to for insight on this issue. There are a number of different possible channels through which monitoring might increase worker effort but also some evidence suggesting that monitoring could be problematic. One possible explanation for how monitoring could improve worker effort is that there is substantial prior evidence that individuals have a strong desire to conform to expected behavioral norms (Benabou and Tirole, 2012; Andreoni and Bernheim, 2009). Frequent monitoring could be a way for a firm to remind workers of a "norm" in which the worker should be working for the firm. Frequent reminders about progress could make this norm more salient or just make sure workers fully understand their current productivity. There is also prior evidence suggesting that individuals like doing things that benefit others due to the fact that these actions can provide an individual "warm-glow" utility (Andreoni, 1990). It is certainly possible that workers could possess such preferences towards their employer and again frequent reporting could remind an individual of their progress towards achieving such a goal. Either of these types of motivations would represent purely intrinsic motivations in which the individual seeks to satisfy their inner preferences regarding putting in a high effort.

It is alternatively possible that intrinsic motivation is not the channel through which reporting affects behavior but extrinsic motivation through how an individual perceives others to think about their actions. Previous studies have shown people care about their social image or how others view them, so they behave more prosocially when observed by other individuals (Andreoni and Petrie, 2004; Karlan and McConnell, 2014; Salmon and Serra, 2017; Bursztyn and Jensen, 2017). This holds purely on the basis of observation in some cases, without requiring any feedback from the observer. Therefore, if workers are concerned about how they are perceived by their employers, they might increase their work effort simply due to the knowledge that their employer is shown information on their productivity. And again, not because they think this will lead to a future raise or monetary benefit but rather because they are concerned about how the boss will judge them as a person. Of course, pure observation may not be enough to alter behavior, feedback may be required. There are multiple prior papers showing that individuals respond to a situation knowing that not only will their behavior be observed by others, but those others will have the chance to provide feedback (Masclot et al., 2003; Xiao and Houser, 2011; Salmon and Serra, 2017). Note the important issue is the knowledge that feedback will be provided and the behavior can be changed simply on the basis of that rather than on the basis of what the actual feedback is. If this is an operative motivation in this worker context, then an employer providing frequent feedback could certainly boost productivity.

There are also indications in prior work that requiring workers to report frequently may be counterproductive as it could drive them to work less. There are studies which show that monitoring individuals can cause the person being monitored to see the monitoring as a form of distrust and this could in turn possibly diminish their self-esteem (Frey, 1993; Dickinson and Villeval, 2008; Falk and Kosfeld, 2006; Belot and Schröder, 2015). Workers who feel distrusted by an employer may work with less effort rather than more. Also, frequent reporting may have an impact on what type of tasks a worker performs for the firm. If a worker has short-term or easy projects which they expect will return positive results quickly,

they may be more likely to shift towards these tasks and away from ones which require more of a time investment before observing a success. This would be due to perceived pressure to produce positive results in relatively narrow time windows essentially de-incentivizing working on long-run projects which take longer to mature. There is prior evidence of similar behavior in other settings which make it seem plausible that such behavior could occur in the workplace (Merlo and Schotter, 1999; Gneezy et al., 2003; Manso, 2011; Ederer and Manso, 2013; Tian and Wang, 2011; Azoulay et al., 2011).

The empirical question is whether or not any of these behavioral motivations apply in a worker setting and if so, how they balance out. We use controlled experiments to investigate these issues. In the experiment, we simulate the relationship between a worker and an employer. The worker is endowed with a certain amount of time to divide between working for the firm and earning utility only for themselves, that is, we fully implement a standard labor supply model. The worker can choose between two different firm tasks which model the types of tasks described above that either payoff quickly or only after more time is invested. In an important methodological innovation, we also include a real leisure option as described in Dutcher et al. (2018) so that we know that our subjects value their outside option of not working on the firms' behalf. The value of this outside option starts out generating more utility to the worker than the firm tasks but declines with time spent on it which mimics the standard assumption of a declining marginal utility of leisure. This gives us an interior equilibrium in terms of effort provision. We will then have the worker report to their manager their progress with varying frequency and styles of reporting regimes. By having a real leisure outside option, we can measure how much utility a worker is willing to give up to expend additional effort on the behalf of the employer under these different conditions.

The experiment is designed to allow us to identify which of the previously described behavioral motivations could be affecting worker effort. We have one regime in which the worker is reminded of their recent productivity knowing that the employer will not see it. A second treatment involves the employee seeing their productivity and knowing that it

will be shown to the employer though with the employer having no opportunity to respond. Then we have a full reporting regime in which the employers observe and can send feedback regarding their level of satisfaction. We compare the behavior observed in these treatments to a baseline with no reporting. By examining how time spent on firm tasks changes between treatments, we can identify if intrinsic motivation, observation or feedback is sufficient to improve effort provision. In each of these treatments, we vary the frequency of reporting to determine if that can impact behavior. Finally, given the multiple tasks, the workers can choose, we can also observe whether the type of monitoring mechanism or the reporting frequency affects the types of tasks the worker chooses to work on.

We find that increasing the frequency of reporting can increase worker effort on firm tasks but this can come at a cost. When set too high, workers shift to spending more time on the task which generates immediate payoffs despite the fact that this task yields lower expected earnings relative to the task which requires more time investment. While there is substantial literature suggesting that the level of observability of employee's actions should impact their behavior, we find a limited impact of differing levels of observation. In the conclusion, we discuss why this might be the case and discuss the possibility of future research to better identify the situations in which observability does and does not have strong impacts on behavior.

## **2 Experiment Design**

The focus of this study is on determining what element of a worker reporting their progress to a supervisor can lead to increased work activity and whether the frequency with which these reports are made can impact worker effort. As described before, there are many indications from prior literature about how reporting mechanisms could lead to increased work activity but there are also some indications that the policy may be counter productive. Our goal is to design an experiment which will enable us to examine all of these issues.

In the experiment, we simulate the relationship between a worker and an employer. The worker will essentially be making labor supply decisions in which they can choose how to allocate their time between work and leisure so as to maximize their utility. Our experiment will allow the worker to choose between spending their time on multiple work related tasks and a real leisure option so that it is clear that the worker receives positive utility from that leisure choice. The choice environment is designed such that we expect the worker to prefer to devote most of their time to leisure and that when choosing to work for the firm, the worker should spend all of that time on a task that requires substantial time commitment to generate a positive outcome. We will then conduct multiple treatments in which we vary the manner in which a worker will report their earnings on firm tasks to their employer and we will vary how frequently these reports are made. This will allow us to identify any changes in worker behavior due to frequency of reporting or to the nature of the reports.

## 2.1 Tasks

As in a standard labor supply model, our workers are endowed with a fixed time budget which they can choose to allocate between multiple options. In our experiment the time endowment is 300 seconds. The workers will experience this as actual time so they will spend 300 seconds on trials of the different tasks with the ability to switch between tasks as often as they like. Each trial of a task that a worker engages in has the possibility to result in a successful outcome or in a failure. The probability of a worker experiencing a success depends on how much time allocated to that specific trial with the probability of success increasing in the amount of time spent on a trial. The specific functional form for this success function is  $prob = 1 - e^{-\lambda t^2}$ , where  $t$  refers to the amount of time spent on the trial and  $\lambda$  captures the curvature of the curve.<sup>5</sup> We use different values of  $\lambda$  for different tasks as shown in Table 1 and Figure 1. The worker will choose how long to spend on a trial, stop that trial and find out the outcome and then start a new trial of whatever task

---

<sup>5</sup>The probability curve is convex when the amount of time is small and then concave as more time spent. This ensures the existence of an interior solution for the optimal amount of time on a trial.

they like until their 300 seconds expires.

The worker has three available tasks; the Firm Challenging Task, the Firm Standard Task and the Own Standard Task. The first is designed to mimic a task which has high value to a firm but which can involve substantial time investment to generate a successful outcome. The Firm Standard task is a more mundane project with middling rewards but that can generate successes in relatively short amount of time. The Own Standard Task is the real leisure option. As seen in Table 1, the Firm Challenging tasks pays 30 ECUs per success to both the Employer and the worker while the Firm Standard Task only pays 10. By examining Figure 1, you can see that the probability of success in the Challenging task is lower for any time expenditure than the Standard task, or rather to achieve any particular probability of success, more time is required in the Challenging task than the Standard task. These two tasks have been specifically constructed such that both the Worker and the Employer should prefer the worker choosing to work on the Challenging task.

The success function for the Own Standard Task is the same as the Firm Standard Task, simply to make the decision problem easier for the subjects. The earnings per success of the Own task are, however, not constant as they are for the Firm tasks. In the Own task, the earnings per success are decreasing with each success achieved. This function is shown in Figure 1. This element is included to satisfy the standard assumption in a labor supply model that there are diminishing returns from leisure and to guarantee an interior solution for how much time the worker should allocate to his or her various tasks.

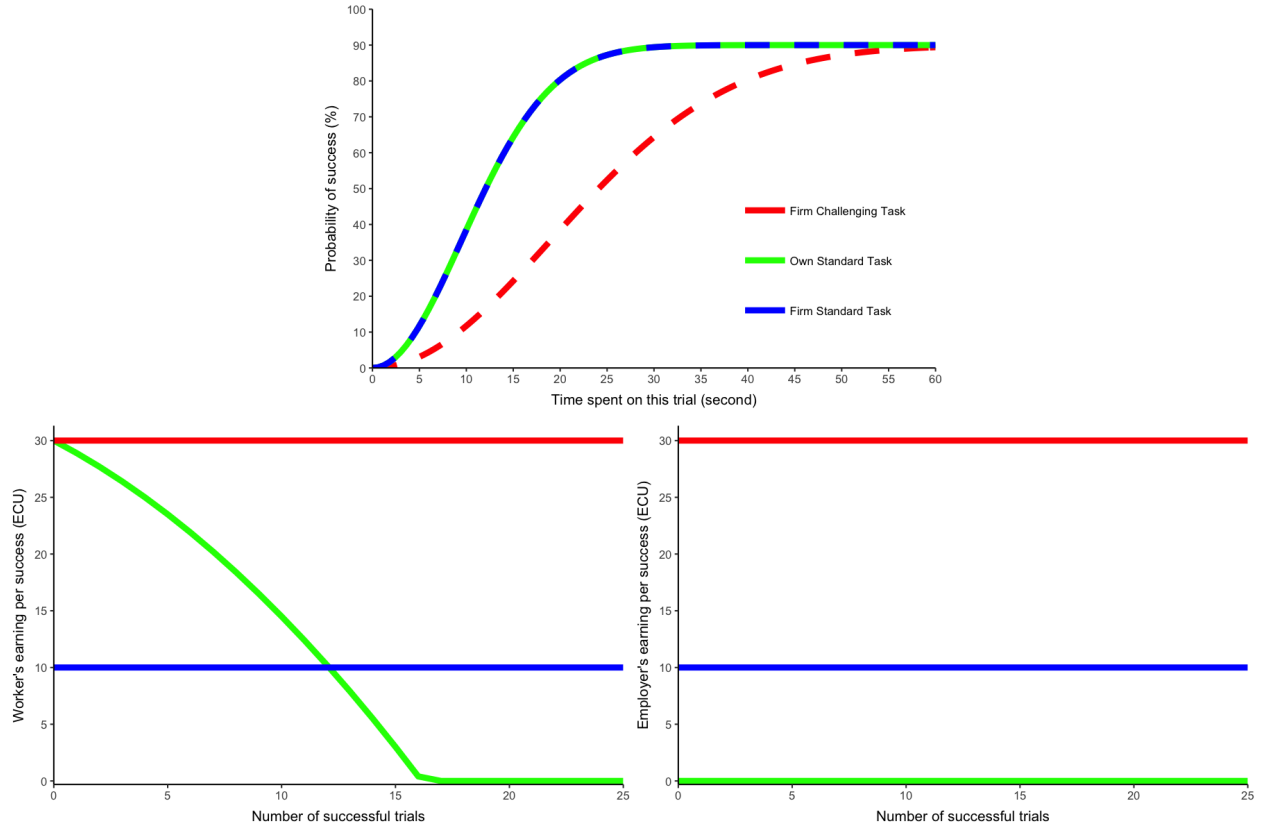
Table 1: Success function and earnings per success for each task

	Success Function	Earnings per Success	
		Employer	Worker
	$1 - e^{-\lambda t^2}$		
Firm Challenging Task	$\lambda_{fc} = 1.395 * 10^{-3}$	30	30
Firm Standard Task	$\lambda_{fs} = 5.582 * 10^{-3}$	10	10
Own Standard Task	$\lambda_{fs} = 5.582 * 10^{-3}$	0	$w_{os}(n_{os})$

To see how a worker should allocate time between tasks, we can look at Figure 2. The first



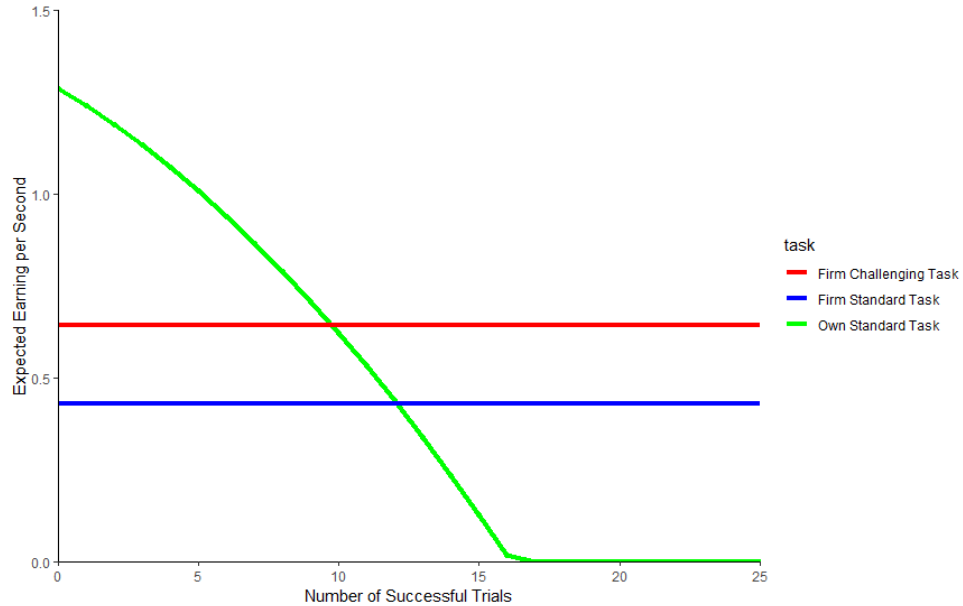
Figure 1: The Difference among Workers' Available Tasks



step in determining optimal allocation of time between tasks requires determining optimal time spent on each task per trial, since for any given amount of time on a task they need to maximize its expected earnings per second. For each trial of either the Own Standard Task or the Firm Standard task, the worker should spend 15 seconds. For each trial of the Firm Challenging task, the worker should spend 30 seconds, or double the amount of time as the Standard task. In all cases, this works out that the Worker will have a 64.37% chance of success for each trial regardless of which task it is from.

Given optimal time spent per trial, it becomes easy to see why the Firm Challenging Task dominates the Firm Standard Task. The Firm Challenging Task pays off three times the rate of the Standard Task but only requires double the time investment. As Figure 2 shows, this leads to the expected earnings per successful trial of the Firm Challenging task always dominating the expected earnings from the Firm Standard Task. For the Own Standard

Figure 2: The worker's expected earnings per second for different tasks



task, the initial per success earnings are the same as the Firm Challenging task but require half of the time investment meaning that it is clearly optimal to begin on Own tasks. With cumulative successes, the expected earnings from the own task drops off until the expected earnings are equal for the Own Standard task and the Firm Challenging task at 10 successful trials of the Own Standard task. After a worker has accumulated 10 successful trials in the Own Standard task, they should clearly devote any remaining time to the Firm Challenging Task. Following this strategy, a worker is expected to spend 233 out of 300 seconds on the Own Standard Task with the remaining 77 seconds being spent on the Firm Challenging task. This should generate an expected 281.61 ECUs to the worker and 43.11 ECUs to his employer.

This baseline theoretical prediction suggests that a standard worker should spend most of their time on their Own task leaving little time for work on tasks for the employer. The preferences of the employer are of course that the Worker would spend more time on the Firm Challenging task. Thus we have the tension faced in a normal workplace and can investigate how an employer might go about increasing the time the worker spends on the Firm Challenging Task. As noted, we will specifically investigate using different treatments

whether asking workers to report their earnings for the Firm can do that and if so what is the effect of different reporting frequencies and what behavioral channel is responsible for the affect.

## **2.2 Procedure**

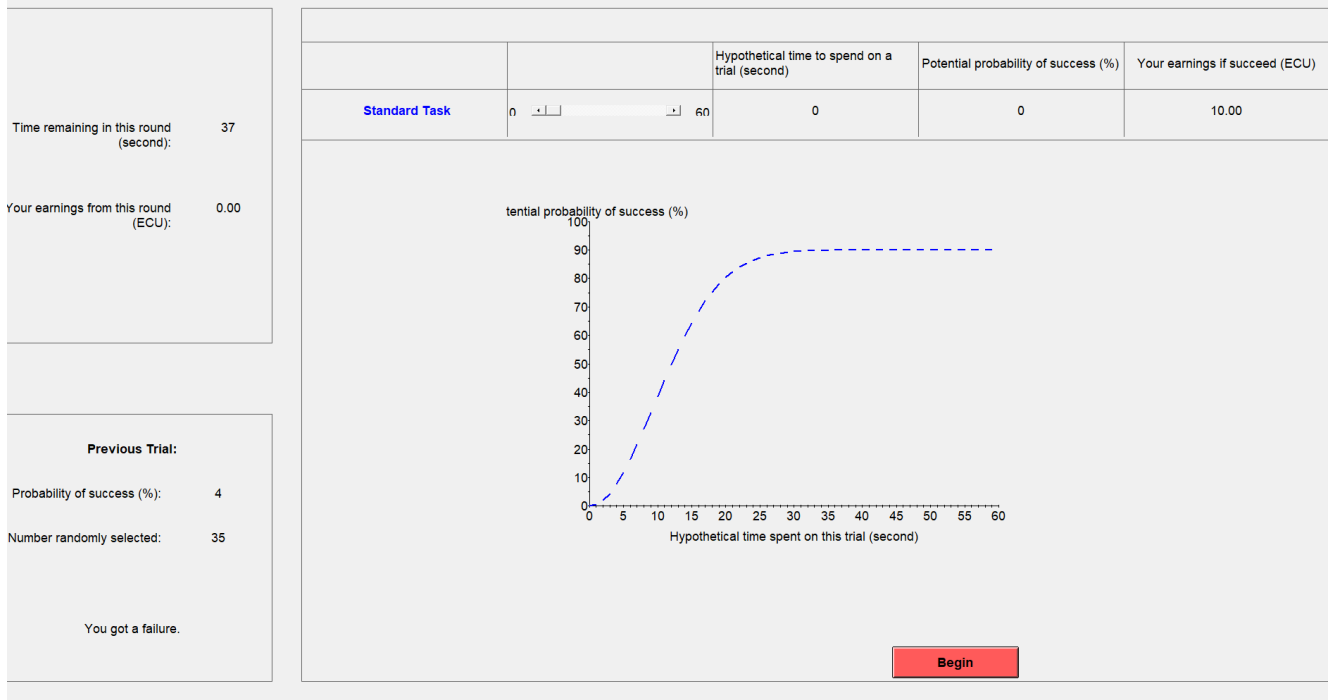
We have two phases in our experiment. The second phase is designed to investigate how the reports would affect workers' time allocation among these different tasks. However, all experiment sessions begin with an identical initial phase to get both workers and employers familiar with the tasks and to make sure that they understand the difference across the tasks. Prior to phase one, subjects are given complete instructions about it and are told that there would be a second phase and they would be given instructions for it after the first phase is complete. Their earnings from the experiment will be equal to the sum of the earnings from both phases.

### **2.2.1 Phase One**

There is no role assignment in the first phase, and all the subjects are treated identically. They act independently, and their actions only affect their own earnings. There are two rounds in the first phase and in each phase, a subject will engage in trials for a single type of task without the ability to choose among multiple tasks. In one round, the subjects will work on trials of the Challenging Task and the other round they will work on the Standard task with the order of the two randomized for each subject. These two tasks are the same as the firm tasks described before except they will only generate earnings to the subject as there is no employer in this phase. When a subject works on the Challenging Task and gets a success, she receives 30 ECUs; when she works on the Standard Task and gets a success, she receives 10 ECUs. In each round, subjects are given 300 seconds to be spent on trials of the assigned task.

In Figure 3, we have shown the sample screen where the subject begins a trial. Before

Figure 3: The screen worker a trial of the Standard Task



the subject starts a trial, she can explore its details on this screen. The success function for that task is shown in the middle. A slider in the table on the top right allows the subject to see how increasing or decreasing the hypothetical amount of time she might spend on the trial will affect her probability of success. As she moves the slider bar, there will be a dot tracking the curve. The subject can start a trial by clicking “Begin”. Then, she will be on another screen where she will spend actual time on the trial and her success probability is tracked along the success function as time progresses. She can submit the trial when she has achieved the success probability she desires. After the submission, the round timer will be paused and the outcome of the trial will be determined. She will be back on the screen in Figure 3 to start another trial. She will continue making these decisions until she uses up all 300 seconds for that round.

There are several purposes for this phase. Firstly, this phase gives subjects time to understand how much time they need to spend on a trial of a task to get certain chance of success and what earnings they would receive if they get a success. Secondly, the experience

with the two tasks in this phase helps subjects understand the difference among the tasks that will be available to workers in the second phase. Thirdly, this phase would help employers in the second phase form reasonable expectations on the earnings their workers can generate to them. The reason that this is important will become clear as we explain the reporting treatments in phase 2.

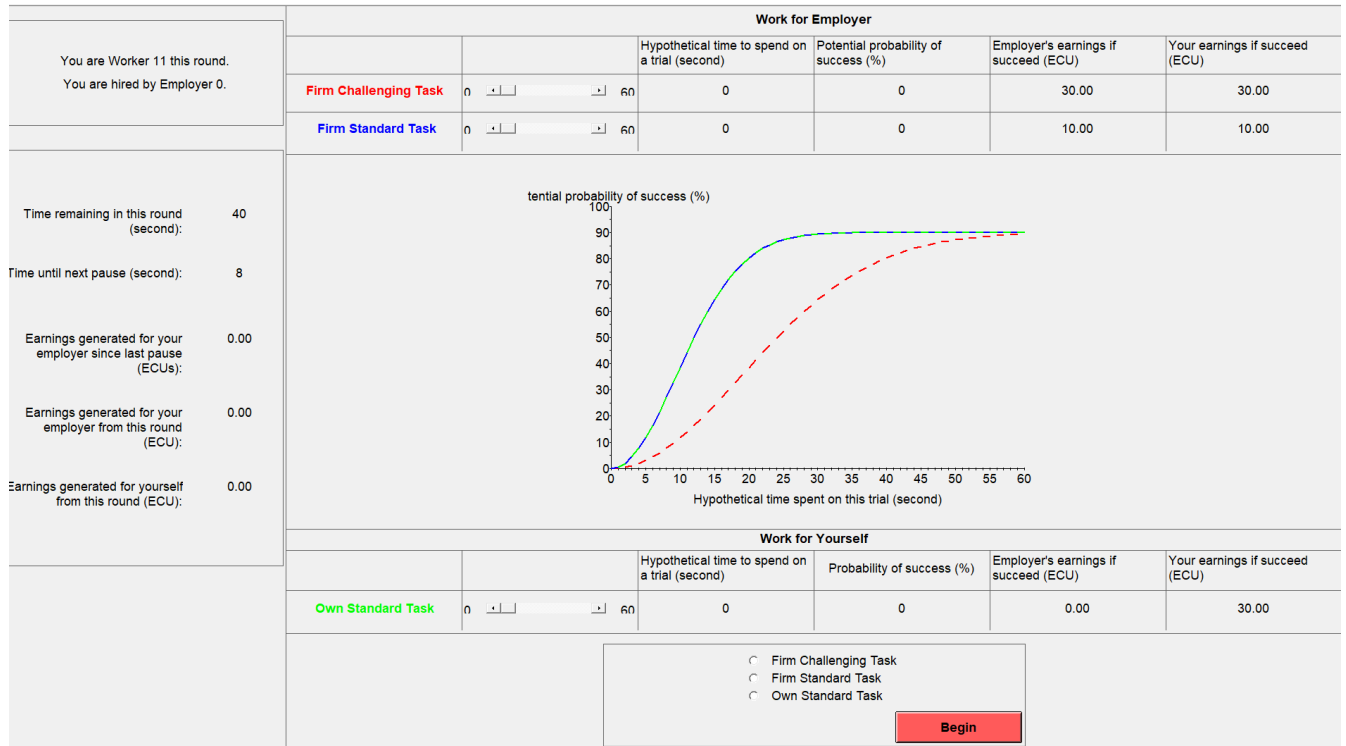
### **2.2.2 Phase Two**

In the second phase of the experiment, we assign half of the participants in an experiment session a role as an employer and the other half a role as a worker. Their roles are assigned at the beginning of the second phase and stay fixed throughout the second phase. There are four rounds in the second phase. At the beginning of each round, we randomly match every worker with an employer to establish the employment relationship. Every employer is required to pay a 10 ECU salary to her worker. We use this salary payment as a way of setting up the employment situation and to potentially activate a norm in which the worker should be working on behalf of their employer. Workers are re-matched to a new employer in every round. The only interaction between subjects in this phase is between each worker-employer pair.

Workers are endowed with 300 seconds (5 minutes) in each round to spend on the three available tasks, the Firm Challenging Task, the Firm Standard Task and the Own Standard Task. In the instructions for this phase, we explain the differences between these tasks to the subjects by showing Figure 1 on their computer screen and explaining the characteristics of each task. As they have just spent phase 1 performing trials of the two Firm tasks, the main elements to explain to them are the Own Standard task and how they will be able to choose among tasks.

In Figure 4, we have shown a sample of the screen where a worker will choose between the three available tasks and then start a trial of the chosen task. The graph and tables on the right of the screen can help the worker examine the potential earnings and probability

Figure 4: The screen worker chooses the task and starts a trial of the chosen task



of success of a trial for each of the three tasks. In those tables, the worker can see not only his earnings from a success but also his employers earnings. The last column of the earnings table shows how much the worker can earn from the next success of the Own Standard task given how many successes he already achieved from that task so far in the current round. This is the key information the worker needs in making his choice among tasks. The worker can then choose a task and start a trial of the chosen task on the bottom of the screen. He will move to the screen where he spends time on the trial, which will be a screen almost identical to the one used in the first phase. As before, a worker can wait until the success probability hits the desired level and submit the trial. After the submission, he will be back on the screen in Figure 4 to observe the outcome and then choose a task for which to start another trial. He will continue making these decisions until he uses up all 300 seconds for that round.

Employers will not be sitting idly but instead they will also be working on a similar

basis. We have them work on two tasks, known as Employer Challenging Task and Employer Standard Task. These two tasks have the same success functions as the tasks the worker will engage with but they generate much lower earnings. Employers only receive 3 ECUs per success from the Employer Challenging Task and 1 ECU from the Employer Standard Task. Workers receive no earnings from the success of their employer. The point of this activity is simply to give the employer some activity to engage in to prevent boredom and to prevent workers from identifying who in the room is an employer. Given that they will not generate much in earnings from their own work, this also makes it clear that the employers are highly dependent on the workers to generate their earnings. Employers are not allowed to choose between the tasks but rather which task they work on is randomly determined after each trial they submit.

## **2.3 Treatments**

The first intervention we perform to determine its impact on worker effort is that we have the workers report on the earnings they have generated for the firm since their last report. To examine how the frequency of reporting affects behavior, we vary the frequency of reporting across rounds. Workers will be asked to report 1, 2, 5 or 10 times in a round. The ordering of these four reporting frequencies across the four production rounds is randomized across workers. A worker reporting once, implies that they only report at the end of the production round. If a worker is asked to report twice, then he will report after he has spent 150 seconds working and then report again after he has spent 300 seconds. The time spent on the reports is excluded from the 300 seconds. This removes the penalty to productivity that would come from extra time spent on the reporting itself. We do not believe such a reporting cost to be negligible in actual workplace settings, rather we wish to remove it here just to focus on the marginal effects of the reporting itself. The only information employers observe about worker productivity comes through this reporting mechanism.

In the introduction we explained that the expected impact of reporting frequency on

employee effort is ambiguous based on prior research. Firms obviously use this reporting mechanism as an attempt to increase productivity but we noted several prior studies which cast doubt on whether that would occur. In particular, there is evidence that frequent monitoring could be seen as distrustful behavior by an employer which could de-motivate a worker. Further, if asked to report too frequently, this could shift behavior towards the Firm Standard task. If you consider the case of reporting 10 times, this means reporting every 30 seconds. That allows for only one trial of the challenging task but two from the standard task. This means that the probability of achieving a success between reports is quite low for the challenging task. A worker concerned about reporting a lack of results may well choose to spend time on the standard task as reporting windows narrow to increase the probability of achieving at least some positive results between reports. These claims form the basis for our first two hypotheses.

**Hypothesis 1. (*Frequency - Total Effect*)** *Increased frequency of reports will lead to an increase in the amount of time spent on the firm tasks.*

**Hypothesis 2. (*Frequency - Substitution Effect*)** *Increased frequency of reports will lead to a worker substituting trials of the standard task in place of the challenging task.*

The next issue we wish to investigate is what aspect of the reporting mechanism drives any increase in effort. While one might expect that having workers report and receive feedback would achieve an increase in worker effort, it isn't clear what aspect of the process is responsible for such a reaction. It is possible that the worker possesses intrinsic motivation for effort and simply reminding them of their production so far would spur them on to greater work effort. On the other hand it is possible that an individual only has such concerns when they know the affected party can observe what they have done and so perhaps simply knowing that the employer will view the results will lead to a productivity increase. Or, perhaps the only way to generate a response is for the worker to know that the employer will not only observe but also be able to provide feedback. We will construct four different reporting



mechanisms that will allow us to differentiate between these channels. Each session of the experiment will use only one of these reporting treatments and it will hold for all of phase 2.

1. *Baseline:* Workers do not report. This is the baseline of the experiment. As a control for the other treatments, workers and employers are asked to pause for several seconds 1, 2, 5 or 10 times per round. The number of seconds of the pause is determined by the average time taken in the reporting stage in the Feedback treatment which is the longest among the three treatment. These pauses are included to help identify that any effort change observed between this treatment and the others is due to the reporting element and not the interruption in work activity caused by the reporting.
2. *Intrinsic Motivation:* Workers are required to acknowledge the earnings they have generated for their employers, but their employers do not see any information about worker productivity. Workers are told this explicitly that while they will see reports on their productivity, their employer will not. Consequently, all the reporting element of this treatment consists of is to remind a worker how much money they have generated for their employer over the previous reporting period.
3. *Observability:* Workers are required to report the earnings they have generated for their employers, and these reports will be sent to and seen by the employers. Again, it is clearly explained to workers that when they send in an earnings report, their employer will see it but the employer has no way of communicating with the worker or responding to the report. This reporting mechanism achieves common knowledge between worker and employer regarding the employer earnings generated by the worker meaning the worker can expect that their employer may be judging them, but that judgment cannot be communicated to the worker. This element indicates why it was necessary for employers to participate in phase 1 as doing so should allow them to get some idea of how much earnings it might be reasonable or possible for the worker to produce for the employer.

4. *Feedback*: This is the full reporting regime in which workers send in their earnings to the employer and the employer can respond with feedback. For feedback, the employer can choose from a scale of 1 to 7 to express their level of satisfaction with the achieved earnings. On this scale, 1 represents strongly dissatisfied while 7 means strongly satisfied. In this treatment, workers might care about not only their employers' judgment but also the feedback sent by employers.

Each of these treatments was designed to test why having employees report might affect their productivity. We have previously discussed the possibility that individuals could possess intrinsic motivation, could be concerned about judgment or that they might be concerned about potentially negative feedback. Of course it is also possible that individuals are concerned about all three and each of these effects could add on to each other. By conducting the baseline and then these three other treatments, we can separate between these issues. This leads to our next three hypotheses.

**Hypothesis 3. (*Intrinsic Motivation*)** *Having workers acknowledge the earnings generated to their employers will increase the time spent on firm tasks compared to when they are not asked to acknowledge those earnings.*

**Hypothesis 4. (*Observability*)** *Having workers report the earnings generated to their employers and know that the report will be seen by their employers will increase time spent on the firm tasks compared to the non-reporting case.*

**Hypothesis 5. (*Feedback*)**: *Having workers report the earnings generated to their employers and know that the report will be seen and commented on by their employers will increase time spent on the firm tasks compared to the non-reporting case.*

## 2.4 Implementation

All of our experiment sessions were conducted at Southern Methodist University. Subjects were recruited from a university-wide subject pool using a computerized recruitment system

Table 2: Experimental Design and Data Points

	Sessions	Subjects
Baseline	2	22
Intrinsic Motivation	3	36
Observability	3	32
Feedback	3	32

based on h-root(Bock et al., 2014). The pool consists of a mix undergraduate and graduate students who had indicated a willingness to be paid volunteers in decision-making experiments. The interactive software system is programmed using z-Tree(Fischbacher, 2007). Subjects’ final payoff is equal to a \$10 show-up fee plus the sum of earnings from both phases. For each phase, we randomly select one round for payment. We translate ECUs into dollars at the rate of 20.00 ECUs = \$1.00. Payments including the show-up fee ranged from a minimum of \$15.59 to a maximum of around \$39.2, with an average of \$26.0. Each session lasted an hour and a half to two hours.

## 3 Experiment Results

### 3.1 Overall View

We will begin presenting the results of the experiments by providing a set of summary statistics to provide an overview of the data. We will not conduct tests on these simple summary statistics as these tests are mis-specified given the nature of the data. Formal tests of the hypotheses will be conducted using properly specified regressions in the next section but having an understanding of these summary statistics can be helpful in properly interpreting the regressions.

An initial question to examine in the data is the degree to which subjects could solve the basic problem of figuring out how much time to spend on a trial of each task. Table 3 shows the average time the subjects spend on a trial by type of task in both phases of the experiment and indeed on average the subjects spent almost the exact optimal amount of

Table 3: Average time spent on a trial by task in both phases.

	Phase One	Phase Two	
		Employer	Worker
Challenging Task	26.92	25.21	29.82
Standard Task	15.95	14.56	15.94
Own Standard Task	-	-	16.67

Notes: standard errors clustered at the individual subject level. p-value in parentheses.  
 \*\*\*  $p < 0.01$ . \*\*  $p < 0.05$ . \*  $p < 0.1$ .

time per trial or 15 seconds per trial on the standard task and 30 seconds per trial on the challenging task. Of course, while on average they chose correctly, this doesn't mean all choices were exactly at the optimal. Figure 5 and Figure 6 show the density plots of time spent on trials to reveal the full distribution. These again show that the mean is very close to the prediction with roughly normal distributions around that mean. Establishing that subjects on average figured out the correct amount of time to spend per trial is useful to make it clear that they understood the relevant incentives and the differences between the tasks. One point to note is that the largest errors as indicated by the plots are for those individuals choosing the standard task in Phase 2. For this group, the average is a little under the optimal choice and there is a fair amount of spread to the choices. Since these subjects are already making a mistake by choosing this task, it stands to reason that they would also make the most mistakes in regard to time spent per trial.

One of the main research questions of the paper is the degree to which different reporting frequencies might impact behavior and the summary statistics breaking down work time allocation by frequency is shown in Table 4. Table 5 then shows similar summary statistics by reporting treatment. As a reminder, the worker is endowed with 300 seconds to be divided among three available tasks. If the worker only cares about his own earnings, he is expected to spend only 67 seconds on the Firm Challenging Task with the rest of the time, 233 seconds, spent on the Own Standard Task. A worker should spend no time on the Firm Standard Task. From these tables, we found the workers on average spend more than 150 seconds on the two firm tasks combined, which is a bit more than twice prediction.

Figure 5: Average time spent on a trial for the two tasks in phase one

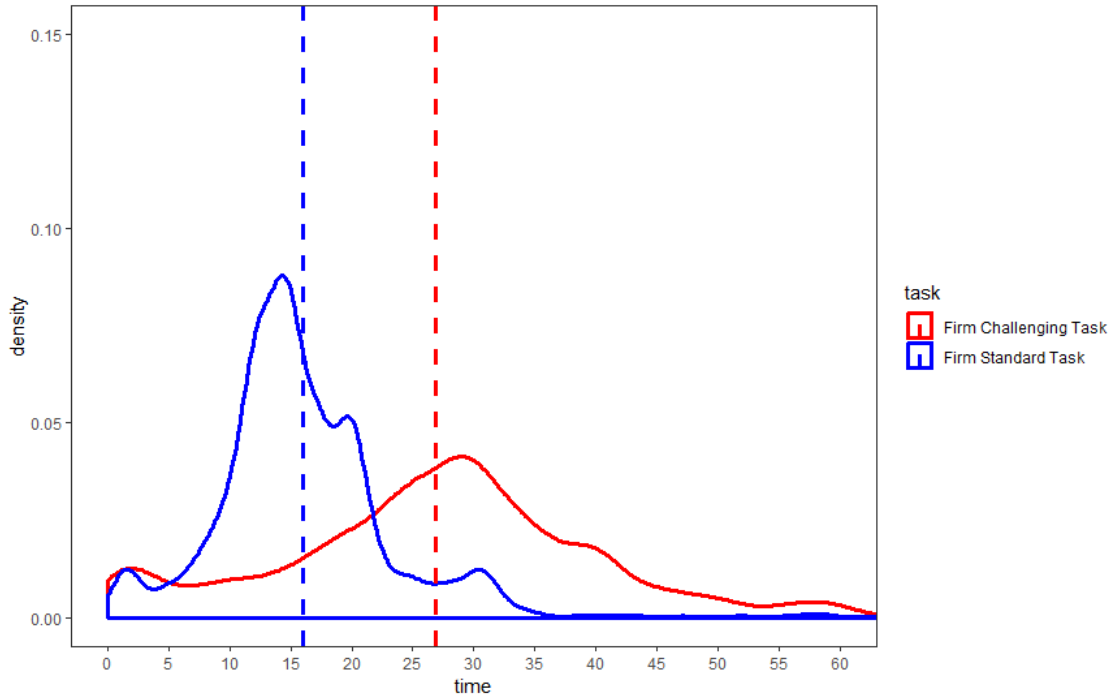


Figure 6: Average time spent on a trial for the workers' three tasks in phase two

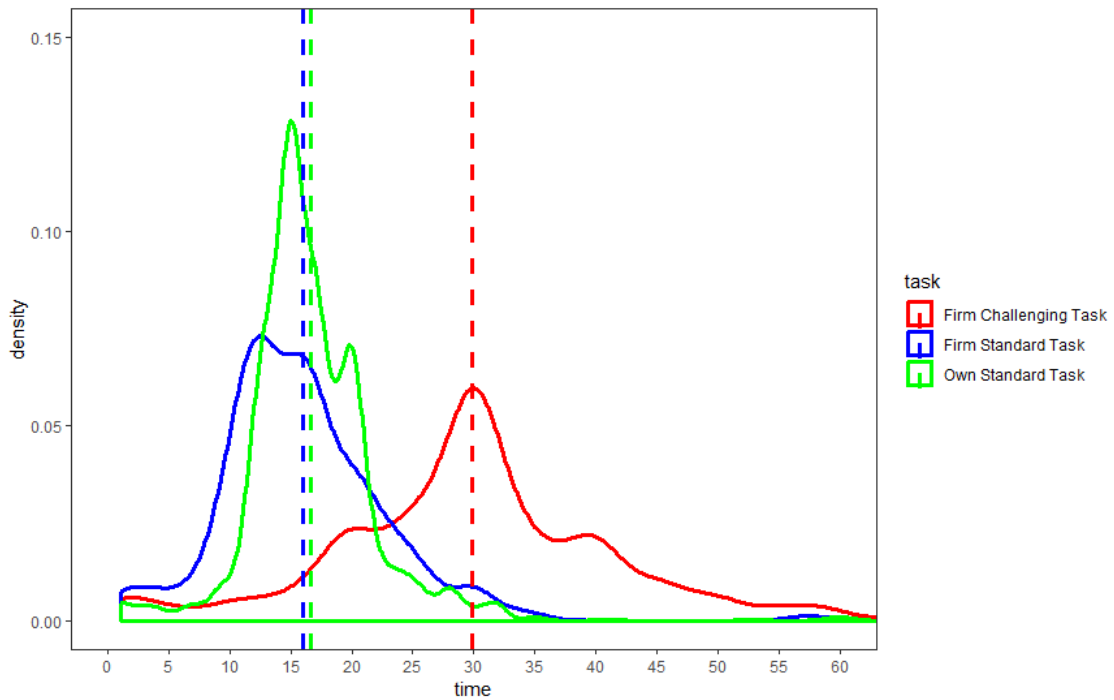


Table 4: Average time allocation of workers by frequency.

	Firm	Firm Chal	Firm Std
Freq = 1	171.52	154.56	16.97
Freq = 2	182.38	159.92	22.36
Freq = 5	196.61	175.38	21.23
Freq = 10	184.41	148.07	36.34

Notes: Firm refers the total time spent on the two firm tasks. Firm Chal refers the Firm Challenging Task. Firm Std refers to the Firm Standard Task. Own Std refers to the Own Standard Task.

Table 5: Average time allocation of workers by treatment.

	Firm	Firm Chal	Firm Std
Baseline	175.31	148.40	26.92
Intrinsic Motivation	188.71	166.89	21.82
Observability	168.74	151.31	17.33
Feedback	197.33	158.25	39.08

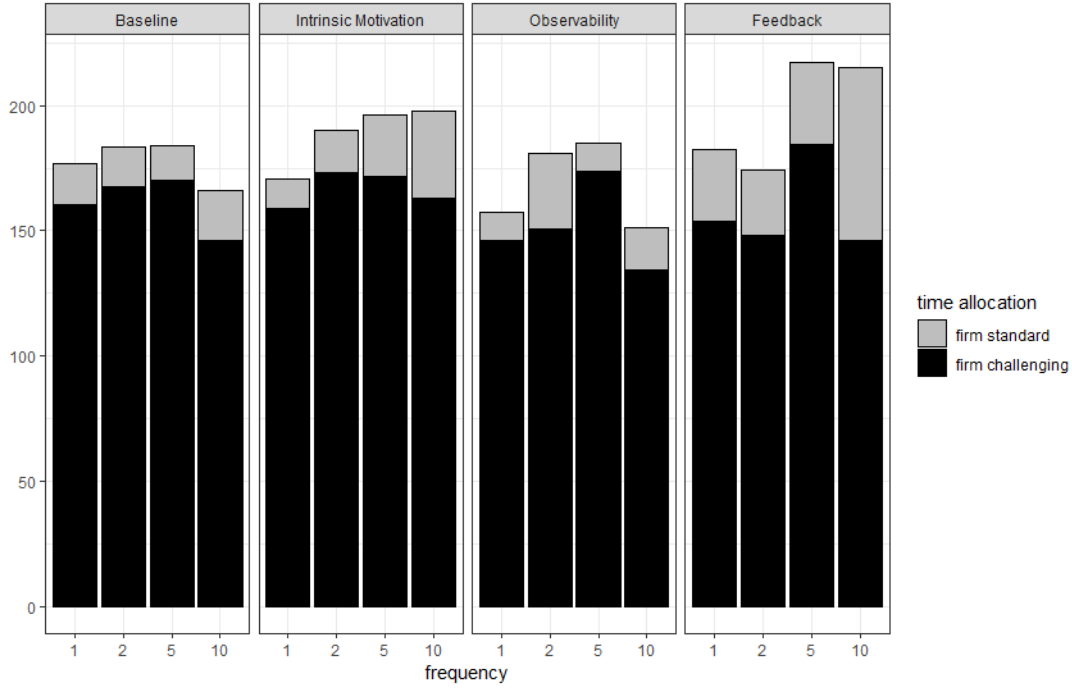
Notes: Firm refers the total time spent on the two firm tasks. Firm Chal refers the Firm Challenging Task. Firm Std refers to the Firm Standard Task. Own Std refers to the Own Standard Task.

This could indicate a willingness on the part of the workers to sacrifice their own welfare to generate earnings for the employer. We also find that while they should spend no time on the Firm Standard Task, workers actually spend around 10% of their time on this task despite it being dominated by the Challenging Task with respect to earnings to the worker and the employer. We will examine possible reasons for this behavior in the next section.

If we look at how worker behavior varies with the reporting frequency, on average there seems to be an increase in time spent on firm tasks as frequency rises up to a peak at the 5 report frequency. Going all the way up to 10 reports seems to pull the total back down while also leading to a substantial increase in time spent on the Firm Standard task. Looking next at the data broken down by reporting treatment shows that there could be an effect from the Feedback treatment of increasing total time spent on firm tasks relative to the no reporting treatment but neither the Intrinsic Motivation or Observability treatments seems to have improved worker effort by much. This is quite surprising given all the prior work demonstrating that in other contexts, interventions like this seem to have been enough to shift behavior.

Figure 7 shows another look at this data as it presents the time allocation split out by both

Figure 7: Time allocation for different settings and different frequencies



reporting treatment and by reporting frequency. Again, it looks like the most substantial productivity improvement was for the 5 report regime in the Feedback treatment and there is a general trend in all reporting treatments of moderate productivity improvements up to the 5 report condition followed by a decline in the 10 report condition.

### 3.2 Effect of Reporting Frequency on Time Allocation

We will now present a series of regressions to examine each of our hypotheses in order beginning with the hypotheses concerning the effect of changing the frequency of reporting. For an initial test, we will examine how reporting frequency might affect behavior differently between the baseline treatment without reporting and all of the reporting treatments pooled together. Table 6 contains a set of random effects panel regressions with standard errors clustered at the subject level to examine how worker behavior might vary with the frequency of reports. We examine in one specification their choice of total time on firm tasks and then we break that out into two other specifications for the time spent on the challenging task

Table 6: Test how the frequency of reports/pauses affect workers' time allocation in the **Baseline** and **Non-Baseline Treatments**.

	Baseline			Non-Baseline		
	Firm	Firm Chal	Firm Std	Firm	Firm Chal	Firm Std
Cons.(Freq = 1)	227.5*** (0.000)	182.9*** (0.000)	44.5*** (0.009)	200.6*** (0.000)	161.7*** (0.000)	38.8*** (0.000)
Freq = 2	9.90 (0.302)	8.96 (0.437)	0.95 (0.908)	12.2* (0.072)	5.04 (0.541)	7.14 (0.176)
Freq = 5	-5.02 (0.791)	4.38 (0.818)	-9.40 (0.104)	29.6*** (0.009)	23.4* (0.051)	6.20 (0.212)
Freq = 10	-19.6 (0.377)	-18.2 (0.409)	-1.36 (0.892)	20.2* (0.071)	-4.19 (0.721)	24.3** (0.016)
Round	-18.5* (0.058)	-8.27 (0.323)	-10.2** (0.013)	-12.4*** (0.000)	-3.46 (0.421)	-8.94*** (0.002)
<i>N</i>	44	44	44	200	200	200

Notes: standard errors clustered at the individual subject level. p-value in parentheses.

\*\*\*  $p < 0.01$ . \*\*  $p < 0.05$ . \*  $p < 0.1$ .

and time spent on the standard task. These regressions provide support for our first result.

**Result 1. (*Frequency Total Effect*)** *Increasing the frequency of reports generates a statistically significant increase in the amount of time spent on firm tasks with a maximum effect at the reporting frequency of 5.*

As Table 6 shows, for the non-baseline treatments or the treatments where subjects are actually reporting to one degree or another, we find that all of the indicator variables for each frequency of reporting to be at least marginally significant in the regression examining total time spent on the firm tasks. The coefficient on the 5 report condition is the largest. The  $\chi^2$  tests show that this coefficient is significantly different than the coefficients on the 2 report condition ( $Prob > \chi^2 = 0.0562$ ) while not significantly different from the 10 report condition ( $Prob > \chi^2 = 0.3722$ ). Thus productivity does increase as hypothesized though there does not seem to be a positive effect at very high levels of reporting

Of course it is possible that this effect is simply due to subjects being asked to pause during their production phase. This possibility can be eliminated by examining the results of this same regression on the Baseline data. In that regression we find the indicator variables for frequency of pauses to be insignificant. Thus it is clear that it is some aspect of the



monitoring combined with the increased reporting frequency that leads to the performance improvement. What is not clear though is which aspect of reporting is required to generate the effect.

Our next hypothesis deals with a possible negative aspect of frequent reporting and that is the possibility that the worker might substitute away from the more lucrative challenging task for the standard task which is more likely to give them positive results that they can include in their reports over short time intervals. Our next result finds in favor of this hypothesis.

**Result 2. (*Frequency - Substitution Effect*)** *Increasing reporting frequency past 5 reports leads to workers decreasing their time spent on the Firm Challenging Task while increasing the time spent on the Firm Standard Task indicating that at high frequencies they are substituting time spent on the Challenging Task for time spent on the Standard Task.*

To examine this we can look at the fifth and sixth columns of Table 6 showing how the time spent on the two types of tasks changes as reporting frequency increases. The key is that at a reporting frequency of 5, the coefficient on the challenging task is large and significant while the coefficient on the 10 report period is actually negative though insignificant. For the standard task regression, we find that this relationship is essentially reversed with the coefficient on the 5 report condition being insignificant but the coefficient on the 10 report condition is large and significant. It is in fact essentially the same as the coefficient as the one for the 5 report condition on the challenging task regression. This shows clear evidence that between the 5 and 10 report conditions, the workers are substituting their effort towards the standard task and away from the challenging task as hypothesized.

Result 1 establishes that increased frequency of reporting can increase effort on firm tasks but it does not tell us which reporting regime or regimes are necessary to deliver that effect. Table 7, provides similar regressions on worker time allocation but this time all of the reporting treatments are considered separately to identify which of them deliver the effect. What we find is that each treatment is delivering behavior shifts in the manner we identified

Table 7: Test how the frequency of reports affect workers' time allocation in **Each of the Non-Baseline Treatments**.

	Intrinsic Motivation			Observability			Feedback		
	Firm	Firm Chal	Firm Std	Firm	Firm Chal	Firm Std	Firm	Firm Chal	Firm Std
Constant	229.7*** (0.000)	181.5*** (0.000)	48.2*** (0.008)	167.4*** (0.000)	137.3*** (0.001)	30.0** (0.024)	190.7*** (0.000)	156.2*** (0.000)	34.6 (0.144)
Freq = 2	16.7 (0.135)	13.5 (0.162)	3.27 (0.683)	23.1* (0.060)	4.82 (0.689)	18.3*** (0.008)	-7.48 (0.472)	-5.14 (0.795)	-2.34 (0.853)
Freq = 5	27.7 (0.125)	13.8 (0.523)	13.9 (0.248)	28.4 (0.190)	27.1 (0.205)	1.30 (0.691)	34.4 (0.118)	30.6 (0.185)	3.83 (0.622)
Freq = 10	12.0 (0.380)	-1.51 (0.948)	13.5 (0.426)	0.87 (0.954)	-18.0 (0.166)	18.8** (0.012)	32.1 (0.232)	-7.68 (0.714)	39.8 (0.120)
Round	-22.0*** (0.008)	-8.43 (0.417)	-13.6** (0.027)	-4.73 (0.326)	4.20 (0.401)	-8.93*** (0.001)	-3.27 (0.498)	-0.95 (0.890)	-2.32 (0.671)
<i>N</i>	72	72	72	64	64	64	64	64	64

Notes: standard errors clustered at the individual subject level. p-value in parentheses.  
\*\*\*  $p < 0.01$ . \*\*  $p < 0.05$ . \*  $p < 0.1$ .

in the pooled data but due to substantial variance in the data, few of these effects manage to obtain significance. The effects are closest to significance in the Observability and Feedback treatments, but even there p-values are in the 10-20 percent range. The indication here is that we are finding relatively weak effects in each treatment which only pass the significance test in the pooled data. This is an interesting finding to which we will return later.

An important consequence of the workers choosing to spend more time on firm tasks as reporting frequency rises is that they should expect to make more money for the employer but less for themselves. We can examine these effects as a way of determining how consequential are the shifts in time allocation that we observe. Table 8 presents random effects panel regressions with standard errors clustered at the subject level of both employer and worker earnings in the baseline and non-baseline data subsets. While the shift in worker behavior is enough to harm their own expected earnings at the 5 and 10 reporting frequency levels, the effect on the employer is not quite significant for any of the reporting levels. Of course it is important to note that what we were investigating here is whether reporting leads to a worker being willing to sacrifice their own welfare to work more for their employer. We find that the workers are actually sacrificing up to 10 percent of their earnings for their employer. Given the variability in the actual task, this doesn't necessarily translate into substantial

Table 8: Test how the frequency of reports/pauses affect the earnings (ECUs) workers generate to their employers and to themselves in the **Baseline** and **Non-Baseline Treatments**.

	Baseline		Non-Baseline	
	Employer	Worker	Employer	Worker
Constant	146.4*** (0.000)	241.7*** (0.000)	109.5*** (0.000)	209.2*** (0.000)
Freq = 2	4.49 (0.816)	-13.4 (0.540)	11.5 (0.117)	0.95 (0.905)
Freq = 5	-6.61 (0.789)	-2.40 (0.882)	11.3 (0.132)	-17.3* (0.058)
Freq = 10	-24.9 (0.230)	-23.9 (0.198)	0.80 (0.922)	-20.1** (0.033)
Round	-14.7* (0.095)	-1.48 (0.810)	-3.72 (0.166)	8.56*** (0.007)
<i>N</i>	44	44	200	200

Notes: standard errors clustered at the individual subject level. p-value in parentheses.  
 \*\*\*  $p < 0.01$ . \*\*  $p < 0.05$ . \*  $p < 0.1$ .

earnings increase for the employer but that is simply due to the parameterization of this task. The key to the finding is that requesting more frequent reports can indeed induce workers to give up their own welfare to work more for the firm. Of course in the 10 report case, the workers are giving up both their own and their employer's potential earnings due to the fact that they are shifting to time spent on the wrong task.

### 3.3 Effect of Reporting Treatments on Time Allocation

We now turn to examining the effect of the different reporting regimes on worker behavior to try to identify what aspect of the reporting task can lead to increases in worker effort. We again conduct a series of random effects panel regressions with standard errors clustered at the subject level with the dependent variables being time spent on firm tasks in total and then split out to the two types of tasks but this time with independent variables being the reporting regime. In these regressions, we are comparing the time allocation in each of these reporting treatments to the baseline without reporting. These regressions are shown in Table 9 and provide the basis for our next three results.

Table 9: Test how the nature of the reports affects workers' time allocation.

	Firm	Firm Chal	Firm Std
Constant	209.9*** (0.000)	171.7*** (0.000)	38.2*** (0.000)
Intrinsic	11.1 (0.734)	5.87 (0.867)	5.23 (0.552)
Observability	-8.97 (0.798)	-9.71 (0.796)	0.74 (0.944)
Feedback	19.7 (0.543)	-2.77 (0.937)	22.5* (0.071)
Round	-12.9*** (0.000)	-4.27 (0.258)	-8.64*** (0.000)
<i>N</i>	244	244	244

Notes: standard errors clustered at the individual subject level. p-value in parentheses.  
 \*\*\*  $p < 0.01$ . \*\*  $p < 0.05$ . \*  $p < 0.1$ .

**Result 3. (*Intrinsic Motivation*)** *Having workers acknowledge the earnings generated to their employers does not lead to a statistically significant increase in the time spent on firm tasks compared to when they are not asked to acknowledge those earnings.*

**Result 4. (*Observability*)** *Having workers report the earnings generated to their employers and know that the report will be seen by their employers does not lead to a statistically significant increase in time spent on the firm tasks compared to the non-reporting case.*

**Result 5. (*Feedback*)** *Having workers report the earnings generated to their employers and know that the report will be seen and commented by their employers does not lead to a statistically significant increase in time spent on the firm tasks compared to the non-reporting case.*

What we find is that all of these treatment coefficients are not significantly different from 0. This might be considered puzzling given the prior results showing that reporting frequency can improve worker effort. What we are essentially observing here is that the reporting regime on it's own is not sufficient to drive effort but interacting it with frequent reporting can do so. In these regressions we are pooling all of the reporting frequencies and given that we found offsetting results for the 5 and 10 report frequencies and small effect at

the 2 report frequencies, this is perhaps less surprising than it more otherwise seem. What these regressions show is that there is not on overall level effect of the reporting regime that occurs at any and all reporting frequencies. One should not conclude that this means that the different types of reporting are ineffective at generating increased worker effort. Rather the indication is that one has to be careful in pairing the reporting regime with the right frequency to get the desired effect.

### **3.4 Effect on Timing and Trials**

In addition to the main questions regarding how increased reporting frequency might impact overall effort provision, there are also important questions regarding other aspects of worker effort that might be impacted by reporting. One of these important questions has to do with the timing with which firm effort is provided. Theoretically, a worker should work on the Own Task for most of the time and then towards the end switch to the Firm Challenging Task. In addition to increasing the amount of time spent on the Firm Challenging Task, a firm might also want to have the worker start on that task earlier in the production period. This might be one important impact of asking for interim reports. While given the production function in this experiment, the timing of the effort isn't important, the timing might be important in many field situations and so it is worth looking at whether these reports can get workers to start on their firm tasks earlier.

In order to see the effect of reports on the timing of tasks, we divide each 300 second round into ten 30-second. We then conduct a series of regressions examining how the amount of time a worker spends on firm tasks is affected by the frequency of monitoring in each of these 10 intervals separately. What these regressions will show is whether any of our frequencies of monitoring increased the time spent on the firm tasks in each of these 10 intervals. Table 10 contains a set of these random effects panel regressions with standard errors clustered at the subject level to examine how the workers' time investment in firm for each interval affected by the changing frequency of reports. What we find is that for the 5 and 10 report conditions,

Table 10: Test how the frequency of reports affect workers' time on firm tasks in each 30-second interval

	1	2	3	4	5	6	7	8	9	10
Cons(Freq=1)	20.0*** (0.000)	19.1*** (0.000)	19.9*** (0.000)	15.8*** (0.000)	17.9*** (0.000)	21.3*** (0.000)	23.8*** (0.000)	21.7*** (0.000)	21.1*** (0.000)	20.0*** (0.000)
Freq=2	-1.38 (0.411)	0.80 (0.579)	0.84 (0.513)	0.88 (0.501)	0.83 (0.592)	-0.47 (0.759)	0.61 (0.735)	1.96 (0.340)	4.68** (0.020)	3.43 (0.112)
Freq=5	2.00 (0.277)	4.54*** (0.010)	3.52* (0.054)	5.54*** (0.004)	2.45 (0.209)	2.01 (0.251)	-0.17 (0.927)	1.84 (0.291)	3.64** (0.026)	4.25** (0.029)
Freq=10	3.25** (0.024)	4.97*** (0.002)	2.48 (0.143)	5.01*** (0.005)	2.26 (0.205)	-0.86 (0.644)	-1.45 (0.485)	0.50 (0.781)	2.66 (0.246)	1.36 (0.517)
Round	-1.95*** (0.000)	-1.43** (0.013)	-1.48*** (0.008)	-0.92 (0.107)	-0.75 (0.177)	-1.24** (0.032)	-1.81*** (0.002)	-0.98 (0.117)	-1.10 (0.150)	-0.73 (0.255)
<i>N</i>	200	200	200	200	200	200	200	200	200	200

Notes: standard errors clustered at the individual subject level. p-value in parentheses.  
\*\*\*  $p < 0.01$ . \*\*  $p < 0.05$ . \*  $p < 0.1$ .

there is a significant effect in several of the first 4 intervals while there is no systematic impact in later intervals. This indicates that the overall time increase on firm tasks we observed previously is primarily coming in the first half of the production period. Thus there is some indication that increased reporting frequency can induce workers to begin their work on firm tasks than without reporting or with lower frequency of reporting.

We can engage in the same exercise to determine if there is a systematic effect of the type of reporting mechanism in use on the timing of effort provision. It isn't clear that this should have an impact separate from the frequency of the reports but we can still examine the issue. Table 11 contains a set of random effects panel regressions with standard errors clustered at the subject level to investigate this. The dependent variables of these regressions are the same as those in Table 10, but the explanatory variables here are dummy variables for reporting treatments. What we find is that all of these treatment coefficients are not significantly different from 0. Thus the reporting regime alone does not lead to any time shifting absent the effect from the monitoring frequency.

A final issue to examine is another possible negative effect of frequent monitoring and this is the possibility that frequent monitoring could cause workers to become impatient or sloppy on their individual trials by submitting them too soon. Given that we've shown

Table 11: Test how the report regimes affect workers' time on firm tasks in each 30-second interval.

	1	2	3	4	5	6	7	8	9	10
Cons(Baseline)	16.8*** (0.000)	19.3*** (0.000)	21.2*** (0.000)	21.2*** (0.000)	19.6*** (0.000)	21.4*** (0.000)	24.2*** (0.000)	22.6*** (0.000)	22.7*** (0.000)	20.8*** (0.000)
Intrinsic	5.57 (0.208)	2.78 (0.526)	1.80 (0.694)	-0.66 (0.875)	-0.35 (0.924)	1.19 (0.755)	-1.81 (0.645)	-0.53 (0.885)	1.05 (0.787)	2.05 (0.627)
Observability	-0.21 (0.965)	0.18 (0.971)	0.095 (0.984)	-3.85 (0.386)	-0.90 (0.815)	-0.35 (0.928)	-2.88 (0.455)	-1.49 (0.689)	-0.26 (0.944)	0.70 (0.866)
Feedback	5.02 (0.270)	4.36 (0.336)	1.20 (0.790)	-2.52 (0.540)	1.90 (0.611)	-0.11 (0.975)	1.35 (0.721)	1.67 (0.643)	2.17 (0.556)	4.66 (0.224)
Round	-1.71*** (0.001)	-1.48*** (0.007)	-1.75*** (0.001)	-1.03* (0.073)	-0.97* (0.076)	-1.33** (0.023)	-1.62*** (0.001)	-0.88 (0.112)	-1.05 (0.138)	-1.10* (0.064)
<i>N</i>	244	244	244	244	244	244	244	244	244	244

Notes: standard errors clustered at the individual subject level. p-value in parentheses.  
 \*\*\*  $p < 0.01$ . \*\*  $p < 0.05$ . \*  $p < 0.1$ .

Table 12: Average time on a trial of workers' tasks by frequency.

	Firm Chal	Firm Std	Own Std
Freq = 1	30.91	16.69	16.78
Freq = 2	30.02	14.83	16.55
Freq = 5	29.23	14.55	16.73
Freq = 10	29.23	15.95	16.25

Notes: Firm Chal refers the Firm Challenging Task. Firm Std refers to the Firm Standard Task. Own Std refers to the Own Standard Task.

that overall subjects did an amazing job of spending the correct amount of time per trial, it is worth investigating whether frequent monitoring led to a decrease in the time per trial as workers may have been rushing to get some positive results. As a simple check on this we can examine the summary statistics of the time spent per trial on the different tasks broken down by the reporting frequency. This is shown in Table 12 and then Table 13 shows similar summary statistics by reporting treatment. All of these summary statistics are practically identical making it clear that even asking subjects to report 10 times in a 300 second production period did not lead to the workers decreasing the time spent per trial.

Table 13: Average time on a trial of workers’ tasks by treatment.

	Firm Chal	Firm Std	Own Std
Baseline	30.80	15.21	15.17
Intrinsic Motivation	27.43	14.55	17.93
Observability	30.84	16.07	15.95
Feedback	31.36	15.93	17.16

Notes: Firm Chal refers the Firm Challenging Task. Firm Std refers to the Firm Standard Task. Own Std refers to the Own Standard Task.

## 4 Conclusion

Many firms have employees report to their supervisors on the status of current projects at specified intervals. An important purpose of these reports is to ensure their employees are working on appropriate tasks and putting sufficient effort in them, since the employees usually prefer shirking instead of working on firm projects as indicated by the standard principal agent problem. We use controlled experiments to examine how the frequency and style of these reports affect workers’ work effort. We find that increasing the frequency of reporting improves workers’ total effort on firm tasks but when set too high, the reporting frequency can have a less beneficial effect. The frequency also affects their choices on what firm tasks to put effort in. If workers are asked to report too often, they shift to performing less lucrative tasks which have more near term payoffs but lower payoffs over all. We also examine what elements of the reporting system, like the observability or feedback, are responsible for any improvement in employees’ effort. While there is substantial literature suggesting that the observability of workers’ actions or the employers’ comments on workers’ actions should impact their behavior, we find at best limited impact of these elements unless they interact with the appropriate frequencies.

Of course one still might be surprised that there was not a stronger overall effect from these different reporting regimes as there is much prior work showing that behavior shifts substantially between cases without observability and with (Masclot et al., 2003; Andreoni and Petrie, 2004; Karlan and McConnell, 2014; Xiao and Houser, 2011; Salmon and Serra, 2017; Bursztyn and Jensen, 2017). There are many possible reasons why the effect of observ-



ability may be weaker here than in some of these prior papers. The issue with observability in these prior papers is that when the amount given or taken from another party is observed versus not, an individual can not escape moral blame for their actions. They know they have chosen the “wrong” action and they know that someone else has clearly observed their choice of “incorrect” behavior. In the environment for this experiment and importantly for many workplaces, individual actions are not what is observable; only outcomes are observable. Given that a particular outcome could occur based on a variety of different actions, there is no longer a clear implication about what actions a worker has taken when a bad outcome is observed. This fact may be enough for individuals to not feel quite the same pressure from observability on their outcomes as they would on their actions. This is consistent with (Dana et al., 2006, 2007; Andreoni and Bernheim, 2009) as in these studies, the authors find generally that individuals are not concerned with doing the right thing but rather with not being seen to do the wrong thing. Further, if they can find a way to essentially “blame” the bad outcome on another actor, then this essentially indemnifies them to be able to engage in more selfish behavior. That may be why in the current environment, the effect of these different reporting regimes is relatively weak. Given that this issue of observability being only possible on outcomes rather than actions is so important to many workplaces, this is an issue that warrants future research to determine whether or not there are other ways might want to design reporting regimes and frequencies to deliver even stronger results.

On the other hand, we do still observe an effect on behavior at certain reporting frequencies indicating that the effect is not actually zero. There are two possible reasons for this. Firstly, When workers are asked to report frequently, the time window and the number of possible trial attempts between two reports becomes smaller, so workers have less wiggle room in terms of the outcomes or earnings generated to their employers. Secondly, increasing the reporting frequency increases number of times of being observed or commented. It is possible that these elements have impact only if the intensity of these elements achieve certain level.

## References

- Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The economic journal*, 100(401):464–477.
- Andreoni, J. and Bernheim, B. D. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, 77(5):1607–1636.
- Andreoni, J. and Petrie, R. (2004). Public goods experiments without confidentiality: a glimpse into fund-raising. *Journal of public Economics*, 88(7-8):1605–1623.
- Azoulay, P., Graff Zivin, J. S., and Manso, G. (2011). Incentives and creativity: evidence from the academic life sciences. *The RAND Journal of Economics*, 42(3):527–554.
- Balcazar, F., Hopkins, B. L., and Suarez, Y. (1985). A critical, objective review of performance feedback. *Journal of Organizational Behavior Management*, 7(3-4):65–89.
- Belot, M. and Schröder, M. (2015). The spillover effects of monitoring: A field experiment. *Management science*, 62(1):37–45.
- Benabou, R. and Tirole, J. (2012). Laws and norms. IZA discussion papers 6290. *Institute for the Study of Labor (IZA)*.
- Bock, O., Baetge, I., and Nicklisch, A. (2014). hroot: Hamburg registration and organization online tool. *European Economic Review*, 71:117–120.
- Bursztyn, L. and Jensen, R. (2017). Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure. *Annual Review of Economics*, 9:131–153.
- Chhokar, J. S. and Wallin, J. A. (1984). A field study of the effect of feedback frequency on performance. *Journal of Applied Psychology*, 69(3):524.
- Dana, J., Cain, D. M., and Dawes, R. M. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and human decision Processes*, 100(2):193–201.
- Dana, J., Weber, R. A., and Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, 33(1):67–80.
- Dickinson, D. and Villeval, M. C. (2008). Does monitoring decrease work effort? *Games and Economic Behavior/Games and Economic Behaviour*, 63(1):pp–56.
- Dutcher, G., Salmon, T., and Saral, K. J. (2018). Is 'real' effort more real?
- Ederer, F. and Manso, G. (2013). Is pay for performance detrimental to innovation? *Management Science*, 59(7):1496–1513.
- Falk, A. and Kosfeld, M. (2006). The hidden costs of control. *American Economic Review*, 96(5):1611–1630.
- Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2):171–178.
- Frey, B. S. (1993). Does monitoring increase work effort? the rivalry with trust and loyalty. *Economic Inquiry*, 31(4):663–670.
- Gneezy, U., Kapteyn, A., and Potters, J. (2003). Evaluation periods and asset prices in a market experiment. *The Journal of Finance*, 58(2):821–837.

- Karlan, D. and McConnell, M. A. (2014). Hey look at me: The effect of giving circles on giving. *Journal of Economic Behavior & Organization*, 106:402–412.
- Lurie, N. and Swaminathan, J. (2007). Is timely information always better? the effect of feedback frequency on performance and knowledge acquisition. Technical report, Working Paper, Georgia Institute of Technology.
- Manso, G. (2011). Motivating innovation. *The Journal of Finance*, 66(5):1823–1860.
- Masclot, D., Noussair, C., Tucker, S., and Villeval, M.-C. (2003). Monetary and nonmonetary punishment in the voluntary contributions mechanism. *American Economic Review*, 93(1):366–380.
- Merlo, A. and Schotter, A. (1999). A surprise-quiz view of learning in economic experiments. *Games and Economic Behavior*, 28(1):25–54.
- Salmon, T. C. and Serra, D. (2017). Corruption, social judgment and culture: An experiment. *Journal of Economic Behavior & Organization*, 142:64–78.
- Tian, X. and Wang, T. Y. (2011). Tolerance for failure and corporate innovation. *The Review of Financial Studies*, 27(1):211–255.
- Xiao, E. and Houser, D. (2011). Punish in public. *Journal of Public Economics*, 95(7-8):1006–1017.