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

Tanjim Hossain and John A. List*

October 31, 2009

Abstract

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 that spans a 6-month period, we report several insights. For example, 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 the 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 importance over time; rather the effects are observed over the entire sample period.

JEL Codes: C93 (Field Experiments), J24 (Labor Productivity), D03 (Behavioral Economics)

Key words: framing, field experiment, worker productivity

---

*Hossain: Rotman School of Management, University of Toronto and List: Department of Economics, University of Chicago. We thank Fuhai Hong for tremendous help in the data collection. Trevor Gallen and Yana Peysakhovich provided research support. Discussions with Omar Al-Ubaydli, Fahad Khalil, Marc Nerlove, Yaron Raviv and Lan Shi led to insights that improved the study, as did comments from several seminar participants. We are indebted to Mr. Sean Wong, Managing Director of Wanlida Group, for allowing us to use his factory as our experimental lab. We gratefully acknowledge the financial support of Hong Kong Research Grant Council. Please direct all correspondence to Tanjim Hossain (tanjim.hossain@utoronto.ca).

One of the pillars within an entrenched branch of behavioral research is the power of framing: the manner in which a decision is presented has been found to affect individual actions considerably. Such effects are commonly classified under the rubric of "anchoring" in the psychology literature, and include famous examples 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). Such empirical anomalies 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.

Although considerable laboratory evidence consonant with dramatic anchoring effects has accumulated in the literature,[1] 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 a mistake made by inexperienced laboratory subjects who through time learn to overcome such biases (see, e.g., List, 2003, 2004). While work has begun to extend the empirical results from the lab to the field, there is limited evidence on first-order questions such as: 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. When such data are

[1] The interested reader should see Epley (2004) for an excellent overview of the literature and the theories underlying how and why anchoring influences decisions. For a recent clever example of how framing can influence choice in the lab, please see Ellingsen et al. (2008).

[2] One notable exception is 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).

available, it is difficult to separate out the consequences of factors of primary interest from the host of simultaneously occurring stimuli.

In this study, we report data from a natural field experiment executed with Wanlida Group Co., 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 revolved 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 six-month long experiment, subjects engaged in their regular tasks, and work schedules within their normal environments. As per company policy, the bonus incentives were paid in addition to the base income, and employees were notified of treatments 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. In this way, the bonus schemes are isomorphic, except for the frame. Nevertheless, prospect theory conjectures that since losses loom larger than gains, the punishment treatment should outperform the reward variant. Alternatively, if workers are

more invigorated by *positive* incentive schemes, the reward treatment should lead to a higher level of productivity.

We report some interesting data patterns. First, incentives increase productivity for bonuses framed as either reward or punishment for both groups of workers and individual workers. Second, the punishment frame outperforms the bonus 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 much less robust. Thus, our experiment provides example of a real world scenario where behavioral biases are stronger among groups than among individuals. Finally, we observe such effects over the entire sampling period, suggesting the power of simple framing manipulations in enhancing productivity. A sustained 1% increase in productivity, purely due to the framing of the incentive scheme, implies a large long-term growth of the economy, thus making this impact economically quite significant.

We view these results as potentially speaking to several diverse research areas. First, within economics, they highlight the power of incentives and illustrate how an important insight from behavioral economics can be useful in the workplace. Second, they complement 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. Finally, they speak to the literature in the broader social sciences on how social structure and institutions serve as important constraints influencing behavior (see, e.g., Landa and Wang, 2001).

II. 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, while our individual treatments pertained to inspections of some of these products.

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 3 unique teams of workers whose task it 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, from July 28-August 22, Team *A* of set G-1 was in the Reward treatment, Team *B* was in the Punishment treatment, and Team *C* began in the baseline. In Round 2, which started on August 25, the teams changed treatments for a four week period. We then observed workers for at least one week after the experimental treatments were terminated.[3]

The other five group sets were conducted similarly, with the main difference being that treatments of these sets were completed with 2 teams. Hence, a comparison of pre- and post-experimental productivity with productivity under treatment is the information used to measure the overall effect of bonuses on productivity for these sets.

[3]Team *C* of set G-1 was terminated during the first round of the 4-week treatment because of a pre-planned re-structuring of the production process.

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, because Wanlida was also interested in unconditional bonuses, we also included a Gift treatment, where the inspector received an unconditional bonus for 4 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 4-week round.[4] In the inspector treatments, the periods consisted of three or four 4-week rounds. The other main difference between the individual and team 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.

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 product to inspect at any given time period, therefore the inspectors can move at their own rate. 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.[5]

---

[4] 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, they worked on other jobs assigned by Wanlida. As a result, the number of observations across rounds is different, as seen on Table 3. Moreover, the set I-2 received 12 weeks, or 3 rounds of treatment, as the management could offer us only 12 weeks of treatment due to a reduction in the production of the P-720 mainboard.

[5] 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 product is completed rather than passed along. In terms of management, there is a manager in charge of a

A. Treatments

Since 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. For example, in the punishment treatment, the relevant portion of the letter read:

> "for every week in which the weekly production average of your team is below 400 units/hour, the salary enhancement will be reduced by RMB 80……"

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

> "you will receive an RMB 80 bonus for every week the weekly production average of your team is above or equal to 400 units/hour……"

Thus, the punishment treatment is not a particularly powerful variant, but one the firm felt was appropriate and natural for this environment.

Further, we intentionally did not call the reduction in payment in the punishment treatment a fine or punishment to reduce potential negative (emotional) connotations. Instead, 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 differences are, therefore, extremely thinly veiled as there is no difference in the timing or method of the payment to the workers.

---

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.

Second, while our experimental design relies on both between- and within-unit variation, the power of our 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 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 this passage replacing the appropriate treatment language above:

> "For the next 4 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 paid 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. As a result, such an incentive is not an alien concept to the workers. Furthermore, workers in the baseline treatments did not receive a letter when they were working within the baseline weeks. Third, at the spot exchange rate during the weeks of the experiment, RMB 80 equaled roughly USD 11.72.[6] Since the average weekly salary of the workers is between RMB 290-375, this represents more than 20% of the weekly salary of the highest-paid worker.

[6] The average exchange rate during the experiment was RMB 1 = USD 0.1465.

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. 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 variables of interest were the average hourly production for a given week as the incentive schemes were specified for weekly per hour productivity rate. This 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 for a week which equals the number of defected products divided by the number of total products the group or individual inspector produced in that week. 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 in minimizing the information transmission between groups under different treatments. Different teams within a set in the group setting were located in separate rooms, if not floors. We also asked the production managers to take steps in reducing comparison of treatments between workers under the individual inspector setting. Furthermore, all teams and individuals ultimately experienced all of the treatments by the end of the 6-month long experiment. The production managers were unaware of our direct research hypothesis related to framing, rather they were informed that the test revolved around understanding 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 post-treatment

control periods, holidays and occasional suspension of work, the entire experiment lasted roughly 6 months, and 165 Wanlida workers participated in our experiment.

B. Experimental Results

Tables 2 and 3 contain a summary of the raw data—weekly per-hour productivity and defect rates, with Table 2 (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 4 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 treatment had 429 units produced whereas the punishment treatment had 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).

The raw data suggest that there might be some important differences across treatments, and that the incentive schemes might be 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 powerful for groups than for individuals

One approach to provide empirical support for our first result is to compute a difference-in-differences estimate at the set level. 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. Here $\kappa_P - \kappa_R$ will quantify 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,

$$P_{iA1} = \left(1+\kappa_R\right)\mu_{iA} + \eta_{i1},\ P_{iB1} = \left(1+\kappa_P\right)\mu_{iB} + \eta_{i1},$$
$$P_{iA2} = \left(1+\kappa_P\right)\mu_{iA} + \eta_{i2},\ \text{and } P_{iB2} = \left(1+\kappa_R\right)\mu_{iB} + \eta_{i2}.$$

This implies that

$$P_{iB1} - P_{iA1} = \left(1+\kappa_P\right)\mu_{iB} - \left(1+\kappa_R\right)\mu_{iA} \text{ and } P_{iA2} - P_{iB2} = \left(1+\kappa_P\right)\mu_{iA} - \left(1+\kappa_R\right)\mu_{iB}$$
$$\Rightarrow \left(P_{iB1} - P_{iA1}\right) + \left(P_{iA2} - P_{iB2}\right) = \left(P_{iB1} - P_{iA1}\right) - \left(P_{iB2} - P_{iA2}\right) = \left(\kappa_P - \kappa_R\right)\left(\mu_{iA} + \mu_{iB}\right).$$

Hence, we need to 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 in 5 of 6 sets the punishment treatment outperformed the reward treatment.

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 in that we learn that the punishment treatment tends to increase productivity on average, where the effect is driven by set I-1. Dividing the difference-in-differences measures by the target productivity level, we find a measurement of the

treatment effect in percentage terms. As a non-parametric estimate of the treatment effect across the eight sets (taking groups and individuals together), we can calculate the Wilcoxon test-statistic and reject the null hypothesis of no treatment effect at 2% significance level.

To complement the ocular summary, we use the raw data to estimate a model in which we regress the logarithm of weekly average per-hour productivity rate on dummy variables for the reward and punishment treatments. We also include a dummy variable for the gift treatment for the individual inspector models. Because we have this extra treatment for individuals, we examine group and individual data separately. Since it is possible that sets had unique time-specific productivity shocks—productivity depends on factors such as product specific deadlines or supply of components which vary across sets—we control for temporal heterogeneity by including set by week fixed effects. We also experiment with using group or individual fixed effects.

Table 4 provides the empirical estimates. In these regressions and throughout the paper, we estimate the following equation or a variation of it:

$$\log\left(\text{Prod}_{ijt}\right) = \alpha_{ij} + \eta_{it} + \beta_{1i}\text{Reward}_{ijt} + \beta_{2i}\text{Punish}_{ijt} + \varepsilon_{ijt} \qquad (1).$$

Here Prod_{ijt} denotes the average per-hour production of team j of set i on week t. The dummy variables Reward_{ijt} and Punish_{ijt} denote whether team j of set i was under Reward or Punishment treatment on week t. Both these dummies equal zero for baseline treatments and pre or post-treatment weeks. The error term is denoted by ε_{ijt}. Using 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 effect. The statistical significance levels of the coefficients do not change if we use absolute

productivity as the dependent variable instead. The first column of Table 4 presents baseline regressions for groups with only set-specific fixed effects, but no group or time-specific 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 and week-specific fixed effects; that is, we include η_{it} to specification (1). We further include group-specific fixed effects in specification (3), which can be exactly described by equation 1.[7] 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.

Columns 4 to 6 present similar regressions for individual inspectors. Here we also include a dummy variable for whether the individual worker was under the Gift treatment. These results are mixed. Rather than finding a significant effect, as Figure 2 would have suggested, even though the baseline estimates in column 4 suggest a similar framing effect, this result is not statistically significant. In addition, it is not robust to inclusion of set-specific time fixed effects or group fixed effects. In fact, estimates in columns 5 and 6 suggest that the reward treatment is more effective than the punishment treatment, but this cannot be distinguished from noise. Results stay unchanged

[7] Panel data models using random effects instead of fixed effects yield similar insights, both quantitatively and qualitatively. These empirical results are available upon request.

qualitatively if we examine other specifications or use robust standard errors.[8] If we look at between-group variation exclusively in the first round when each worker has experienced only one treatment, we find exactly the same qualitative result. For groups, productivity increase in punishment treatment is statistically significant with time fixed-effects while the framing effect is never statistically significant for individuals.

Beyond treatment comparisons of framing manipulations, we can also explore the effect of incentives in our data. While 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 post-experimental periods for all 6 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, a second result emerges:

Result 2: There is evidence that our pecuniary incentives considerably enhanced productivity for both teams and individuals

Evidence to support this result can be found in Table 5, where we summarize the raw data by comparing productivity when incentives are in place versus when they are not in place.[9] That is, we pool the incentive treatments for this ocular comparison. Overall, incentive treatments increased productivity for 7 out of the 8 sets, and on the top end, productivity was almost 12% and 18% higher under the incentive treatments compared to the baseline for sets G-3 and I-2. The baseline includes both pre and post-treatment

[8] 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.

[9] For economists, this might seem like a rather mundane result, but the literature contained in our sister fields might find this result rather surprising (see the discussion of incentives in the workplace in Kohn 1993), for example).

periods. Potential inertia in productivity rate, thus, makes this estimate a conservative estimate of incentive effects.

Table 5, of course, does not account for time-specific productivity shocks that sets may endure. For that, we return to Table 4. Recall that the treatment coefficients provide an estimate of the incentive effect, thus, the regression results complement the insights gained from Table 5, but permit the two 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. Importantly, they are robust to inclusion of fixed effects. For individual inspectors, the Gift treatment allows us to explore the impact of an incentive scheme that is not dependent on productivity. Recall that workers in this treatment received an unconditional gift as a one-time salary enhancement of RMB 320 for a 4-week round. The coefficient estimates suggest that even when the incentive is unconditional, it increases productivity compared to the baseline.

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 pre-experiment average productivity and in consultation with 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 level, then the incentive and framing effect results remain unchanged.

C. Discussion

Several features of these results merit further consideration. First, given the results in the literature that report individual level experience attenuates the effects of

certain anomalies such as loss aversion, 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 in our experiment we are only implementing a one-time change in the frame, and that 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. We cannot test the other part of the experience hypothesis directly, but note that we view this result as highlighting that even in environments with experienced agents, if that experience does not revolve around 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. 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, for example, 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). Intuitively, these models suggest that workers in a group setting are concerned about letting fellow

team members down. 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 6 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. However, individual inspectors had slightly lower levels of education than workers in groups. In Table 7, we explore the framing effect while controlling for worker characteristics. To execute a clean analysis of the framing effect, we examine only reward and punishment treatments to reduce confounds. 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. Alternatively, none of the demographic characteristics has a significant effect on productivity difference between punishment and reward treatments for individual sets. The sign of the coefficients in column 1 of Table 7 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. This result is consonant with the literature regarding the effects of experience on market anomalies.

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 manipulations. We view this area as ripe for future research, as 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. Yet, Wanlida did not "inspect the inspectors" formally prior to our experiment, as 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 and week-specific fixed effects. Table 8 summarizes these results, and shows that in neither the group nor the individual sets, the productivity level or the treatments have any statistically significant impact on the quality of the product. 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 quality, the next issue pertains to whether the increased productivity materially affected Wanlida's bottom line. Our incentive schemes increased total labor costs by RMB 64,960. This compares favorably to the increased labor bill that would have resulted if the company desired to increase output under their old incentive regime. To estimate that number, we make use of the pre-treatment average productivity and the low estimate of the average cost of hiring an experienced worker of RMB 7, as Wanlida management suggested. We find that under these assumptions, to match the extra quantity that our incentive scheme induced, the labor bill would have increased by more than RMB 69,900. Thus, marginal production costs were roughly reduced by 7% with our bonus treatments. This estimate does not even include related costs of employment such as additional benefit payments, taxes 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. As 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 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).

Since 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 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 9, we only include observations from weeks 1 through 5 and the pre- and post-experiment periods. In column 2, we only include observations from weeks 1 through 4, week 6, and the pre and post-experiment 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 the four columns. As a robustness check, if we examine two weeks of Round 2 along with Round 1, that is,

5th and 6th weeks with Round 1 and 7th and 8th weeks 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 tth 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. As these results are qualitatively the same as those in Table 9, we do not present them here but 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 (2008), as well as Hennig-Schmidt, Rockenbach and Sadrieh's (2006) 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 10 shows that there is little evidence of a time trend in the observed incentive or framing effects 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 round. Thus, effects of any of the incentive schemes ─ Reward, Punishment, and Gift ─ does not wane over time. Since 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. Using the array of tests discussed above, we observe little evidence of waning in the impact of the Gift treatment (see also, Al-Ubaydli et al., 2008). We, therefore, conclude that reputational considerations are important when adopting unconditional reward structures, and that the two may serve as complements.

III. 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 4-digit SIC industries. Syverson (2008) provides a discussion of the determinants of productivity and the underlying productivity differences observed at the micro-level, 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. A persistent increase in productivity, even by 1%, will have a large impact on economic growth in the long run. 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 investigations of framing effect in labor productivity in the private sector. In a methodological sense, it 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 that simple modification of contractual language can play a significant role on the outcomes of incentive schemes. This is another area of research that we plan on tackling in the future.

References

Al-Ubaydli, Omar, Steffen Andersen, Uri Gneezy, and John A. List, 2008. "Incentive Schemes to Promote Optimal Work Performance: Evidence from a Multi-Tasking Field Experiment," working paper, University of Chicago.

Cheng, Pi-Yueh and Wen-Bin Chiou. 2008. "Framing effects in group investment decision making: Role of group polarization," *Psychological Reports*, 102(1): 283-292.

Ellingsen, Tore, Magnus Johannesson, Sara Munkhammar, and Johanna Möllerström. 2008. "Why Labels Affect Cooperation," working paper, Department of Economics, Stockholm School of Economics.

Epley, N. 2004. "A Tale of Tuned Decks? Anchoring as Adjustment and Anchoring as Activation," In *The Blackwell Handbook of Judgment and Decision Making*, eds. D.J. Koehler & N. Harvey, Oxford, UK: Blackwell Publishers.

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.

Hanneman Michael. 1991. "Willingness to Pay and Willingness to Accept: How Much Can They Differ?" *American Economic Review*, 81(3): 635-647.

Hennig-Schmidt Heike, Bettina Rockenbach, and Abdolkarim Sadrieh. 2006. Forthcoming. "In Search of Workers' Real Effort Reciprocity – A Field and a Laboratory Experiment," *Journal of the European Economic Association.*

Kahneman, Daniel and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47(2): 263-292.

Kahneman, Daniel. 1986. 'Comments by Professor Daniel Kahneman,' In *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method*, ed. In R.G. Cummings, D.S. Brookshire and W.D. Schulze, 185-193 Totowa, N.J. Rowman and Allanheld,

Knetsch, Jack L. 1989. "The Endowment Effect and Evidence of Nonreversible Indifference Curves," *American Economic Review*, 79(5): 1277-1284.

Kohn, Alfie, 1993. "Punished by Rewards: The Trouble with Gold Stars, Incentive Plans, A's, Praise, and Other Bribes," Boston: Houghton Mifflin.

Landa, Janet T. and Xiao T. Wang. 2001. 'Bounded rationality of economic man: Decision making under ecological, social, and institutional constraints,' *Journal of Bioeconomics*, 3(2): 217-235.

Lee, Darin and Nicholas G. Rupp. 2007. "Retracting a Gift: How Does Employee Effort Respond to Wage Reductions?" *Journal of Labor Economics* 25(4):725-62.

Levitt, Steven and John A. List. 2007. "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?" *Journal of Economic Perspectives*, 21(2): 153-174.

List, John A. 2003. "Does Market Experience Eliminate Market Anomalies?" *Quarterly Journal of Economics*, 118(1): 41-71.

List, John A. 2004. "Neoclassical Theory Versus Prospect Theory: Evidence from the Marketplace," *Econometrica*, 72(2): 615-625.

Loewenstein, George. 2005. "Hot-cold empathy gaps and medical decision-making," *Health Psychology*, 24(4): S49-S56.

MacCrimmon, Kenneth R. and Stig Larsson, "Utility Theory: Axioms versus Paradoxes," In *The Expected Utility Hypothesis and the Allais Paradox*, eds. M. Allais and O. Hagen, 333-409. Dordrecht, The Netherlands: D. Riedel.

McElroy, Todd and John J. Seta. 2006. "Does it matter if it involves my group? How the importance of collective-esteem influences a group-based framing task," *Social Cognition*, 24(4): 496-510.

Paese, Paul W., Mary Bieser and Mark E. Tubbs. 1993. "Framing Effects and Choice Shifts in Group Decision Making," *Organizational Behavior and Human Decision Processes*, 56(1): 149-165.

Samuelson, William and Richard Zeckhauser. 1988. "Status Quo Bias in Decision Making," *Journal of Risk and Uncertainty*, 1(1): 7-59.

Stoddard, James E. and Edward F. Fern. 2002. "Buying Group Choice: The Effect of Individual Group Member's Prior Decision Frame," *Psychology and Marketing*, 19(1): 59-90.

Syverson, Chad. 2006. Forthcoming. "What Determines Productivity at the Micro Level?" *Journal of Economic Literature.*

Thaler, Richard. 1980. "Toward a Positive Theory of Consumer Choice," *Journal of Economic Behavior & Organization*, 1(1): 39-60.

Appendix: 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 4 weeks starting from July 28, in addition to your standard salary, you will receive an RMB 80 bonus for every week the weekly production average of your team is above or equal to *K* units/hour.[10] 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 bonus 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 4 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 paid 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 4 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 paid 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.

[10] 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.

Table 1: Experimental Design

Set	Job	Number of Groups	Group Size	Target	Group	Week 1 - Week 4 (Round 1)	Week 5 - Week 8 (Round 2)	Week 9 - Week 12 (Round 3)	Week 13 - Week 16 (Round 4)
G-1	DVD player MD Chip production	3	14	400	Team *A* Team *B* Team *C*	reward punishment baseline	punishment reward		
G-2	P720 main-board plug-in	2	10	500	Team *A* Team *B*	reward punishment	punishment reward		
G-3	Digital photo frame bracket production	2	7	900	Team *A* Team *B*	reward punishment	punishment reward		
G-4	Digital photo frame packaging	2	7	900	Team *A* Team *B*	reward punishment	punishment reward		
G-5	Adapter plug-in	2	12	550	Team *A* Team *B*	reward punishment	punishment reward		
G-6	Adapter joining	2	15	900	Team *A* Team *B*	reward punishment	punishment reward		
I-1	DVD player main-board inspection	11	1	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	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	

Table 1 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 pre-experiment baseline observations and one week of post-experiment baseline observation.

Table 2: Productivity & 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%
	N	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%
	N	3	4	4	4	4	4	4	4	4	4	4	4
Baseline (Pre and Post-Treatment 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%	
	N	6		4		3		3		4		4	

Table 2 reports team average weekly per-hour productivity and weekly defect rate by round, set, and treatment for groups.

Table 3: Productivity & 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
	N	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
	N	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
	N	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
	N	8	3	8	3	8	6	5
Baseline (Pre and Post-Treatment Periods)	Weekly Productivity	95.960				41.663		
	(SD)	(18.420)				(5.178)		
	Defect Rate	0.001%				0.010%		
	N	28				23		

Table 3 reports individual average weekly per-hour productivity and weekly defect rate by round, set, and treatment for inspectors.

Table 4: Treatment Effects on Producitivity

Dependent Variable: Log of Per-hour Productivity on a Given Week

	Groups			Individuals		
	(1)	(2)	(3)	(4)	(5)	(6)
Reward	0.0365**	0.0864***	0.0846***	0.1178***	0.0561***	0.0495***
	(0.0158)	(0.0293)	(0.0252)	(0.0221)	(0.0155)	(0.0160)
Punishment	0.0470***	0.0969***	0.0951***	0.1354***	0.0439***	0.0308**
	(0.0158)	(0.0293)	(0.0251)	(0.0209)	(0.0150)	(0.0155)
Gift				0.1259***	0.0385***	0.0339**
				(0.0203)	(0.0146)	(0.0146)
Set-Specific Time Fixed Effects	No	Yes	Yes	No	Yes	Yes
Group/Individual-Specific Fixed Effects	No	No	Yes	No	No	Yes
N	118	118	118	249	249	249
Adjusted R-squared	0.9655	0.9950	0.9964	0.9050	0.9685	0.9708
F-Statistic Reward = Punishment	0.66	4.53**	6.44**	0.51	0.67	1.61

Table 4 reports empirical estimates of punishment and reward treatment effects using pre- and post-treatment 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.

Table 5: Incentive Effects on Productivity

	Set G-1	Set G-2	Set G-3	Set G-4	Set G-5	Set G-6	Set I-1	Set I-2
Average Productivity under Baseline	415.120	429.883	817.214	803.318	526.769	831.631	100.089	45.082
(SD)	(29.801)	(10.120)	(86.154)	(131.993)	(38.388)	(93.615)	(16.503)	(7.188)
N	6	4	3	3	4	4	50	42
Average Productivity with Incentives	417.529	424.951	914.680	884.064	558.319	843.866	108.452	53.151
(SD)	(16.591)	(17.700)	(14.311)	(67.741)	(4.054)	(66.029)	(11.806)	(2.709)
N	14	16	16	16	16	16	91	66
Increase under Incentives	0.58%	-1.15%	11.93%	10.05%	5.99%	1.47%	8.36%	17.90%

Table 5 compares average per-hour productivity for the baseline treatment (including pre and post-treatment periods) against the incentive treatments for each set.

Table 6: 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	0.286	0.100	0.143	0.286	0	0.133	0	0.409	0.155	0.184
	(0.074)	(0)	(0)	(0)	(0)	(0)	(0)	(0.497)	(0.105)	(0.389)
Age (in Years)	21.811	21.475	23.214	20.250	22.208	22.967	24.852	20.523	21.991	22.908
	(0.760)	(1.329)	(0.074)	(0.863)	(0.904)	(0.780)	(3.400)	(2.816)	(1.311)	(3.810)
Education	0.484	0.450	0.286	0.357	0.333	0.167	0.056	0.386	0.343	0.204
	(0.057)	(0.155)	(0.148)	(0.074)	(0)	(0.034)	(0.231)	(0.493)	(0.141)	(0.405)
Tenure (in Months)	23.445	37.360	39.486	28.943	47.95	36.380	80.933	27.591	35.852	56.984
	(3.613)	(14.908)	(5.134)	(10.595)	(15.130)	(6.144)	(49.918)	(15.110)	(12.740)	(46.625)

Table 6 reports average demographic data for all sets separately and also the aggregates for group sets and individual sets. Standard deviations are in parentheses. For eduction: primary school=-1, junior middle school=0, high school or polytechnic school=1. Age and tenure are as of year 2008 and July 2008, respectively.

Table 7: Effect of Worker Characteristics on the Framing Effect Dependent Variable: Log of Per-hour Productivity on a Given Week		
	Groups	Individuals
Punishment	0.3185***	-0.1265*
	(0.1045)	(0.0748)
Gender * Punishment	-0.0745	0.04300
	(0.0451)	(0.0562)
Age * Punishment	-0.0128***	0.0054
	(0.0043)	(0.0032)
Education * Punishment	0.0144	-0.0383
	(0.0392)	(0.0611)
Tenure * Punishment	-0.0006*	-0.0002
	(0.0003)	(0.0004)
Set-Specific Time Fixed Effects	Yes	Yes
N	94	98
Adjusted R-squared	0.9968	0.9748

Table 7 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%,

Table 8: Effect of Productivity on Defect Rates Dependent Variable: Defect Rate in a Given Week		
	Groups	**Individuals**
Log of Hourly Productivity	0.0053	0.00003
	(0.0032)	(0.0004)
Reward	-0.0005	0.00003
	(0.0008)	(0.0001)
Punishment	-0.0004	0.0001
	(0.0008)	(0.0001)
Gift		0.0001
		(0.0001)
Set-Specific Time Fixed Effects	Yes	Yes
N	118	249
Adjusted R-squared	0.8982	0.1301

Table 8 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. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels

Table 9: 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**	0.0830**	0.0829**	0.0837**
	(0.0317)	(0.0310)	(0.0315)	(0.0324)
Punishment	0.0998***	0.1003***	0.1004***	0.0996***
	(0.0317)	(0.0310)	(0.0315)	(0.0324)
Week Included from Round 2	Week 5	Week 6	Week 7	Week 8
Set-Specific Time Fixed Effects	Yes	Yes	Yes	Yes
N	82	82	82	82
Adjusted R-squared	0.9939	0.9942	0.9940	0.9937
F-Statistic Reward = Punishment	5.86**	6.82**	6.84**	5.34**

Table 9 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.

Table 10: Framing Effect Over Time for Individuals

Dependent Variable: Log of Per-hour Productivity on a Given Week for Groups

	(1)	(2)	(3)	(4)
Reward	0.0845***	0.0963***	0.0901***	0.0961***
	(0.0213)	(0.0226)	(0.0254)	(0.0234)
Punishment	0.0689***	0.0820***	0.0646***	0.0789***
	(0.0214)	(0.0220)	(0.0240)	(0.0230)
Gift	0.0689***	0.0742***	0.0700***	0.0716***
	(0.0220)	(0.0224)	(0.0243)	(0.0232)
Weeks Included from Rounds 2, 3, and 4	5, 9, & 13	6, 10, & 14	7, 11, & 15	8, 12, & 16
Set-Specific Time Fixed Effects	Yes	Yes	Yes	Yes
N	150	141	137	136
Adjusted R-squared	0.9616	0.9636	0.9615	0.9602
F-Statistic Reward = Punishment	5.86**	6.82**	6.84**	5.34**

Table 10 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 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.

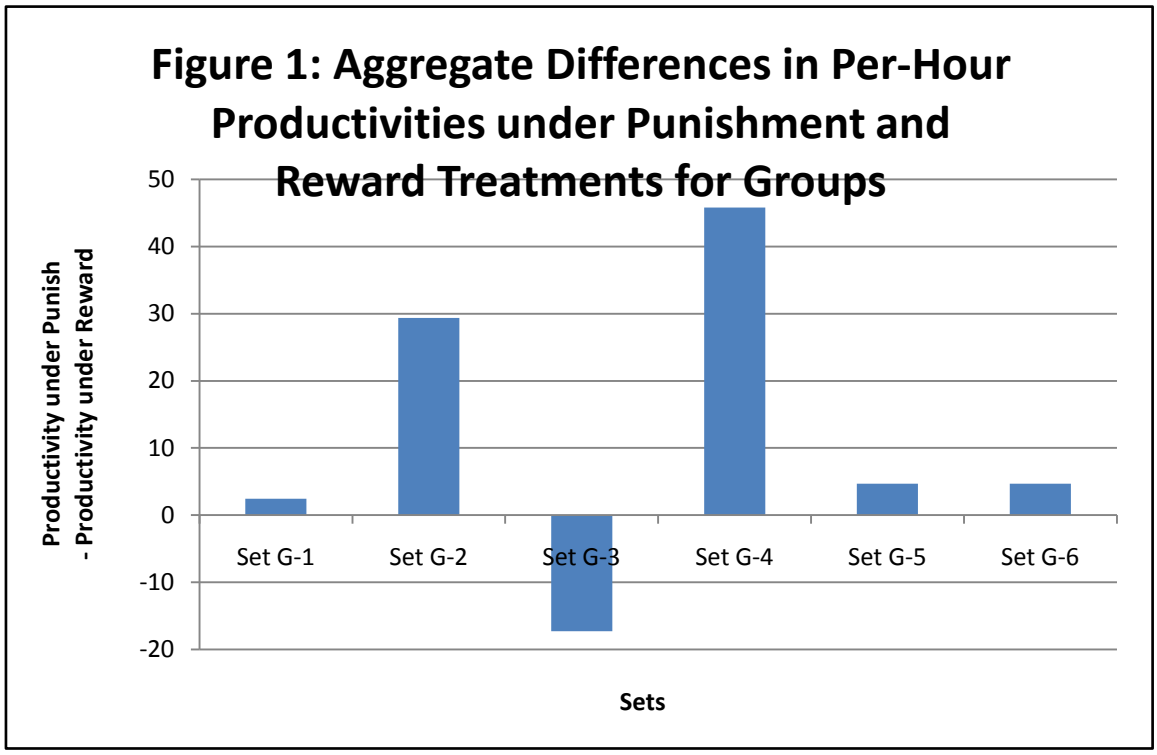

Figure 1 displays the aggregated differences in productivity between punishment and reward treatments within a set for teams. See Table 2 for absolute productivity levels of each treatment.

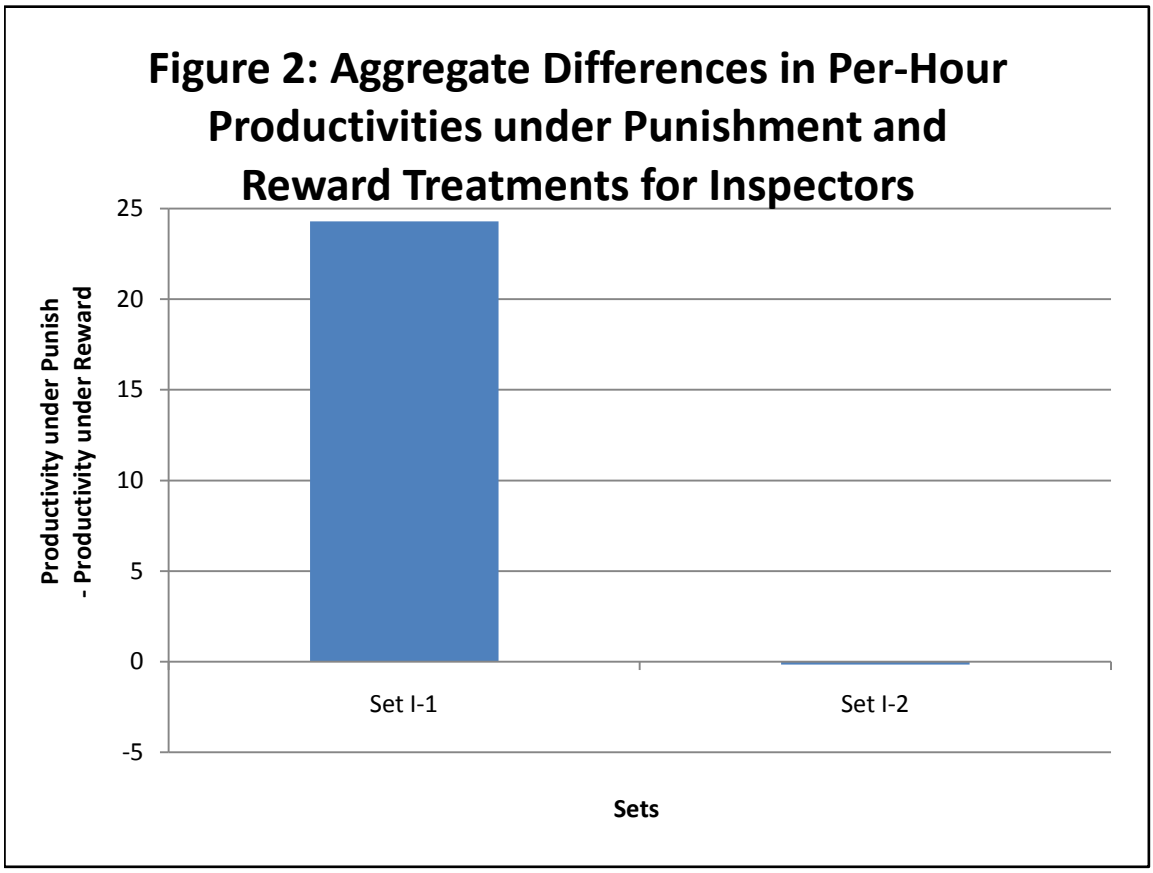

Figure 2 displays the aggregated differences in productivity between punishment and reward treatments within a set individual inspectors. See Table 3 for absolute productivity levels of each treatment.