



Management Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

The Behavioralist Visits the Factory: Increasing Productivity Using Simple Framing Manipulations

Tanjim Hossain, John A. List,

To cite this article:

Tanjim Hossain, John A. List, (2012) The Behavioralist Visits the Factory: Increasing Productivity Using Simple Framing Manipulations. Management Science 58(12):2151-2167. <http://dx.doi.org/10.1287/mnsc.1120.1544>

Full terms and conditions of use: <http://pubsonline.informs.org/page/terms-and-conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2012, INFORMS

Please scroll down for article—it is on subsequent pages



INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

The Behavioralist Visits the Factory: Increasing Productivity Using Simple Framing Manipulations

Tanjim Hossain

Rotman School of Management, University of Toronto, Toronto, Ontario M5S 3E6, Canada, tanjim.hossain@utoronto.ca

John A. List

Department of Economics, University of Chicago, Chicago, Illinois 60673, jlist@uchicago.edu

Recent discoveries in behavioral economics have led to important new insights concerning what can happen in markets. Such gains in knowledge have come primarily via laboratory experiments—a missing piece of the puzzle in many cases is parallel evidence drawn from naturally occurring field counterparts. We provide a small movement in this direction by taking advantage of a unique opportunity to work with a Chinese high-tech manufacturing facility. Our study revolves around using insights gained from one of the most influential lines of behavioral research—framing manipulations—in an attempt to increase worker productivity in the facility. Using a natural field experiment, we report several insights. For example, conditional incentives framed as both “losses” and “gains” increase productivity for both individuals and teams. In addition, teams more acutely respond to bonuses posed as losses than as comparable bonuses posed as gains. The magnitude of this framing effect is roughly 1%: that is, total team productivity is enhanced by 1% purely due to the framing manipulation. Importantly, we find that neither the framing nor the incentive effect lose their significance over time; rather, the effects are observed over the entire sample period. Moreover, we learn that repeated interaction with workers and conditionality of the bonus contract are substitutes for sustenance of incentive effects in the long run.

Key words: framing effect; natural field experiment; worker productivity; loss aversion

History: Received July 6, 2011; accepted February 3, 2012, by Gérard P. Cachon, decision analysis. Published online in *Articles in Advance* July 18, 2012.

1. Introduction

One of the pillars within an entrenched branch of behavioral research is the power of framing: The manner in which a choice set is presented has been found to affect individual decision making and actions considerably. Such effects are closely related to other behavioral anomalies, such as the endowment effect (Thaler 1980), status quo bias (Samuelson and Zeckhauser 1988), and observed divergences of willingness-to-pay and willingness-to-accept measures of value (Hanneman 1991). These examples of reference-dependent decision making are broadly consistent with a notion of *loss aversion*, an insight gained from Kahneman and Tversky's (1979) prospect theory, which surmises that carriers of utility are changes relative to a neutral reference point rather than absolute levels.¹

Although considerable laboratory evidence of framing effects has accumulated in the literature, a natural inclination for many economists is to discount such results on the grounds that they reflect either poorly

designed experiments (e.g., they lack sufficient incentives for meaningful response) or are merely the result of mistakes made by inexperienced laboratory subjects who through time learn to overcome such biases (see, e.g., List 2003, 2004). Although work has begun to extend the empirical results from the lab to the field, there is limited evidence on first-order questions such as the following: Can behavioral insights, such as simple framing manipulations, have economically significant effects in the field?² This is not surprising in light of the difficulties associated with executing a clean empirical test of such phenomena in the field. When such data are available, it is difficult to separate the consequences of factors of primary interest from the host of simultaneously occurring stimuli without using randomization.

In this study, we make this step by using a natural field experiment to explore how incentive schemes can be used to increase labor productivity in a real marketplace. We specifically investigate whether workers are susceptible to framing manipulations. We

¹ Pope and Schweitzer (2011) provide an illustration of loss aversion adversely affecting even the world's best golfers. Mas (2006) shows how reference points affect workplace behavior among police officers.

² One notable exception is the work on the status quo effect, which reveals the power of the status quo when agents make retirement allocations or insurance decisions (see Samuelson and Zeckhauser 1988).

report data from a natural field experiment executed with Wanlida Group Company, a high-tech Chinese enterprise engaged in the production and distribution of consumer electronics. Wanlida is one of the top 100 electronics enterprises in China, with centers located in Nanjing, Zhangzhou, and Shenzhen, and employs over 20,000 employees. The experiment revolves around using different bonus schemes with a subset of Wanlida employees to learn if simple incentives and their concomitant frames influenced productivity, both among teams (groups) of workers and among individual workers.

During our natural field experiment, subjects engaged in their regular tasks and had standard work schedules. As per company policy, the bonus incentives were paid in addition to the base income, and employees were notified of the bonuses via personal letters. The main insights gained in the experiment come from a comparison of productivity measures across a baseline and two treatments: In the positively framed bonus ("Reward") treatment, employees are notified that if the week's average per-hour production reaches a certain threshold, a bonus is paid at the end of the pay period. In the negatively framed bonus ("Punishment") treatment, employees are *provisionally* given the bonus before the work week begins, but are notified that if the average per-hour production does not reach a certain threshold, it is retracted at the end of the pay period. Thus, in contrast with studies such as Bandiera et al. (2005) and Shi (2010), where reward and punishment treatments within field experiments were different in a real sense, the two schemes in our experiment are isomorphic, except for the frame. If loss aversion is strong among the workers, then losses will loom larger than gains and the Punishment treatment should outperform the Reward variant. Alternatively, if the workers are more invigorated by *positive* incentive schemes, then the Reward treatment should lead to a higher level of productivity.

There have been several laboratory experiments that compare incentive schemes framed as reward and punishment where subjects are not involved in a real production process. The incentive schemes dressed as carrots and sticks in Dickinson (2001) are economically different. More relevant is the experiment conducted by Hannan et al. (2005), who compare economically isomorphic incentive schemes. They find that although subjects viewed positively framed incentive schemes to be fairer than negatively framed incentive schemes (bonus and penalty contracts, respectively, in their terminology), they expended more effort under the penalty contract. Reciprocity would predict that subjects would put in more effort under the fairer contract. To explain their results, they surmise that loss aversion is stronger than reciprocity leading to the greater effectiveness

of penalty contracts. Our experiment is also designed similarly and asks similar questions. However, as we use real factory workers engaged in their usual work in a field experiment, our setting is more natural. Hence, we expect the findings to be more applicable to the labor market and in determining economic policy in general. Moreover, with real production, we can investigate impact of the treatments also on the quality, in addition to the quantity, of production. Furthermore, because our field experiment spans several months, we can go beyond an experimental design that delivers only short-run substitution effects to focus attention on both short-run and long-run effects. In addition, in a complementary set of treatments, we compare the conditional reward (performance pay) schemes with unconditional reward schemes, wherein the workers receive the reward regardless of whether they achieved their production goal. In this way, we are able to link efficiency wage theory with performance pay schemes.

We report some interesting data patterns. First, although incentives increase productivity for bonuses framed as either reward or punishment for both groups of workers and individual workers, the Punishment frame outperforms the Reward frame in both the individual and group treatments, with observed differences slightly above 1%. That is, total productivity increases by 1% when moving from the Reward to the Punishment treatment. The differences for the group treatments are statistically significant and robust to various controls, whereas the individual differences are not statistically significant at conventional levels. Hence, within this particular labor market, the overall framing effect is stronger among groups than among individuals.

Second, efficiency wages are as successful as performance pay in increasing productivity of individual workers. Third, our observed treatment effects persist over the entire sample period. For example, the productivity-enhancing effects of incentives framed as punishment persist over the entire experiment, suggesting the power of simple framing manipulations in enhancing productivity. The 1% increase in productivity purely due to the framing of incentive schemes implies that there are economically significant levers that can be pulled that can materially influence the long-term growth of firms. Finally, combining our results with the extant literature reveals that conditionality of the reward and worker reputation are substitutes: if either is in place, the worker is sufficiently incentivized. Alternatively, an unconditional reward coupled with a worker who has little reputational capital at stake leads to an unproductive work environment.

We view these results as potentially speaking to several diverse research areas. First, within economics

and management, they highlight the power of incentives and illustrate how an important insight from behavioral economics can be useful in the workplace. For instance, our experiment can be considered an example of the prevalence of loss aversion in a natural labor market using a field experiment where the treatment period lasted two months for groups and four months for individuals.³ Second, our study complements the burgeoning field of industrial psychology by expanding the available tool kit that scholars and practitioners might wish to consider to enhance plant-level productivity. In this sense, the finding that worker reputation and the conditionality of the bonus contract are substitutes is an important consideration both normatively and positively. Finally, our results speak to the literature in the broader social sciences on how social structure and institutions serve as important constraints that influence behavior (see, e.g., Landa and Wang 2001).

2. Experimental Design and Results

The experiment was conducted over the months of July 2008 to January 2009, in Wanlida's factory in Nanjing. Wanlida focuses on consumer electronics and specializes in digital AV products, notebook PCs and peripherals, GPS navigation devices, car multimedia electronics, small home appliances, communication devices, and lithium polymer batteries. Our subjects included both groups of workers producing as a team and individual inspectors working independently. The group treatments pertained to production of DVD players, digital photo frames, and associated parts, whereas our individual treatments pertained to inspections of some of these products. Within a specific set of work, all teams or workers were included in the experiment.

Table 1 provides a summary of our experimental design. The table can be read as follows: In row 1 we summarize set G-1 (denoting group 1), which includes three unique teams of workers whose task is to produce chips for DVD players. All three teams had group sizes of 14. We observed workers for one or two weeks before the experiment, and then we initiated the experiment with round 1. In round 1, teams A and B of set G-1 were in the Reward and Punishment treatments, respectively, and team C was in the Baseline. In round 2, the teams changed treatments for a four-week period. We then observed workers for one or more weeks after the experimental treatments were terminated.⁴ Average

productivity for this set prior to our experiment (base productivity) was 350 units per hour, and we set a target of 400 units per hour for receiving the bonus under the Reward and Punishment treatments.

The other five group sets were conducted similarly, with the main difference being that treatments of these sets were completed with two teams. Hence, a comparison of pre- and postexperimental productivity with productivity under treatment is the information used to measure the overall effect of bonuses on productivity for these sets. The teams were formed by Wanlida management before our experiment was run, and the teammates had been working together for some time. Moreover, compositions of workers in the teams were kept unchanged prior to, during, and after the experiment. Our main research question concerns the comparison of outcomes from the Reward and Punishment treatments. Hence, even if the incentive schemes led to workers believing that they were being observed or were *chosen* workers, that effect will be present under both Reward and Punishment treatments.

The two sets of individual inspectors had 11 and 10 workers. For sets I-1 and I-2, we had Baseline observations throughout the sample period. Moreover, to analyze the impact of unconditional bonuses to workers with long-term contracts, we also included a Gift treatment, where the inspector received an unconditional bonus for four weeks without any productivity requirement. Although not all inspectors spent time on our target work in every week, each inspector participated in the target work in at least one week of a four-week round.⁵ The inspector treatments consisted of three or four four-week rounds.

The number of workers in a set allows us to run the Gift treatment with individuals, which we could not run with groups. Another main difference between the individual and group sets is that the individual bonuses depend only on one's own productivity, whereas all members of a team have the same rate of productivity. Interaction between coworkers is not important for individuals, whereas it is potentially of high import for groups. Differences in the work structure between producers and inspectors also imply that the means of altering productivity are different for groups and individuals, as the

³ See Goette et al. (2004) and Haigh and List (2005) for other examples of loss aversion among professionals using experiments of shorter duration.

⁴ Team C of set G-1 was terminated during the first round of the four-week treatment because of a preplanned restructuring of the production process.

⁵ The amount an individual worker devoted to target work, the work for which we were paying the bonus, depended on the demand for different jobs within the factory. When an individual was not working on our target work, she worked on other jobs assigned by Wanlida. As a result, the number of observations varies across rounds on Table 3. Which worker participated in our target work was chronologically decided by the management, alleviating any concern about selection issues. The set I-2 received 12 weeks, or three rounds of treatment, because the management could offer us only 12 weeks of treatment due to a reduction in the production of the P-720 main board.

Table 1 Experimental Design

Set	Job	Number of groups	Group size	Base productivity	Target	Group	Weeks 1–4 (Round 1)	Weeks 5–8 (Round 2)	Weeks 9–12 (Round 3)	Weeks 13–16 (Round 4)
G-1	DVD player MD chip production	3	14	350	400	Team A Team B Team C	Reward Punishment Baseline	Punishment Reward		
G-2	P720 main-board plug-in	2	10	390	500	Team A Team B	Reward Punishment	Punishment Reward		
G-3	Digital photo frame bracket production	2	7	780	900	Team A Team B	Reward Punishment	Punishment Reward		
G-4	Digital photo frame packaging	2	7	780	900	Team A Team B	Reward Punishment	Punishment Reward		
G-5	Adapter plug-in	2	12	490	550	Team A Team B	Reward Punishment	Punishment Reward		
G-6	Adapter joining	2	15	780	900	Team A Team B	Reward Punishment	Punishment Reward		
I-1	DVD player main-board inspection	11	1	90	110	Inspector 1–Inspector 3 Inspector 4–Inspector 6 Inspector 7–Inspector 8 Inspector 9–Inspector 11	Reward Punishment Gift Baseline	Punishment Gift Baseline Reward	Baseline Reward Punishment Gift	Gift Baseline Reward Punishment
I-2	P720 main-board inspection	10	1	38	50	Inspector 1–Inspector 3 Inspector 4–Inspector 6 Inspector 7–Inspector 8 Inspector 9–Inspector 10	Reward Punishment Gift Baseline	Punishment Gift Baseline Reward	Baseline Reward Punishment Gift	

Notes. This table reports experimental design by sets. Each set was broken up into a number of groups (teams), each of the same group size. “Target” denotes the team’s target goal for per-hour productivity. Treatments are broken down by week number. All sets included one or two weeks of preexperiment baseline observations and one week of postexperiment baseline observation.

following paragraph shows. Hence, we analyze data from groups and individuals separately in our empirical analysis.

At this point, it is important to consider how an individual or group can alter productivity in the plant. In the case of individual inspectors in sets I-1 and I-2, there is a ready balance of products to inspect at any given time period; therefore, the inspectors can move at their own pace. Among the groups, set G-6 and a portion of set G-1 have belt lines. For these two sets, workers may adjust the speed of lines to accommodate an increase of productivity. For sets G-2 and G-5, there are guide rails that run automatically, but the pace of work is flexible in that workers can move items by hand to accommodate their working pace. Finally, there are no lines for sets G-3 and G-4, permitting workers to adjust their working pace in a flexible manner.⁶

2.1. Treatments

Because our goal was to execute a natural field experiment, we worked closely with Wanlida management

in making the treatments follow company guidelines. Under this approach, there are two reasons why our particular framing treatments might not produce results that are significantly different from one another. First, the framing treatment is a passive one. For instance, in the Punishment treatment, rather than actually giving the employees the bonus money before the work week commenced, we provisionally allocated them the bonus, to be paid at the end of the pay period. In the Punishment treatment, the relevant portion of the letters to workers read as follows:

[F]or every week in which the weekly production average of your team is below K units/hour, the salary enhancement will be reduced by RMB 80”

Conversely, in the Reward treatment, the relevant description was changed to

[Y]ou will receive a salary enhancement of RMB 80 for every week the weekly production average of your team is above or equal to K units/hour

Thus, the Punishment treatment is not a particularly powerful variant, but one the firm felt was appropriate and natural for this environment. In this manner, the effect of framing might be viewed as a lower-bound estimate.

Furthermore, in an attempt to reduce potential negative (emotional) connotations, we intentionally did not label the reduction in payment in the Punishment treatment as a “fine” or “punishment.” Instead,

⁶ Even for sets G-1 and G-6, the conveyor belt runs continuously, and worker productivity is not rigidly related to the belt speed. For instance, if all workers move faster, their productivity increases because the product is completed rather than passed along. In terms of management, there is a manager in charge of a set. They are “management officials” and do not set the pace of production in a micro sense and therefore are not included in our incentive schemes.

we were interested in making the Reward and Punishment treatments merely different framings of the same incentive program. As such, the payments were made at the same time for all teams or individuals within a set, thus eliminating any credibility or time discounting issue. The only difference between the schemes is, the workers start with a bonus of zero and can potentially earn RMB 80 in each of the four weeks under the Reward treatment, and they start with a bonus of RMB 320 and can potentially lose RMB 80 in each week under the Punishment treatment. Thus, the starting reference points are different in the two treatment letters.⁷ Still, the workers are paid only at the end of the four weeks under both treatments, and there is no difference in the timing or method of the payment to workers. The difference in framing, therefore, is extremely thinly veiled.

Second, although our experimental design relies on both between- and within-unit variations, the power of our experimental design is derived from comparing within-unit data. This is because there is heterogeneity in production both within and across sets. In light of the fact that one might consider our treatments quite transparent, our within-unit experimental design is a highly demanding test to detect significant treatment differences because workers might readily deduce that the two frames yield isomorphic payoff schedules (see MacCrimmon and Larsson 1979 for a broader discussion of this issue).

Before moving to the results summary, we should note a few other experimental particulars of interest. First, the Gift treatment followed the other treatments, but the letter contained the following passage, replacing the appropriate treatment language above:

For the next four weeks from July 28 to August 23, in addition to your standard salary, you will receive a one-time salary enhancement of RMB 320. This payment will be made on August 25.

Second, workers were never aware that an experiment was taking place, and they did not know that a treatment change would occur. The source of the salary enhancements in the letter to subjects was intentionally kept vague, and workers were not asked to do any unusual work. The electronics manufacturer itself has been casually analyzing incentive schemes to improve productivity to maintain its competitive edge. The incentives in our design are natural for our workers, and given that they were not aware that they were taking part in an experiment, our approach is considered a natural field experiment in the parlance of Harrison and List (2004). Furthermore, workers in the Baseline treatments did not receive a letter.

Third, at the spot exchange rate during the weeks of the experiment, RMB 80 equaled roughly USD 11.72.⁸ Because the average weekly salary of these workers, who are under fixed wage contracts, is between RMB 290 and RMB 375, this represents more than 20% of the weekly salary of the highest-paid worker.

Fourth, we set the targets based on the observational data that we collected before the experiment, and our conversations with management, who desired targets to be achieved in 60%–80% of cases. The base productivity and the targets are presented in Table 1. The targets were well above the average pretreatment productivity rates for all groups and individuals. At the end of each week, we received a detailed report on daily production, number of actual hours worked, and the number of units produced that were found defective for each team or individual. The main variable of interest is the average hourly production for a given week because the incentive schemes were specified for weekly per-hour productivity rate. The average productivity rate equals the total production by a group or individual inspector in a week divided by the number of hours they worked in that week. Another variable of interest is the defect rate, which equals the number of defected products divided by the number of units produced on a given week expressed in percentage terms. The subjects were officially informed of their per-hour productivity rate for the week only at the end of the week.

Fifth, we were careful to minimize information transmission between groups under different treatments. For example, different teams within a set in the group treatments were located in separate rooms, if not floors. We also asked the production managers to take steps to reduce comparison of treatments across workers under the individual inspector setting. Furthermore, all teams and individuals ultimately experienced all of the treatments by the end of the experiment, affording us both between- and within-worker variation. Importantly, production managers were unaware of our direct research hypothesis related to framing; rather, they were informed that the test was designed to understand the impact of incentives. A Mandarin-speaking representative of our research team also periodically visited the factory to ensure proper execution of the experiment, smooth transition of the rounds, and to oversee the payment to the workers after the end of a round. Finally, including the pre and posttreatment control periods, holidays, and occasional suspension of work, the experiment lasted roughly four months for groups and six months for individuals. In total, 165 Wanlida workers participated in our experiment.

⁷ English translations of sample letters to the workers are provided in Appendix A.

⁸ The average exchange rate during the experiment was RMB 1 = USD 0.1465.

2.1.1. Loss Aversion and Reward vs. Punishment Frames. The theory underlying why there might be behavioral differences between the gain and loss frame is due to Kahneman and Tversky's prospect theory. Prospect theory conjectures that an agent's net utility can be described by a value function that is (i) measured over deviations from a reference point that is assessed over some narrowly bracketed time-frame, (ii) convex for losses and concave for gains, and (iii) initially steeper for losses than for gains (Tversky and Kahneman 1991). For our purposes, consider a representative agent who derives benefits and costs as follows:

$$V(w, e, w^r) = u(w) + v(e) + R(w, w^r),$$

where u is utility over income, v is utility over effort, and r is the value function of prospect theory. Let $u(\cdot)$ be increasing and concave in w and $v(\cdot)$ decreasing and concave in e . Furthermore, note that in certain treatments of our intervention, w is a function of e , $w(e)$, where w is increasing in effort.⁹ This gives us $de^*/dw > 0$.¹⁰ We define utility derived in relation to a reference point, $R(\cdot)$:

$$R(w, w^r) = \begin{cases} r(w - w^r) & \text{if } w \geq w^r, \\ s(w - w^r) & \text{if } w < w^r, \end{cases}$$

where r is increasing and concave and s is increasing and convex. Estimates of the ratio of r and s (when linearity is assumed) have found $-s(-x)/r(x) \cong 2$ (see Tversky and Kahneman 1991). In the spirit of this finding, if workers are loss averse, then the negative utility a worker receives from a loss of x is greater in magnitude than the positive utility she receives from a gain of x for any positive x .

We compare the net utility under the two frames for a loss-averse worker who earns a base (reference) salary of RMB w per four weeks. Her net pay-offs from the incentive schemes when her productivity reaches the target in $k \in \{0, 1, \dots, 4\}$ weeks are $u(w + 80k) + r(80k)$ and $u(w + 80k) + s(80k - 320)$, respectively. Thus, her net payoff difference between the two schemes will be zero when her productivity reaches the target in all four weeks and the largest when it never reaches the target. This implies that reaching the target more frequently is more profitable, in a relative sense, under the punishment scheme. Hence, a loss-averse worker will expend more effort, leading to a higher productivity rate on average,

⁹ We assume no income effect because the effort decision is made as part of a dynamic labor supply decision where any transitory incentive to increase effort will lead to negligible changes in lifetime income.

¹⁰ We can also think of reciprocity in the effort provision decision by thinking of effort as a function of gifts from the employer, $e(g)$, where $e'(g) < 0$.

under the Punishment treatment compared to the Reward treatment.

For individual sets, we include the Gift treatment where workers receive an unconditional bonus of RMB 320 for a four-week period. This enters worker utility through e , in $v(\cdot)$. A gift, g , lowers effort cost, increasing productivity through feelings of reciprocity. This can be viewed as an example of an efficiency wage. Because the bonus is guaranteed, the presence of loss aversion should not have an impact here.

2.2. Experimental Results

Tables 2 and 3 contain a summary of the raw data—weekly per-hour productivity and defect rates, with Table 2 (Table 3) summarizing the team (individual) data at the set level. The tables can be read as follows. In Table 2 in set G-1, in the first four weeks, the Reward treatment had an average of 401 units produced, whereas the Punishment treatment had an average of 402 units produced per hour. In weeks 5–8, the Reward and Punishment treatments, respectively, had 429 and 430 units produced per hour. For a within-team assessment, one needs to compare numbers from each set diagonally. For example, for set G-1, the Punishment treatment induced 30 more units of production (430–400) per hour from one team, and for the other team the Reward treatment outperformed the Punishment treatment (429–402). In these tables, the number of observations (N) denotes the number of weeks a group or individual was subject to that particular incentive scheme.

The raw data suggest that there are important differences across treatments and that the incentive schemes are working, but a more rigorous data analysis is necessary. Upon doing so, a first result emerges:

RESULT 1. *There is evidence that framing can be used to enhance productivity, but it is much more robust for groups than for individuals.*

To provide empirical support for our first result, we employ a difference-in-differences empirical approach. Suppose the average per-hour production of team j of set i under treatment $k \in \{R, P\}$ at time t is $P_{ijt} = \mu_{ij} + \kappa_k \mu_{ij} + \eta_{it}$, where μ_{ij} is the inherent productivity level of team j in set i and η_{it} is a time-specific productivity shock to all teams in set i . We assume that the productivity increase due to an incentive scheme depends on a group or individual's inherent productivity μ_{ij} . If, instead, we assume that the incentive leads to an absolute increase, so that $P_{ijt} = \mu_{ij} + \kappa_k + \eta_{it}$, the difference-in-differences analysis, as explained below, is still applicable.

Nevertheless, we believe that it is more appropriate to allow the incentive effects to be potentially heterogeneous for different teams or individuals in an absolute sense, but the same across sets in a percentage

Table 2 Productivity and Defect Rates for Groups

	Set G-1		Set G-2		Set G-3		Set G-4		Set G-5		Set G-6	
	Round 1	Round 2	Round 1	Round 2	Round 1	Round 2	Round 1	Round 2	Round 1	Round 2	Round 1	Round 2
Reward												
Weekly productivity	400.869	428.935	402.004	433.202	909.520	928.490	830.192	915.039	558.391	555.917	791.686	893.715
(SD)	(1.393)	(7.137)	(5.530)	(15.545)	(4.561)	(22.913)	(85.637)	(13.836)	(2.334)	(3.089)	(7.760)	(54.827)
Defect rate (%)	0	0	0.507	0.313	0	0	0.004	0.006	0.141	0.163	0.088	0.066
<i>N</i>	3	4	4	4	4	4	4	4	4	4	4	4
Punishment												
Weekly productivity	401.944	430.308	424.407	440.189	911.921	908.788	930.599	860.428	562.292	556.679	901.260	788.802
(SD)	(2.701)	(14.971)	(8.844)	(10.638)	(8.298)	(7.949)	(19.328)	(79.147)	(3.381)	(4.841)	(56.561)	(12.462)
Defect rate (%)	0	0	0.642	0.369	0	0	0.005	0.014	0.121	0.136	0.092	0.073
<i>N</i>	3	4	4	4	4	4	4	4	4	4	4	4
Baseline (pre- and posttreatment periods and set G-1 team C)												
Weekly productivity	415.120		429.883		817.214		803.318		526.769		831.631	
(SD)	(29.808)		(10.120)		(86.154)		(131.993)		(38.388)		(93.615)	
Defect rate (%)	0		0.395		0		0.010		0.300		0.096	
<i>N</i>	6		4		3		3		4		4	

Note. This table reports team average weekly per-hour productivity and weekly defect rate by round, set, and treatment for groups.

sense. Because it is possible that sets had unique time-specific productivity shocks—for example, productivity depends on factors such as product-specific deadlines or supply of components that vary across sets—we allow the time-specific productivity shock to depend on the set (η_{it}).

Under this specification, $\kappa_P - \kappa_R$ quantifies the framing effect. Suppose team *A* of set *i* is under Reward and Punishment treatments in rounds 1 and 2, respectively, and the treatment sequence is reversed for team *B* of the same set. Then,

$$P_{iA1} = (1 + \kappa_R)\mu_{iA} + \eta_{i1}, \quad P_{iB1} = (1 + \kappa_P)\mu_{iB} + \eta_{i1},$$

$$P_{iA2} = (1 + \kappa_P)\mu_{iA} + \eta_{i2}, \quad \text{and} \quad P_{iB2} = (1 + \kappa_R)\mu_{iB} + \eta_{i2}.$$

This implies that

$$P_{iB1} - P_{iA1} = (1 + \kappa_P)\mu_{iB} - (1 + \kappa_R)\mu_{iA} \quad \text{and}$$

$$P_{iA2} - P_{iB2} = (1 + \kappa_P)\mu_{iA} - (1 + \kappa_R)\mu_{iB}$$

$$\Rightarrow (P_{iB1} - P_{iA1}) + (P_{iA2} - P_{iB2})$$

$$= (P_{iB1} - P_{iA1}) - (P_{iB2} - P_{iA2})$$

$$= (\kappa_P - \kappa_R)(\mu_{iA} + \mu_{iB}).$$

Hence, we must compute the across-rounds difference of the productivity differences between teams *B* and *A* to estimate the framing effect. For example, for set G-2 we find that the Punishment treatment yielded 29 more units of product $((424 - 402) - (433 - 440))$. This is identical to summing the differences in productivity between Punishment and Reward treatments across the two rounds. Results from this exercise for each of the six team sets are summarized in Figure 1. Interestingly, the figure shows that the Punishment treatment outperformed the Reward treatment in five of six sets.

For the individual inspector sets, treatments are not flipped in two consecutive rounds as was done for the group experiment. Rather, the four treatments were assigned cyclically to individuals over the four rounds. As a result, a parallel difference-in-differences analysis of the individual data is not obvious. Nevertheless, it can be easily shown that we can estimate the framing effect in exactly the same manner: sum the differences in average productivity from Punishment and Reward treatments within a round, over all rounds. This exercise produces Figure 2, which reveals a similar behavioral pattern: the Punishment treatment tends to increase productivity on average, where the effect is driven by set I-1.

To complement this ocular summary, we use the raw data to estimate the above statistical model by regressing the logarithm of weekly average per-hour productivity rate on dummy variables for the Reward and Punishment treatments with set-specific fixed effects. We also include a dummy variable

Table 3 Productivity and Defect Rates for Individuals

	Set I-1				Set I-2		
	Round 1	Round 2	Round 3	Round 4	Round 1	Round 2	Round 3
Reward							
Weekly productivity	106.033	111.771	109.729	93.677	50.463	55.862	55.990
(SD)	(6.100)	(8.926)	(5.188)	(6.328)	(0.742)	(4.063)	(0.480)
Defect rate (%)	0	0	0.023	0.129	0.010	0	0
<i>N</i>	7	4	11	3	12	3	5
Punishment							
Weekly productivity	100.274	113.221	108.689	123.316	50.855	55.096	56.205
(SD)	(10.378)	(8.254)	(6.798)	(22.954)	(0.544)	(1.025)	(0.697)
Defect rate (%)	0.036	0.006	0.050	0	0.018	0.005	0
<i>N</i>	8	11	6	4	12	9	3
Gift							
Weekly productivity	102.883	109.356	105.477	115.707	50.454	55.083	56.417
(SD)	(4.513)	(8.232)	(10.009)	(24.001)	(0.315)	(1.093)	(0.307)
Defect rate (%)	0.061	0.025	0.016	0.031	0	0.005	0
<i>N</i>	7	11	12	7	8	10	4
Baseline (treatment periods)							
Weekly productivity	103.771	105.898	105.231	109.283	41.056	54.849	55.532
(SD)	(12.169)	(4.054)	(4.485)	(30.175)	(0.916)	(0.900)	(0.611)
Defect rate (%)	0.030	0.032	0.003	0.025	0.006	0	0
<i>N</i>	8	3	8	3	8	6	5
Baseline (pre- and posttreatment periods)							
Weekly productivity		95.960				41.663	
(SD)		(18.420)				(5.178)	
Defect rate (%)		0.001				0.010	
<i>N</i>		28				23	

Note. This table reports individual average weekly per-hour productivity and weekly defect rate by round, set, and treatment for inspectors.

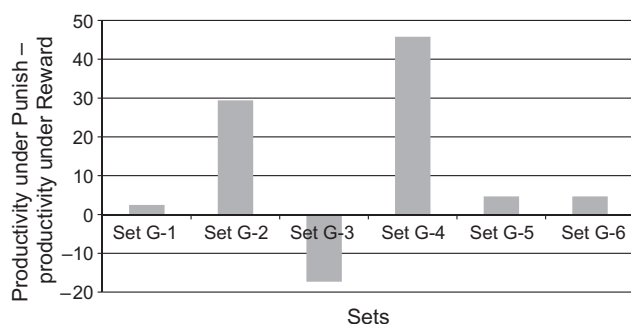
for the Gift treatment for the individual inspector models. Because of the difference between sets with teams of workers and individual workers, including this extra treatment for individuals as discussed in the introduction, we examine group and individual data separately. Moreover, we control for temporal heterogeneity by including set by week fixed effects.

Table 4 provides the summary empirical estimates. In these fixed-effects regressions and throughout the

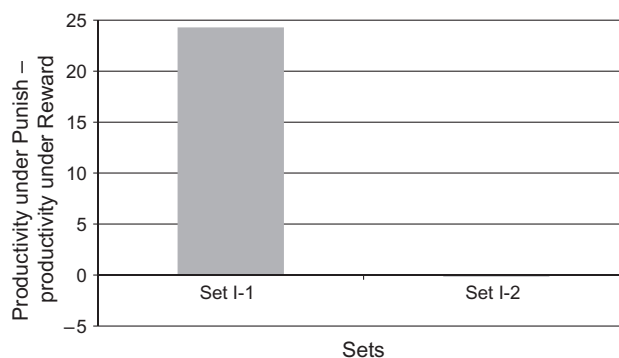
paper, we estimate the following equation or an appropriate variant:

$$\log(\text{Prod}_{ijt}) = \alpha_{ij} + \eta_{it} + \beta_1 \text{Reward}_{ijt} + \beta_2 \text{Punish}_{ijt} + \varepsilon_{ijt}. \quad (1)$$

In Equation (1), Prod_{ijt} denotes the average per-hour production of set i , team j , in week t . The dichotomous

Figure 1 Aggregate Differences in Per-Hour Productivities Under Punishment and Reward Treatments for Groups

Notes. This figure displays the aggregated differences in productivity between Punishment and Reward treatments within a set for teams. See Table 2 for absolute productivity levels of each treatment.

Figure 2 Aggregate Differences in Per-Hour Productivities Under Punishment and Reward Treatments for Inspectors

Notes. This figure displays the aggregated differences in productivity between Punishment and Reward treatments within a set individual inspectors. See Table 3 for absolute productivity levels of each treatment.

Table 4 Treatment Effects on Productivity

Dependent variable: Log of per-hour productivity on a given week						
	Groups			Individuals		
	(1)	(2)	(3)	(4)	(5)	(6)
Reward	0.0365** (0.0158)	0.0864*** (0.0293)	0.0846*** (0.0252)	0.1178*** (0.0221)	0.0561*** (0.0155)	0.0495*** (0.0160)
Punishment	0.0470*** (0.0158)	0.0969*** (0.0293)	0.0951*** (0.0251)	0.1354*** (0.0209)	0.0439*** (0.0150)	0.0308** (0.0155)
Gift				0.1259*** (0.0203)	0.0385*** (0.0146)	0.0339** (0.0146)
Set-specific time fixed effects	No	Yes	Yes	No	Yes	Yes
Team/individual-specific fixed effects	No	No	Yes	No	No	Yes
<i>N</i>	118	118	118	249	249	249
Adjusted <i>R</i> ²	0.9655	0.9950	0.9964	0.9050	0.9685	0.9708
<i>F</i> -statistic Reward = Punishment	0.66	4.53**	6.44**	0.51	0.67	1.61

Notes. This table reports empirical estimates of Punishment and Reward treatment effects using pre- and posttreatment periods as a baseline. Standard errors are displayed in parentheses below. Specifications (1) and (4) include set-specific fixed effects. Specifications (2) and (5), for groups and individuals, respectively, include time and set fixed effects, which are specific to a set and week. Specifications (3) and (6) also include group/individual-specific fixed effects.

***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

variables, Reward_{ijt} and Punish_{ijt} , denote whether team j of set i was in the Reward or Punishment treatment for week t . Both of these dichotomous variables equal zero for Baseline weeks and pre- or posttreatment weeks. The error term is denoted by ε_{ijt} . Using a log of hourly productivity as the dependent variable, we can interpret the coefficient of the treatment dummies as the percentage change in the productivity due to treatment. Here β_1 and β_2 denote the incentive effects of the Reward and Punishment treatments, respectively, and $\beta_1 - \beta_2$ denotes the framing effect.

The first column of Table 4 presents a regression estimate for groups with only set-specific fixed effects, but no group or time fixed effects. For this specification, we can replace α_{ij} with α_i and exclude η_{it} in Equation (1). In specification (2), we include set-specific time fixed effects; that is, we add η_{it} to specification (1). We further include team-specific fixed effects in specification (3), which can be exactly described by Equation (1).¹¹ The first three columns reveal that the Punishment treatment increases productivity over the Reward treatment by roughly 1% for groups. Using an *F*-test, we find that this impact is statistically significant when we include the set by time fixed effects under specifications (2) and (3). This suggests that, upon controlling for team heterogeneity and week-specific productivity shocks, framing an incentive scheme as punishment rather than as a reward induces higher productivity. The fact that the

treatment effect (of Punishment over Reward) stays the same in size but becomes statistically significant when we include time fixed effects suggests that heterogeneity in productivity shocks across weeks for a given set were quite severe.¹²

To ensure that the order of presentation of the two treatments to the same group and ultimate exposure to both treatments do not affect the result, we examine between-group variation exclusively in the first round when each team has experienced only one treatment. In that case, the treatment effect doubles in size and remains statistically significant at the $p < 0.05$ level. Because the regressions in Table 4 involve panel data, the R^2 in all the regressions are quite high.

Columns (4)–(6) in Table 4 present similar regressions for individual inspectors. Here, we also include a dummy variable for whether the individual worker was under the Gift treatment. Even though the initial estimates in column (4) suggest a similar framing effect, it is not robust to inclusion of time or group fixed effects. In fact, estimates in columns (5) and (6) suggest that the Reward treatment is more effective than the Punishment treatment. However, none of these can be distinguished from noise. Results stay unchanged qualitatively if we examine other specifications or use robust standard errors.¹³

Beyond treatment comparisons of framing manipulations, we can also explore the effect of incentives

¹¹ Panel data models using random effects instead of fixed effects yield similar insights, both quantitatively and qualitatively. These empirical results are available upon request. These results are also robust to the inclusion of lag productivity and interaction of that with the treatment dummies.

¹² This intuition is illustrated with simulated data in Appendix B.

¹³ A point to note is that treatment effects on variances in productivity is not systematic, and parametric *F*-tests of equality of variances for all eight sets together suggests that the Reward and Punishment treatments yield similar variances.

in our data. Whereas we have clean comparisons for our individual inspectors during the actual treatment period, the relevant comparison is more difficult in the group-level data. However, using the observations from pre- and postexperimental periods for all six sets, and the few weeks of Baseline treatment of team C in set G-1, we can compare the effects of merely having incentives available on productivity. Upon doing so, the next result emerges:

RESULT 2. *Our pecuniary incentives considerably enhanced productivity for both teams and individuals even when the incentives were provided unconditionally.*

First evidence to support this result can be gleaned from the raw data in Tables 2 and 3 by comparing productivity when incentives are in place versus when they are not in place. However, such comparisons do not account for time-specific productivity shocks that sets may endure. Moreover, the baseline observations in the posttreatment periods are usually higher because of some inertia from the productivity increase during the treatment periods. This leads to a number of observations where the productivity rate is lower under an incentive scheme compared to the baseline. Thus, controlling for time-specific heterogeneity is important in measuring any incentive effects.

For a more reliable measurement of the incentive effects, we return to Table 4. Recall that the treatment coefficients provide an estimate of the incentive effect while permitting the incentive treatments to vary in their success. Table 4 reveals that productivity increases in the bonus treatments for both individual and group data are sizable.¹⁴ Importantly, they are robust to inclusion of fixed effects. Specifically, for group sets, the Reward and Punishment treatments increase by 3.7%–8.6% and 4.7%–9.7%, respectively. For individual inspectors, the Reward, Punishment, and Gift treatments increase productivity by 5%–11.8%, 3.1%–13.5%, and 3.4%–12.6%, respectively. Although similar results have been found by a number of studies, it is important to quantify the impact of the bonus in terms of productivity enhancement within our experiment.¹⁵ The Gift treatment, where workers receive an unconditional gift as a one-time bonus of RMB 320, significantly increases productivity compared to the Baseline treatment. This is consistent with the findings of Bellemare and Shearer (2009). This is also supportive of the efficiency wage hypothesis in this market. Another

thing to note is that, because posttreatment weeks are included in these regressions, potential inertia in productivity makes the incentive effect estimates relatively conservative.

Note that we chose a specific target for a set and used the same target throughout our experiment. As mentioned earlier, this target was chosen based on the preexperiment average productivity and in consultation with Wanlida management. Exactly where the target is set does not seem to affect productivity; the target level has a statistically insignificant coefficient when we include it in productivity regressions. Hence, we do not include it in the regression models above. Nevertheless, if we do include the target levels, the results on incentive and framing effects remain unchanged.

For the group sets, teams reached their targets in 67% of the weeks. For individuals, the number is 69.4%. If we analyze the framing effect on the probability of meeting the target, we find a statistically insignificant positive impact of the Punishment treatment. A team and an individual were 2% and 9% more likely to reach the target, respectively, under the Punishment treatment. However, both of these coefficients are statistically insignificant. If we include a dummy variable indicating whether a team or individual had reached the target in the previous week in the productivity regressions, as in Table 4, we find no statistically significant impact. Even though there are significant productivity shocks across weeks, most of the variance in productivity arises at a level well below or well above the target level. As a result, whether the team or individual reaches the target has less of an impact on productivity. This is important because it means that our treatments are cost neutral.

2.3. Discussions

Several features of these results merit further consideration. First, given the results in the literature that report individual-level experience attenuates the effects of behavioral anomalies, it is important to consider why we find treatment effects amongst this group of seasoned workers. A key result within the previous research is that amongst agents in the field who are inexperienced, behavior varies little between them and students in lab experiments (see List 2003, 2004). Given that our experiment only implements a one-time change in the framing, and that most workers likely have little experience with treatments such as our Punishment treatment, the empirical results herein are consistent with this aspect of the previous literature that finds dramatic effects of experience.¹⁶ We cannot test the other part of the experience hypothesis directly, but note that we view this result

¹⁴ For economists, this might seem like a rather mundane result, but scholars in sister fields might find this result rather surprising—see the discussion of incentives in the workplace in Kohn (1993), for example.

¹⁵ See Bloom and Van Reenen (2010) for a survey of the empirical literature.

¹⁶ We return to the experience hypothesis later in the paper.

as highlighting that even in environments with experienced agents, if that experience does not involve the manipulation itself, it might not affect the power of that manipulation.

Second, the result that our framing manipulation is much more powerful across groups of workers than individuals merits more patient discussion. One may expect the framing effect to be stronger for individuals because free riding may be prevalent in groups. However, results in Hamilton et al. (2003) suggest that free-riding may not be a major issue in practice in such a setting. Our results are consistent with their findings. Moreover, Mas and Moretti (2009) show that peer monitoring reduces free-riding in teams. Informal monitoring by teammates can be one reason for free-riding not to be an issue among group sets in our experiment.

Whereas an individual worker works for herself, a worker in the group setting may work for her teammates in addition to herself. Thus, if a worker in the group setting identifies strongly with her team, peer effects can overcome the temptations to free-ride. Moreover, a group can become more loss averse than an individual because one may feel more strongly that she has let the teammates down when not reaching the target results in a fine. Such forms of group identity, which may explain the divergence of results from groups and individuals, can be considered as examples of identity affecting economic outcomes, as suggested by Akerlof and Kranton (2000 and 2008).

Using laboratory and field experiments, Chen et al. (2011) find that when workers are familiar with each other, workers in team-based contests exert more effort than workers in individual-based contests. They attribute this to guilt aversion in teams, which is absent when workers are rewarded for individual performance. In our setting, workers in group sets may feel a strong sense of guilt if their team fails to reach the target. An individual worker does not suffer from such guilt and, as other workers do not necessarily know whether their productivity reached the target or not, they are unlikely to suffer from social loss aversion. These aspects of social behavior in the workplace may also explain why the Punishment treatment leads to higher productivity for groups, but not for individuals. Another possibility is that because team production is some weighted mean of individual productivities, the variance of productivity is lower for groups, as suggested by van de Ven (2012).

In general, we view this result as fitting in well with the broader literature on the important role that salient properties of the situation can play. For instance, it has been shown that environmental variables such as social structure (group size, group composition, etc.) and institutional infrastructure (the

formal and informal “rules of the game”) can importantly influence behavior (see, e.g., Paese et al. 1993, Landa and Wang 2001, Stoddard and Fern 2002, and in economics, Levitt and List 2007).

Several models have been proposed to explain such data patterns, ranging from simple economic models to models of “group polarization” in psychology (Cheng and Chiou 2008) and “collective esteem” in sociology (McElroy and Seta 2006). To provide some concrete insight into why loss aversion may not be as strong among individuals in a Chinese work environment, we look at the sociocultural context of the Chinese workplace. In that quest, we turn to the cross-cultural study by Weber and Hsee (1998). They report that Chinese subjects have a lower perception of risk compared to Western subjects because of the phenomenon they termed as the “cushion hypothesis.” Because people in a socially collectivist society, such as China, are more likely to receive help if they are in financial need, losses may be less threatening because of the social financial “cushion.” Because loss aversion is closely related to risk perception, this may explain the absence of the framing effect in the individual inspector setting. Alternatively, Weber and Hsee hypothesized, and found, that the cushion hypothesis is only relevant to financial losses, not to “social losses,” presumably because a social network can provide more financial help than social help. Social loss such as “losing face,” in fact, is stronger in a collective culture than an individualistic one.

As a result, the shame of potentially publicly being regarded to be the reason why the team is punished may loom larger than being complimented for working extra hard to contribute to the team’s success in achieving the bonus. In a team setting, social standing may become more important than money, and that can lead to the frame being more effective in groups. Moreover, a loss-averse worker might be more vigilant in making sure that his or her team does not incur a “fine.” Clearly, larger groups are more likely to contain at least one highly loss-averse worker than smaller groups; *ceteris paribus*, effectively making teams more susceptible to loss aversion than individuals.

Although our experimental design cannot parse such differences directly, we can delve deeper into this result by exploring whether there are observable differences between workers in the group and individual treatments that might explain the robustness of the framing effect for only the groups. Table 5 presents the average gender composition, age, education level, and tenure for workers across the various sets.

Individual sets had a higher percentage of male workers, and that workers in the individual sets were relatively older and had longer tenure at Wanlida.

Table 5 Demographic Data for All Sets

	Set G-1	Set G-2	Set G-3	Set G-4	Set G-5	Set G-6	Set I-1	Set I-2	Group sets	Individual sets
Percentage of male (%)	28.6 (0.074)	10.0 (0)	14.3 (0)	28.6 (0)	0 (0)	13.3 (0)	0 (0)	40.9 (0.497)	15.5 (0.105)	18.4 (0.389)
Age (in years)	21.811 (0.760)	21.475 (1.329)	23.214 (0.074)	20.250 (0.863)	22.208 (0.904)	22.967 (0.780)	24.852 (3.400)	20.523 (2.816)	21.991 (1.311)	22.908 (3.810)
Education	0.484 (0.057)	0.450 (0.155)	0.286 (0.148)	0.357 (0.074)	0.333 (0)	0.167 (0.034)	0.056 (0.231)	0.386 (0.493)	0.343 (0.141)	0.204 (0.405)
Tenure (in months)	23.445 (3.613)	37.360 (14.908)	39.486 (5.134)	28.943 (10.595)	47.95 (15.130)	36.380 (6.144)	80.933 (49.918)	27.591 (15.110)	35.852 (12.740)	56.984 (46.625)

Notes. This table reports average demographic data for all sets separately and also the aggregates for group sets and individual sets. Standard deviations are in parentheses. For education, we assign a value of -1 , 0 , or 1 depending on whether the worker attended primary, junior middle, or high or polytechnic schools. Specifically, we use primary school = -1 , junior middle school = 0 , and high school or polytechnic school = 1 . Age and tenure are as of year 2008 and July 2008, respectively.

However, individual inspectors had slightly lower levels of education than workers in groups. In Table 6, we explore the framing effect while controlling for worker characteristics. To execute a clean analysis of the framing effect, we present the results with only Reward and Punishment treatments. Similar results hold when we include the entire sample. For groups, the productivity increase in Punishment treatments (compared to the Reward treatment) is tempered as average age or tenure of workers in a group increases. On the other hand, none of the demographic characteristics has a significant effect on productivity difference between Punishment and Reward treatments for individual sets. The signs of the coefficients in the first column of Table 6, along with the demographic differences between group and individual sets, provide some support for the hypothesis that the difference in the framing effect between groups and individuals

might be due to younger and less-experienced workers in group sets.

If longer-tenured workers indeed have had more experience with loss frames in the workplace, then this result is consistent with the literature regarding the effects of experience on market anomalies. For example, in a pair of early field experiments exploring the effects of market experience on behavioral anomalies, List (2003, 2004) reports that market experience measured by activity in naturally occurring trading markets for pins, sports card, and sports memorabilia leads to behavior that is not significantly different from neoclassical expectations. Even though the thrust of List's experience results have been broadly replicated in both the lab and the field (see, e.g., Feng and Seasholes 2005, Kermer et al. 2006, Dhar and Zhu 2006, Greenwood and Nagel 2009, Gächter et al. 2009, Choe and Eom 2009, Engelmann and Hollard 2010, Seru et al. 2010, Munro and Ferreira De Sousa 2012), one nagging issue is that individual market experience is endogenous. Even though these studies attempt to parse market experience from selection, the main empirical results rely on certain modeling assumptions. List (2011) rectifies this issue by making market experience exogenous, and finds similar results of experience: experience, by itself, attenuates behavioral anomalies.

Another potential reason for not finding framing effect for individuals can be due to small sample size. However, given that the data set is from controlled experiments, the numbers of observations are not extremely small for either the group or individual sets. Moreover, there are more observations for the individual sets compared to the group sets for which we find evidence of framing effect. The results are very robust of many different specifications that we used. Thus, small sample size is less likely to be the driving force for our results. Whatever the mechanism at work in our data is, it is important for future empirical work to more fully understand the dynamics of individuals versus groups when presented with such

Table 6 Effect of Worker Characteristics on the Framing Effect

Dependent variable: Log of per-hour productivity on a given week		
	Groups	Individuals
Punishment	0.3185*** (0.1045)	−0.1265* (0.0748)
Gender × Punishment	−0.0745 (0.0451)	0.04300 (0.0562)
Age × Punishment	−0.0128*** (0.0043)	0.0054 (0.0032)
Education × Punishment	0.0144 (0.0392)	−0.0383 (0.0611)
Tenure × Punishment	−0.0006* (0.0003)	−0.0002 (0.0004)
Set-specific time fixed effects	Yes	Yes
<i>N</i>	94	98
Adjusted <i>R</i> ²	0.9968	0.9748

Notes. This table reports the effect of worker characteristics on framing effect for both groups and individuals. The estimates include time and set fixed effects, which are specific to a set and week.

***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

manipulations. We view this area as ripe for future research, because worker teams are quite common in practice.

A third area worthy of further inquiry is whether our incentives were profitable for the firm. Clearly, the framing effects are “free” in the sense that once an optimal scheme and reward amount is determined, framing can be used to induce a greater level of achievement, yielding greater profits conditional on similar success rates. We can go further by computing back-of-the-envelope numbers to determine whether our particular incentive scheme was profitable.

A first consideration is that even though workers increased productivity, they might have produced more defects or missed important defects in the case of the individual inspectors. Such a result can potentially limit, or reverse, our measured productivity gains. Wanlida had in place rigorous quality checks for the group-level production, and workers were well aware of such checks. However, Wanlida did not “inspect the inspectors” formally prior to our experiment, because the inspected products would be tested when they are used in the next step, but the identity of the component and the inspectors were not clearly mapped. Thus, one would not be able to determine the exact inspector who had allowed a faulty component to continue in the production process. With our help before the experiment, Wanlida commenced keeping records that link an inspected component with its inspector, allowing us to measure the missed defects of each inspector precisely. Inspectors were made aware of this change in company policy before the experiment began.

Raw defect rates summarized in Tables 2 and 3 provide the percentage of faulty production. A quick look at the default rates does not suggest a large difference across treatments, suggesting that observed productivity increases were not importantly limited by quality deficiencies. To formally test this hypothesis, we regress the defect rate on the log of the hourly productivity and the treatments, controlling for the set-specific time fixed effects. Table 7 summarizes these results, and shows that in neither the group nor the individual sets does the productivity level or the treatments have any statistically significant impact on the quality of the product.¹⁷ This leads to our third result:

RESULT 3. *There was no discernable change in product defects or faulty inspections associated with the change in incentives.*

Given that the observed productivity increase was not accompanied by a perverse change in product

Table 7 Effect of Productivity on Defect Rates

Dependent variable: Defect rate (in percentage terms) in a given week		
	Groups	Individuals
Log of hourly productivity	0.5269 (0.3241)	0.0034 (0.0389)
Reward	−0.0525 (0.0769)	0.0026 (0.0091)
Punishment	−0.0441 (0.0782)	0.0113 (0.0087)
Gift		0.0090 (0.0085)
Set-specific time fixed effects	Yes	Yes
<i>N</i>	118	249
Adjusted <i>R</i> ²	0.8982	0.1301

Notes. This table reports the effect of productivity and treatment on quality (defect rates) for both groups and individuals. These estimates include set- and week-specific fixed effects. None of the above coefficients is statistically significant at even the 10% level.

quality, the next issue pertains to whether the increased productivity materially affected Wanlida’s bottom line. We offer a simple estimate of profitability of the scheme assuming that the productivity increase compared to preexperiment productivity is sustained if the workers participate on the target projects full time. Our incentive scheme increased total labor costs by RMB 64,960. That is the amount of bonus payments we made in total to group and individual workers. This compares favorably to the increased labor bill that would have resulted if the company desired to increase the output by the same amount under their old incentive regime using additional hours of work. To estimate that number, we make use of the pretreatment average productivity and the lower bound of the average cost of hiring an experienced worker of RMB 7, as Wanlida management suggested. Under these assumptions, to hire additional workers to match the extra quantity that our incentive scheme induced, the labor bill would have increased by more than RMB 69,900. Thus, our bonus treatments reduced the marginal production costs by almost 7%. This very rough estimate should be considered conservative because it does not include related costs of employment extension or new hiring, such as additional benefit payments, taxes, recruitment costs, etc.

This leads us to our next question: Would it make sense to permanently adopt an incentive structure such as the one imposed in our experiment? It is important to first determine whether the incentive and framing effects are persistent or temporary. A temporary increase in productivity may not be worth the increased cost even if the initial spike is large. Because the framing effect was significant for the groups but not individuals, we first explore the persistence of the framing effect in the group sets. Recall that in all weeks within a round, a team in a

¹⁷ Surprisingly, the *R*² for regressions with individuals is much lower than that concerning groups, unlike in other regressions presented in this paper. This may have resulted from the defect rate usually being smaller for individual sets than the group sets.

Table 8 Framing Effect over Time for Groups

Dependent variable: Log of per-hour productivity on a given week for groups				
	(1)	(2)	(3)	(4)
Reward	0.0835** (0.0317)	0.0830** (0.0310)	0.0829** (0.0315)	0.0837** (0.0324)
Punishment	0.0998*** (0.0317)	0.1003*** (0.0310)	0.1004*** (0.0315)	0.0996*** (0.0324)
Week included from round 2	Week 5	Week 6	Week 7	Week 8
Set-specific time fixed effects	Yes	Yes	Yes	Yes
<i>N</i>	82	82	82	82
Adjusted <i>R</i> ²	0.9939	0.9942	0.9940	0.9937
<i>F</i> -statistic Reward = Punishment	5.86**	6.82**	6.84**	5.34**

Notes. This table reports the effect of framing over time for groups with baseline and pre- and post-treatment periods included. The sample in specification (*t*) includes round 1 and the *t*th week of round 2 with *t* from 1 to 4. Standard errors are displayed in parentheses below the coefficients. These estimates include set and week-specific fixed effects.

***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

set was under the same treatment, and the treatments switched from Reward to Punishment or vice versa in round 2 (for sets G-1 to G-6).

Because there is heterogeneity in productivity across teams within a set, simply comparing productivities within sets over time is not useful. Instead, to investigate the incentive and framing effects over time for groups, we examine both within-set productivity changes over time as well as within-group productivity differences. From these exercises, the final result emerges:

RESULT 4. *Neither the incentive nor the framing effect wanes through time for groups.*

As a first test of whether the treatment effects wane over time, we compare results from the regression models presented above with models that exclude weeks of data. For example, in column (1) of Table 8, we only include observations from weeks 1–5 and the pre- and postexperiment periods. In column (2), we only include observations from weeks 1–4, week 6, and the pre and postexperiment periods. Similarly, we use data from weeks 7 and 8 along with round 1 data in columns (3) and (4), respectively. If there is significant waning of the treatment effect over time, the regression results in the four columns should lead to a systematic pattern in the coefficients.

We find no trend in the coefficients and the framing effect (the difference in the Punishment and Reward coefficients) stays remarkably unchanged in all four columns. As a robustness check, if we examine two weeks of round 2 along with round 1, that is, weeks 5

and 6 with round 1 and weeks 7 and 8 with round 1, we again find that the framing and incentive effects are equally strong in both regressions. Similar results are observed when we include group-level dummy variables.

Another approach to investigate the path of the treatment effects across time is to interact the treatment dummy with a dummy for the *t*th week within a round (rounds 1 or 2) where *t* equals 1 through 4. If treatment effects wane over time, the incentive and framing effects in weeks 1 and 5 will be larger than those in weeks 2 and 6, which, in turn, will be larger than those in weeks 3 and 7. Again, we do not observe any time trend on the treatment effects, although sometimes we do not get statistically significant treatment effects as number of observations becomes small in certain cases. Because these results are qualitatively the same as those in Table 8, we do not present them here, but we make them available upon request.

Given the results summarized in Loewenstein (2005) and, more recently, the labor market results in Gneezy and List (2006) and Lee and Rupp (2007), as well as the Hennig-Schmidt et al. (2010) field experiment, one might have suspected that our treatment effect would wane over time. Importantly, these labor market results are completed in *unconditional* rather than *conditional* reward/punishment space, and they are typically within one-shot work environments or weaker reputational environments than our repeated setting. We suspect that each of these features alone has the power to attenuate the waning effect observed in the literature, and together they are particularly powerful. Accordingly, we view this final result as providing a boundary condition on the insights gained in this literature.

We can provide further insights on this boundary condition by exploring the time path of productivity under the Gift treatment for individuals. We execute similar tests as in Table 9 using the individual inspector data. Table 9 shows that there is little evidence of a time trend in the observed treatment effect for individual workers. Coefficients of the three treatment dummies show no systematic trend whether we look at the first round and the first, second, third, or fourth weeks of the following rounds. Thus, effects of any of the incentive schemes do not wane over time. Because the Gift treatment uses unconditional rewards but takes place in a repeated game setting, these data are unique in the sense that we can test for a waning effect in the typical work setting over an unconditional bonus.¹⁸ Using the array of tests discussed above, we observe little evidence of waning

¹⁸ This is in contrast to the data from Gneezy and List (2006), where temporary workers participate in a gift exchange game over only a two-day period.

Table 9 Framing Effect over Time for Individuals

Dependent variable: Log of per-hour productivity on a given week for individuals				
	(1)	(2)	(3)	(4)
Reward	0.0845*** (0.0213)	0.0963*** (0.0226)	0.0901*** (0.0254)	0.0961*** (0.0234)
Punishment	0.0689*** (0.0214)	0.0820*** (0.0220)	0.0646*** (0.0240)	0.0789*** (0.0230)
Gift	0.0689*** (0.0220)	0.0742*** (0.0224)	0.0700*** (0.0243)	0.0716*** (0.0232)
Weeks included from rounds 2, 3, and 4	5, 9, and 13	6, 10, and 14	7, 11, and 15	8, 12, and 16
Set-specific time fixed effects	Yes	Yes	Yes	Yes
<i>N</i>	150	141	137	136
Adjusted <i>R</i> ²	0.9616	0.9636	0.9615	0.9602
<i>F</i> -statistic Reward = Punishment	5.86**	6.82**	6.84**	5.34**

Notes. This table reports the effect of framing over time for groups with baseline and pre-and posttreatment periods included. The sample in specification (*t*) includes round 1 and the *t*th week of rounds 2 to 4 with *t* from 1 to 4. Standard errors are displayed in parentheses below the coefficients. These estimates include set and week-specific fixed effects.

***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

in the impact of the Gift treatment. The workers in our experiments have long-term relationships with the factory, and this alone may have induced the persistence of incentive effect of even unconditional bonuses. We therefore conclude that reputational and relational considerations are important when adopting unconditional reward structures, and that conditionality and reputation serve as substitutes.

3. Conclusions

Understanding the sources of productivity differences across space and time remains an important task. Interestingly, total factor productivity ratios of 3:1 or more are not unusual across 90th percentile to 10th percentile producers within four-digit SIC industries. Syverson (2011) provides a discussion of the determinants of productivity and the underlying productivity differences observed at the microlevel, but a missing component of the vast productivity literature is a causal test of the effects of what behavioral economists might deem as first order. At the same time, whether and to what extent observations from the lab spill over to the field remains a central issue within the experimental sciences.

In this paper, we combine the literatures on understanding productivity enhancements with behavioral economics to explore whether a foundational insight gained from the latter literature can speak to the former. We find that it can: A simple framing manipulation changed productivity by roughly 1% for teams of workers. Economic significance of this difference

is clearer when we recall that this increase in higher productivity comes at no extra cost, rather, only from the language of the contract. Of course, there is much productivity variation not accounted for by such simple manipulations, but the study showcases that productivity gains can be had in the workplace by recognizing insights gained within the experimental and behavioral communities.

This study presents one of the first natural field experimental investigations of framing effects in labor productivity in the private sector. One limitation of our experiment is relatively small sample size. Similar field experiments in the labor market but on a much larger scale will be an obvious candidate for future research. Moreover, the finding that the framing effect is much stronger in groups than in individuals can be more thoroughly investigated with a larger data set. Another question that arises from our paper is how much individual contributions within a team are affected by framing of incentives. We are hopeful that future research with carefully designed field experiments to understand this issue will shed light on this issue. In a methodological sense, this paper showcases how field experimental evidence can supplement insights gained from the lab to further our understanding of important economic issues in a more practical context. Our field experiments also illustrate how simple modifications to contractual language can play a significant role on the outcomes of incentive schemes. This is another area of research that merits serious consideration.

Acknowledgments

The authors thank the editor, associate editor, and three anonymous reviewers for insightful comments that led to many improvements. Fuhai Hong provided tremendous help in the data collection. Trevor Gallen and Yana Peysakhovich provided research support. Discussions with Omar Al-Ubaydli, Charles Bellemare, Daniel Kahneman, Dean Karlan, Fahad Khalil, Marc Nerlove, Yaron Raviv, Stephan Schott, Lan Shi, and Rami Zwick led to insights that improved the study, as did comments from several seminar participants. The authors are indebted to Sean Wong, managing director of Wanlida Group, for allowing them to use his factory as their experimental lab. They gratefully acknowledge the financial support of the Hong Kong Research Grant Council [Grant 641808].

Appendix A. Summary of Letter Contents to Workers

English Translations of Sample Letters to Workers in the Different Treatments

Reward

Dear —,

We are glad to let you know that your team has been chosen into a short-term program. For the next four weeks starting from July 28, in addition to your standard salary,

you will receive a salary enhancement of RMB 80 for every week the weekly production average of your team is above or equal to K units/hour.¹⁹ This program will continue until the end of the week starting on August 18 and end on August 23. On August 25, you will receive your salary enhancement according to the above criterion.

For example, if your team produces at a rate above K units/hour in two weeks, you will receive RMB 160 on August 25.

Warm regards.

Punishment

The relevant description of the treatment was changed to:

For the next four weeks starting from July 28 to August 23, in addition to your standard salary, you will receive a one-time salary enhancement of RMB 320. This payment will be made on August 25. However, for every week in which the weekly production average of your team is below K units/hour, the salary enhancement will be reduced by RMB 80.

For example, if your team fails to produce at a rate of K units/hour in two weeks, your salary enhancement will be reduced by RMB 160. Then on August 25, you will only receive RMB 160.

Gift

The description of the treatment was changed to:

For the next four weeks from July 28 to August 23, in addition to your standard salary, you will receive a one-time salary enhancement of RMB 320. This payment will be made on August 25.

Note that the subjects received letters written in traditional Chinese, and the letters were appropriately edited for individual inspectors. Here K denotes the target level of per-hour productivity, which was the same for all teams or individuals within a set.

Appendix B. An Illustration of the Effect of Across-Time Heterogeneity in Productivity with Simulated Data

Results from Table 4 shows that the framing effect on groups is statistically significant only if we include set-specific time fixed effects. This occurs because of large across-time heterogeneity in productivity that is different for different sets. Both Reward and Punishment treatments are run in the same week. Thus, when we include set-specific time fixed effects, they absorb the across-time heterogeneity. This leads to smaller standard errors. For example, the standard error of the difference in log productivity under the two treatments goes from 0.011 to 0.005 when we include set-specific time fixed effects. The standard error falls further when team-specific fixed effects are included in specification (3). The intuition can be nicely illustrated by a simulated data set. We generated 400 observations from a uniform distribution on [500, 700]. Then,

¹⁹ Please note that a week is counted from Monday to Saturday, and we will use weekly production average within your real working hours on the target work.

for each of them, we generated another observation that equaled the original value plus an error term drawn from a uniform distribution on $[-5, 15]$. These can be thought of as productivity data from 400 weeks where the first data point corresponds to the control and the second one to the treatment data. If we regress this data without any fixed effect, we find a positive treatment effect that is not statistically significant. However, the treatment effect becomes significant (a p -value of less than 0.0005) if we include week-specific fixed effects. This illustrates how high correlation between productivities under the Reward and Punishment treatments on a given week can lead to the framing effect becoming significant for groups when set-specific time fixed effects are included under specifications (2) and (3) in Table 4. We find the same results even if we use 47 weeks of simulated observations (as is the case for our data set) instead of 400 weeks.

References

- Akerlof GA, Kranton RE (2000) Economics and Identity. *Quart. J. Econom.* 115(3):715–753.
- Akerlof GA, Kranton RE (2008) Identity, supervision, and work groups. *Amer. Econom. Rev.* 98(2):212–217.
- Bandiera O, Barankay I, Rasul I (2005) Social preferences and the response to incentives: Evidence from personnel data. *Quart. J. Econom.* 120(3):917–962.
- Bellemare C, Shearer B (2009) Gift giving and worker productivity: Evidence from a firm-level experiment. *Games Econom. Behav.* 67(1):233–244.
- Bloom N, Van Reenen J (2010) Human resource management and productivity. Ashenfelter O, Card D, eds. *Handbook of Labor Economics*, Vol. 4B (Elsevier Amsterdam), 1697–1767.
- Chen H, Lim N, Ahearne M (2011) Should managers use team-based contests? An experimental study. Working paper, University of Wisconsin–Madison, Madison.
- Cheng P-Y, Chiou W-B (2008) Framing effects in group investment decision making: Role of group polarization. *Psych. Rep.* 102(1):283–292.
- Choe H, Eom Y (2009) The disposition effect and investment performance in the futures market. *J. Futures Markets* 29(6):496–522.
- Dhar R, Zhu N (2006) Up close and personal: Investor sophistication and the disposition effect. *Management Sci.* 52(5):726–740.
- Dickinson D (2001) The carrot vs. the stick in work team motivation. *Experiment. Econom.* 4:107–124.
- Engelmann D, Hollard G (2010) Reconsidering the effect of market experience on the “endowment effect.” *Econometrica* 78(6):2005–2019.
- Feng L, Seasholes MS (2005) Do investor sophistication and trading experience eliminate behavioral biases in financial markets? *Rev. Finance* 9(3):305–351.
- Gächter S, Orzen H, Renner E, Starmer C (2009) Are experimental economists prone to framing effects? A natural field experiment. *J. Econom. Behav. Organ.* 70(3):443–446.
- Gneezy U, List JA (2006) Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. *Econometrica* 74(5):1365–1384.
- Goette L, Huffman D, Fehr E (2004) Loss aversion and labor supply. *J. Eur. Econom. Assoc.* 2(2):216–228.
- Greenwood R, Nagel S (2009) Inexperienced investors and bubbles. *J. Financial Econom.* 93(2):239–258.
- Haigh MS, List JA (2005) Do professional traders exhibit myopic loss aversion? An experimental analysis. *J. Finance* 60(1):521–534.

- Hamilton BH, Nickerson JA, Owan H (2003) Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *J. Political Econom.* 111:465–497.
- Hannan RL, Hoffman VB, Moser DV (2005) Bonus versus penalty: Does contract frame affect employee effort? Rapoport A, Zwick R, eds. *Experimental Business Research*, Vol. II (Springer, Dordrecht, The Netherlands), 151–169.
- Hanneman M (1991) Willingness to pay and willingness to accept: How much can they differ? *Amer. Econom. Rev.* 81(3):635–647.
- Harrison GW, List JA (2004) Field experiments. *J. Econom. Literature* 42(4):1009–1055.
- Hennig-Schmidt H, Rockenbach B, Sadrieh A (2010) In search of workers' real effort reciprocity—A field and a laboratory experiment. *J. Eur. Econom. Assoc.* 8(4):817–837.
- Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263–292.
- Kermer DA, Driver-Linn E, Wilson TD, Gilbert DT (2006) Loss aversion is an affective forecasting error. *Psych. Sci.* 17(8):649–653.
- Kohn A (1993) *Punished by Rewards: The Trouble with Gold Stars, Incentive Plans, A's, Praise, and Other Bribes*. (Houghton Mifflin, Boston).
- Landa JT, Wang XT (2001) Bounded rationality of economic man: Decision making under ecological, social, and institutional constraints. *J. Bioeconomics* 3(2):217–235.
- Lee D, Rupp NG (2007) Retracting a gift: How does employee effort respond to wage reductions? *J. Labor Econom.* 25(4):725–762.
- Levitt S, List JA (2007) What do laboratory experiments measuring social preferences reveal about the real world? *J. Econom. Perspect.* 21(2):153–174.
- List JA (2003) Does market experience eliminate market anomalies? *Quart. J. Econom.* 118(1):41–71.
- List JA (2004) Neoclassical theory versus prospect theory: Evidence from the marketplace. *Econometrica* 72(2):615–625.
- List JA (2011) Does market experience eliminate market anomalies? The case of exogenous market experience. *Amer. Econom. Rev.* 101(3):313–317.
- Loewenstein G (2005) Hot-cold empathy gaps and medical decision-making. *Health Psych.* 24(4):S49–S56.
- MacCrimmon KR, Larsson S (1979) Utility theory: Axioms versus paradoxes. Allais M, Hagen O, eds. *The Expected Utility Hypothesis and the Allais Paradox* (D. Riedel, Dordrecht, The Netherlands), 333–409.
- Mas A (2006) Pay, reference points, and policy performance. *Quart. J. Econom.* 121(3):783–821.
- Mas A, Moretti E (2009) Peers at work. *Amer. Econom. Rev.* 99(1):112–145.
- McElroy T, Seta JJ (2006) Does it matter if it involves my group? How the importance of collective-esteem influences a group-based framing task. *Soc. Cognition* 24(4):496–510.
- Munro A, Ferreira De Sousa Y (2012) Truck, barter and exchange versus the endowment effect: Virtual field experiments in an online game environment. *J. Econom. Psych.* 33(3):482–493.
- Paese PW, Bieser M, Tubbs ME (1993) Framing effects and choice shifts in group decision making. *Organ. Behav. Human Decision Processes* 56(1):149–165.
- Pope DG, Schweitzer ME (2011) Is Tiger Woods loss averse? Persistent bias in the face of experience, competition, and high stakes. *Amer. Econom. Rev.* 101(1):129–157.
- Samuelson W, Zeckhauser R (1988) Status quo bias in decision making. *J. Risk Uncertainty* 1(1):7–59.
- Seru A, Shumway T, Stoffman N (2010) Learning by trading. *Rev. Financial Stud.* 23(2):705–739.
- Shi L (2010) Incentive effects of piece rate contracts: Evidence from two small field experiments. *Topics Econom. Anal. Policy* 10(1):Article 61.
- Stoddard JE, Fern EF (2002) Buying group choice: The effect of individual group member's prior decision frame. *Psych. Marketing* 19(1):59–90.
- Syverson C (2011) What determines productivity? *J. Econom. Literature* 49(2):326–365.
- Thaler R (1980) Toward a positive theory of consumer choice. *J. Econom. Behav. Organ.* 1(1):39–60.
- Tversky A, Kahneman D (1991) Loss aversion in riskless choice: A reference-dependent model. *Quart. J. Econom.* 106(4):1039–1061.
- Van de Ven J (2012) Framing contracts. *J. Institutional Theoret. Econom.* 168(1):89–93.
- Weber EU, Hsee C (1998) Cross-cultural differences in risk perception, but cross-cultural similarities in attitudes towards perceived risk. *Management Sci.* 44(9):1205–1217.