Testing the Theory of Multitasking: Evidence from a Natural Field Experiment in Chinese Factories

Fuhai Hong, Tanjim Hossain, John A. List, Migiwa Tanaka*

May 2015

Abstract

Using a natural field experiment where we observe factory workers under varying incentive schemes, we quantify the impact of one-dimensional monetary incentives on both incentivized and non-incentivized dimensions of output. While the management observes only the incentivized output (quantity), we also observe the non-incentivized output (quality) by hiring inspectors who secretly inspected the quality. Unlike most prior studies, we find sharp evidence that workers trade off quality for quantity when a quantity-based performance-pay scheme is introduced. Moreover, consistent with our theoretical model, treatment effects are much stronger for workers whose base salary structure is a flat wage compared to those under a piece-rate base salary.

Keywords: Natural Field Experiment, Multitasking, Agency Theory, Hawthorne Effect

JEL Codes: C93 (Field Experiments), D86 (Economics of Contract), J24 (Labor Productivity)

---

* Hong: Division of Economics, Nanyang Technological University, Hossain: Department of Management, University of Toronto Mississauga and Rotman School of Management, List: Department of Economics, University of Chicago, and Tanaka: Department of Economics, Wilfrid Laurier University. We thank Raj Chetty, Aureo de Paula, Florian Englmaier, Ying Fan, April Franco, Avi Goldfarb, Matthew Grennan, Nicola Lacetera, Magne Mogstad, Stephanie Rosch, Justin Sydnor, and conference and seminar participants at the Industrial Organization Society Winter Meetings, the 12th International Industrial Organization Conference, Ryerson University, the Singapore Economic Review Conference, University College London, University of Konstanz, University of Toronto, University of Western Ontario, Wilfrid Laurier University, and York University for helpful comments. We gratefully acknowledge the financial support of SSHRC Grant No. 489160 and from Shanghai University of Finance and Economics. All correspondence should be directed to Tanjim Hossain (tanjim.hossain@utoronto.ca).

I. Introduction

One ubiquitous feature of modern economies is the importance of principal-agent relations. Be it at home, at school, in the board room, or in the doctor's office, each contains significant components of the principal-agent relationship. The general structure of the problem is that the agent has better information about her actions than the principal and, without proper incentives, inefficient outcomes are obtained. For instance, a worker in a firm usually knows more about how hard she is working and how such effort maps into productivity than does the owner. A particularly important and relatively complex problem arises when output has multiple dimensions that vary in their quantifiability.

The core principle of the multitasking theory initiated by Holmstrom and Milgrom (1991) is that agents will focus their effort on measurable and rewarded tasks at the expense of other tasks (when higher effort on one task raises the marginal cost of effort on other tasks), potentially adversely influencing the principal's benefits. Therefore, it is desirable for the principal to keep a balance between incentives across tasks to avoid this form of "task arbitrage" by the agent. Ever since the seminal work of Holmstrom and Milgrom (1991), theorists have made important advances related to the multitasking problem and its relation to contract theory (see Prendergast, 1999 for an excellent review). Moreover, many multitasking problems are policy relevant. For example, performance-pay to doctors in New York City public hospitals (see Hartocolis, 2013 and Keller, 2013) or under the Affordable Care Act in the United States (James, 2012) may have adverse effect on unrewarded quality dimensions of healthcare if predictions from the multitasking theory bear out in practice.

Understanding how incentive contracts affect agents' effort choice in rewarded and unrewarded dimensions is a necessary first step in the study of contract design under multitasking. While there have been a number of empirical tests of the multitasking theory, the results are mixed and the magnitudes of the impacts are small as described later in this section. An underlying assumption of the multitasking theory is that the interaction is not repeated or the unrewarded dimension of effort is unobserved by the principal. In most existing studies, however, while output levels in certain dimensions are not contracted upon, the agents know that those output levels are observed by the principal. Moreover, the principal and the agents are engaged in long-term contracts. As a result, the agents may have incorporated that into their objective function even if those dimensions are not incentivized in the contract.[1] Thus, absence of large trade-offs between the incentivized and non-incentivized outputs in these studies may not imply that in reality such trade-offs are of no concern. Our study circumvents this issue by observing production quality when it is not observed by the principal.

We conduct a natural field experiment to investigate how agents substitute between efforts in different dimensions of a task when the non-incentivized dimensions are not systematically observed by the principal. By overlaying a field experiment in a natural setting with factory workers, we can explore how workers actually respond to incentives in a multitasking environment and quantify the effect of such incentives on incentivized and non-incentivized output dimensions. Specifically, we choose work where the quality of the produced goods is usually lightly inspected. More importantly, the settings of these inspections are such that the exact worker who produced the product is not identified during the sample inspection. During our experiment, we hired inspectors who *secretly* inspected the quality of each produced unit while identifying who produced each product. Even though we, the researchers, have perfect measure for the quality of production, from the workers' and the principal's points of view, quality is unobservable to the principal. Reputation concerns arising from observability of outputs and long-term interactions between the principal and the agent are, hence, unlikely to affect an agent's effort choice in the quality dimension. Thus, we create an ideal setting to test predictions from the multitasking theory where production quality is not observable. As such, our main objective is to estimate how workers respond to a one dimensional incentive scheme in a natural multitasking environment while staying as close to the theoretical model as possible.

Our experimental testing ground is five firms located on the southern side of Fujian, a southeastern coastal province of China with a high concentration of manufacturers of electronics and clocks and watches. These five firms allowed us to introduce treatments to induce greater production levels of GPS devices, alarm devices, and clocks. Importantly, before our intervention, the base salary structure of workers in some of the factories is a flat per hour wage, whereas in some of the factories the workers are paid with piece rates. Our key experimental treatment revolves around workers' pay: during the incentive treatment, workers received

[1] James (2012) suggests that the findings that pay for performance schemes in Medicare-certified hospitals do not have a significant impact may have resulted from the mandatory public reporting of quality-of-care measures.

monetary incentives based on their observed productivity in addition to their base salary, whereas under the control they did not receive any additional monetary incentive. The monetary incentives were approximately 40% of the base salary. Using a data set with more than 2200 observations across 126 workers, we report several insights.

First, we find that our incentives worked: compared to the control periods, the workers increased their productivity by 16.2%, on average, when they received monetary incentives. Second, we also observe a difference in defect rates between treatments: although workers increased productivity as a result of the monetary inducement, their quality of production decreased, as predicted by the multitasking theory. Specifically, we find that the workers increase their defect rates by 72.4%, under the bonus scheme over the control. Analyzing the relation between productivity and the defect rate, we find that a 1% increase in the hourly productivity comes with a 1.87% increase in the defect rate. This result means that even though workers know that quality was important, they reduce their effort on that dimension to earn greater financial rewards when we offered a generous bonus scheme.

There is, however, an important caveat to our first finding: incentives work, but they are much stronger for workers who were not incented on the margin in their everyday jobs. In fact, those workers whose base salaries are paid with fixed hourly wages showed very large incentive effects (29.0%) whereas those whose base salaries are paid by piece rates showed smaller effects (3.7 %). This result is consistent with the notion that the workers who are incented on the margin under their typical (base) salary produced near their individual production frontiers before the experiment whereas those not incented on the margin were far from their personal frontiers. Further analysis of the data suggests that the quantity-quality trade-off is present only for workers under a flat-rate base salary. The bonus scheme is associated with a 97.0% increase in the defect rate for workers under a flat-rate base salary. However, the increase in the defect rate is not statistically significant for workers under a piece-rate base salary. While workers under a flat-rate base salary clearly substituted effort in the unobservable dimension with effort in the observable dimension, such effects seem much weaker for workers under a piece-rate base salary. This result is consistent with predictions from our simple theoretical model.

Unlike most existing studies, we find strong evidence that workers trade off an incentivized output at the expense of a non-incentivized one when the non-incentivized output is unobservable to the management.[2] Moreover, the magnitude of the substitution between tasks critically depends on the structure of the workers' everyday incentive scheme. Our data, therefore, support the predictions of standard contract theory highlighting that large reductions in non-incentivized output dimensions at the introduction of a performance-pay measure can be an important concern when it is unobserved and the agents are not already producing the incentivized output at the margin. An implication for researchers and policymakers is that great care must be taken when generalizing results on introduction of incentives because the extant economic environment can greatly influence observed treatment effects.

Finally, our data from the control groups show the importance of a Hawthorne-type effect. Even though data from the original Hawthorne experiments do not stand up to closer scrutiny (see Levitt and List, 2011), the data from Chinese factories do: we find a robust and economically significant Hawthorne effect (broadly defined) in our data—an encouraging letter that drew the workers' attention but provided no monetary incentive increased productivity by 9% relative to the baseline. This effect is temporally resilient: it lasts the entire experimental session.

Now we briefly discuss the extant literature that analyze the impact of marginal incentives in a multitasking environment, using observational and experimental data. The results from these studies are rather mixed with relatively weak, if any, evidence of quantity-quality, or similar, trade-offs. Using observational data, Marschke (1996), Paarsch and Shearer (2004), Dumont et al. (2008), Johnson, Reiley, and Munoz (2012), and Lu (2012) find some support for the received theory. Mullen, Frank, and Rosenthal (2010), on the other hand, does not find any evidence that pay for performance has noticeable impact on either rewarded or unrewarded dimensions of quality of care by HMOs. It is noteworthy that in many empirical setting with observational data, quality is systematically monitored and workers are penalized for lowered quality or defects. For example, in the seminal study by Lazear (2000) while workers received a

[2] The importance of observability for multitasking issues to arise is also nicely illustrated in a recent study by Chetty, Saez, and Sandor (2014). They find that when referees for the *Journal of Public Economics* were paid for submitting reports before the deadline, the speed of report submission increased significantly. However, it did not affect the quality of their recommendations to the editor which are not anonymized, but slightly reduced the length of their reports to the authors which are anonymized. We also note that Hossain and List (2012) ran their experiments in a factory that is one of the factories we used, but with tasks for which the quality is observable to the management. They do not find any evidence of workers trading off quality for quantity as a result of a quantity-based bonus.

piece-rate for installation of auto windshields, they had to reinstall any defected windshields at their own cost.

Although not aiming to test the multitasking theory directly, field experiments of Shearer (2004), Bandiera, Barankay, and Rasul (2005), and Hossain and List (2012) do not find that the quality of work is affected by the incentives in the quantity of production. Similarly, Englmaier, Roider, and Sunde (2014) finds that making the piecewise wage rate more salient to the workers increases productivity but does not affect the quality of production significantly. On the other hand, Al-Ubaydli *et al* (2012) find that workers under a piece-rate wage produce work with a higher quality than do workers under a flat wage rate. In Hossain and Li (2014), workers reciprocate to a high piece-rate wage by increasing the quality of work even though their income does not depend on the quality. In a gift exchange situation, Kim and Slonim (2012) finds that workers reciprocate to a high flat wage by improving the quality of data entry without changing the quantity of data entered. Some of these studies involve temporary or irregular workers. Practical relevance and policy implications of the multitasking theory, hence, cannot be clearly determined from them.

Educational experiments in Kenya and India (Kremer, Glewwe, and Ilias (2010) and Muralidharan and Sundararaman (2013)) show that performance-pay to teachers increases student performance in the dimensions along which teachers are incentivized, without causing adverse effects in the unrewarded measures. In an experiment where students are paid for performing well academically, Fryer and Holden (2012) finds adverse multitasking effects from low-achieving students but not high-achieving students. They assume that students do not know their learning production functions.[3] With a somewhat different objective, Manthei and Sliwka (2014) finds the use of objective performance measures to be successful in increasing profits in a multitasking environment using data from an experiment run by a retail bank.

There have been very few empirical or experimental studies on how multitasking issues affect contracts. One notable exception is Slade (1996), who studies how complementarity between different tasks affects incentive contracts in a vertical relationship. In this paper, we do not investigate the profitability of specific performance-pay schemes, as in Jensen (2001) and

[3] The relevance of a strict principal-agent setting is, however, unclear as studying is intrinsically good for the "agents."

Griffith and Neely (2009). Rather, we ask the more general question of whether performance-pay based on observed effort dimensions leads to substitution of unobserved efforts.

The remainder of our paper proceeds as follows. The next section outlines the theoretical framework. Sections III and IV describe the experimental design and the main results, respectively. Section V discusses the extent of the Hawthorne effect in our data and Section VI concludes.

II. Theoretical Framework

In what follows, we outline a basic form of the multitasking theory based on Holmstrom and Milgrom (1991) and Baker (1992). In this model, an agent chooses a level of effort $e = (e_1, e_2)$ to provide to a task given by a principal, where e is two-dimensional. Intuitively, we can think that the first dimension of effort affects the quantity of output and the second dimension affects the quality of output. The quantity and quality are, respectively, given by the production functions $f(e_1)$ and $d(e_2)$, which are both strictly increasing.[4] However, incentive contracts can be based only on $f(e_1)$ as the principal only observes the quantity of production. Without loss of generality, we assume that $f(e_1)=e_1$ and $d(e_2)=e_2$. To provide effort level of e, we assume that the agent faces an effort cost of $C(e)$. We impose a standard set of regularity conditions on $C(e)$, namely that:

- $C(e)$ is strictly convex and is continuously differentiable on its domain and
- $C_{12}(e)$ is strictly positive on its domain. Here subscripts of 1 and 2 denote partial derivatives with respect to e_1 and e_2, respectively.

As in Holmstrom and Milgrom (1991), $C(e)$ attains an interior minimum at some finite, strictly positive vector $\bar{e} = (\bar{e_1}, \bar{e_2})$, representing the effort choice of the agent when neither dimension of effort is incentivized. As in Holmstrom and Milgrom (1991), we assume that $\bar{e}$ is strictly positive due to concerns for the employer or the firm, disutility from boredom if the agent exerts no effort, etc. The assumption that C_{12} is strictly positive implies that increasing effort in one dimension increases the marginal cost in the other dimension of effort. Thus, increasing effort in one dimension leads to some negative externality on the other dimension.

[4] Theory models solving for optimal contracts often assume that there are unobservable shocks in the production functions. However, for our purpose of understanding agent behavior, deterministic production functions suffice.

In this context, the principal offers the agent a wage of $\gamma + \alpha_0 e_1$ where γ is a fixed payment and $\alpha_0 \geq 0$ is a piece-rate payment on the observable component of effort. In this paper, we do not investigate how the principal chooses such a contract. Rather, we focus solely on how the agent responds to a given wage contract. Given the wage offer, the agent will choose her effort level e^* such that:

$$e^* = \max_e \gamma + \alpha_0 e_1 - C(e).$$

From this model, we would predict that an agent who is offered a flat rate of pay, $\alpha_0 = 0$ (i.e. not incented on the margin), would choose the interior minimum level of effort $\bar{e}$. In addition, since we assumed that e_1 and e_2 are substitutes in her effort, we predict that when facing a piece rate $\alpha_0 > 0$, the agent will increase effort on the observable component of effort while reducing effort on the unobservable component of effort. This follows from the fact that reducing e_2 *reduces* the marginal cost of increasing e_1. These results are reported in Lemma 1.

Lemma 1 An agent facing a contract where α_0 equals 0 will respond by choosing $e^* = \bar{e}$. Moreover, the larger the value of α_0 is, the larger will be her choice of e_1 and the smaller will be her choice of e_2 when she maximizes her payoff.

Proof: If $\alpha_0 = 0$, then the agent will choose e^* such that $C_1(e^*) = C_2(e^*) = 0$. However, that implies that e^* is the unique interior minimizer for the function *C*. In other words, $e^* = \bar{e}$. In general, the agent solves the following maximization problem: $\max \alpha_0 e_1 - C(e)$. First order conditions imply that at the optimal level of effort e^*, $C_1(e^*) = \alpha_0$ and $C_2(e^*) = 0$. Total differentiation of these two conditions yields $C_{11}de_1 + C_{12}de_2 - d\alpha_0 = 0$ and $C_{21}de_1 + C_{22}de_2 = 0$. The second condition implies that $de_2 = -\frac{C_{12}}{C_{22}}de_1$. Inserting this in the first condition, we get $C_{11}de_1 - \frac{C_{12}^2}{C_{22}}de_1 = d\alpha_0 \Rightarrow \frac{de_1}{d\alpha_0} = \frac{C_{22}}{C_{11}C_{22}-C_{12}^2}$. Now, $C_{22} > 0$ and $C_{11}C_{22} - C_{12}^2 > 0$ because *C* is strictly convex. Hence, $\frac{de_1}{d\alpha_0} > 0$. Moreover, $\frac{de_2}{d\alpha_0} = \frac{de_2}{de_1}\frac{de_1}{d\alpha_0} = -\frac{C_{12}}{C_{22}}\frac{de_1}{d\alpha_0} < 0$. Therefore, as α_0 increases, the chosen level of effort increases in the first dimension and decreases in the second dimension.▪

Lemma 1 implies that compared to an agent receiving a flat wage rate, an agent under a piece-rate contract will put in more effort in the quantity dimension and exert less effort in the quality dimension.[5] We see the multitasking problem that the principal faces from this result. On one hand, offering no incentive at the margin may lead the agent to choose effort level $e^* = \bar{e}$ which is suboptimal for the principal. On the other hand, if the second component of effort is sufficiently important, the principal may not be able to improve upon the agent's choice of $\bar{e}$.

Now we analyze the impact of a piece-rate bonus on top of the base salary of $\gamma + \alpha_0 e_1$. Specifically, an agent who produces e_1 units receives a bonus of $\alpha_1(e_1 - e_{1t})$ when she produces $e_1 > e_{1t}$ for a set target level $e_{1t} \geq 0$ with $\alpha_1 > 0$. That is, the agent producing e_1 units will earn a wage of w where $w = \begin{cases} \gamma + \alpha_0 e_1 & \text{if } e_1 \leq e_{1t}, \\ \gamma + \alpha_0 e_1 + \alpha_1(e_1 - e_{1t}) & \text{if } e_1 > e_{1t}. \end{cases}$

For agents with a flat-rate base salary, $\alpha_0=0$ and for agents with a piece-rate base salary, $\alpha_0>0$. We use Lemma 1 to show the first prediction of the multitasking theory that we test. The introduction of a piece-rate bonus will increase effort in the quantity dimension and decrease effort in the quality dimension under both flat-rate and piece-rate base salaries. For all further analyses in this section, we will focus on the scenario that the bonus piece rate is high enough and the target is low enough such that the agent chooses e_1 larger than e_{1t} when she is offered the bonus scheme as offering the bonus scheme will be meaningless otherwise. Under this scenario, the agent will increase effort in the quantity dimension and reduce effort in the quality dimension.

Proposition 1 The agent will increase e_1 and decrease e_2 under the bonus scheme.

Proof: Suppose that the agent chooses the effort vectors e' and e'' under the base salary and the bonus scheme, respectively. Since we assumed that the chosen level of e_1 is larger than e_{1t} under the bonus scheme, the first order conditions are: $C_1(e') = \alpha_0, C_1(e'') = \alpha_0 + \alpha_1$ and $C_2(e') = C_2(e'') = 0$. Here $\alpha_0 \geq 0$ and $\alpha_1 > 0$. Thus, one can view the bonus scheme to be an incentive scheme that offers a larger piece rate at the margin. Following Lemma 1, $e_1'' > e_1'$ and $e_2'' < e_2'$.▪

We can further show that under suitable conditions, the increase in effort in the quantity dimension will be smaller as α_0 increases for a given α_1. For this result, we make a number of additional assumptions on the marginal cost functions C_1 and C_2. First, the cross partial

[5] We get the same theoretical prediction if we assume that the production functions f and d, instead of the cost function C, depend on efforts in both dimensions. Specifically, suppose that fixing e_1, increasing e_2 reduces quantity while increasing quality. On the other hand, if e_2 is fixed, increasing e_1 increases quantity and decreasing quality. The agent's strategic trade-offs in this model are fundamentally the same as those in our model above. In both cases, agents strategically increase e_1 and reduce e_2 as a result of a quantity-based bonus when quality is unobserved.

derivatives of C_1 and C_2 equal zero; that is $C_{112} = C_{221} = 0$. Thus, C_{11} and C_{22} are independent of the value of e_2 and e_1, respectively. Moreover, we assume that $C_{111} > 0$ and $C_{222} > 0$ and $\frac{C_{111}}{C_{222}} > \frac{C_{11}C_{12}}{C_{22}^2}$ for all e. Thus, the marginal costs are convex and, in some sense, C_1 is sufficiently more convex than C_2. These assumptions lead to the following proposition.

Proposition 2 For a given α_1, the larger the value of α_0 is, the smaller will be the magnitudes of the increase in e_1 and the decrease in e_2 as a result of the bonus.

Proof: Suppose that the agent chooses the effort vectors e' and e'' under the base salary and the bonus scheme, respectively. The first order conditions are: $C_1(e') = \alpha_0, C_1(e'') = \alpha_0 + \alpha_1$ and $C_2(e') = C_2(e'') = 0$. We need to show that $\frac{d(e_1''-e_1')}{d\alpha_0} < 0$ and $\frac{d(e_2'-e_2'')}{d\alpha_0} < 0$. Recall that, Lemma 1 implies that the larger α_0 is, the larger will be e_1' and the smaller will be e_2'. Therefore, it is sufficient to show that $\frac{d(e_1''-e_1')}{de_1'} < 0$ and $\frac{d(e_2''-e_2')}{de_2'} < 0$. The first order conditions imply that $C_1(e'') - C_1(e') = \alpha_1$. Using total differentiation and the fact that α_1 is kept unchanged, we can show that $C_{11}(e'')de_1'' + C_{12}(e'')de_2'' - C_{11}(e')de_1' - C_{12}(e')de_2' = 0$. (1)

Total differentiations of the first order conditions on C_2 gives us $de_2 = -\frac{C_{12}}{C_{22}}de_1$ for both e' and e''. Inserting these values in equation (1), we get

$$C_{11}(e'')de_1'' - \frac{C_{12}^2(e'')}{C_{22}(e'')}de_1'' = C_{11}(e')de_1' - \frac{C_{12}^2(e')}{C_{22}(e')}de_1'$$

$$\Rightarrow \frac{de_1''}{de_1'} = \frac{\frac{C_{11}(e')C_{22}(e') - C_{12}^2(e')}{C_{22}(e')}}{\frac{C_{11}(e'')C_{22}(e'') - C_{12}^2(e'')}{C_{22}(e'')}} \qquad (2).$$

By strict convexity of C, both the numerator and denominator in the right side of Equation (2) are positive. Then, showing that $C_{11}(e'') - \frac{C_{12}^2(e'')}{C_{22}(e'')} > C_{11}(e') - \frac{C_{12}^2(e')}{C_{22}(e')}$ will be sufficient for showing that $\frac{de_1''}{de_1'} < 1$ and the increase in effort in dimension 1 due to the bonus scheme is decreasing in α_0. Let us define $G(e) = C_{11}(e) - \frac{C_{12}^2(e)}{C_{22}(e)}$. Then,

$$dG = C_{111}de_1 + C_{112}de_2 - \frac{2C_{12}C_{112}C_{22} - C_{12}^2C_{122}}{C_{22}^2}de_1 - \frac{2C_{12}C_{122}C_{22} - C_{12}^2C_{222}}{C_{22}^2}de_2$$

$$= C_{111}de_1 + \frac{C_{12}^2 C_{222}}{C_{22}^2} de_2 = \left(C_{111} - \frac{C_{12}^3 C_{222}}{C_{22}^3}\right) de_1.$$

Note that the above uses our assumption that the third-order cross partial derivatives (e.g. C_{112}, C_{221}, etc.) equal zero. Now, strict convexity of C implies that $C_{11}C_{22} > C_{12}^2 \Rightarrow C_{11}C_{12} > \frac{C_{12}^3}{C_{22}}$ and we assumed that $\frac{C_{111}}{C_{222}} > \frac{C_{11}C_{12}}{C_{22}^2}$. Therefore, $\frac{C_{111}}{C_{222}} > \frac{C_{12}^3}{C_{22}^3}$. Since $e_1'' > e_1'$, this implies that $G(e'') > G(e')$ and $\frac{de_1''}{de_1'} < 1 \Rightarrow \frac{d(e_1''-e_1')}{de_1'} < 0$.

To prove the part for e_2, we substitute $de_1 = -\frac{C_{22}}{C_{12}} de_2$ into Equation (1), which yields

$$C_{12}(e'')de_2'' - \frac{C_{11}(e'')C_{22}(e'')}{C_{12}(e'')} de_2'' = C_{12}(e')de_2' - \frac{C_{11}(e')C_{22}(e')}{C_{12}(e')} de_2',$$

and therefore

$$\frac{de_2''}{de_2'} = \frac{C_{12}(e') - \frac{C_{11}(e')C_{22}(e')}{C_{12}(e')}}{C_{12}(e'') - \frac{C_{11}(e'')C_{22}(e'')}{C_{12}(e'')}}.$$

Define $H(e) = C_{12}(e) - \frac{C_{11}(e)C_{22}(e)}{C_{12}(e)}$. Given our assumptions,

$$dH = C_{121}de_1 + C_{122}de_2 - \frac{C_{111}C_{22}C_{12} + C_{11}C_{221}C_{12} - C_{11}C_{22}C_{121}}{C_{12}^2} de_1$$
$$- \frac{C_{112}C_{22}C_{12} + C_{11}C_{222}C_{12} - C_{11}C_{22}C_{122}}{C_{12}^2} de_2 = \left(\frac{C_{111}C_{22}^2}{C_{12}^2} - \frac{C_{11}C_{222}}{C_{12}}\right) de_2.$$

Our assumption $\frac{C_{111}}{C_{222}} > \frac{C_{11}C_{12}}{C_{22}^2}$ implies that $\frac{C_{111}C_{22}^2}{C_{12}^2} > \frac{C_{11}C_{222}}{C_{12}}$. Then, $0 > H(e') > H(e'')$ since $e_2' > e_2''$. As a result, $\frac{de_2''}{de_2'} < 1$ and $\frac{d(e_2''-e_2')}{de_2'} < 0$. This completes the proof of the proposition.▪

Our assumption that the marginal cost functions are convex implies that the increase in the marginal cost resulting from a unit increase in the effort level along a dimension is greater when the base effort level along that dimension is higher. Thus, for a given bonus piece rate α_1 (which denotes the change in the marginal cost in the quantity dimension), the associated change in the effort level is smaller when we start from a higher base effort level along the quantity dimension. As a result, for a given agent, the higher the piece-rate component of her base salary is, the smaller is the change in her effort level in the quantity dimension as a result of a given piece-rate bonus scheme. A corollary of this result is that, for a given bonus scheme, the increase in production quantity will be greater when the base salary structure is flat rate.

III. Experimental Design

To provide empirical insights into the theory, we ran nine experimental sessions in five Chinese firms between April 2009 and July 2012. All of the firms are located on the southern side of Fujian, a southeastern coastal province of China. The five firms (Hengli, Jiali, Heyu, Wanlida, and Shike) allowed us to introduce treatments pertaining to the production of GPS devices, alarm devices, and clocks. Among the five firms, Hengli, Jiali, and Heyu are clock or clock module manufacturers located in Zhangzhou Prefecture. Wanlida is a large electronics manufacturing company whose production center is located in Nanjing County. Shike, a relatively small electronics manufacturer located in Quanzhou City, mainly produces alarm devices.

In our sample, the base salary structure of workers in some of the factories is flat rate—workers receive a fixed hourly wage in Hengli, Wanlida, and Shike. In Jiali and Heyu, on the other hand, workers are paid a piece-rate salary possibly with a small flat component. Within a factory, the base salary scheme of most non-administrative workers follows the same format—either flat rate or piece rate. A factory usually chooses a base salary format based on the production process of the main products the factory manufactures. For example, if the production processes are team or product line based, the factories usually choose flat-rate salary schemes. If the main products are manufactured individually with stable demands for the products, base salaries are likely to be piece rate. For our experiment, we chose works that are not mainstays of the factories, rather supporting works such as packaging or simple maneuvering that require relatively little training or human capital. Thus, it is unlikely that workers in our experiments have sorted themselves out to a specific kind of base salary structure by their choice of profession. While we cannot rule out that different wage structures condition workers differently over time making them incomparable, our discussions with the management in the factories and other industry participants indicate that there should not be any fundamental difference between workers under flat and piece-rate base salary structures in our experiment. We also note that the piece-rate component of the base salary is the same for all workers within a session with a piece-rate base salary. Table 1 provides a summary of the nine sessions including the description of the work place and the experimental design described below. The sessions are numbered chronologically in terms of the date of running the session.

The workers in our experiment performed their tasks individually and independent of the other workers within the session. The tasks included product packaging or finishing (wall clocks, alarm devices, or attachments to GPS devices), wedging components into clock modules, and twining metallic threads for clock modules. Within each session, all subjects performed the same task and all workers involved in that specific task were included as subjects in our experiment. For all of the tasks, the quantity produced is regularly measured even in the factories where the base salary scheme is a fixed hourly wage. The workers are aware that the management records the quantity produced by each worker. However, quality is not individually recorded. The usual quality control process is sampling inspection. The sampling rates vary from firm to firm, but none is more than 5%. Moreover, the factories do not record the exact mapping between a sample product and the worker who produced it. Thus, from the perspective of a worker, the extent of the effort she expends to control the quality of production is virtually unobservable to the managers. Note that the quality control process depends on the job description of a worker. Within a factory, quality of some tasks are heavily monitored and some are lightly monitored, as is the case with the tasks that we chose for this experiment. Whether the base salary structure is piece-rate or flat-rate, on the other hand, depends on the main products produced in the factory. The structure is the same for all workers in the factory, not only the tasks chosen here. Thus, quality inspection process and the base salary structure of the workers are typically independent of each other.

During our experimental sessions, the workers engaged in tasks within their natural work environment during their regular work hours, unaware that an experiment was taking place. Accordingly, we denote our experiment as a natural field experiment, following the terminology of Harrison and List (2004). One session typically lasted three days, or around 15 to 24 work hours. For each work hour, we recorded the production and the non-work minutes (if any) for each worker. We then projected the worker's production to her *hourly productivity* by taking the non-work minutes within the hour into account.

In our setting, a test of the multitasking theory requires the availability of data on the *quality* of the workers' productions as well. However, if the workers know that the quality of their production is to be observed, a compounding repeated-game effect would emerge. Our experimental design resolves this concern. We chose works that were lightly inspected in the usual production procedure, and hired inspectors to *secretly* inspect all of the workers' productions and record the number of defects made by each worker for each hour. To achieve confidentiality, inspection was conducted either in an isolated space away from the workers' workplaces or when all the workers were off duty. Some of the inspections were done after the experimental session was over. We ensured that the inspectors and the managers kept this secret. As noted above, the regular rate of sampling inspection is very low. The workers were unaware of the heightened quality control measures or any change in the inspection process during our experimental sessions.

Depending on the nature of the work, a worker may make more than one defect in a produced unit. For example, each worker in our session at Shike Alarm System Electronic Co. Ltd. was responsible for plugging a circuit wafer into a plastic shell, placing a label on the shell, brushing some items, folding the instruction sheet, and then putting all of these items together into a paper box. Workers can potentially commit an error in any of these steps, and each mistake is counted separately. We can calculate the maximum number of possible mistakes or defects a worker can make in a given hour by multiplying the number of units she produces in that hour by the maximum number of defects that can be made for each unit produced. We define the *defect rate* to be ratio of the number of defects the worker made in that hour and the maximum number of possible defects. Thus, the defect rate takes a value between 0 and 1.

A session is divided into three rounds: a pre-intervention *baseline* round was followed by two rounds of interventions. The interventions include the *bonus* round and the *control* round. Workers received monetary incentives conditional on their productivity during the bonus round. During the control round, they did not receive any additional monetary incentive. Baseline rounds lasted three to eight hours. The bonus and control rounds lasted six to eight hours each. Each round typically corresponds to one day. After the baseline round, half of the workers were randomly selected into the bonus group while the others were in the control group. The management notified the subjects of the bonus treatment via personal letters. At the beginning of a round, each worker in the bonus group received a letter saying that she had been selected into a short-term program which lasted for t work hours. For each of the hours, if her productivity exceeded a target of y units, she would receive a bonus at the rate of RMB α_1 per unit for each unit she produced beyond the target of y units. Thus, if a worker's productivity for a given hour was x units then she would receive no bonus if $x \leq y$ and would receive a bonus of RMB $\alpha_1(x - y)$ otherwise. Bonuses, if any, were paid in addition to the workers' base salary. Some hours of the day include regularly scheduled break periods (typically 10 minutes long). Workers were told that these non-work minutes would be taken into account when measuring hourly productivity to calculate the bonus. Most workers in our experiments were familiar with such a bonus. In peak seasons when they need workers to work harder, the factories sometimes provide piece-rate or flat-rate bonuses if production exceeds a certain target. Thus, the bonus scheme was not particularly surprising to the workers. The targets, however, are usually based on daily production, not hourly production.

We chose the target y and the piece rate α_1 based on the baseline data of productivity. Generally speaking, we wanted the target to be difficult to achieve for a worker with average productivity, but she would have a reasonable chance to achieve it if she tried hard. We followed the same formula to choose the parameters so that the target and the piece rate across sessions are somewhat comparable. For a session, let M and SD, respectively, denote the mean and standard deviation of the baseline productivity across all workers in the session. Generally, we chose the target y, in consultation with the management in the factories, to be a round number close to M + SD that could be considered as a natural target level.[6] On average, the target was 24% higher than M and we observed about 12% of worker-period combinations in which a worker met the target during a baseline round. In these factories, the managers typically are erstwhile "regular" workers who were promoted to a management position. As a result, they have clear insights on productivity and difficulty of various tasks.

With respect to the setting of the piece rate α_1, we chose a number that we thought was attractive enough to effectively incentivize even workers with average productivity and would provide almost twice the amount of their average base income as a bonus if they produced at a very high rate, e.g., at two standard deviations above the target level y. For example, in session 8, the salary for a worker-subject with average productivity was around RMB 60-65 per day (which

[6] We made slight adjustments to the target level in order to take outliers in terms of productivity in the baseline hours, if any, into account.

approximately equaled USD 10 at the exchange rate during the period of the experiment). Under the bonus round, which lasted for one work day, a worker obtained a bonus of RMB 25, on average, which is approximately 40% of her base salary. The highest bonus paid in this session was RMB 111, almost double of the average daily salary.

In order to control for the potential effects of receiving a letter, we also sent each worker in the control group a letter which only encouraged her to work hard.[7] In the second round, we switched the bonus to those subjects who were in the control group in the first round. Those who received the bonus scheme in the first round were in the control group in the second round. The reason for exposing each worker to both bonus and control was twofold. This allows us to identify a worker's effort choices under multitasking consideration using variations within that particular worker. Moreover, as our incentive schemes potentially provided a substantial amount of additional payments to workers, the factories required that we offer the bonus treatments to all workers within a particular set of work. To control for the impact of the treatment sequence, we offered the bonus first to half of the workers and control first to the rest. As all workers within a given set experience both treatments, we can control for peer effects as in Mas and Moretti (2009) using a difference-in differences approach.[8]

A few other experimental particulars of interest are worth noting before we move to the experimental results. First, the source of the bonus was intentionally kept vague. Second, the letters clearly mentioned that this was a short-term incentive program and the workers were likely to assume that the incentive schemes were one-shot opportunities. Third, during the baseline round, the workers were unaware of the fact that they may receive bonus schemes in the following days; during the first round of experiment, workers in the control group were unaware that they would receive bonus schemes in the next day.

IV. Experimental Results

In total, the nine experimental sessions included 126 workers. As we collected data on productivity and defects every hour, we have an observation for each hour. We have a total of 2272 observations – 653, 812, and 807 observations under the baseline, bonus, and control rounds, respectively. Table 2 presents summary statistics of hourly productivity, defect rate, and

[7] The Appendix includes, as an example, English translation of the letters used in session 5.
[8] Workers who received the same treatment at a given time where located in the same room.

the percentage of hours in which the worker reached the target for the baseline, control, and bonus rounds, for all sessions and sessions with flat-rate and piece-rate base salary structures separately. Recall that hourly productivity is calculated by projecting a worker's production in a work hour net of the non-work minutes within the hour. Hourly defect rate is defined as the number of defects divided by the maximum possible number of defects for a given hour. Percentage of periods meeting the target reports the ratio of observations in which the worker's productivity was at least as high as the target level set during the bonus round for the session.[9] As the base productivity and defect rates and the number of hours under each round vary across sessions, the summary statistics aggregating observations from all sessions are not extremely informative. When we look at the sessions under the same base salary structures, the treatment effects are somewhat clearer. Nevertheless, we need to control for session and individual specific heterogeneity to truly identify the treatment effects.

Effect on Productivity

First, we analyze hourly productivity under the bonus and control rounds compared to the baseline round graphically. Recall that the bonus and control rounds were between six to eight hours long and we have baseline production data for all workers. For each worker, we calculated the percentage deviation in productivity compared to her average productivity in the baseline round for each hour during the bonus and control rounds. Figure 1 presents the percentage deviation in productivity from the mean baseline productivity, averaged across all workers, in each of the first six hours of the bonus and control rounds.[10] The shaded areas correspond to 95% confidence intervals. The figure suggests that both control and bonus rounds increase productivity with the increase being much greater under the bonus scheme. Moreover, the treatment effects do not decrease over time. The figure does not change qualitatively if we present the deviation from the productivity in the last hour of the baseline round instead of the deviation from average productivity in the baseline round.

[9] Of all the workers, 28.6% did not meet the target in any period. The results presented in this paper remain qualitatively unchanged if we exclude these workers from the regressions in Tables 3 to 8.
[10] Treatment effects vary across experimental sessions. As a result, when we aggregate the deviation of productivity from the baseline round productivity across workers, we need to ensure that the composition of workers is the same between different hours to keep them comparable. Each of the bonus and control rounds was at least six hours long. Hence, we restrict attention only to the first six hours within a bonus or control round in Figure 1.

Next, we analyze the data more closely exploiting panel structure of the data set as we have observations over time for each worker. The base estimation model can be described as:

$$\log(\text{productivity}_{it}) = \beta_1 \times \text{bonus}_{it} + \beta_2 \times \text{control}_{it} + \beta_3' T_{it} + c_i + \varepsilon_{it} \quad (3),$$

where productivity$_{it}$ denotes hourly productivity of worker i in hour t, bonus$_{it}$ and control$_{it}$ are dichotomous variables indicating the treatment worker i experienced in hour t, and T_{it} is a vector of variables to control for time effects. The error term consists of time invariant individual specific term c_i which controls for heterogeneity among workers, and time variant idiosyncratic individual specific error term ε_{it}. We allow the term c_i to be correlated with other independent variables, in particular T_{it}. Therefore, we present estimates of this model under the fixed effects framework. Nevertheless, we also estimated the model under the random effects framework in the cases where c_i is unlikely to be correlated with other independent variables. None of our results change either qualitatively or quantitatively under the random effects specifications. We assume that ε_{it} is independently but not necessarily identically distributed across i and t. In all of our regression specifications reported below, we use heteroskedasticity-robust standard errors. Moreover they are clustered at the worker level, addressing the concern that observations for a worker are not independent across time.[11]

In Table 3, we report the results of this regression under the fixed effects specification. In column (1) of Table 3, we do not control for any time effect. The column shows that compared to the baseline hours, the bonus scheme increased productivity by 25.6%. Interestingly, the data also show that in the control round, when the workers received an encouraging letter but no monetary incentive, productivity increased by 9.4% compared to the baseline. Thus, the bonus increased productivity by 16.2% over the control.

Each of our experimental sessions was run over three days and the workers had already experienced the production processes prior to our experiment. Thus, learning by doing should not be important during our experimental sessions. Nevertheless, as the bonus scheme was new to them, there may still be some learning about how fast they can produce and how to adjust production to this incentive program. In addition, production of any worker may vary over time

[11] Even though the variation in worker productivity over time across workers within a given session is likely to be independent, for robustness, we also calculated standard errors while clustering at the session level. The main results do not change qualitatively in that case.

due to fatigue within a day. As a round typically corresponds to one day, while the n^{th} work hour within a round may have a different impact on morale and fatigue from the $n–1^{\text{th}}$ work hour in the same round, the impact should be similar to that in the n^{th} work hour in another round—effort costs are likely to be separable across days. To control for this, we create a variable that counts the number of hours under a specific round. That is, this variable starts from one in the baseline round and gets reset to one every time the worker enters a new round—bonus or control. To allow for non-linear time effect within a round, we also create a dummy variable for each work hour within a round. Column (2) includes these time dummy variables as regressors.

Although this approach serves to reduce the coefficient sizes slightly, the bonus and control rounds continue to increase productivity significantly: by 24.8% and 8.6%, respectively. Importantly, F-tests show that these two coefficients are significantly different from each other with a p-value below 0.0001. Thus, we conclude that our high-powered incentive scheme had a very large impact in increasing productivity. Specifically, compared to the control, the bonus scheme increased productivity by 16.2%.[12] These results are in line with the summary statistics of the proportion of periods in which a worker reached the target, reported in Table 2. This proportion increased to 58.5% in the bonus round and 36.9% in the control round from 11.8% in the baseline round. We also found that the number of workers who met the target more frequently in the bonus round compared to the baseline round is greater than comparable figures for the control round. Likewise, the percentage of workers who always met the target throughout a round indicates a similar effect of treatments: in the baseline round, 3.2% of workers met the target in all periods, while 46% and 23% of workers did so in the bonus and control rounds, respectively. Also, the percentage of workers who never met the target in a round decreases: 73% of workers never met the target in the baseline round while 31% and 44% of workers never met the target in the bonus and control rounds, respectively.

Relationship between Productivity and Defect Rate

Next we analyze how the defect rate is affected by the increase in productivity as reported above. In this way, we can conduct a test of Proposition 1 which is based on the theoretical predictions of Holmstrom and Milgrom (1991) and Baker (1992). First, we test

[12] We have also allowed different sessions to have different time effects, by including session specific time variables but the results did not change significantly.

whether the average defect rates in the bonus and control rounds are different from that in the baseline round. Figure 2 presents the average defect rates across all sessions for baseline, control, and bonus rounds: the mean defect rates are 0.009, 0.012, and 0.016, respectively. The difference in defect rates between the control and baseline rounds is statistically significant at the 5% level, and that between bonus and baseline rounds is statistically significant at the 1% level. Moreover, the difference in defect rates under bonus and control rounds is significant at 10% level (the p-value is 0.060). While the workers increased productivity as a result of the quantity-based bonus scheme, they also reduced the quality of production, as the theory predicts.

Table 4 confirms this result by presenting fixed effects panel regressions of log of the defect rate on dummy variables for bonus and control rounds. The base estimation model is described as:

$$\log(\text{defectrate}_{it}) = \Gamma' X_{it} + \beta' T_{it} + c_i + \varepsilon_{it} \qquad (4),$$

where defectrate_{it} denotes hourly defect rate of worker i at time t.[13] Moreover, X_{it} is either a vector containing dummy variables indicating bonus and control rounds or is the log of hourly productivity for worker i at time t, T_{it} is the time variable vector defined earlier, c_i is time invariant individual specific effect for worker i, and ε_{it} is an idiosyncratic error term. The vectors Γ and β contain the coefficients associated with X_{it} and T_{it}, respectively.

Column (1) of the Table 4 shows that, relative to the baseline round, the bonus scheme statistically significantly increased the defect rate by 61.3%. However, the coefficient for the dummy variable for the control round is negative and statistically insignificant. The difference between the two coefficients is significant at the 1% level. Relative to the baseline, workers increase productivity during control round (perhaps as a result of the encouraging letter they receive) without increasing the defect rate as they have no additional incentive to shift effort from the quality dimension to the quantity dimension.[14] On the other hand, the piece-rate bonus for productivity above the target level provides incentives for shifting effort from quality to the

[13] Since the defect rate for about 25% of the observations is zero, we added a small value (10^{-7}) to all observations before taking the natural logarithm of defect rate. The results are robust to the magnitude of this adjustment.
[14] This also suggests that there is no inherent constraint in the production process that makes it impossible to increase productivity at all without decreasing the quality of output.

quantity dimension. As a result, the bonus scheme leads to a very large increase in quantity but a sharp decrease in quality.

Another way of testing the theoretical prediction that workers reduce effort in terms of quality while increasing effort in terms of quantity is to analyze whether the defect rate increases when the worker produces more as is done in Bandiera, Barankay, and Rasul (2005) and Hossain and List (2012). Column (2) of Table 4 presents fixed effects panel regressions of the hourly defect rate on the log of productivity in that hour as well as time of the day dummy variables. A 1% increase in productivity is associated with an increase in the defect rate by around 1.87%. One might worry about the endogeneity of productivity in this model. For example, if highly productive workers inherently have higher defect rates, that may lead to similar results without implying a substitution of efforts from quality to quantity dimension. To control for endogeneity, we ran the same regression as in column (2) while instrumenting the log of productivity with bonus and control dummy variables. These treatments are exogenously assigned and thus not correlated with the error term, ε_{it}, of equation (4). The estimate is reported in column (3). It is similar to column (2) but greater in magnitude: a 1% increase in productivity increases the defect rate by 2.86 %. For all tests of the incentive effects on the defect rate, we get qualitatively the same results if we use the defect rate itself instead of the log of it.

Together, the two results presented so far are consistent with Proposition 1. In sum, our data provide strong evidence that workers trade off quality for quantity when their income does not depend on quality and they believe that quality is not carefully monitored.[15] According to the best of our knowledge, this is the first natural field experiment conducted in a regular workplace to find such stark evidence of the most basic theoretical predictions of Holmstrom and Milgrom (1991) and Baker (1992).[16]

Varying the Treatment Effects Depending on the Base Salary Structure

The base salary structure is different for different sessions within our experiment. Workers in sessions 1, 3, 4, 6, and 7 are paid by fixed hourly wage rates and workers in sessions 2, 5, 8, and 9 are paid piece-rate salaries by their employers. Such richness of our data allows us

[15] Recall that the workers were completely unaware of the heightened inspection rate and, from their perspectives, there was no change in the monitoring of quality.
[16] Note that the increase in quantity is relatively large compared to the decrease in quality. As a result, the overall effect of incentives on net productivity (productivity net of defects) is positive under all specifications.

to investigate whether the base salary structure of the worker has an impact on the treatment effects on quantity and quality of production. With that goal, we run regressions similar to equations (3) and (4), but decompose the treatment dummies by whether the base salary structure is flat rate or piece rate.

Table 5 presents the regression results of log productivity on four dummy variables: bonus and control dummies interacted with the base salary structure of flat rate and piece rate. We present regressions with and without time variables under the fixed effects specification. Column (1) of Table 5 suggests that while the bonus round increases productivity by 50.1% for the sessions under a flat-rate base salary, it increases productivity by only 4.9% for sessions under a piece-rate base salary. Moreover, the control round for sessions with a flat-rate base salary increases productivity by 21.1%. Interestingly, the increase in productivity under control round for the sessions under a piece-rate base salary is small (1.2%) and not statistically significant. All of these coefficients are significantly different from each other at conventional significance levels.[17] Column (2) includes time variables, as in Table 3. The results remain virtually unchanged. Overall, these results are consistent with the prediction regarding the observable effort dimension in Proposition 2. Even though we cannot compare their wage rates directly as the sessions are for different kinds of tasks, the presence of sessions with flat-rate and piece-rate base salaries allows us to test the predictions of Proposition 2 quite cleanly.

On Table 6, we report results from regressions similar to those in Table 4, but decompose the treatment effects and log productivity based on the base salary structure. In column (1), we find that while the bonus round has a significantly positive effect on the defect rate when the base salary is flat rate, there is no treatment effect for workers under piece-rate base salaries. Columns (2) and (3) suggest similar results in terms of the effect of productivity on the defect rate. Note that in column (3), interaction terms with the log of productivity are instrumented with interactions of base salary structure dummies and control or bonus dummy. For sessions with a piece-rate base salary, there is no statistically significant impact of productivity on the defect rate. For sessions with a flat-rate base salary, however, a 1% increase in productivity increases the

[17] When this specification is estimated under a random effects framework, we obtain both quantitatively and qualitatively similar results but the coefficient for the dummy variable for the control round under a piece-rate base salary also becomes statistically significant at 1%. The heteroskedasticity and cluster-robust Hausman test rejects the null hypothesis that the random effects estimator is consistent and efficient at significance levels less than 1%.

defect rate by 2.20%. This effect is statistically significant. The result that incentive effects in terms of quantity is large and there is a quantity-quality trade-off only for flat-rate sessions is generally supported even if we allow the impacts to be different for each session.

The above results lead us to qualify the results presented in Tables 3 and 4 by noting that the quantity-quality trade-off occurs only when the workers are not, in the status quo, incented at the margin. A speculative interpretation of these results is that the workers under piece rates might have already been producing near their productivity frontiers, while there is a lot more room for productivity increase under flat rates. Hence, incentives succeed in increasing productivity, but the magnitude is much greater for workers who are currently not incented on the margin. An implication for the body of research that explores incentive effects is that great care must be taken when generalizing empirical results because the extant economic environment (in this case, status quo wage contracts) can greatly influence observed treatment effects.

V. The Hawthorne Effect

Table 3 shows that productivity of workers increased by at least 8.6% during the control round when subjects received no monetary incentive. The workers under the control round received an encouraging letter that accords attention to them. Such a non-monetary change in the work environment may make them feel that they are being observed by the management even though they were unaware of an experiment being run. While the Hawthorne effect has come to take on a very broad definition (see Levitt and List, 2011), we consider the effect in the control round to be broadly consonant with a potential "Hawthorne type" of effect. [18]

In our experiment, the control round came after the bonus round for half of the workers and a strong inertia in productivity may lead to higher productivity in the control periods, relative to the baseline, for these workers. As these workers were exposed to a set target level already in the bonus round, that may have also contributed to high productivity during the control round afterwards. To investigate this possibility, we examine the treatment effect for workers who were in the bonus group first and those who were in the control group first, separately. In column (1)

[18] As defined by the Oxford English Dictionary, the Hawthorne effect means "an improvement in the performance of workers resulting from a change in their working conditions, and caused either by their response to innovation or by the feeling that they are being accorded some attention."

of Table 7, we decompose the two treatment dummy variables in terms of whether they received the bonus first or second.

We find that both the workers who received the bonus first and also those who received the control first had productivity increase during the control round: the former group's productivity increases by 8.2 % and the latter group's productivity increases by 9.1 % in the control round relative to the baseline round. Moreover, the differences in the productivity increase between the two groups are not statistically significant. Hence, the productivity increase under the control round cannot be attributed to positive inertia of the workers who received the bonus first.[19]

In Table 5, we found that the control round led to productivity increase only for workers under the flat-rate base salary structure. We present regressions with only the workers under flat-rate and piece-rate base salaries in columns (2) and (3) of Table 7, respectively. The coefficients for the control round dummies are statistically significant only in column (2): under the flat-rate base wage, we find 21.6% and 16.4% productivity increase, relative to the baseline round, for workers who received control in the first and the second round, respectively. Therefore, we find a Hawthorne effect for workers under a flat-rate base salary structure, but not for those under a piece-rate base salary structure.[20] Although close inspection of the original data suggests that there was no Hawthorne effect in the Western Electric's Hawthorne Plant in Cicero, IL, USA (Levitt and List, 2011 and the cites therein), we find a strong effect among factory workers under a flat-rate base salary structure in Fujian, China. While Leonard (2008) and Leonard and Masatu (2006 and 2010) find significant evidence of the Hawthorne effect in terms of the quality of health service and Attari et al (2014) find such evidence in household electricity use, as far as we know, our paper presents the first verified evidence of a large Hawthorne-type effect among factory workers.

Persistence of the Effect

[19] We ran similar regressions for the log defect rate. We did not find evidence to suggest that workers changed their production quality statistically significantly during the control round relative to the baseline round for either the workers who received bonus first or those who received control first.

[20] This is likely to be related to the fact that the productivities of the workers under a flat-rate base salary structure in the baseline round seem to have been far below their production frontiers. As a result, when a worker under flat rate receives an encouraging letter and feels that she might be observed, she may respond by increasing her effort substantially. On the other hand, a worker under a piece-rate salary is more likely to feel that she is working hard enough and the possibility of being observed does not alter her behavior substantially.

We use a conditional bonus scheme in this experiment. Unlike the experiments where the incentive effect of an unconditional bonus is short lived (see, Gneezy and List, 2006, for example), we expect the impact of our bonus to persist. We can test whether that is indeed the case. One may also wonder whether the Hawthorne-type effect of productivity increase in the control group is short lived, as the positive impact on productivity received from the encouraging letter may die down quickly.

The first evidence of persistence of this effect is seen in Figure 1, where we do not see any evidence of waning in the incentive effect or the effect of the encouraging letter over time. We also devise an additional test taking advantage of the panel structure of our data set. We divide the bonus and control rounds into two halves. For example, both bonus and control rounds for session 1 were seven hours long. Then, the first four hours are considered to be in the first half and the following three hours are in the second half. We look at the impact of the bonus and control rounds dividing each of them into two halves in column (1) of Table 8. If any of the treatment effects is short-lived then the coefficient for that treatment dummy will be much smaller in the second half of that treatment. Our regressions find that not to be the case. The productivity increase during the first half of bonus round relative to the baseline round is 25.1%, while during the second half it is 26.2%. Productivity increases relative to the baseline round are 8.6% and 10.2% in the first and second halves of the control round, respectively. Coefficient differences between the two halves are, however, not statistically significant. The encouraging letter provided a nudge to increase productivity that was persistent throughout the entire day of the control round.

VI. Conclusions

Principal-agent models have become the workhorse framework for modeling asymmetric information settings. In the field, when the agent cannot be certain that a dimension of output is closely monitored, the classic multitasking theory applies. This study provides empirical insights into the multitasking problem by making use of a unique naturally-occurring setting: incentive contracts for regular workers on the floor of various factories. Through our interactions with managers at these factories, we are able to implement a natural field experiment to explore basic questions within the classic principal-agent setting. Specifically, by secretly inspecting the quality of products which are not typically closely inspected, we provide a clean test of the multitasking theory in a natural setting.

Our main results paint an intriguing picture. First, the first order predictions of the theory are found in our data: as we incent workers on the margin, they move their effort to the incented activity to the detriment of the non-incented one. But, there is an important caveat to this result: we only find this result amongst workers who were previously working under a fixed wage scheme. For those workers previously under a piece-rate scheme, their output moves by a small magnitude when we introduce a supplementary performance incentive. Our results suggest that, when a task is better described by multiple dimensions, simplifying it into a single task dimension may lead to inaccurate implications. For example, in structural estimations of principal-agent models, such simplifications may yield biased estimates of structural parameters.

Second, our study underscores the importance of observability of the unrewarded task in mitigating multitasking issues. Our work taken together with the existing body of literature on multitasking, including James (2012) and Chetty, Saez, and Sandor (2014), suggests that while an agent may shirk on an unrewarded task if it is unobservable, she may not do so for observable unrewarded tasks. A policy implication is that if it is difficult to pay agents for performance in many different dimensions of a task, the principal should at least try to observe and perhaps publicize the agents' performances in the unrewarded dimensions of the task.

Further, we report an interesting Hawthorne-type effect. We find that a simple reminder letter to workers leads to a robust and economically significant increase in worker productivity. While data from the original Hawthorne experiments do not stand up to closer scrutiny, data from the Chinese factories do. Finally, an overarching lesson learned from this exercise is that one can gain enough control in a field environment to test important theories of multitasking incentive schemes. This allows us to gain invaluable insights on how theoretical models can be used in designing optimal contracts and useful public policy.

Reference

Al-Ubaydli, Omar, S. Anderson, Uri Gneezy and J.A. List (2012): "Carrots That Look Like Sticks: Toward an Understanding of Multitasking Incentive Schemes," NBER working paper no. 18453.

Attari, Shahzeen, Gautam Gowrisankaran, Troy Simpson, and Sabine Marx (2014): "Does Information Feedback from In-Home Devices Reduce Electricity Use? Evidence from a Field Experiment," NBER Working Paper No. 20809.

Baker, George P. (1992): "Incentive Contracts and Performance Measurement," *Journal of Political Economy*, 100 (3), 598-614.

Baker, George P., Michael C. Jensen, and Kevin J. Murphy (1988): "Compensation and Incentives: Practice vs. Theory," *Journal of Finance*, 43 (3), 593-616.

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

Chetty, Raj, Emmanuel Saez, and Laszlo Sandor (2014): "How Can We Increase Prosocial Behavior? An Experiment with Referees at the *Journal of Public Economics*," forthcoming, *Journal of Economic Perspectives*.

Dumont, Etienne, Bernard Fortin, Nicolas Jacquemet, and Bruce Shearer (2008): "Physician's Multitasking and Incentives: Empirical Evidence from a Natural Experiment," *Journal of Health Economics*, 27 (6), 1436-1450.

Englmaier, Florian, Andreas Roider, and Uwe Sunde (2014): "The Role of Communication of Performance Schemes: Evidence from a Field Experiment," working paper, LMU Munich.

Gneezy, Uri and John A. List (2006): "Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments," *Econometrica*, 74 (5), 1365-1384.

Griffith, Rachel, and Andrew Neely (2009): "Performance Pay and Mangerial Experience in Multitask Teams: Evidence from within a Firm," *Journal of Labor Economics*, 27 (1), 49-82.

Fryer, Roland and Jr. Richard Holden (2012): "Multitasking, Learning, and Incentives: A Cautionary Tale," NBER Working Paper 17752.

Harrison, Glenn W., and John A. List (2004): "Field experiments," *Journal of Economic Literature*, 42 (4), 1009-1055.

Hartocolis, Anemona (2013): "New York City Ties Doctor's Income to Quality of Care," *The New York Times*, January 11, 2013.

Holmstrom, Bengt and Paul Milgrom (1991): "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership and Job Design," *Journal of Law, Economics and Organization*, 7, 24-52.

Hossain, Tanjim and King King Li (2014): "Social Preference as an Incentive Device: Some Hidden Costs," *Management Science*, 60 (5), 1148-1160.

Hossain, Tanjim and John A. List (2012): "The Behavioralist Visits the Factory: Increasing Productivity Using Simple Framing Manipulations," *Management Science*, 58 (12), 2151-2167.

James, Julia (2012): "Health Policy Brief: Pay-For-Performance," *Health Affairs*, October 11, 2012.

Jensen, Michael (2001): "Value Maximization, Stakeholder Theory, and the Corporate Objective Function," *Journal of Applied Corporate Finance*, 14 (3), 8-21.

Johnson, Ryan M., David H. Reiley, and Juan Carlos Munoz (2012): "The War for the Fare: How Driver Compensation Affects Bus System Performance," working paper, University of Arizona.

Keller, Bill (2013): "Carrots for Doctors," *The New York Times*, January 27, 2013.

Kim, Min-Taec and Robert Slonim (2012): "The Multi-Dimensional Effects of Reciprocity on Worker Effort: Evidence from a Hybrid Field-Laboratory Labor Market Experiment," IZA discussion paper no. 6410.

Kremer, Michael, Paul Glewwe, and Nauman Ilias (2010): "Teacher Incentives," *American Economic Journal: Applied Economics*, 2 (3), 205-227.

Lazear, Edward P. (2000): "Performance Pay and Productivity," *American Economic Review*, 90 (5), 1346-1361.

Leonard, Kenneth L. (2008): "Is Patient Satisfaction Sensitive to Changes in the Quality of Care? An Exploitation of the Hawthorne Effect," *Journal of Health Economics*, 27 (2), 444-459.

Leonard, Kenneth L. and M.C. Masatu (2006): "Outpatient Process Quality Evaluation and the Hawthorne Effect," *Social Science & Medicine*, 63 (9), 2330-2340.

Leonard, Kenneth L. and M.C. Masatu (2010): "Using the Hawthorne Effect to Examine the Gap between a Doctor's Best Possible Practice and Actual Practice," *Journal of Development Economics*, 93 (2), 226-243.

Levitt, Steve and John A. List (2011): "Was There Really a Hawthorne Effect at the Hawthorne Plant? An Analysis of the Original Illumination Experiments," *American Economic Journal: Applied Economics*, 3 (1), 224-238.

Lu, Feng (2012): "Multitasking, Information Disclosure and Product Quality: Evidence from Nursing Homes," *Journal of Economic and Management Strategy*, 21 (3), 673-705.

Manthei, Kathrin and Dirk Sliwka (2014): "Multitasking and Subjective Performance Evaluations – Evidence from a Field Experiment," working paper, University of Cologne.

Mas, Alexandre and Enrico Moretti (2009): "Peers at Work," *American Economic Review*, 99 (1), 112-45.

Marschke, J. (1996): "Incentives in a Government Bureaucracy," mimeo, University of Chicago.

Mullen, Kathleen J., Richard G. Frank, and Meredith B. Rosenthal (2010): "Can You Get What You Pay for? Pay-for-Performance and the Quality of Healthcare Providers," *RAND Journal of Economics*, 41 (1), 64-91.

Muralidharan, Karthik and Venkatesh Sundararaman (2013): "Contract Teachers: Experimental Evidence from India," NBER Working Paper 19440.

Paarsch, Harry and Bruce Shearer (2000): "Piece Rates, Fixed Wages and Incentives Effects: Statistical Evidence from Payroll Records," *International Economic Review*, 41 (1), 59-92.

Prendergast, Canice (1999): "The Provision of Incentives in Firms," *Journal of Economic Literature*, 37 (1), 7-63.

Shearer, Bruce (2004): "Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment," *Review of Economic Studies*, 71 (2), 513-34.

Slade, Margaret (1996): "Multitask Agency and Contract Choice: An Empirical Exploration," *International Economic Review*, 37 (2), 465-486.

Appendix

Translation of Letters Sent to the Workers in Session 5

The letter used for the treatment group

Hello, *Name of the worker.*

Thanks for your unceasing hard work. We are glad to let you know that you have been chosen into a short-term program. In the following *7* working hours today, for the current clock-wedging job, we will count your production after each working hour. In any of the working hour, for the part of your production exceeding 700 units, you will receive a reward at the rate of RMB 0.2 per unit. We will take into account the non-work minutes when calculating your productivity. The payment will be made in early June.

For example, during the hour of 9-10 am, you produce 720 units with 3 non-work minutes. Then in this hour, you will obtain a reward payment for the following amount:
$[720\times 60/(60–3) – 700]\times 0.2$ = RMB 11.6

The reward for each working hour today will be calculated in the similar way. This reward scheme only lasts for today.

Warm regards.

The letter used for the control group

Hello, *Name of the worker.*

Thanks for your unceasing hard work.

Warm regards.

Figure 1: Percentage Deviation from Mean Baseline Productivity during Bonus and Control Rounds

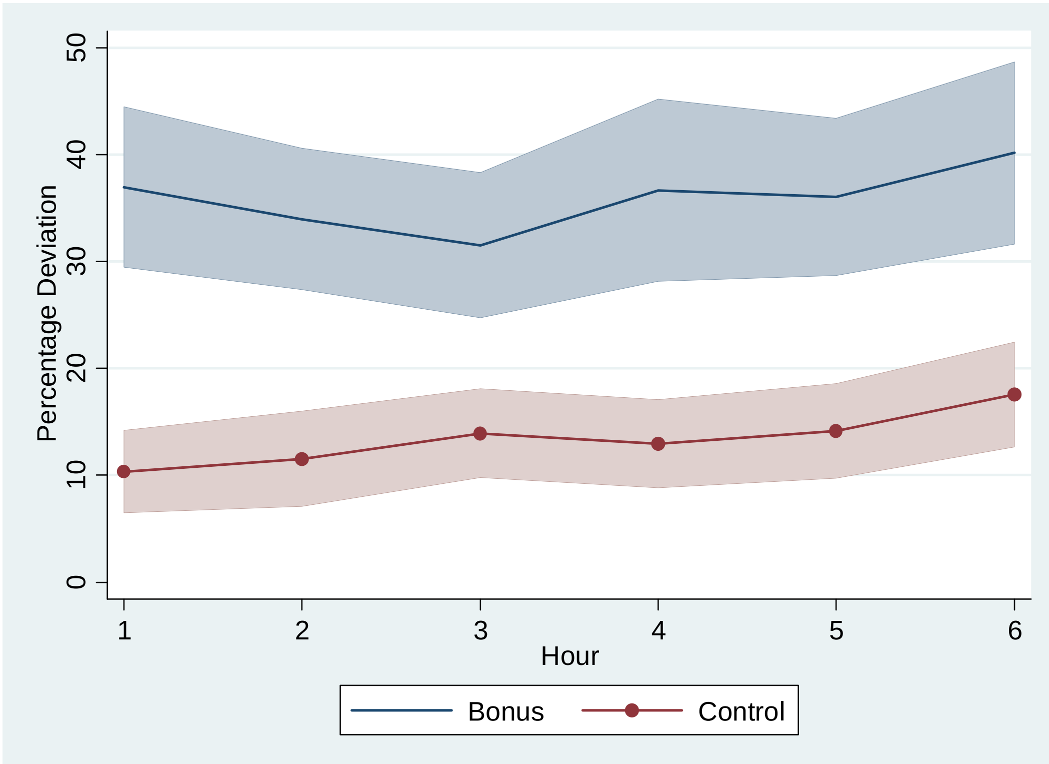

Figure 2: Average Defect Rates across All Sessions under Different Treatments

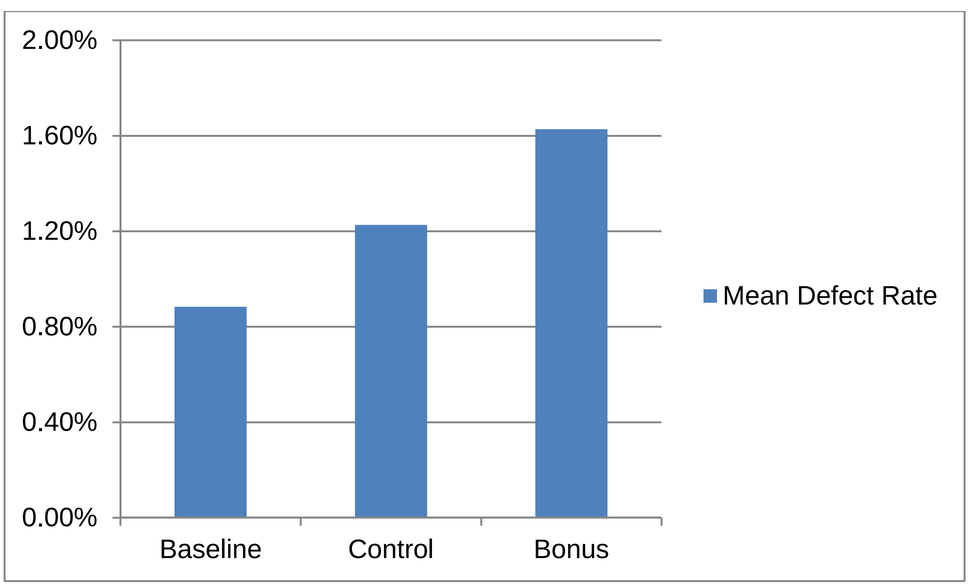

Table 1: A Summary of Experimental Design

Session	Company	Task	Base Salary Structure	Number of Subjects	Baseline Round Duration (Hours)	Bonus/Control Round Duration (Hours)	Mean Baseline Productivity	SD Baseline Productivity	Bonus Scheme	
									Target (y)	Piece Rate (α_1) (RMB/unit)
1	Wanlida (Wanlida Group Co.)	GPS attachment packaging	Flat Rate	15	5	7	95.1	18.74	120	1.25
2	Jiali (Zhangzhou Jiali Electronic Co., Ltd.)	Clock component wedging	Piece Rate	14	8	8	608	96.5	700	0.2
3	Shike (Shike Alarm System Electronic Co., Ltd.)	Alarm device circuit plugging and labeling	Flat Rate	10	4	6	24	6.9	33	1
4	Wanlida	GPS attachment packaging	Flat Rate	14	3	6	121.8	17.9	150	0.6
5	Jiali	Clock component wedging	Piece Rate	17	6	7	597	113	700	0.2
6	Hengli (Zhangzhou Hengli Electronic Co., Ltd.)	Wall clock packaging	Flat Rate	10	6	7	30.6	7.7	40	0.8
7	Hengli	Wall clock safeguarding	Flat Rate	15	4	6	62.1	10.9	85	0.5
8	Jiali	Clock component wedging	Piece Rate	21	6	6	503.4	98	620	0.3
9	Heyu (Zhangzhou Heyu Electronic Co., Ltd)	Twining metallic threads in a clock module	Piece Rate	10	5	6	569	97	700	0.15

Table 2: Summary Statistics

	All Sessions			Flat Rate Base Salary			Piece Rate Base Salary		
	Baseline	Control	Bonus	Baseline	Control	Bonus	Baseline	Control	Bonus
Hourly Productivity	355.9 (262.5)	336.4 (257.5)	358.3 (253.3)	67.8 (37.7)	89.4 (51.0)	117.1 (61.8)	568.1 (110.6)	575.5 (107.7)	594.7 (101.4)
Hourly Defect Rate	0.884% (1.770%)	1.227% (3.324%)	1.628% (5.073%)	1.904% (2.353%)	2.367% (4.460%)	3.161% (6.881%)	0.132% (0.187%)	0.123% (0.177%)	0.125% (0.171%)
Percentage of Periods Meeting the Target	11.8%	36.9%	58.5%	7.9%	51.1%	87.1%	14.6%	23.2%	30.5%
Number of Workers	126	126	126	64	64	64	62	62	62
Observations	653	807	812	277	397	402	376	410	410

Note: The top number in each cell denotes the mean and the bottom number, in parentheses, denotes standard deviation.

Table 3: Treatment Effect on Productivity

	Dependent Variable: log of Hourly Productivity		
	(1)	**(2)**	**(3)**
Bonus	0.256***	0.248***	0.238***
	(0.025)	(0.024)	(0.024)
Control	0.094***	0.086***	0.076***
	(0.017)	(0.016)	(0.015)
F -test statistic: Coefficients for Bonus and Control Dummies are Equal	71.25	71.22	
p -value	< 0.0001	< 0.0001	
Time Variables Included	No	Yes	No
Observations	2272	2272	2272
R^2	0.271	0.285	0.325

Notes: This table presents fixed effects panel regressions of log hourly productivity on dummy variables denoting bonus and control rounds. In column (2), dummy variables denoting the hour within a round are included to control for time effects. Hetero-skedasticity robust standard errors are presented inside parentheses. *, **, and *** represent significance at 10%, 5%, and 1% levels, respectively.

Table 4: Treatment Effect on the Defect Rate

	Dependent Variable: log of Hourly Defect Rate		
	(1)	**(2)**	**(3)**
Bonus	0.613**		
	(0.299)		
Control	-0.111		
	(0.307)		
log(Productivity)		1.870***	2.862***
		(0.494)	(0.864)
F -test statistic: Coefficients for Bonus and Control Dummies are Equal	10.17		
p -value	0.0018		
Instrument		No	Yes
Observations	2272	2272	2272
R^2	0.008	0.0095	0.0072

Notes: This table presents fixed effects panel regressions of the hourly defect rate on dummies denoting bonus and control rounds or log of hourly productivity. Under both specifications, dummy variables denoting the hour within a round are included to control for time effects. Hetero-skedasticity robust standard errors are presented inside parentheses. *, **, and *** represent significance at 10%, 5%, and 1% levels, respectively.

Table 5: Varying Treatment Effect According to the Base Salary Structure

	Dependent Variable: log of Hourly Productivity		
	(1)	**(2)**	**(3)**
Bonus × The Base Salary is Flat Rate	0.501***	0.492***	0.475***
	(0.028)	(0.028)	(0.029)
Control × The Base Salary is Flat Rate	0.211***	0.202***	0.185***
	(0.027)	(0.027)	(0.027)
Bonus × The Base Salary is Piece Rate	0.049***	0.047***	0.049***
	(0.013)	(0.013)	(0.013)
Control × The Base Salary is Piece Rate	0.012	0.010	0.012
	(0.011)	(0.011)	(0.011)
Time Variables Included	No	Yes	No
Observations	2272	2272	2272
R^2	0.476	0.482	0.502

Notes: This table presents fixed effects panel regressions of log hourly productivity on dummy variables denoting bonus and control rounds interacted with dummy variables indicating the base salary structure. In column (2), dummy variables denoting the hour within a round are included to control for time effects. Hetero-skedasticity robust standard errors are presented inside parentheses. *, **, and *** represent significance at 10%, 5%, and 1% levels, respectively.

Table 6: Impact of Base Salary Structure on the Defect Rate

	Dependent Variable: log of Hourly Defect Rate		
	(1)	**(2)**	**(3)**
Bonus × The Base Salary is Flat Rate	0.972***		
	(0.359)		
Control × The Base Salary is Flat Rate	0.002		
	(0.444)		
Bonus × The Base Salary is Piece Rate	0.312		
	(0.450)		
Control × The Base Salary is Piece Rate	-0.173		
	(0.416)		
log(Productivity) × The Base Salary is Flat Rate		1.696***	2.199***
		(0.485)	(0.655)
log(Productivity) × The Base Salary is Piece Rate		3.840	8.045
		(2.369)	(5.779)
Instruments		No	Yes
Observations	2272	2272	2272
R^2	0.009	0.010	0.007

Notes: This table presents fixed effects panel regressions of the hourly defect rate on bonus and control dummies and log productivity decomposed with respect to the base salary structure. Under both specifications, dummy variables denoting the hour within a round are included to control for time effects. Hetero-skedasticity robust standard errors are presented inside parentheses. *, **, and *** represent significance at 10%, 5%, and 1% levels, respectively.

Table 7: The Hawthorne Effect

	Dependent Variable: log of Hourly Productivity		
	(1)	**(2)**	**(3)**
Bonus × The Worker Received Bonus First	0.207***	0.347***	0.079***
	(0.023)	(0.022)	(0.020)
Control × The Worker Received Bonus First	0.082***	0.164***	0.015
	(0.023)	(0.039)	(0.018)
Bonus × The Worker Received Control First	0.290***	0.622***	0.021
	(0.044)	(0.038)	(0.015)
Control × The Worker Received Control First	0.091***	0.216***	0.010
	(0.022)	(0.036)	(0.014)
Sessions Included	All	Flat rate Base Salary Only	Piece rate Base Salary Only
Observations	2272	1076	1196
R^2	0.295	0.574	0.117

Notes: This table presents fixed effects panel regressions of log hourly productivity on dummy variables denoting bonus and control rounds interacted with dummies to indicate whether the bonus treatment was used first or second. Column (1) presents result with all sessions. Columns (2) and (3), respectively, use sessions with flat rate and piece rate base salary only. Under all specifications, dummy variables denoting the hour within a round are included to control for time effects. Hetero-skedasticity robust standard errors are presented inside parentheses. *, **, and *** represent significance at 10%, 5%, and 1% levels, respectively.

Table 8: Persistence of the Hawthorne Effect

	Dependent Variable: log of Hourly Productivity
First Half of Bonus	0.251***
	(0.025)
Second Half of Bonus	0.262***
	(0.027)
First Half of Control	0.086***
	(0.017)
Second Half of Control	0.102***
	(0.018)
F -test statistic: Bonus in Both Halves Have Equal Coefficients	1.10
p -value	0.2956
F -test statistic : Control in Both Halves Have Equal Coefficients	2.63
p -value	0.1075
Observations	2272
R^2	0.272

Notes: This table presents fixed effects panel regressions of log hourly productivity on dummies denoting bonus and control rounds and those dummies divided into two halves. Hetero-skedasticity robust standard errors are presented inside parentheses. *, **, and *** represent significance at 10%, 5%, and 1% levels, respectively.