



ELSEVIER

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo



Framing manipulations in contests: A natural field experiment

Fuhai Hong^a, Tanjim Hossain^{b,*}, John A. List^c

^a Division of Economics, Nanyang Technological University, Singapore

^b Institute for Management & Innovation and Rotman School of Management, University of Toronto, Toronto, ON, Canada

^c Department of Economics, University of Chicago, Chicago, IL, United States

ARTICLE INFO

Article history:

Received 14 February 2014

Received in revised form 26 October 2014

Accepted 22 February 2015

Available online xxx

Keywords:

Framing effect

Field experiment

Worker productivity

Contests

Loss aversion

Incentive contracts

JEL classification:

C93 (field experiments)

J24 (labor productivity)

D03 (behavioral economics)

L11 (production)

ABSTRACT

Exploiting findings that losses loom larger than gains, studies have shown that framing manipulations can increase productivity of workers. Using a natural field experiment that exogenously manipulates wage bonuses within contests in a Chinese high-tech manufacturing facility, we show that how loss aversion affects worker behavior critically depends on the incentive scheme as well as the framing manipulation. Four sets of two identical teams competed against each other to win a bonus given to the team, within a set, with the higher average hourly productivity over the week. In each set, the bonus was framed as a *reward* or *gain* for one team and as a *punishment* or *loss* for the other. Average weekly productivity was slightly higher under the *loss* treatment, but this increase was statistically insignificant. However, the team under the *loss* treatment was at least 35% more likely to win the contest. As teams' payoffs are based on relative productivity under a contest, framing effect is much stronger in terms of relative productivity. Finally, workers seemingly responded to the bonus by increasing the quality of production as well as quantity—defect rate fell as productivity increased.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Loss aversion, which suggests that losses loom larger than gains, is one of the central features of prospect theory proposed by Kahneman and Tversky (1979). Exploiting loss aversion, Hossain and List (2012) recently showed that framing manipulations can be used to increase productivity of workers even in the field with regular workers under long-term contracts. Specifically, incentive schemes that are framed as punishments increase productivity over isomorphic schemes framed as rewards. While that and other studies present ample evidence of factory and farm workers, students, teachers, and laboratory subjects being loss-averse, there is relatively little work on how variations in the incentive schemes interact with framing manipulations designed to exploit loss aversion. In this paper, we ran a natural field experiment in a Chinese high-tech manufacturing facility where sets of two identical teams competed against each other to win a bonus given to the team with the higher average hourly productivity. The bonus was framed as a *reward* or *gain* for one team and as a *punishment* or *loss* for the other. While productivity increased by a greater degree under the latter framing, this effect was

* Corresponding author. Tel.: +1 416 978 0823.
E-mail address: tanjim.hossain@utoronto.ca (T. Hossain).

statistically insignificant. However, framing the incentive scheme as loss made a team at least 35% more likely to produce at a level of productivity higher than that of a team whose incentive was posed as gain. This contrasts Hossain and List's finding that similar framing manipulations, but with bonuses based on absolute productivity, resulted in a significant framing effect in terms of absolute, but not relative, productivity. This suggests that the type of incentive scheme seems to matter as much as framing in determining how incentive contracts affect worker behavior.

A central question in industrial organization is how incentive contracts can be used to affect agents' effort choice. Standard literature typically focuses on the economic contents of an incentive contract rather than the language of the contract¹. Evidence from laboratory experiments regarding the power of framing—however, suggest that even apparently innocuous differences in the presentation of a contract may significantly affect agents' behavior. In recent years, work has begun to extend the empirical results from the lab to the field, although most studies focus on individual outcomes². In line with this strand of literature, we are particularly interested in the following first-order question: what are the effects of simple exogenous framing manipulations among competing teams of workers? Difficulties in answering this question are associated with implementing a clean empirical test of such phenomena, and separating out the consequences of factors of primary interest from the host of simultaneously occurring stimuli.

Rather than relying on observational data, we approach this question by executing a natural field experiment in partnership with the Wanlida Group Company, a large-sized Chinese manufacturing company based in the Fujian province of China. With more than 20,000 employees, it produces a variety of consumer electronics and ranks as one of the “top 100 electronics enterprises” in China. We use a subset of Wanlida employees in their production center in Nanjing in the Fujian province. The goal of our experiment is twofold. First, we aim to analyze how simple framing manipulations of a bonus scheme influence productivity of *competing teams of workers*. Second, comparing our results with that of Hossain and List (2012), we can investigate how the impact of an incentive contract on worker behavior depends on the interaction of the economic (financial incentive) and noneconomic (framing manipulation) contents of the contract.

During our 8-week long experiment, subjects engaged in their regular tasks and work schedules within their normal work environments. Our experiment included four different sets of work with each set consisting of two teams of workers of identical composition. For each set of work, we provided a weekly bonus to the team with the higher per-hour productivity. The bonus depended only on relative performances of the two competing teams. The team under the positively framed reward treatment was notified that for each week in which the team's per-hour production was higher than that of the competing team, a bonus of RMB 80 would be paid at the end of a 4-week long pay period. The team under the negatively framed punishment treatment was promised a *provisional* bonus of RMB 320 before the 4-week long pay period began, but was notified that for each week in which its per-hour production was lower than that of the competing team, the bonus at the end of pay period would be reduced by RMB 80.

In contrast with studies like Dickinson (2001) where the underlying incentive schemes for the reward and punishment treatments within a laboratory experiment were different, the two schemes in our experiment were isomorphic, except for the frame³. Unlike Lazear (2000), we do not compare inherently different types of incentive schemes in this paper and restrict attention to contests between two teams. While standard theory suggests that these incentive contracts should lead to identical outcomes, insights gained from Kahneman and Tversky's (1979) prospect theory suggest otherwise. If losses loom larger than gains to our factory workers, as prospect theory conjectures, the punishment treatment should outperform the reward variant. Alternatively, if workers get more invigorated by positive incentive schemes, the reward treatment should lead to a higher level of productivity. This experiment provides us with insights on how productivity, both in absolute and relative sense, of teams competing in a contest is affected by framing of the incentive scheme.

In our experiment, we find that framing effects display different and more subtle patterns compared to those found in Hossain and List. Furthermore, an additional psychological effect from the framing manipulation might arise due to the competitive nature of the game. First, both bonus schemes succeeded in increasing productivity just as Dickinson (2001) and Hossain and List (2012) found. Average weekly productivity increased by almost 14% across the four sets. We find some more interesting data patterns on the impact of framing of the schemes. There is no statistically significant evidence that teams competing under the punishment frame outperform teams competing under the reward frame in terms of average weekly productivity. Yet, the result on which team wins the contest is very robust. The team in the punishment treatment is at least 35% more likely to produce at a rate higher than the team under the reward treatment. Hence, while framing has a very strong effect in determining the winner of the contests, the variance in absolute productivity is too large leading to a statistically insignificant framing effect in terms of productivity. These results contrast Hossain and List's (2012) findings. They may have resulted from the fact that winning or losing, not the absolute productivity level, is the sole determinant of the bonus under our incentive scheme. Workers may focus their efforts and adjust it inter-temporally to win the contests most number of times instead of maximizing their productivity. These insights are important for the design of incentive

¹ Prendergast (1999) presents an excellent survey of the literature on incentives to agents in firms.

² For an extensive survey on evidence of framing effects from the field, see DellaVigna (2009), especially Section 4.1. For a recent illustration, see Hossain and Li (2014).

³ For some other examples of field experiments comparing economically different incentive schemes, see Bandiera et al. (2005) and Shi (2010). Our paper is closely related to recent work by Armantier and Boly (2013) who, using laboratory and framed field experiments, show that bonus schemes that combine reward and punishment conditional on exogenously set targets increases effort provision.

Table 1
 Experimental design by sets.

Set	Job	Number of teams	Group size for teams	Pre-treatment average productivity	Team	Treatments	
						Round 1 (Week 1–week 4)	Round 2 (Week 5–week 8)
1	DVD player MD Chip production	2	16	350	Team A Team B	Punishment Reward	Reward Punishment
2	P720 main-board plug-in	2	10	390	Team A Team B	Punishment Reward	
3	Adapter plug-in	2	12	490	Team A Team B	Punishment Reward	Reward Punishment
4	Adapter joining	2	15	780	Team A Team B	Punishment Reward	Reward Punishment

Each set consists of two independent teams of the same group size and composition. Pre-treatment average productivity denotes the approximate average hourly productivity before treatments. Treatments are broken down by week number.

contracts and can also be applied to contests in sports, academia, and innovation. While contests avoid the arbitrariness of target setting and serves as means of filtering common shocks in homogenous production teams, it may inadvertently shift the focus from absolute to relative productivity reducing the loss frame’s ability to increase absolute productivity.

Another important result is that the increase in productivity due to incentive effects is not met with higher defects⁴. In fact, production quality, quantified by the percentage of products without any defect, is increasing in the team’s per-hour productivity rate. Overall, our results complement the existing literature and point to the importance of behavioral factors such as the notion of loss aversion in writing incentive contracts in labor environments such as ours. These insights can contribute to the literature on incorporating behavioral issues in contract theory as done in Hart and Moore (2008). Our main lesson from this paper is that while behavioral biases are persistent even in the field, how they affect workers’ behavior depends very much on the economic incentives they receive. Thus, there is much to be learned from going into the workplace and using exogenous variation in important variables via field experiments. The rest of the paper is designed as follows: the following section presents the experimental design. Section 3 presents the results and discusses their implications. Section 4 concludes.

2. Experimental design

The experiment was conducted over the months of October, 2008 to February, 2009. Wanlida focuses on consumer electronics and specializes in digital AV products, notebook PCs and peripherals, GPS navigation devices, car multimedia electronics, small home appliances, communications, and lithium polymer batteries. Our treatments pertained to the production of parts for DVD players and included four randomly selected sets of work each consisting of two independent production teams of the same number of workers. Workers in each team worked together in an assembly line. The two teams in a set worked on exactly the same tasks following identical procedures. The teams were not formed for the purpose of the experiment, but were indigenous in the factory. Table 1 provides a summary of the experimental design. The table can be read as follows: row 1 summarizes Set 1 which includes two unique teams, or lines of workers, whose task is to produce chips for DVD players. Both teams had a group size of 16 and produced 350 units per hour on average during the pre-treatment period. In Round 1, Team A of Set 1 was under the Punishment treatment and Team B was under the Reward treatment for 4 weeks. In Round 2, the teams changed treatments. The same was true for the teams in the other three sets⁵. In total, 106 Wanlida employees were part of this experiment.

The teams in our experiment had been in place for quite some time at Wanlida factory. These teams are organically formed by the management for regular production in the factory. We did not manipulate the teams in any way and team compositions were kept unchanged prior to, during, and after the experiment. Generally, the management tries to keep the teams similar to keep the chain of production smooth. Table 2 presents summary statistics of some demographic variables for each of the 8 teams. Overall, 86.8% of the workers were female and 66.0% of the workers finished at most junior high school. The average worker was 23.16 years old with 3.67 years of tenure at Wanlida. The differences in demographic variables between two teams within a set are statistically insignificant in most cases⁶. Nevertheless, each team experiencing both treatments helps address potential variations between teams in estimating the treatment effects. Moreover, as there can be

⁴ The absence of a multitasking tradeoff is also reported in Duflo and Hanna (2005), a field experiment conducted in schools in rural India.

⁵ Set 2 did not produce any more after three weeks of treatment because of a reduction in demand for the product. Thus, we do not have any data from Round 2 for that set.

⁶ Only mean tenure at Wanlida is different for Sets 2 and 3 at the 5% level. The tenure is higher for the team receiving punishment first for Set 2 and for the team receiving reward first for Set 3.

Table 2
 Demographical summary statistics for each of the eight teams within the four sets.

Set	Team	Percentage of male workers (%)	Age (in years)	Percentage of workers with senior high school education (or above) (%)	Tenure at Wanlida (in years)
1	Team A	31.3	23.19	43.8	3.07
	Team B	25.0	22.13	50.0	2.28
2	Team A	10.0	23.50	30.0	4.63
	Team B	10.0	21.20	50.0	2.60
3	Team A	0.0	22.75	33.3	3.28
	Team B	0.0	24.08	33.3	5.88
4	Team A	13.3	23.60	20.0	4.17
	Team B	6.7	24.47	13.3	3.88
All		13.2	23.16	34.0	3.67

Tenure at Wanlida is defined as the years a worker had spent working in Wanlida when our experiment was conducted.

significant time trend in production, we assign the two teams opposite frames in a given week to compare across teams. These between and within-team variations of treatments allow us to pursue out any heterogeneity in productivity across time or across teams within a set by means of difference-in-differences analysis as outlined in the following section.

Our framing manipulation is embedded within contests between two teams. This is an example of a tournament-styled incentive scheme. Among various incentive schemes, relative performance evaluation is used as means of filtering out common shocks from compensation packages. Tournament competition, as a special case of relative performance evaluation, has theoretically advantageous features (see, for example, Lazear and Rosen, 1981) as well as wide applications in sports (Ehrenberg and Bognanno, 1990 analyze data from golf tournaments), business (Knoeber and Thurman, 1994 study the broiler chicken industry), and even political turnovers (see Li and Zhou, 2005). In our experiment, the homogeneity of team structure and job characteristics for teams within a set makes contest an appropriate scheme to incentivize agents.

It is important to point out that there are four reasons why our particular way of framing might not be strong enough to produce results that are significantly different from one another. First, the treatment is a passive one. In the punishment treatment, rather than actually giving the employees the bonus before the work commenced, we provisionally allocated them the bonus, to be paid at the end of the pay period⁷. For example, in the reward treatment, the relevant portion of the letter read:

For the next 4 weeks starting from *Day 1*, in addition to your standard salary, you will receive an RMB 80 bonus for every week the weekly average per-hour production of your team is above the weekly average per-hour production of (*name of the competing team within the same set*). This program will continue until the end of the week starting on *Day 22* and end on *Day 28*. On *Day 29*, you will receive your bonus according to the above criterion.

For example, if your team produces at a rate above the weekly average per-hour production of (*name of the competing team within the same set*) in two weeks, you will receive RMB 160 on *Day 29*.

Conversely, in the punishment treatment, the relevant description was changed to:

For the next 4-week period starting from *Day 1*, you will receive a one-time bonus of RMB 320. This payment will be made on *Day 29*. However, for every week in which the weekly average per-hour production of your team is below the weekly average per-hour production of (*name of the competing team within the same set*), the bonus will be reduced by RMB 80.

For example, if your team produces at a rate below the average per-hour production of (*name of the competing team within the same set*) in two weeks, your bonus will be reduced by RMB 160 and you will receive RMB 160 on *Day 29*.

Thus, the punishment treatment is not a particularly powerful variant, but one the firm felt was appropriate and natural for this environment. Since our goal was to execute a natural field experiment, we followed the direction of Wanlida in making the treatments follow company guidelines. Further, we intentionally did not call the reduction in payment in the punishment treatment a fine or punishment to reduce potential negative (emotional) connotation associated with a fine. Instead, we were interested in making the reward and punishment treatments merely different framings of the same incentive contract. As such, the payments were made at the same time for both teams within a set, thus eliminating any credibility or time discounting issue. The differences are, therefore, extremely thinly veiled.

Second, because there is sufficient heterogeneity in production across time, sets, and, potentially, teams, we designed the treatments to control for both between and within-team variations as explained above. The power of our experimental

⁷ English translations of sample letters to the workers are available in an online appendix.

design comes from aggregating within and across-team data. In light of the fact that some might consider our treatments transparent, it might be considered a demanding test to find differences within a team that ultimately experiences both treatments (MacCrimmon and Larsson, 1979). Since any statistically significant finding of the framing effect is robust to the double exposure to the framings, this can be considered as strength of the experimental design.

Third, it is worth discussing the amount of communication that took place between teams within a set. For some sets, the two teams are located in the same room while for others they are located in different rooms within the factory. In either case, workers from the two teams within a set knew each other and might have had opportunities to interact with each other. Formally, we informed the workers about their performance relative to the other team only at the end of four weeks when they received the payments. Nevertheless, it is conceivable that they were aware of each other's progress to some extent during the contest. Potential communication between teams makes our experimental design even more demanding in terms of finding a framing effect.

Fourth, although Wanlida is a large manufacturer and we tried hard to maximize our sample size, its scale limited our sample size. According to the rules of thumb provided by List et al. (2011), in our experimental design, our sample size is sufficient to detect about a 1/3 difference of winning probability and about 0.7-standard deviation difference of productivity, with power of 0.8 and significance level of 0.05. This means that our test can only detect framing effects with sufficient economic significance. This further makes our design a demanding test of framing effects in both absolute and relative productivity dimensions.

A few other experimental particulars of interest are worth noting before we move on to the experimental results. First, in all sets, workers were never aware that an experiment was occurring, and they did not know that a new treatment would begin after four weeks. Management notified subjects of treatments via personal letters. In accordance with company policy, the bonus incentives were paid in addition to the base income. The source of the bonus was intentionally kept vague in the letters to subjects and the workers were not asked to do any unusual work. The electronics manufacturer itself had been analyzing schemes to improve the productivity of the workers to maintain its competitive edge in a global market. A change in incentive schemes was not an alien concept to the workers. Second, however, using competition between two teams within a set of work for incentive payments was a new concept for these workers. Third, the letters clearly mentioned that this was a short-term incentive program and the workers were likely to assume that the incentive schemes were one-shot opportunities. The language of the letters were kept identical (other than the part containing the framing manipulation) so that workers under different treatments were less likely to have different expectations about the impact of their performances during this program on their future income potential. Fourth, RMB 80 equaled slightly below USD 12 at the exchange rate during the experiment. The workers received fixed baseline weekly salary, which ranged between RMB 300–375. Thus, this bonus was non-trivial—more than 20% of the salary of the highest-paid worker in the subject pool. Fifth, workers were aware that the quality of their production would be checked before they are used in the following step of production. Finally, each week we received detailed reports on daily production, number of hours worked, and the number of defective units produced for each team. Note that the production managers were unaware of our direct research hypothesis related to framing. They were only informed that the test revolved around understanding incentives. One of the coauthors periodically visited the factory to ensure proper execution of the experiment and smooth transition of the rounds, and to oversee the payment to the workers after the end of a round.

In many naturally occurring tournaments, the players are teams instead of individuals. However, theoretical and experimental literature on tournaments and contests mostly analyze contests among individuals (e.g. Davis and Reilly, 1998; Potters et al., 1998; Onculer and Croson, 2005). In contrast, teams of workers in production lines participated in contests in our experiment. While one may worry that free-riding can be a problem for teams, Hamilton et al. (2003) empirically suggest that free-riding may not be a major issue in practice. Abbink et al. (2010) experimentally analyzed contests between groups and individuals and found that contests involving groups did not lead to free-riding problems. Rather, group members contributed more than what theory predicts. Nalbantian and Schotter (1997) found that competition between teams generated the highest level of efficiency among the four incentive schemes they used in their lab experiment.

Our main variable of interest is average hourly production for a given week as the incentive schemes were specified for weekly per hour productivity rate—the total production by a team in a week divided by the number of hours they worked in that week. We chose to pay for hourly productivity instead of total production as the number of hours worked may be different for the two teams within a set. Another variable of interest is the weekly defect rate which equals the number of defective products divided by the number of total products produced by the team in that week expressed in percentage terms. We focus on weekly data as the bonus was defined for each week within a 4-week treatment round.

We now briefly discuss how such framing may affect worker behavior when they are loss-averse following the utility specifications in Section 2.1 of Hossain and List (2012)⁸. The model suggests that a bonus framed as punishment will have a greater impact on productivity than a bonus framed as reward. This implies that, holding everything else identical, workers under the punishment frame will have a higher productivity than workers under the reward frame. Moreover, the team under the punishment treatment will win the weekly contests over the team under the reward treatment. Given our experimental design, even when teams are potentially heterogeneous, we expect the teams to have a higher productivity level and win

⁸ We do not formally discuss a model here due to a limit on the length of the paper.

Table 3
 Incentive effects in the four sets.

Set	Pre-treatment average productivity	During-treatment average productivity	Productivity increase (%)
Set 1	350	431.33	23.2
Set 2	390	469.47	20.4
Set 3	490	462.90	-5.5
Set 4	780	904.82	16.0

Pre-treatment productivity denotes the approximate average number of units produced per hour before treatments began. During-treatment productivity denotes the average hourly productivity during both the incentive schemes averaged across the two teams within a set. Set 3 used a different production method for both teams during the first round of incentive schemes.

Table 4
 Productivity and defect rates for teams.

		Set 1		Set 2	Set 3		Set 4	
		Weeks 1–4	Weeks 5–8	Weeks 1–4	Weeks 1–4	Weeks 5–8	Weeks 1–4	Weeks 5–8
Reward	Ave Prod	461.974	405.750	464.666	404.859	495.493	822.784	965.880
	SD Prod	(47.175)	(4.77)	(18.706)	(19.454)	(80.075)	(113.223)	(84.977)
	Defect Rate	0	0	0.461%	0.333%	0.457%	0.087%	0.151%
	N	4	4	3	4	4	4	4
Punishment	Ave Prod	457.075	400.52	474.263	435.377	515.864	844.872	985.747
	SD Prod	(62.226)	(16.35)	(28.079)	(27.667)	(58.895)	(100.380)	(69.134)
	Defect Rate	0	0.004%	0.392%	0.374%	0.421%	0.104%	0.178%
	N	4	4	3	4	4	4	4

Team average weekly per-hour productivity, weekly defect rate, and the number of weeks in which the team won the contest within a 4-week round during the treatment period by set and treatments are reported. Note that for Set 3, the production methods were different in the two rounds.

the contest more frequently when they are under the punishment frame. We will test these two hypotheses in the first two subsections of the following section.

3. Experimental results

To begin our discussion, we first consider Table 3 that summarizes the pre-treatment and during-treatment weekly productivity rates for the four sets. A comparison of average productivity data during treatment with the pre-treatment data shows that the incentive schemes seem to have raised overall productivity of the workers in three of the four sets. Next, we present weekly per-hour productivity, defect rates, and contest outcomes for each treatment for the four sets in Table 4. This suggests that the reduction in productivity for Set 3 comes from low productivity during the first round of incentives. In the first four weeks of the incentive scheme, Set 3 used a different production method for both teams, which led to a significantly reduced rate of productivity. Hence, to calculate the overall impact of the bonus (under either framing) we use productivity measures for Set 3 only during the second round of incentives when the production process returned to the pre-treatment one. We surmise that the contest incentives raised productivity by almost 14%, on average, relative to the pre-treatment productivity levels.

A natural follow-up question revolves around whether the framing of incentive schemes *per se* affected productivity and contest outcomes. A more detailed analysis on the data leads us to the following three main results:

- (i) The difference in average weekly productivity between the reward and punishment treatments is not statistically significant.
- (ii) The team in the punishment treatment won significantly more often in the productivity contest.
- (iii) Product defects decreased when per-hour productivity increased.

3.1. Productivity levels under the two frames

We first analyze weekly average productivity, which is the mean of the four observations of weekly per-hour productivity during a round (there were three observations for the first round of Set 2). Take Set 4 for example: the averages of weekly per-hour productivity rate for the first 4 weeks were 845 units and 823 units under the punishment and rewards treatments, respectively. In weeks 5 to 8, the averages were 986 and 966 units, respectively. Similarly, average defect rate is the mean of the four observations of the weekly defect rate.

To compare the levels of productivity under the reward and punishment treatments, we employ a difference-in-differences approach as there can be significant heterogeneity in production across weeks due to change of production plans, supply of material within the assembly line etc. and there may also be heterogeneity in the inherent level of productivity between teams within a set. Suppose the average production of team $j \in \{A, B\}$ of set $i \in \{1, \dots, 4\}$ under

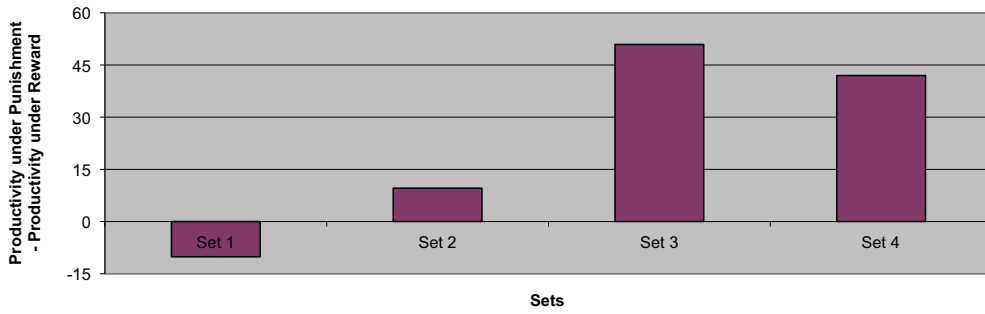


Fig. 1. Displays the aggregated differences in hourly productivities between punishment and reward treatments within a set. See Table 4 for absolute productivity levels from each treatment.

Table 5
 Framing effect on productivity.

Dependent variable: log of average hourly productivity on a given week			
	(1)	(2)	(3)
Punishment	0.0230 (0.016)	0.0234 (0.020)	0.023 (0.018)
Team fixed effects	No	Yes	No
N	54	54	54
Adjusted R-squared	0.980	0.0976	0.980

Empirical estimates of punishment treatment effect on average hourly productivity compared to the reward treatment are reported. Specification (1) includes set-week specific fixed effects and specification (2) further includes team-specific fixed effects. Robust standard errors clustered at the team level are displayed in parentheses for columns (1) and (2), whereas column (3) repeats column (1) with BRL standard error. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels respectively.

treatment *R* or *P* in period *t* is $P_{ijt} = (1 + \kappa_{ijt}) \mu_{ij} + \eta_{it}$, where μ_{ij} is the inherent productivity level of team *j* in set *i*, η_{it} is a time-specific productivity shock to set *i*, and $\kappa_{ijt} \in \{\kappa_R, \kappa_P\}$ quantifies the treatment effect of the relevant bonus scheme.

In this case, $\kappa_P - \kappa_R$ quantifies the framing effect. Notice that, if $\mu_{iA} \neq \mu_{iB}$ or η_{it} is nonzero, within-team or within-round comparison of average productivity from Table 4 alone will not allow us to estimate the framing effect precisely. 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,

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

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

Hence, we can calculate the across-round difference of the within-round productivity differences between Team B and Team A to estimate the framing effect. For example, for Set 4 we find that the punishment treatment yielded 42 more units of product ((845 – 823) + (986 – 966)). These results, as summarized in Fig. 1, show that in 3 of 4 sets, the punishment treatment outperformed the reward treatment⁹.

To complement the raw data summary, we estimate a model in which we regress the logarithm of weekly productivity on a dummy variable denoting the punishment treatment. We control for heterogeneity across teams of workers and time by including time and team-specific fixed effects. To allow for different sets of work to be subject to different shocks in a given week, as represented by η_{it} in the above production equation, the time-specific fixed effects account for each set-week combination. Empirical results are contained in Table 5¹⁰. The first column in Table 5 includes the above mentioned set-week specific fixed effects, but no team fixed effects. That is, we allow the time-specific shock η_{it} to vary across set (*i*) and week (*t*), but both teams are assumed to have the same inherent productivity, $\mu_{iA} = \mu_{iB}$. In the second column, we further control for team-specific fixed effects allowing μ_{iA} and μ_{iB} to be different. The punishment treatment increases the productivity over the reward treatment by 2.3%. However, the coefficient is statistically insignificant at conventional levels. Robust standard errors clustered at the team level are reported in columns (1) and (2)¹¹. We repeat the regression from column (1) with

⁹ We have data only from three weeks of the first round for Set 2. Hence, we calculate the difference in productivity between the team under punishment treatment and the team under reward treatment, which filters out time specific shocks. Similarity of the two teams suggests that cross-team variations are less important than cross-time variations.

¹⁰ Our empirical analysis includes data from Set 3 where production method in the first round is different from that in the second round, but is the same for both teams in a given round. Assuming that the different production methods affect productivity of the two teams by the same amount, this would not bias our estimation. In order to check the robustness of the framing effects, however, we repeat the regressions in Tables 5 and 6 excluding Round 1 of Set 3 and report the results in Tables C1 and C2 in the online appendix. There is no qualitative difference in the results.

¹¹ If we do not control for the potential clustering of error terms at the team level, the coefficient is significant at the 10% level for column (1).

Table 6
 Framing effect on probability of winning.

Dependent variable: dummy variable denoting whether <i>Team A</i> produced at a higher productivity than <i>Team B</i>					
	(1)	(2)	(3)	(4)	(5)
Punishment	0.383*** (0.080)	0.417*** (0.083)	0.356*** (0.067)	0.383*** (0.083)	0.417*** (0.083)
Conditional logit with team fixed effects	No	No	Yes	No	No
Set 2 Excluded	No	Yes	Yes	No	Yes
<i>N</i>	27	24	24	27	24
Pseudo <i>R</i> -squared or adjusted <i>R</i> -squared	0.120	0.146	0.183	0.092	0.104

Empirical estimates of punishment treatment effect on probability of winning in competition compared to the reward treatment are reported. The table presents marginal effect coefficients from logit regression of whether *Team A* of a given set won the contest. Specification (3) runs a conditional logit regression with team fixed effects. Specifications (2) and (3) do not include observations from Set 2. Robust standard errors clustered at the team level are displayed in parentheses for specifications (1) to (3). Specifications (4) and (5) repeat specifications (1) and (2) respectively, using BRL estimation. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

standard errors adjusted using the Bias Reduced Linearization (BRL) estimator following Angrist and Lavy (2009) to address the potential problem of having too few clusters. The result is reported in column (3). The coefficient of the punishment treatment in columns (3) is not significant either.

The evidence above shows that we find no statistically significant difference in average productivity between bonuses framed as reward or punishment during our experiment¹². Absence of any framing effect is certainly one possibility that may account for this evidence. However, we conjecture that the statistical insignificance could also arise even if there is strong framing effect, due to a particular feature of contests or tournaments. In contests such as the ones in our experiment, players care little about their absolute performances (here, the average hourly productivity) and more about winning the game as only the relative performances enter their incentive scheme. Thus, the dependent variable in the above regression (average hourly productivity) may not properly represent the objective of the agents. This leads to our next analysis.

3.2. Winning the contest

If the teams focus on winning the competition and the framing effect is strong, then we should find a large difference between the likelihood of producing at a rate higher than that of the competing team under the two frames. In our data set, the team under the punishment treatment had higher per-hour productivity than the team under the reward treatment in 70.4% of the cases.

Table 6 presents marginal effects coefficients from logit regressions of whether *Team A* had won the incentive competition in the week (by producing at a higher rate than the competing team). As the same variable for *Team B* is just the binary opposite, we only include one data point from each of the 27 week-long contests. The dependent variable indicates whether the productivity of *Team A* of set *i* at time *t* (P_{iAt}) is higher than that of *Team B* (P_{iBt}); that is, whether $P_{iAt} - P_{iBt} = (1 + \kappa_{iAt}) \mu_{iA} - (1 + \kappa_{iBt}) \mu_{iB} > 0$. Thus, it is independent of the time-specific productivity shock η_{it} as it affects both teams in a set equally. The variable herein captures the objective function of the teams in competition and this may be a better measurement of the framing effect. Specification (1) presents logit regression for all four sets without team-specific fixed effects. Specification (3) includes team-specific fixed effects and, hence, we use conditional logit regression and exclude Set 2 as we only have one-round observations from that set. For completeness, specification (2) presents the same logit regression as in specification (1), but excludes the observations from Set 2. In all of these specifications, we again control for the clustering of error terms at the team level to address the potential serial correlation problem. Columns (4) and (5) repeat the regressions in specifications (1) and (2) with BRL adjusted standard errors reported. The estimates in Table 6 confirm the impact of the framing effect: the punishment treatment, compared to the reward treatment, significantly increases the probability of winning the competition by more than 35%. This result suggests that framing effect is present in our field experiment. Further, *winning the competition* may enter more prominently in agents' objective functions when they compete against each other.

We contrast our results with the findings of Hossain and List (2012) who find that team productivity is higher in punishment treatments compared to reward treatments by around 1% when workers receive bonus for meeting a target. In their study, workers are rewarded for absolute performance as teams' bonuses depend only on their own productivity, rather than relative performance, and the increase in hourly productivity due to framing is statistically significant in their case¹³. Even if framing effects are equally strong under both schemes, this difference in incentive structure may explain why we find significant framing effect only for the probability of winning while their study does find that framing manipulation significantly affects average productivity rate. On the other hand, running similar binary choice regressions using the data

¹² Taking into account the power of our test, we are at least sure that the framing effect on productivity, if any, is too small to be detected in our test. This is in contrast with the large and significant framing effect on winning probability that we report in details below.

¹³ One drawback of setting a target for the bonus is that the outcomes may critically depend on the chosen target level. The contest format, as in this paper, does not suffer from such an issue.

Table 7
 Association of quality with productivity for teams.

Dependent variable: percentage of products with defects on a given week			
	(1)	(2)	(3)
Log of hourly productivity	−0.706** (0.289)	−0.719* (0.340)	−0.706* (0.393)
Team fixed effects	No	Yes	No
N	54	54	54
Adjusted R-squared	0.862	0.852	0.862

Specification (1) includes set-week fixed effects, and specification (2) further includes team fixed effects. Robust standard errors clustered at the team level are displayed in parentheses for columns (1) and (2), whereas column (3) repeats column (1) with BRL standard error.

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

from Hossain and List, we find little framing effect in terms of producing more than the other team in their data, unlike in our study.

Finally, we test whether the punishment treatment had a relatively stronger effect in the first round as teams might be more likely to recognize that the two bonus schemes are monetarily equivalent when they are exposed to both treatments. To test this hypothesis, we run a logit regression of whether the team under the punishment treatment won the contest on a dummy variable that denotes the observations from the first round. Running such analysis, although we find that the punishment treatments in the first round had a slightly stronger effect, the effect is not statistically significant even at 10% level. Moreover, we do not see any trend in the contest outcome across different weeks. We, however, analyze dynamics of production during a week and discuss it in the following paragraph as that helps in explaining our main findings.

In our experiment, the framing effect is very strong in terms of relative performance between two teams but virtually non-existent in terms of absolute performance of the teams. As workers knew their own productivity and might have, informally, received some information about the other team’s productivity during a week’s contest, their effort choice could be dynamically adjusted although they knew the official contest outcomes only at the end of the round. Suppose at some point during a week, members of a production team feel that they are unlikely to win in that week given the current gap in productivity rates between the two teams. Then they may give up and their productivity may decline to the usual non-incentivized level or lower¹⁴. Casas-Arce and Martinez-Jerez (2009) theoretically show that when participants in a tournament perceive the ability gap between players to be too large to overcome, they tend to exert less effort. This could make the average productivity highly variant resulting in the statistical insignificance of the framing effect in Table 5. Following this intuition, we find that the productivity gap between the two teams within a set at the midpoint of a week has significant impact on the teams’ final productivity levels for the week. Specifically, the larger the midway productivity gap is, the lower is the final productivity of the team that is behind at the middle of the week. As losses loom larger than gains, the productivity of the team that is behind may be reduced more if the team is under the punishment treatment¹⁵.

3.3. Quality of production

Although our bonus scheme with framing manipulation motivates workers to produce at a faster rate, the productivity increase may come at the cost of lowered quality of production. Fortunately, all products are tested by quality control inspectors before they are sent to the next level of production. Nevertheless, incentives are imposed on the quantity dimension while the quality issue is never mentioned in the personal letters to the subjects in our experiment. If quality control is of second degree of importance to the workers, they may put more effort in producing at a higher rate and shirk in the quality dimension as Holmstrom and Milgrom (1991) suggest.

Alternatively, as quality is measured thoroughly, there is likely to be an impact on the future income prospects of the workers. The workers in our experiment have long-term relation with the firm and that may deter them from shirking in quality if they are sufficiently forward-looking. A casual look at the average defect rates in Table 4 does not suggest a strong pattern in quality. However, upon closer inspection, we find an interesting result. Following Bandiera et al. (2005), in Table 7, we regress the defect rate on the logarithm of the hourly productivity rate, with robust standard errors clustered at the team level¹⁶. Specification (1) controls for set-week specific fixed effects, and specification (2) further includes team-specific fixed effects. We report BRL estimation results of specification (1) in column (3). The table suggests that a 1% increase in productivity is associated with a reduction of the defect rate by around 0.007 percentage points and this effect is statistically significant. Given that average defect rate is rather small to begin with, this represents a substantial impact.

¹⁴ “Giving up” when a team is behind by an insurmountable margin is not uncommon in sports. Brown (2011) showed that professional golfers expended less effort while competing against Tiger Woods. Gill and Prowse (2012) showed that agents who were behind in a tournament exerted less effort.

¹⁵ These results are presented in detail in an online appendix.

¹⁶ Adding a dummy variable for the punishment treatment does not change the results. The coefficient is insignificant, implying that the framing itself does not directly affect defect rates, which is reassuring.

We ascribe the observed data pattern to the repeated-game effect of long-term work contracts and reciprocity. The letters to the workers mentioned that the incentive schemes were part of a temporary program. Nevertheless, the workers may still conjecture that similar incentive schemes may occur in the future if their current performance is satisfactory to the management. Moreover, even if the bonus scheme is considered to be one-shot, as regular workers in a well-organized workplace, subjects may recognize the importance of reputation in the long-run interaction with the firm management. Either case may lead to the increase in production quality.

Reciprocity may also play a role here. Meidinger and Villeval (2003) find laboratory evidence of reciprocity towards the principal in teamwork. In our field context, due to the vagueness of the source of the bonus, workers, by default, may think that the funding is from their management. As the scheme increases the payment to one of the teams for sure, the workers may want to return the favor to the management by improving the quality, especially when they are already quite likely to win. Notice that our bonus contracts are incomplete in the quality dimension while productivity is incentivized. This lends support to the idea that “the norm of reciprocity that leads to gift exchange is an effective contract enforcement device under conditions of contractual incompleteness” as Gächter and Fehr (2001) concludes (see also Gneezy et al., 2000). Our experiment with regular workers in a well-organized workplace shows that some issues rising from incompleteness of contracts may be alleviated in the real world because of long-term optimization and reciprocity.

A natural question on the above interpretation is, whether to build reputation or to reciprocate, why the workers do not simply enhance productivity? A first reason is that the management cares about both quantity and quality of production. Second, we refer to Irlenbusch and Sliwka (2005, 2006) who show that transparency posits positive effects on reciprocity: transparency makes it easier to build a reputation of reciprocity and reciprocity may be stronger if it is directly perceived by the person with whom interaction takes place¹⁷. In our setting, the management may view an increase in productivity solely motivated by the desire to win the weekly prize while an increase in quality clearly signals an intention to benefit the management. Thus, to build reputation or reciprocate, subjects may want to exert more efforts into the more transparent quality dimension. Al-Ubaydli et al. (2008) find that workers focus on observable effort when reciprocating if one dimension of effort is observable and another is not. Our result offers a natural extension of their finding. Given that both dimensions of effort are observable in our setting, workers focus on the more transparent dimension of effort when they reciprocate.

4. Conclusion

Modern contract theory literature is developed under the assumption of classical rational agents. In our natural field experiment, we find that even the language and framing of incentive contracts affect behavior in the workplace. Specifically, loss aversion seems to play a role in workers' effort choice in a specific way that depends on the fact that the underlying incentive scheme is a contest. While contests framed as loss does not increase per-hour productivity relative to contests framed as gain, the probability of winning the contest is significantly higher under the loss framing. Framing effect is both economically and statistically significant for relative productivity, but not absolute productivity even though the power of test is similar for both productivity dimensions.

The above result may have resulted from the fact that winning or losing, not the absolute productivity level, was the sole determinant of the bonus under our incentive scheme. Workers may focus primarily on winning and adjust their efforts intertemporally to win the contests most number of times instead of maximizing their productivity. Our paper thus contributes to the emerging field of behavioral industrial organization by demonstrating that incorporating behavioral issues in contract writing and, especially, tournament-styled incentive schemes is particularly important. Taking the results of Hossain and List (2012) and this paper together, one lesson is that to affect absolute performance using framing effect, incentive design should be based on absolute performance. Moreover, to affect relative performance, incentives should be based on relative performance. In this study, framing within a contest does not seem to be very effective in increasing productivity, the main objective of factory managers. However, when the objective of managers is to increase “wins,” framing can be used to enhance the efficacy of a contest-based incentive scheme. Thus, insights gained from this study may have implications for situations beyond the labor market. For example, the results suggest that team executives in professional sports leagues should offer bonuses to players for winning playoff series in terms of loss instead of gain.

A lesson from the field experiment literature is that experience usually attenuates behavioral biases prevalent in laboratory experiments (see List, 2003 and Levitt and List, 2007 for example). While we find framing affecting behavior of seasoned workers, these workers are likely to have little experience of participating in a contest with another team in the factory and the punishment treatment. Our results highlight that even in environments with experienced agents, novel framing manipulations still can make a difference if the agents are inexperienced with that particular manipulation. While laboratory evidences on labor behavior abound (see Charness and Kuhn, 2010, for a recent survey), there are fewer experimental studies using real workers from a large organization participating in their usual work in their natural workplace. In a methodological sense, this study thus showcases how field experiments can be used in providing insights on labor behavior in practice and in identifying the directions future theoretical research should pursue.

¹⁷ In their language, reciprocity essentially contains both strategic reciprocity (repeated-game effects) and non-strategic reciprocity (the “pure” reciprocity). To avoid confusion, we make a distinction between the two effects following the majority of literature.

Acknowledgements

We are extremely grateful to Mr. Sean Wong, Managing Director of Wanlida Group, for allowing us to use Wanlida factories as a real-world laboratory. We thank participants in the WISE experimental workshop, Qu Feng, and Xun Lu for helpful comments. We gratefully acknowledge the SSHRC Grant No. 489160.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jebo.2015.02.014>.

References

- Abbinck, Klaus, Brandts, Jordi, Herrmann, Benedikt, Orzen, Henrik, 2010. Intergroup conflict and inter-group punishment in an experimental context game. *Am. Econ. Rev.* 100, 420–447.
- Al-Ubaydli, Omar, Anderson, Steffen, Gneezy, Uri, List, John A., 2008. For love or money? Comparing the effects of non-pecuniary and pecuniary incentive schemes in the workplace. In: *Working Paper*. University of Chicago, Chicago, IL.
- Angrist, Joshua, Lavy, Victor, 2009. The effects of High Stakes High School Achievement Awards: evidence from a group-randomized trial. *Am. Econ. Rev.* 99 (4), 1384–1414.
- Armantier, Olivier, Boly, Amadou, 2013. Framing of incentives and effort provision. *Int. Econ. Rev.* (Forthcoming).
- Bandiera, Oriana, Barankay, Iwan, Rasul, Imran, 2005. social preferences and the response to incentives: evidence from personnel data. *Q. J. Econ.* 120, 917–962.
- Brown, Jennifer, 2011. Quitters never win: the (adverse) incentive effects of competing with superstars. *J. Polit. Econ.* 119, 982–1013.
- Casas-Arce, Pablo, Martínez-Jerez, F. Asís, 2009. Relative performance compensation, contests, and dynamic incentives. *Manage. Sci.* 55, 1306–1320.
- Charness, Gary, Kuhn, Peter, 2010. Lab labor: what can labor economists learn from the lab? *Handbook of Labor Economics*, vol. 4. Elsevier, pp. 229–331.
- Davis, Douglas D., Reilly, Robert J., 1998. Do too many cooks always spoil the stew? An experimental analysis of rent-seeking and the role of a strategic buyer. *Public Choice* 95, 89–115.
- DellaVigna, Stefano, 2009. Psychology and economics: evidence from the field. *J. Econ. Lit.* 47, 315–372.
- Dickinson, David, 2001. The carrot vs. the stick in work team motivation. *Exp. Econ.* 4, 107–124.
- Dufló, Esther, Hanna, Rema, 2005. Monitoring works: getting teachers to come to school. In: *NBER Working Paper 11880*.
- Ehrenberg, Ronald, Bognanno, Michael, 1990. Do tournaments have incentive effects? *J. Polit. Econ.* 98, 1307–1324.
- Gächter, Simon, Fehr, Ernst, 2001. Fairness in the labour market—a survey of experimental results. In: Bolle, F., Lehmann-Waffenschmidt, M. (Eds.), *Surveys in Experimental Economics: Bargaining, Cooperation and Election Stock Markets*. Physica Verlag, Berlin, Germany, pp. 95–132.
- Gill, David, Prowse, Victoria, 2012. A structural analysis of disappointment aversion in a real effort competition. *Am. Econ. Rev.* 102, 469–503.
- Gneezy, Uri, Guth, Werner, Verboven, Frank, 2000. Presents or investments? An experimental analysis. *J. Econ. Psychol.* 21, 481–493.
- Hart, Oliver, Moore, John, 2008. Contracts as reference points. *Q. J. Econ.* 123, 1–48.
- Hamilton, Barton H., Nickerson, Jack A., Owan, Hideo, 2003. Team incentives and worker heterogeneity: an empirical analysis of the impact of teams on productivity and participation. *J. Polit. Econ.* 111, 465–497.
- Holmstrom, Bengt, Milgrom, Paul, 1991. Multitask principal-agent analyses: incentive contracts, asset ownership, and job design. *J. Law, Econ. Organ.* 7, 24–52.
- Hossain Tanjim, List, John A., 2012. The behavioralist visits the factory: increasing productivity using simple framing manipulations. *Manage. Sci.* 58, 2151–2167.
- Hossain, Tanjim, Li, King King, 2014. Crowding out in the labor market: a pro-social setting is necessary. *Manage. Sci.* 60, 1148–1160.
- Irlenbusch, Bernd, Sliwka, Dirk, 2005. Transparency and reciprocal behavior in employment relations. *J. Econ. Behav. Organ.* 56, 383–403.
- Irlenbusch, Bernd, Sliwka, Dirk, 2006. Career concerns in a simple experimental labour market. *Eur. Econ. Rev.* 50, 147–170.
- Kahneman, Daniel, Tversky, Amos, 1979. Prospect theory: an analysis of decision under risk. *Econometrica* 47, 263–292.
- Knoeber, Charles, Thurman, Walter, 1994. Testing the theory of tournaments: an empirical analysis of broiler production. *J. Labor Econ.* 12, 155–179.
- Lazear, Edward P., 2000. Performance pay and productivity. *Am. Econ. Rev.* 90, 1346–1361.
- Lazear, Edward, Rosen, Sherwin, 1981. Rank-order tournaments as optimum labor contracts. *Journal of Political Economy* 89, 841–864.
- Levitt, Steven, List, John A., 2007. What do laboratory experiments measuring social preferences reveal about the real world? *J. Econ. Perspect.* 21, 153–174.
- Li, Hongbin, Zhou, Li-An, 2005. Political turnover and economic performance: the incentive role of personnel control in china. *J. Public Econ.* 89, 1743–1762.
- List, John, Sadoff, Sally, Wagner, Mathis, 2011. So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design. *Exp. Econ.* 14, 439–457.
- List, John A., 2003. Does market experience eliminate market anomalies? *Q. J. Econ.* 118, 41–71.
- MacCrimmon, Kenneth R., Larsson, Stig, 1979. Utility theory: axioms versus paradoxes. In: Allais, M., Hagen, O. (Eds.), *The Expected Utility Hypothesis and the Allais Paradox*. D. Riedel, Dordrecht, The Netherlands, pp. 333–409.
- Meidinger, Jean-Louis Rullière, Villeval, Marie-Claire, 2003. Does team-based compensation give rise to problems when agents vary in their ability? *Exp. Econ.* 6, 253–272.
- Nalbantian, Haig R., Schotter, Andrew, 1997. Productivity under group incentives: an experimental study. *Am. Econ. Rev.* 87, 314–341.
- Onculer, Ayse, Croson, Rachel, 2005. Rent-seeking for a risky rent: a model and experimental investigation. *J. Theor. Polit.* 17, 403–429.
- Potters, Jan, de Vries, Casper G., van Winde, Frans, 1998. An experimental examination of rational rent-seeking. *Eur. J. Polit. Econ.* 14, 783–800.
- Prendergast, Canice, 1999. The provision of incentives in firms. *J. Econ. Lit.* 37, 7–63.
- Shi, Lan, 2010. Incentive effects of piece rate contracts: evidence from two small field experiments. *B.E.J. Econ. Anal. Policy* 10 (Article 61).