THE MORALE EFFECTS OF PAY INEQUALITY*

EMILY BREZA[†], SUPREET KAUR[‡], AND YOGITA SHAMDASANI[∓]

DRAFT: PRELIMINARY AND INCOMPLETE

November 30, 2015

Abstract. The idea that worker utility is affected by co-worker wages has potentially broad implications for the labor market—for example, through wage compression, wage rigidity, firm boundaries, and the distribution of earnings. We use a month-long field experiment with Indian manufacturing workers to test whether relative pay comparisons affect effort and labor supply. In our setting, workers are paid a flat daily wage and organized into distinct product teams. We randomize teams to receive either compressed wages (where all teammates earn the same wage) or heterogeneous wages (where each team member is paid a different wage according to his baseline productivity rank). For a given absolute wage level, workers reduce output by 0.36 standard deviations if they are on a team where they are paid less than their peers. They are also less likely to come to work—giving up 9% of their earnings. These effects strengthen in later weeks. In contrast, workers do not increase output when they are paid more than their peers. The perceived justification for pay differences mediates negative morale effects. Specifically, lower relative pay does not cause effort reductions when co-worker output is highly observable, or when one's higher-paid co-workers are substantially more productive than oneself (in terms of baseline productivity). Finally, performance on endline games indicates that pay disparity reduces team members' ability to cooperate in their own self-interest.

JEL Classification Codes: J3, O15, D03, E24

Keywords: Reference dependence, peer comparisons, wage compression

---

* We thank James Andreoni, Dan Benjamin, Stefano DellaVigna, Pascaline Dupas, Edward Glaeser, Robert Gibbons, Uri Gneezy, Seema Jayachandran, Lawrence Katz, David Laibson, Ulrike Malmendier, Bentley MacLeod, Sendhil Mullainathan, Mark Rosenzweig, Bernard Salanie, and Eric Verhoogen for their helpful comments. Arnesh Chowdhury, Mohar Dey, Piyush Tank, and Deepak Saraswat at JPAL South Asia provided outstanding research assistance. We gratefully acknowledge financial support from the National Science Foundation, the IZA Growth and Labor Markets in Low Income Countries (GLM-LIC) program, and the Private Enterprise Development for Low-Income Countries (PEDL) initiative.

[†]Columbia Business School. Email: ebreza@columbia.edu.

[‡]Department of Economics and SIPA, Columbia University. Email: kaur@columbia.edu.

[∓]Department of Economics, Columbia University. Email: yjs2112@columbia.edu.

1. Introduction

In traditional agency models, workers care about only their own wage levels when making effort and labor supply decisions. However, a long tradition in economic thought—as well as in psychology, sociology, and human resource management—has advanced the notion that individuals also care about their pay *relative* to that of their co-workers.[1] This implies that relative pay may be a compensating differential—affecting utility and therefore the willingness to accept work at a given absolute pay level. In addition, these utility effects could translate into changes in worker effort, and therefore output. A growing literature emphasizes the potential importance of such "morale" effects in the presence of incomplete contracting (Bewley 1999, Fehr et al. 2009).[2]

If relative pay concerns affect effort and labor supply, this could influence many features of the labor market. For example, it could help explain why wage compression—when wages vary less than the marginal product of labor—appears prevalent in both poor and rich countries (Dreze and Mukherjee 1989, Fang and Moscarini 2005, Charness and Kuhn 2007). Such considerations have also been proposed as a micro-foundation for wage rigidity, with consequences for unemployment and volatility (Akerlof and Yellen 1990). In addition, relative pay concerns could affect how heterogeneous workers are sorted into firms, possibly leading some firms to specialize in higher or lower ability workers (e.g., Frank 1984). They could also influence whether labor is contracted through the external market or organized within firm boundaries (e.g., Nickerson and Zenger 2008). Finally, such effects could themselves depend on features of production—for example, if the observability of co-worker output affects whether pay differences are perceived as justified (e.g., Bracha et al. 2015).[3]

These potential implications rest on the presumption that workers do indeed care about relative pay. In this paper, we empirically test the validity of this view using a field experiment with manufacturing workers. We conceptualize relative pay concerns as reference dependence in co-worker wages, with effort effects coming from reciprocity under incomplete contracting. Under most prominent formulations of reference dependence, being paid less than one's reference point will decrease utility and effort. Earning more than one's reference point may weakly increase effort—for example, under loss-aversion (Kahneman and Tversky 1979). We develop a design to enable comparisons of workers who earn the same absolute wage, so that potential reference dependence effects arise only from changes in co-worker wages. We formulate primary tests of our predictions by constructing cases where the relationship between own wage and the reference point is clear: when a worker

[1]See, for example, Marshall (1890), Veblen and Almy (1899), Hicks (1932),Duesenberry (1949), Easterlin (1974), Hamermesh (1975).

[2]Many employment arrangements are characterized by some degree of incomplete contracting; very few occupations are solely governed by explicit performance incentives such as piece rates (MacLeod and Parent 1999). Bewley (1999) documents that firm managers consider relative pay concerns to be important for worker motivation.

[3]See Fehr et al. (2009)for an overview of how fairness considerations may affect the labor market.

earns strictly more or less than all of his peers. We use additional tests to better understand the nature of reference dependence. For example, we examine what happens when a worker earns the median wage. In addition, we explore whether perceived justifications for pay differences affect whether fairness violations are triggered.

In the experiment, 378 workers in Odisha, India are employed full-time for one month in seasonal manufacturing jobs—a prominent source of employment in the area. They work in small factories, where they are organized into distinct teams, with three workers per team. All team members produce the same exact product (e.g. rope), while every team within a factory produces a different product (e.g. rope vs. brooms). One's teammates therefore constitute a natural and salient reference group for pay comparisons.[4] Note that there is no joint production in teams—production is an individual activity. We measure output by hiring extra staff to measure each worker's individual production at the end of each day. All workers are paid a flat daily wage, in accordance with the typical pay structure in the area.

To induce exogenous variation in co-worker pay, each team is randomized into one of two pay structures. In the *Heterogeneous* pay condition, each team member is paid a different wage—w_{High}, w_{Med}, or w_{Low}—in accordance with his respective productivity rank within the team (as determined by baseline productivity levels). These relative pay differences are fairly modest: the difference between each wage level is less than 5%. In the *Compressed* pay condition, all team members are paid the exact same wage; we randomly assign the level of this wage to be w_{High}, w_{Med}, or w_{Low}. Each team is assigned to either the *Heterogeneous* or one of the three *Compressed* wage treatments.[5] This allows us to compare, for example, two workers with the same baseline productivity level who both earn an absolute wage of w_{Low}, but differ in whether they are paid less than their peers (under the *Heterogenous* treatment) or the same as their peers (under the *Compressed* low wage treatment).

To test whether perceived justifications mediate morale effects, we incorporate two additional sources of variation into our design. First, while wage differentials are fixed at 5%, underlying baseline productivity is continuous. This induces variation in "actual fairness": the extent to which pay differences among co-workers overstate productivity differences. Second, we generate variation within and across treatments in "perceived fairness": whether a team engages in a production task for which it is easy to observe co-worker output.[6]

[4]This is consistent with Card et al. (2012), for example, who found that relative pay comparisons were stronger within university departments than across departments.

[5]At the beginning of the baseline (i.e. "training") period, workers are told that they will receive a wage increase on a pre-specified date, and that the size of this increase may depend on their baseline productivity. Once they are randomized into their wage treatment on this date, no additional future wage changes are possible. This shuts down dynamic incentive effects; see below for a discussion of this.

[6]We quantified the observability of each of the ten possible production tasks ex ante using a pilot. A different sample of workers–all of whom were on teams with *Compressed* wages–were asked after three weeks of working together to rank their output relative to that of their teammates. We use the mean accuracy of these responses for a given task as the observability value for that task. In the experiment, we stratify wage treatments by production task, ensuring variation in task observability within each wage treatment cell.

We find that for a given absolute pay level, output declines by 0.36 standard deviations (approximately 22%) on average when a worker is paid less than both his co-workers.[7] This is accompanied by a 11.8 percentage point decrease in attendance. Using endline data on outside earnings on absences, we estimate that employees give up approximately 9% of their earnings to avoid a workplace where they are paid less than their peers. These negative effects persist over the duration of the employment period, with some evidence that they become stronger in later weeks. In contrast, we find little evidence that performance improves if a worker is paid more than both his peers: average effects on output are statistically indistinguishable from zero, and attendance actually declines. In addition, we similarly cannot reject that there is no impact on the output or attendance of *Heterogeneous* pay workers who receive the median wage on their team.

Perceived justifications play an important role in mitigating the negative treatment effect of low relative pay. We examine two sources of variation in perceived justifications, both of which yield the same pattern of effects. First, we exploit variation in the ratio of pay differences to productivity differences. When teammates' baseline productivity levels are farther apart—so that differences in productivity swamp differences in wages—we find no evidence for negative effects of being paid less than one's peers. Second, we exploit variation in the observability of co-worker output. In production tasks where workers can easily see that their higher paid peers are more productive than themselves, there is no negative effect of being paid less than one's peers.

In contrast, perceived justifications have no differential effect on workers who are not aggrieved—for example, those who receive the median wage on their team. These findings suggest that in our particular setting, the reference point violation does not come from simply comparing a worker's own ratio of $^{pay}/_{productivity}$ relative to that of referent others (Adams 1963). Rather, workers appear to compare differences in pay in levels. When lower relative pay levels trigger a potential fairness violation, this is mitigated if lower pay is clearly justified by relative productivity.

Finally, we use endline activities to examine an additional dimension of morale. If relative pay effects operate through emotions such as resentment or envy of co-workers, this could undermine social cohesion or cooperation among team members. On the final day of the experiment, workers play cooperative games that require teamwork to solve picture-based puzzles. In playing the games, workers randomly are reshuffled into pairs of two—with variation in whether they are paired with someone from their own product team or from another team. Pairs receive piece rates for performance on the games, with clearly no benefit to the firm from effort. *Compressed* team workers perform better on the games when they are paired with someone from their own product team than when paired with a stranger (i.e.

[7]This estimate is based on comparing low-rank workers in *Heterogeneous* (who are paid w_{Low}) with low-rank workers in *Compressed* teams where everyone is paid w_{Low}. In general, all relative pay effects are identified off pairwise comparisons of workers in *Heterogeneous* who receive a given wage and those on *Compressed* teams who have the same absolute wage level and same rank (i.e. average productivity).

someone from another team). In contrast, *Heterogeneous* team workers perform 21% worse when they are paired with someone from their own product team than from another team. In addition, when in mixed pairs, there is no evidence that *Heterogeneous* pay workers perform worse than *Compressed* pay workers—their decrease in performance arises only when they work with someone from their own team.

While our findings indicate that in our setting, pay differences can cause substantial reductions in individual effort and team cooperation, one cannot draw conclusions about optimal pay structure. One potential benefit to firms of differential pay is dynamic incentives: workers know that if they work hard now, it could lead to higher pay in the future. Our study design shuts down this channel because after the baseline period, there is no further chance of wage changes. This is important for our specific purpose: cleanly isolating the morale effects of relative pay differences. It is also a realistic feature of seasonal and other contract jobs—a very common form of employment among the workers in our study. More generally, in deciding on optimal pay structure, a firm would weigh any potential costs of differential pay (e.g. morale reductions) against the potential benefits (e.g. dynamic incentive or selection effects). Our findings indicate that workers' relative pay concerns could affect this calculus.

This study builds on the literature on relative pay comparisons in the workplace. Two recent experimental studies with workers have examined relative pay concerns. First, Card et al. (2012) document that University of California employees report higher job dissatisfaction on surveys when they find out that they are paid less than their co-workers. Second, Cohn et al. (2012) show that relative random pay cuts matter more than absolute pay cuts for effort; these effects persist strongly over a six-hour period. Our results are consistent with those in both these studies.[8] In addition, our work relates to the broader literature on the effect of fairness preferences on effort provision under incomplete contracting, such as gift-exchange (Akerlof 1982).[9]

[8]A small number of laboratory studies have explored relative pay comparisons using gift exchange games, with mixed results (Charness and Kuhn 2007, Gatcher and Thoni 2010, Bartling and von Siemens 2011). Notably, Bracha et al. (2015) find support for relative pay concerns, and for the importance of perceived justifications. Related laboratory experiments have examined the effects of rank (Brown et al. 2008, Clark et al. 2010, Kuziemko et al. 2011). In addition, several studies examine relative pay concerns using observational data. Dube, Guliano, and Leonard (2015) document an increase in quitting behavior when an individual's pay increase is lower than that of her co-workers. Mas (2015) offers evidence that mandatory pay disclosure for municipal employees led to pay cuts and subsequent quits for high earners; he interprets this as public aversion to high compensation. A number of other studies are consistent with a relationship between relative pay and worker satisfaction or behavior (Levine 1993, Pfeffer and Langton 1993, Clark and Oswald 1996, Hamermesh 2001, Kwon and Milgrom 2008, Mas 2008, Rege and Solli 2013). Related work has explored links between relative income and other outcomes, such as happiness (Frey and Stutzer 2002, Luttmer 2005), health (e.g. Marmot 2004), and reward-related brain activity (e.g. Fliessback et al. 2007).

[9]These studies test for reference points that are determined by a worker's own absolute past wage (or expected wage). A large number of studies find gift exchange in the lab (e.g., Fehr et al. 1993). Field evidence, however, is more limited; existing field experiments generally find limited support for positive reciprocity (i.e., effort increases from wage increases), while the evidence on negative reciprocity (i.e. effort reductions from wage cuts) is more mixed over a one-day period (Gneezy and List 2006, Kube et al. 2013,

Our study advances the literature by finding substantial impacts of relative pay comparisons on effort and labor supply. In our experiment, wage differences reflect interpersonal differences in worker productivity; this reflects why wage differences may arise in the labor market and is important given laboratory findings that justifications can undo fairness violations (Falk et al. 2008, Bracha et al. 2015). We augment these findings with the first piece of field evidence that perceived justifications play an important role in mediating morale effects. This has bearing on understanding, for example, why wage compression may arise in some settings or occupations and not in others. Finally, workers make decisions for a job from which they derive full-time earnings over the one-month study period. This helps verify that impacts from reference dependence do not disappear once the novelty of treatments wears off (Gneezy and List 2006, Levitt and List 2007).

While our results indicate that relative pay concerns can affect output in large magnitudes, they also suggest that negative morale effects can be mitigated when the justification for differential pay is extremely transparent. These findings suggest that firms may have several potential tools at their disposal to manage morale in the presence of pay dispersion. For example, technologies that make it easier to quantify worker productivity could have aggregate output benefits—not just through a reduction in moral hazard, but also through improved morale. Firms could also potentially alter the organizational structure of the workplace itself—through job titles, physical co-location of similar workers, or the construction of "teams" (as we did in the experiment)—to affect who a worker views as being in her reference group. The extent to which firms can and do make use of such strategies has the potential to affect wage compression, wage rigidity, and firm boundaries. While speculative, such possibilities suggest a variety of ways through which relative pay concerns could affect pay structure, organizational arrangements, unemployment, and other labor market outcomes.

The remainder of the paper proceeds as follows. Section 3 describes our empirical setting and experimental design, Section 4.2 presents our results, Section 5 discusses threats to validity, and Section 6 concludes.

2. Framework

2.1. **The Worker's Maximization Problem.** We adapt the framework presented in DellaVigna et al. (2015), which allows the social preferences of workers to affect their effort decisions. We modify their approach to allow worker morale to be affected by peer wages in addition to a worker's own wage.

We assume that a worker i receives a take-it-or-leave-it wage offer w_i from the firm and makes two decisions: a) whether to work that day $s_i \in \{0, 1\}$ and b) if $s_i = 1$, how much effort $e_i \geq 0$ to expend working. Effort is not contractible by the firm. If the worker chooses

Esteves-Sorenson & Macera 2015, DellaVigna et al. 2015). For reviews of this literature, see Fehr et al. (2009), List (2009), and Charness and Kuhn (2011).

not to work (i.e., $s_i = 0$), then he receives a stochastic outside option $R_i = R + \varepsilon_i$. The worker's payoff from working is

$$V\left(w_i, \mathbf{w_{-i}}, e_i\right) = w_i - c\left(e_i\right) + M\left(w_i, \mathbf{w_{-i}}\right) e_i.$$

This payoff is a function of the worker's own wage, w_i, the wages of the worker's peers, $\mathbf{w_{-i}}$, and the effort choice e_i. We assume that the worker pays a convex effort cost $c\left(e_i\right)$ also experiences a morale effect, $\mathbf{M}\left(\mathbf{w_i}, \mathbf{w_{-i}}\right)$, that scales linearly with effort.

We conceptualize relative pay concerns as reference- dependence in utility, where co-worker pay enters as an argument into the worker's reference wage, $w_R\left(\mathbf{w_{-i}}\right)$. We incorporate this into worker utility in the following way

$$M\left(w_i, \mathbf{w_{-i}}\right) = \alpha 1_{w_i < w_R\left(\mathbf{w_{-i}}\right)} + \beta 1_{w_i > w_R\left(\mathbf{w_{-i}}\right)} + f\left(w_i\right)$$

$f\left(w_i\right)$ captures the non-peer-dependent contributors toward morale such as gift exchange.[10] If the worker is paid less than the reference wage, $1_{w_i < w_R\left(\mathbf{w_{-i}}\right)} = 1$, then there is an effect of α on the worker's utility per unit of effort provided. Conversely, β captures the effect on a worker's utility per unit of effort from being paid more than his reference wage.

Prior work conceptualizing relative pay comparisons predicts that $\alpha < 0$, that is, individuals dislike being paid less than their peers (i.e., Adams 1963, Akerlof and Yellen 1990). The prediction on the sign and relative magnitude of β, however, varies. Preferences for status or advantageous inequality generate $\beta > 0$. Further, under loss aversion $|\alpha| < |\beta|$. Inequality aversion, on the other hand, would lead to $\beta < 0$.

2.2. **Labor Supply and Effort Decisions.** We now consider what happens to effort and labor supply decisions when the employer changes the wages of a worker's peers, holding his own wage fixed. It is helpful to first define some notation. Let $e_0^* = \arg\max_\epsilon V\left(w_i, w_i, \epsilon\right)$. This is the optimal effort chosen by the worker when $w_i = w_R\left(\mathbf{w_{-i}}\right)$, that is when the worker's own wage is equal to the peer reference wage. Let $e_-^* = \arg\max_\epsilon V\left(w_i, \hat{\mathbf{w}}_{\mathbf{-i}}, \epsilon\right)$, where $\hat{\mathbf{w}}_{\mathbf{-i}}$ is set such that $w_i < w_R\left(\hat{\mathbf{w}}_{\mathbf{-i}}\right)$. Finally, let $e_-^* = \arg\max_\epsilon V\left(w_i, \tilde{\mathbf{w}}_{\mathbf{-i}}, \epsilon\right)$, where $\tilde{\mathbf{w}}_{\mathbf{-i}}$ is set such that $w_i > w_R\left(\tilde{\mathbf{w}}_{\mathbf{-i}}\right)$. We define the optimal labor supply decisions under different reference wages similarly, $s_0^* = 1\left(V\left(w_i, w_i, e_0^*\right) > R + \varepsilon_i\right)$, and $s_-^* = 1\left(V\left(w_i, \hat{\mathbf{w}}_{\mathbf{-i}}, e_-^*\right) > R + \varepsilon_i\right)$. s_+^* is defined similarly.

The following proposition is the basis for our empirical tests.

Proposition 2.1. *Let the support of ε_i be unbounded.*

If $\alpha < 0$, then $e_-^* < e_0^*$and $\mathrm{P}_\varepsilon\left(s_-^*\right) < \mathrm{P}_\varepsilon\left(s_0^*\right)$. If $\alpha = 0$, then $e_-^* = e_0^*$and $\mathrm{P}_\varepsilon\left(s_-^*\right) = \mathrm{P}_\varepsilon\left(s_0^*\right)$. If $\alpha > 0$, then $e_-^* > e_0^*$and $\mathrm{P}_\varepsilon\left(s_-^*\right) > \mathrm{P}_\varepsilon\left(s_0^*\right)$.

Similarly, if $\beta < 0$, then $e_+^* < e_0^*$and $\mathrm{P}_\varepsilon\left(s_+^*\right) < \mathrm{P}_\varepsilon\left(s_0^*\right)$. If $\beta = 0$, then $e_+^* = e_0^*$and $\mathrm{P}_\varepsilon\left(s_+^*\right) = \mathrm{P}_\varepsilon\left(s_0^*\right)$. If $\beta > 0$, then $e_+^* > e_0^*$and $\mathrm{P}_\varepsilon\left(s_+^*\right) > \mathrm{P}_\varepsilon\left(s_0^*\right)$.

[10] Alternately, $f\left(w_i\right)$ could capture other drivers of effort including monitoring norms for effort provision. DellaVigna et al. (2015) focus on $f\left(w_i\right)$ to unpack its determinants.

The proposition implies that when $\alpha < 0$, workers respond to small positive deviations in the reference wage relative to own wage by decreasing both effort and labor supply. When $\beta > 0$, workers respond to small negative deviations in the reference wage relative to own wage by increasing both labor supply and effort.

Our primary goal is to identify the signs of α and β using our experiment. Note that if $w_R(\mathbf{w_{-i}})$ were fully observable, then one could fully identify the signs of α and β by observing the responses to small changes in peer wages that trigger $1_{w_i < w_R(\mathbf{w_{-i}})} = 1$ and $1_{w_i > w_R(\mathbf{w_{-i}})} = 1$. While we do not take a strong ex ante stand on the functional form of $w_R(\mathbf{w_{-i}})$, we discuss below how we identify instances where, under most reasonable cases, $1_{w_i < w_R(\mathbf{w_{-i}})} = 1$ and $1_{w_i > w_R(\mathbf{w_{-i}})} = 1$.

3. Experimental Design and Data

3.1. **Experimental Design.** We construct a design to test the above predictions with manufacturing workers employed in small factories (see details below). In this setting, there is incomplete contracting on effort: in accordance with the typical pay structure in the area, all workers are paid a flat daily wage for each day they come to work. This provides them with some latitude to select both attendance (with implications for earnings) as well as effort (with implications for output).

In order to test for reference-dependence in co-worker pay, we must first define, for each worker, a clear reference group of peers. To accomplish this, within each factory, workers are organized into "teams" of three workers each. All team members produce the same exact product (e.g. rope), while every team within a factory produces a different product (e.g. rope vs. brooms). Production is an individual activity—teammates sit together but do not do any work jointly. Because each worker's two teammates are the only other people at the factory making the same product, they are likely the most salient reference group for wage comparisons.

To construct tests for our core predictions, we design wage treatments that allow us to fix workers' absolute pay levels, while creating variation in co-worker pay. Using baseline productivity data, we rank each worker as the lowest, medium, or highest productivity worker within his respective team. Each team is then randomized into one of four wage structures, as shown in Table 1:

- *Heterogeneous*: Each team member is paid according to his productivity rank within the team, where the rank is based on workers' baseline productivity level. The wages for the lowest, middle, and highest productivity workers are w_L, w_M, and w_H, respectively.
- *Compressed_L*: All team members are paid the same daily wage of w_L.
- *Compressed_M*: All team members are paid the same daily wage of w_M.
- *Compressed_H*: All team members are paid the same daily wage of w_H.

These wage differences are fairly modest: the difference between each of the three wage levels is approximately 5%. For each of the three ranks, this design enables us to compare groups of workers who have the same average productivity levels and are paid the same absolute wages, but differ in the distribution of their co-workers' wages.

To test for the role of justifications, we cross-cut the wage treatment with two additional sources of variation. First, we vary actual fairness—the extent to which pay differentials overstate productivity differentials—by randomizing workers into teams. Because output is continuous while productivity rankings are discrete, this generates variation in how much a worker's productivity level differs from that of his teammates. This, in turn, enables us to examine how effects vary with changes in the ratio of {wage difference}/{productivity difference} within and across wage treatments.

Second, we vary perceived fairness—the extent to which workers can observe co-worker productivity. The ten production tasks in the factories differ in how easy it is to observe the output of one's teammates. To ex-ante quantify the observability of each task, we used pilot trials to measure whether workers could accurately rank their output relative to that of their teammates after three weeks of working together. In these trials, all teammates were paid the same wage, so that wage was not a signal of productivity rank. We stratify wage treatments by production task, enabling us to test for the effects of observability within and across wage treatments. The randomization design is summarized in Figure 1.

3.2. **Predictions.** To test our first core prediction—a strict decrease in morale when $w_i < w_R$—we compare outcomes for Low rank workers in *Heterogeneous* with those in *Compressed_L*. Low rank workers in *Heterogeneous* are paid strictly less than all their teammates. Under virtually any reference point that depends on co-worker pay levels, they will feel more aggrieved than their counterparts in *Compressed_L*—who receive the same absolute pay of w_L, but whose teammates earn the same as they do.

To test our second core prediction—asymmetric effects from deviations from the reference point—we compare High rank workers in *Heterogeneous* with those in *Compressed_H*. We predict a weak increase in effort and attendance for High rank workers in *Heterogeneous* with those in *Compressed_H*.

Note that there is no clear ex ante prediction on the behavior of Medium rank workers in *Heterogeneous* relative to those in *Compressed_M*. Examining effects for this group will help provide insight into the nature of the reference point in our setting.

Finally, if the justification for pay differences matters for fairness violations, then the effects will be mediated by the perceived and actual fairness of pay differences. The magnitude of treatment effects will be smaller when differences between a co-worker and his higher paid peers is large, and when these differences are observable.[11]

[11]Note that these predictions are consistent with those of the model in Fang and Moscarini (2005).

3.3. **Time Line, Recruitment and Survey Instruments.** We detail the implementation of our experiment in Figure 2. Workers are males between the ages of 18 and 55. We recruit workers from villages surrounding rural factory sites in Odisha, India. We never recruit from the same village more than once. Each round requires hiring 30 workers; if more than the required number of workers apply for the job within a village, we randomly select among applicants.

After recruitment, workers are randomly assigned into teams of 3 workers each at the start of the round, and each team is assigned to a unique production task. These team assignments are stable for the duration of the employment period. The round begins with training period for each of the tasks. During the first three days of training, factory staff focus on making sure that the workers fully understand how to complete their tasks and how to ensure a baseline level of quality demanded in the market. Typically after day four, output has reached a level of quality that can be sold in the market, and this is the time at which we begin recording individual output per worker.[12]

Although output is salable by day 4, we prolong the "training" period to 14 days to obtain accurate measures of baseline productivity. On day 10, workers are given individual, private feedback by the factory manager on their rank relative to the other members of the team.[13]

On day 14, we randomly assign teams to treatments and inform each worker in private of his new wage. We again deliver this message in private and remind each worker that this is the wage that will be paid for the remainder of the contract period, that there will be no future opportunities for wage changes, and that there will be no future job opportunities after the end of the contract period. The factories then run as usual under the treatment wages until day 34. On day 35, workers participate in endline activities. These include playing incentivized games that measure team cohesion, as described below. On this last day, workers also take an endline survey which includes questions on potential earnings from outside jobs (on days they were absent from work). At the conclusion of the round, we also do household visits to survey all workers who quit their job at the factory in order to obtain information on their earnings.

On the first day of employment, workers are shown calendars (which they see at work everyday) that outline the dates of each of these events. They are also told on the first day that their post-training wage may depend on their baseline productivity.

An implication of our design is that within a factory, different teams have differing pay structures and average pay levels. This is not odd since every team within a factory produces a unique product, in conjunction with a distinct contractor. Also, note that factory

[12] Throughout the contract period, we hire extra workers to maintain accurate records of individual-level output on a twice-daily basis.

[13] This helps underscore to workers that we are paying attention to productivity. It also helps ensure our subsequent wage treatment effects are not confounded by information revelation by the factory.

managers maintain pay secrecy—each individual is privately told only his own wage; to the extent that we observe effects of relative pay differences, it is through self-disclosure among team members.

We collect information about the workers at several points in time. First, when we compile a list of interested workers after the village meeting, we record information about household size and landholdings. Second, once workers have reported to the worksites for the first day of training, we collect a very short baseline survey to capture worker demographic characteristics including age, literacy, employment history, and basic information about household assets. Third, throughout the period of employment, we collect daily measures of worker attendance, production, as well as a subjective measure on the quality of worker output. If workers are absent we record the reason for the absence when they return. Fourth, on the final day of employment, we record information on worker activity and earnings on days for which they were absent, as well as on the labor market activities of the other members in the household. We also use a survey instrument to map out their social networks with other workers at the worksite.

In Panel A of Table 2, we briefly describe the workers employed at our factories.[14] They are all males between the ages of 18-55 who engage primarily in casual labor. 54% of these workers own land, with average landholdings of 0.7 acres. In addition, 70% sharecrop land, with an average land size of 1.2 acres of land. While many workers do own land or sharecrop land, nevertheless, the land holdings are too small to generate year-round income. All of the workers primarily supply their labor to the daily labor market.

In Panel B of Table 2, we briefly describe the workers' collective labor market experiences. 71% of workers have worked on piece rates in the past. While 72% of workers report ever receiving wages that differ from the prevailing agricultural wage in their village (largely due to piece rate work such as stone cutting or non-agricultural work such as construction), only 17% of workers report ever receiving a wage different from that of other laborers in the village for the same task.

3.4. **Regression Specifications.** To test our key predictions, we compare outcomes between individuals in the *Heterogeneous* and *Compressed* teams, holding fixed a worker's production ranking rank and wage. Recall, from Table 1 that the most direct comparisons are between the Low rank *Heterogeneous* worker with the Low rank *Compressed_L* worker, the Medium rank *Heterogeneous* worker with the Medium rank *Compressed_M* worker, and the High rank *Heterogeneous* worker with the High rank *Compressed_H* worker. We refer to this set of six worker treatment types as the "relevant group". To use all of the variation in our experimental data, we use a differences-in-differences strategy to incorporate the pre-period production information.

[14]Table 2 is based on a small subset of responses from a representative sample of workers. The majority of the baseline and endline surveys are still being entered.

The most basic differences-in-differences approach restricts the sample to this "relevant" group of six worker treatment types and estimates the following regression specification:

(3.1)
$$\begin{aligned} y_{it} \;=\; & \alpha_0 \;+\; \alpha_1 Post_t \times Het_i \;+\; \alpha_2 Post_t \times Het_i \times RankM_i \;+\; \alpha_3 Post_t \times Het_i \times RankH_i \\ & + \alpha_4 Post_t \times RankM_i \;+\; \alpha_5 Post_t \times RankH_i \;+\; \lambda_i \;+\; \tau_t \;+\; \varepsilon_{it} \end{aligned}$$

In all of our empirical specifications, i indexes the worker and t indexes the day of each round. In most specifications, we focus on outcomes y_{it}, attendance and production. Our attendance measure is a binary variable capturing whether worker i is present on day t. The production output variable measures standardized production in units of one standard deviation of task-level production. To harmonize production across all ten tasks in the worksites, we use the pre-treatment data for each task to demean and standardize the raw production data. Raw production is coded as a zero when workers are absent.

Turning to the regressors, $Post_t$ is an indicator for the days after the wage treatment, Het_i is an indicator for being a member of a *Heterogeneous* team, while the variables $RankM_i$ and $RankH_i$ indicate Medium and High rank workers, respectively. In this differences-in-differences specification, any time-invariant team or worker characteristics are absorbed by the worker fixed effect, λ_i, while any time trends across the experimental period are captured by the day-by-round fixed effects, τ_t.

The key coefficient of interest is α_1, which measures differences in outcomes for Low rank workers in *Heterogeneous* with those in *Compressed_L* teams. α_2 captures the differential treatment effect for Medium rank workers in *Heterogeneous* vs. *Compressed_M* teams, compared to the treatment effect for the Low rank workers. Similarly, α_3 captures the differential treatment effect for High rank workers in *Heterogeneous* vs. *Compressed_H* teams, compared to the treatment effect for the Low rank workers. Recall that our main prediction is $\alpha_1 < 0$.

Note that by restricting the sample to only the so-called "relevant" workers, the above specification ignores variation that could be helpful in estimating the round-by-day fixed effects. To incorporate the observations from the "irrelevant" workers (the complement of the "relevant" group), we augment Equation 3.1 as follows:

(3.2)
$$\begin{aligned} y_{it} \;=\; & \alpha_0 \;+\; \alpha_1 Post_t \times Het_i \;+\; \alpha_2 Post_t \times Het_i \times RankM_i \;+\; \alpha_3 Post_t \times Het_i \times RankH_i \\ & + \alpha_4 Post_t \times RankM_i \;+\; \alpha_5 Post_t \times RankH_i \\ & + \theta_1 Post_t \times RankM_i \times Irrel_i + \; \theta_2 Post_t \times RankH_i \times Irrel_i + \; \theta_3 Post_t \times Irrel_i \\ & + \lambda_i \;+\; \tau_t \;+\; \varepsilon_{it} \end{aligned}$$

Here, $Irrel_i$ is an indicator for workers who are not among the six "relevant" types.[15] We also include interactions of $Irrel_i$ with both worker rank, including $RankM_i$ and $RankH_i$, and $Post_t$. Again, the time-invariant terms are absorbed by the worker fixed effect, λ_i. Note that the key parameters of interest $(\alpha_1, \alpha_2, \alpha_3)$ are estimated using the same sources of variation as in Equation 3.1. The benefit of this specification is that data from the "irrelevant" workers is used to estimate the round by day fixed effects, τ_t

Finally, we further augment the regression specification in Equation 3.2 to incorporate one additional consideration. Our key predictions require that the production team be the clear reference group against which each worker compares his pay. To ensure that this is the case, we incorporate information about the relative isolation of each team into our empirical specification. Using detailed seating charts from each of the worksites, we identify whether a team had any neighboring teams that could be observed or overheard from that team's location.[16] The researchers assigned teams to locations, so the presence of neighbors is exogenous to both treatment status and worker characteristics. Our main regression specification is therefore:

$$
\begin{aligned}
y_{it} \;=\;& \alpha_0 + \alpha_1 Post_t \times Het_i + \alpha_2 Post_t \times Het_i \times RankM_i + \alpha_3 Post_t \times Het_i \times RankH_i \\
& + \alpha_4 Post_t \times RankM_i + \alpha_5 Post_t \times RankH_i \\
& + \bar{Irrel}'_{it}\theta + \bar{Neigh}'_{it}\gamma + \lambda_i + \tau_t + \varepsilon_{it}
\end{aligned} \tag{3.3}
$$

The vector $Neigh'_{it}$ contains a time-invariant indicator for having a neighbor $Neigh_i$, interacted with $Post_t$, the treatment of worker i, the rank of worker i, and the treatment status of $Neigh_i$. We present the results of estimating Equation 3.3 in Section 4.2.

.

3.5. **Perceived Fairness.** One of our key aims is to understand the conditions under which relative pay differences may be deemed acceptable by workers. As described above in Section 3.1, we consider two sources of heterogeneity that were built into the experimental design.

First, because workers are randomly assigned to teams, the relative differences in productivity between the Low and Medium rank workers and between the Medium and High rank workers vary exogenously. When pre-treatment productivity differences are higher, wage differences may seem more justified. Second, some of the production tasks are more observable than others. For relatively unobservable tasks, it may be harder for workers to see that they are less productive than their higher-paid peers. In order to test these

[15]Recall that the "relevant" workers are all of the members of the *Heterogeneous* teams, along with Low rank workers in *Compressed_L* teams, Medium rank workers in *Compressed_M*, and High rank workers in *Compressed_H*.

[16]Seating charts of each worksite are available upon request.

hypotheses, we estimate heterogeneous effects regressions, fully interacting the variables in Equation 3.3 with measures of productivity differentials and task observability, respectively. Results are presented in Section 4.3

4. Results

In what follows, we present results from fourteen worksite rounds, employing a total of 378 workers.

4.1. **Knowledge of Co-worker Wages.** Given that managers maintained pay secrecy throughout the experiment, the wage treatments should only have power if workers discussed their wages with one another. Using the endline survey, we asked workers about their knowledge of their co-workers' wages. Among *Compressed* teams, approximately 97% of workers reported that they could name the wage of at least one team member, and 98% of those workers correctly reported that both co-workers earned the same wage as themselves. Among *Heterogeneous* teams, however, only 92% of workers reported that they could name the wage of at least one team member, and of those workers, only 82% had the correct beliefs about their co-workers' wages. Overall, this implies that 95% of workers on *Compressed* teams and 76% of workers on *Heterogeneous* teams could correctly report co-worker wages. The accuracy of the beliefs is similar for the Low, Medium, and High rank workers. These findings indicate that overall, there was a substantial amount of discussion of wages within teams. In addition, the difference in information sharing between the *Compressed* and *Heterogeneous* teams, while only suggestive, is consistent with awkwardness of discussing pay with teammates when there are wage differences.[17]

4.2. **Effects of Pay Disparity.** We now turn to our main results. Figure 3 provides an overview of the underlying data. It plots average production on each day for each of the 3 sets of relevant pairwise comparisons. Among Low rank workers, in the baseline period, production in the *Compressed_L* and *Heterogeneous* teams shows a common trend as workers gain experience. The treatment ("post") period begins on day 0, when workers are privately told their individual post-training wage. Within about 5 days (i.e. by the first pay period following the wage change), differences in output start to emerge, with workers on *Heterogeneous* teams (who are paid less than their peers) reducing output relative to the *Compressed_L* teams. This delay in the onset of treatment effects is consistent with non-immediate diffusion of pay information among team members; it potentially suggests that it takes up to a week (one pay period) for workers to become aware of pay differences within the team. In addition, there is no evidence of positive effects of relatively higher pay for High rank workers who are paid more than their peers (Panel C). Finally, the graphs

[17]It is also worth mentioning that only four workers on *Heterogeneous* teams thought that all of the workers earned the same wage.

suggest that the output of Medium rank workers in *Compressed_M* is no different from their counterparts in *Heterogeneous* pay teams.

Table 3 presents the estimation of Equation 3.3 on the full sample of workers in each round. Columns 1-4 measure the effects on standardized production, while Columns 5-8 measure effects on attendance. We find that workers decrease both production and attendance when they are paid less than their team-mates, holding their absolute wage levels fixed ($\alpha_1 < 0$). The output of Low wage workers declines by 0.380 standard deviations in response to the *Heterogeneous* treatment (significant at the 1% level). This effect is robust to the inclusion of individual fixed effects (Col. 2, our preferred specification). The 0.361 standard deviation decrease in Col. 2 is equivalent to about a 22% reduction in output relative to the *Compressed* (control) treatment mean. While overall output declines for these workers, we see little evidence for decreases in quality (Appendix Table 10).[18]

Turning to attendance, we find that on average, relatively lower-paid workers are approximately 11.8 percentage points less likely to come to work (on a base of 93.9% attendance pooled across the *Compressed* groups, Col. 6). Given that overall attendance increases on the weekly paydays for all workers, we should expect this effect to be strongest on non-paydays. Cols. 7-8 indicate large decreases in attendance on non-paydays (14.6 percentage points), but small and insignificant attendance effects on paydays. It is important to note that workers are not compensated for days during which they are absent. Thus, a natural question to ask is how these decreases in attendance map to the worker's overall earnings. Combining our administrative data with endline survey data on workers' outside employment activities and wages when absent, we estimate that these workers give up about 9% of their total earnings to avoid a workplace where they are paid less than their peers.

In contrast, we find little evidence that performance improves when workers are paid more than their peers. In Columns 1-4, we cannot reject that there is zero impact on production for high rank workers in *Heterogeneous* teams relative to their counterparts on *Compressed_H* teams. The point estimates (given by $\alpha_1 + \alpha_3$) are actually negative, though statistically indistinguishable from zero, as reported in the F-test p-values at the bottom of the table. In fact, the effects on attendance for these workers are negative, and significant at the 5% or 10% level depending on the specification (Cols. 5-8). In addition, there is no evidence that medium rank workers on the *Heterogeneous* teams have different average output or attendance than their counterparts on the *Compressed_M* teams (given by $\alpha_1 + \alpha_2$). Our results indeed suggest that the effort and attendance responses from being paid less than one's co-workers are much larger than any positive effects from being paid more than one's co-workers. While we did not have strong ex ante predictions for the behavior of the Medium and High rank types, our results on the High rank types may be

[18]It is difficult to quantify quality levels in our setting. To obtain a rough proxy, for a subset of rounds and days, we had management rate the quality of each worker's output for that day on a scale of 1-5. We do not see evidence for a change in these subjective quality ratings. It is possible, however, that this subjective measure is too crude and noisy to enable us to examine effects.

indicative of a hostile work environment. While only suggestive, exit focus groups suggest that the High rank workers in *Heterogeneous* teams may have felt awkward while at work. This may have results in a need to work even harder to justify the higher wages to their teammates. Such a story may lead to our observed effects of more absences, but no overall decreases in productivity.

One natural question is whether the large negative impacts on the Low rank workers persist or instead wear off over time. In Table 4, we separately estimate Equation 3.3 over three parts of the post-treatment period.[19] Workers were paid their earnings weekly (every Friday). The evidence suggests that differences in relative pay become especially observable or salient after the first post-treatment payday, and then strengthen over time.

Attendance and Effort Decomposition. We have documented that there are large, negative effects of *Heterogeneous* wages on production when a worker is paid less than his peers. This deleterious effect on production can occur through both the extensive (attendance) and intensive margins (effort conditional on attendance). The attendance effect poses problems for identifying any intensive margin effects on effort. If dis-advantageous peer wage comparisons affect some types of workers more than others, then running the regression in Equation 3.3 conditioning on attendance may introduce a potentially severe selection problem. Thus, we do not run the conditional regression, but instead rely on two different strategies that provide suggestive support that both channels are at play.

First, we estimate effects limiting our analysis to only paydays. On paydays, workers almost always come to work in order to collect their payments (overall attendance is 97%); reasons for missing work on those days are usually idiosyncratic such as illness (based on worker self-reports). Consistent with this, treatment status has a small and statistically insignificant effect on attendance on paydays (Table 3, Col. 8). However, we find that negative effects on standardized production hold robustly on paydays as well, with a 0.359 standard deviation decrease (significant at the 10% level) (Table 3, Col. 4). The fact that the effect on output is similar on paydays even though there is no attendance effect could be due to a larger treatment effect on paydays (because pay disparities feel more salient on those days). Alternately, it could arise from a compositional effect, where the workers who are most aggrieved are absent on non-paydays; when they come to work to collect their pay, their lower productivity creates a larger output decrease.

Second, we use a back-of-the-envelope calculation to decompose these effects. The mean output conditional on attendance for Low rank workers in the *Compressed_L* team is 1.86. The effect of *Heterogeneous* wages for the Low rank workers on attendance is -11.8 percentage points. If the full treatment effect on production were coming through attendance, then we would predict an output decrease of $-0.118 \times 1.86 = -0.219$ standard deviations. This corresponds to 60% of the total effect on production. We do acknowledge that both

[19]The full pre-treatment period is included in all regressions.

of these approaches have limitations. However, both findings suggest that in our setting, morale effects operate through both the labor supply and effort channels.

4.3. **The Role of Perceived Justifications.** We now investigate if there are circumstances that might mitigate the negative effects of lower relative pay that we measure above.

We first examine whether morale effects of pay differences are mediated by "actual fairness"–the difference in peers' productivity levels. Results are shown in Table 5. For each low and medium rank worker in each team, we compute the difference in mean baseline productivity between that worker and his next higher-ranked peer. We then add interaction terms of measures of baseline productivity differences into the main specification.[20] In Column 1, the interaction term is the continuous linear productivity difference; in Column 2, it is a binary measure of whether the baseline productivity difference is above the mean. Both columns paint a similar picture. Compared to their counterparts on the *Compressed_L* teams, low rank workers in *Heterogeneous* strongly lower output when they are closer in productivity to their higher-paid medium rank peers. However, when the difference in baseline productivity is large, we cannot reject that there is zero effort reduction for these workers. Column 3 shows that these results are robust to including interactions of treatment status with the worker's own baseline productivity–so that the effects of interest are identified off of changes in co-workers' productivity (rather than one's own level of productivity). Column 4 shows that the results are robust to excluding the bottom decile of low-rank workers (in terms of baseline productivity). This indicates that the results are not driven by the fact that the least productive low-rank workers hit some production floor, and that is what drives the interaction effects of interest. Columns 5-8 show that the effects on attendance follow a similar pattern.

We next check for mediating effects of "perceived fairness"–whether co-worker output is observable. Some of our tasks, by nature, are much more observable than others. We quantify observability using data from 3-week pilot rounds with a different sample of workers (conducted before the start of this experiment). Unlike in our main experimental rounds, in the pilot rounds, workers were never told their productivity rankings. On the last day of these pilot rounds, we instead asked workers to rank their co-workers by productivity.[21] We consider a task to be more observable if the rankings given to us by the workers better match the actual productivity rankings obtained from the administrative data. In Figure 4, we present the correlations between the actual and survey rankings for each production task, which range from negative values (bottom-coded at zero) to 0.87.

[20]Note that this upward productivity comparison is not defined for the high rank workers.

[21]At the end of the pilots, we added two additional tasks. We followed a similar procedure with a separate sample of workers to quantify observability for these two additional tasks, and these were run concurrently with the experiment.

Table 6 presents a pattern of results similar to that in Table 5. Column 1 adds interactions with the continuous measure of observability, and Column 2 adds interactions with a dummy for whether the production task has an observability correlation above 0.7. The results in both columns indicate that negative morale effects for low-rank *Heterogeneous* workers are concentrated in production tasks where it is more difficult to observe co-worker productivity. When productivity is highly observable, we also cannot reject an absence of negative morale effects of being paid less than one's peers. Taken together, both sets of results suggest that the underlying context shapes a worker's response to receiving a lower wage than his peers.[22]

4.4. **Effects on Team Cohesion.** The above results suggest that negative morale effects reduce individual effort provision to the firm. Another potentially important aspect of morale is cooperation among team members—since many jobs involve some degree of team production, either explicitly or implicitly. If the effects of relative pay differences operate through emotions such as resentment or envy of other co-workers, this may erode the ability of peers to cooperate in their own self-interest.

To test for effects on team cohesion, we developed two sets of cooperative games that required teamwork. Workers played these games at endline—on the last day of work, as part of a "fun farewell" day. Importantly, there was clearly no benefit to the firm from worker effort on the games. Workers were paid piece rates for performance on the games, in addition to their usual daily wage.[23]

In the first game, workers had to build a tower with the other members of their assigned product team. Each team was given a set of raw materials (e.g., cardboard, pens, rubber bands, playing cards), and asked to build as high a tower as possible with these materials. Teams were given a 25 minute time limit, though were free to stop earlier if they wanted. The payment schedule for this game was a linear piece rate for the tower's height (measured in cm), paid equally to each of the three team members.

In the second set of games, workers had to solve cooperative puzzles in pairs of two. In the "Spot the Difference" game, each person in the pair received a sheet with similar pictures on both sheets. The workers had to compare their sheets with each other, and circle any difference in the pictures on the two sheets (Appendix Figure 6, Panel A). Payment was a piece rate for every correct difference that was circled on *both* workers' respective sheets. In the "Symbol Matching" game, each pair member was given a sheet with a grid of symbols. Workers had to match symbols—circling all instances where the same symbol appeared in the same grid position in both of their respective sheets (Appendix Figure

[22] Appendix Table 11 verifies that the productivity difference results (Table 5) and observability results (Table 6) are identified off of different sources of variation. There is substantial non-overlap among these two measures in the sample.

[23] At the end of the day, we randomly selected one of the games for each worker, and the worker received his piece rate earnings for that game only. Workers were told we would randomly select one game for payment in advance. Note that these endline games were only played in the final 8 rounds of the experiment (80 teams only).

6, Panel B).[24] Payment was a piece rate for every correct match that was circled on *both* workers' respective sheets. Payment was therefore always the same for members of a pair within each pair-game.

For these games, we constructed pairs by reshuffling workers across product teams. Each worker was paired with 8 different people—playing four iterations each of Spot the Difference and Symbol Matching. Specifically, we randomized pair construction so that in 50% of cases, the two members of the pair were drawn from the same product team; in the other 50% of cases, paired workers were from different product teams. In other words, each worker played each game twice with each of his product teammates, and twice with someone from another team—for a total of 8 games.

Table 7 indicates that *Heterogenous* pay decreases team cooperation. In the tower building game, *Heterogeneous* teams build towers that are 8.974 cm (17%) shorter than the other teams on average (Col. 1). This effect persists when comparing *Heterogeneous* teams to only the *Compressed_Low* and *Compressed_Medium* teams, with a treatment effect of 7.978 cm or 15% (Col. 2).

Table 7, Cols. 3-5 examine effects on the cooperative pairwise puzzles. If both members of a pair are from the same team, they score 0.929 points (21%) lower if that team had *Heterogeneous* pay than if it had *Compressed* pay (Col. 3). Overall, the results indicate that among *Compressed* teams, when both members of the pair are from the same product team, workers perform better than when playing with a stranger (i.e. someone from another team). However, the interaction term indicates that this effect is completely undone in *Heterogeneous* pay teams. In addition, note that *Heterogeneous* team members do not perform worse than *Compressed* team members in general—the point estimate for at least one Heterogeneous worker in pair is actually positive (though insignificant). Rather, *Heterogeneous* pay workers only perform worse when they are paired with another person from their own team.[25]

It is worth noting that the worker productivity rankings in the main experiment have some predictive power for performance on these endline games. Cols. 4-5 indicate that on average, a pair with a Low or Medium rank worker score 18.9% and 13.8% less, respectively, than a pair with a High rank worker. This suggests that the baseline productivity rankings capture, in part, some stable differences in ability or effort across workers.

[24]We thank Heather Schofield for providing us with the Symbol Matching game grids.

[25]Since the endline games were conducted on the last day of work—when all workers received their final pay for the contract job—attendance was high. However, 4.6% of workers were absent on this last day. In the tower game, teams just played with whichever workers were present. In the cooperative puzzles, if a worker was absent, then the person who had been paired with that worker for a pair-game sat out during that round and their score for that pair-game is coded as 0. Appendix Table 14 verifies that the endline game results are not driven by differential absence across teams. First, as expected on the last day, treatment status did not affect worker attendance. For example, the difference in average attendance rates between *Heterogeneous* and *Compressed* teams is -0.0045 (Col. 1). Cols. 3-4 replicate results for the tower game only for teams where all workers in the team were present. Cols. 5-6 replicate results for only those pair-games where both mebers of a pair were present. The results for both sets of games are similar to those in Table 7.

Overall, the endline games suggest a decrease in *Heterogeneous* workers' ability to cooperate in order to earn money for themselves. This decreased performance cannot be due to retaliation against the firm, since the firm gains no benefit from performance on the games. In addition, the cooperative pair games indicate that lower *Heterogeneous* worker performance is not due to general disgruntlement. Low performance only arises when they must work with the other people in their own product team.

4.5. **Other Results.**

Heterogeneous Effects by Caste Composition. We next investigate whether other factors may mitigate or exacerbate the negative productivity effects of earning less than one's peers. Given our setting in rural India, one natural dimension of heterogeneity to explore is the caste composition of the production teams induced by our randomization. In our baseline survey, we recorded information about worker caste. Among the workers in our analysis sample, 72% reported being from one of the traditionally disadvantaged Scheduled Castes or Scheduled Tribes (SCSTs) of Orissa.

We note that we had no strong priors ex ante for how caste composition might interact with upward pay comparisons, and we view this exercise as suggestive. On one hand, if differences in pay happen to align with traditional caste hierarchies, then low caste and low-rank workers may feel especially resentful and may even interpret differences in wages as coming from caste-based discrimination. However, it may be when the traditional caste hierarchies are upended that workers feel most aggrieved – i.e., when the low rank workers happen to be of high castes.

We present results from heterogeneous treatment effects regressions in Table 8. Columns 1-3 present effects on standardized production, while columns 4-6 present effects on daily attendance. We consider three different measures of caste heterogeneity. First, in columns 1 and 4, we ask whether the effects of heterogeneous pay are any different when the high rank worker is not an SCST (i.e., high caste). We find suggestive evidence that the effects of heterogeneous pay are strongest on the low rank worker when the high rank worker is high caste. However, these tests suffer from a lack of statistical power. In columns 2 and 5, we split the teams by whether the low rank worker is high caste. We find suggestive evidence that the largest effects on the low rank worker occur when the low rank worker is SCST (i.e., low caste).

Given these suggestive patterns, we examine the case when the low rank worker is low caste and the high rank worker is high caste in columns 3 and 6. Note that this situation most closely resembles the traditional caste hierarchy. We find that the effect of heterogeneous pay on the low rank worker becomes much stronger when the traditional hierarchy is followed. One interpretation is that breaking from the traditional hierarchy only happens in meritocratic situations. Hewing to traditional caste hierarchies may signal to workers the absence of meritocratic pay.

The specifications in columns 3 and 7 also allow us to shed light on what may be driving the negative effects on productivity and attendance experienced by the high rank workers in the heterogeneous pay teams. When the traditional caste hierarchy is followed, we do not find any negative effects on the high rank worker. However, it is when the traditional hierarchy is not being followed that the high rank workers reduce their attendance. This may come from a hostile work environment or feelings of awkwardness for the low caste, high rank workers. It may also be that high caste, low rank workers are able to intimidate low caste, high rank workers to keep them from coming to work.

These results suggest to us that underlying social relationships may interact with pay policies in important ways. The interaction of workplace dynamics and social hierarchies is a fruitful direction for future work.

Effects of Pay Increases and Gift Exchange. While not our primary goal, we can use our experimental variation to explore whether the *Compressed_M* and *Compressed_H* teams increase their productivity levels relative to the lower compensated *Compressed_L* teams. In Table 9, we present differences-in-difference regressions before and after the wage change, across team wage assignment. In each regression, the omitted category is the *Compressed_L* team in the post wage change period. Columns 1 and 4 present results from the full sample period, Columns 2 and 5 restrict to the first two days of the post-wage change period, while Columns 3 and 6 focus only on the first week post wage change. If teams reciprocate higher wages with higher levels of productivity, then we should expect positive productivity effects from the *Compressed_M* and *Compressed_H* teams. Further, any positive peer effects from harder working teammates should amplify such an effect. However, we find no evidence of positive gift exchange on productivity at any time horizon. It does appear that *Compressed_M* and *Compressed_H* teams may increase attendance, which could be consistent with a simple income effect, though the coefficients are not statistically significant.

5. threats to Validity and Discussion

Internal validity concerns. Could an explanation other than relative pay comparisons deliver our findings? One potential confound is career concerns. Suppose that—even though we stress to workers that this is a one-time seasonal temp job—workers supply effort partly in hopes of increasing the probability of future employment. When a worker in *Heterogeneous* observes he is paid less than his co-workers, he may believe the firm is less likely to hire him in the future and therefore decrease effort. However, our design generates additional predictions that are not consistent with career concerns. First, we find that workers that are close in productivity to their higher paid colleagues—and therefore more valuable to the firm—are more likely to decrease effort. In contrast, in a career concerns model, workers that are relatively further behind their colleagues should be more likely to believe their chances of future employment are low, generating the opposite prediction. Second, given that we find large extensive margin effects on attendance, it is difficult to explain under a career concerns model why workers are willing to give up full-time earnings (due to poor attendance) and sit at home unemployed. Similar arguments apply to the potential concern that lower-paid Low rank workers in *Heterogeneous* decrease effort because the wage is a signal that helps them learn about their own type, affecting their future expectations.

Another potential issue is possible gift exchange effects from the fact that all workers receive a wage increase after training. Such effects should be common across all treatments, since all workers receive a pay raise. If peer effects from more higher-paid (and therefore more productive) peers increase own output, this would make it harder to detect our main effects. Further Table 9 suggests that gift exchange is not very important in our setting.

Our design relies on the presumption that each worker's reference group is comprised of his two teammates. If workers instead compared themselves to those in other teams, it could create contamination across treatment groups. However, this should decrease the potency of our treatments and make it harder to find our hypothesized effects. Given our experimental design, we believe that it is reasonable to expect that for someone making rope, the other 2 people making rope (who sit with and work next to him daily) are a more salient comparison group than those making incense sticks or brooms (who have their own unique seating area and production task). This is consistent with the findings of Card et al. (2012), where workers cared about pay relative to others in their particular departments, and less about other departments in the same workplace.

We can also rule out the hypothesis that our results are driven by an instability in the rankings of workers across time. In Appendix Table 13, we show that our main results are robust to dropping individuals who fell in rank between the training period and the ranking period. While the estimates lose statistical power, they barely change in magnitude.

Finally, our experiment is not well-suited to precisely disentangle the psychological mechanism that drives effort reductions. For example, unfairness and envy are different emotions that could trigger a decrease in morale, and could micro-found reference dependence in utility. We do not take a stance on the underlying psychology—what matters for our interpretation is that the mechanism is something that operates through reference-dependence in co-worker pay.

Humiliation from being identified as a low productivity type in front of one's peers, for example, is one competing explanation. In this world, the effects on output might still operate through a loss of worker morale, but not through reference dependence in co-worker pay. However, this class of mechanism is also unlikely to explain our full results. Under a story of humiliation, workers that are close in productivity to their higher-paid colleagues should experience less shame and motivation to decrease output. We find the opposite result. In addition, note that we maintain a policy of pay secrecy; if workers disclose that they are lower paid, then they do so voluntarily.

External validity concerns. Two important external validity concerns stem from whether the wage treatments appear unusual to the workers. First, since we have selected tasks in which output is measurable, firms could consider paying piece rates or some other form of explicit incentives. However, whether this makes sense will depend on the cost of the monitoring technology. In the experiment, we bear the considerable expense to hire extra staff to measure each worker's output daily. In addition, in the local context in which our experiment takes place, it is common for workers to receive flat wages even when output is measurable. For example, many retail goods are produced under both piece rates and under flat wages by firms in the study region. Similarly, some employers pay piece rates while others in the same village pay fixed daily wages to harvest a given crop. It is also the case that under explicit incentives, quantity may improve but at the expense of quality—such multitasking problems are well documented.

Second, workers may have found it odd that some teams were paid based on baseline productivity while others were paid equal wages. We developed our design to mitigate this concern to the full extent possible. This is one of the driving reasons for having each team produce a unique task, which in turn is associated with its own unique contractor. There was thus no opportunity to compare one's own wages with those of other teams producing the same output in the same worksite.

A related issue is whether it is reasonable for the firm to pay differential wages based on training output (rather than ex-post output). This is also common in many settings. For example, firms usually set the pay of short-term consultants based on expected productivity. Even for salaried workers, pay is usually based on ex-ante expectations, with stickiness throughout a worker's tenure at the firm (Fehr et al. 2009)—this is not adjusted with new information on ex-post performance, but rather re-negotiated at infrequent intervals. More generally, explicit incentives like piece rates based on ex-post output are not that common in poor or rich countries (e.g. Dreze and Mukherjee 1989, MacLeod and Parent 1999).

One potential benefit to firms of differential pay is dynamic incentives: workers know that if they work hard now, it could lead to higher pay in the future.[26] Our study design shuts down this channel since after the training period, there is no further chance of wage changes. However, the objective of our study is not to isolate the optimal pay policy for firms. Rather, our objective is to test whether relative pay comparisons affect effort—a topic on which there is limited field evidence, and which is currently ignored in mainstream agency models of pay structure. The optimal pay policy for a firm would depend on weighing the potential costs of differential pay (e.g. morale reductions) against the potential benefits (e.g. dynamic incentives). In addition, evidence on when differential pay is most likely to damage morale—for example when output is harder to precisely quantify or less observable

[26]For example, pay differences could also affect selection of workers into the firm (Lazear 2000, and Guiteras and Jack 2014).

by co-workers—can enhance our understanding of why we observe differential pay in some occupations and not in others.

6. Conclusion

We find that when workers are paid less than their peers, they reduce output and are willing to give up substantial earnings through decreases in attendance. The perceived justification for these pay differences plays an important role in mediating these effects. Our findings provide support for reference dependence in co-worker pay, and indicate that transparency about the firm's rationale for pay is important for fairness perceptions and output.

The results suggest that optimal pay for a given worker will potentially be a function of co-worker pay. This could help us understand why wage compression—when wages vary less than the marginal product of labor—is so prevalent. For example, in many occupations—from tollbooth attendants to supermarket cashiers—all workers in a firm are paid the same fixed hourly wage even though managers are aware of their productivity differences. In casual daily labor markets—for example among agricultural day laborers in India or California—an employer usually pays all workers the same prevailing daily wage, despite knowing which are more productive than others (e.g. Dreze and Mukherjee 1989). While such behavior is hard to reconcile under neoclassical agency theory, if relative pay is important, then it may be profit maximizing for firms to compress wages. Our results may also have bearing on explaining the conditions under which differential pay will arise—for example, when it is easy to observe and quantify co-workers' relative productivity. This could help explain why workers accept earnings dispersion under piece rates or within sports teams (where performance statistics reflect productivity), but not among clerical workers at the University of California (Card et al. 2012).

Wage compression could have potentially important effects on labor market outcomes. For example, Akerlof and Yellen (1990) tie this to wage rigidity: if firms cannot cut pay after individual adverse shocks and therefore fire workers instead, this will increase unemployment and business cycle volatility. Wage compression may also have distributional consequences. If the wage for all laborers is the same, better quality workers will be hired first and worse quality ones may be more likely to face involuntary unemployment. This implies that small productivity differences may lead to large earnings differences, exacerbating inequality and amplifying the adverse effects of shocks like illness. Thus, the rationing mechanism may hurt the most vulnerable, generating a rationale for targeting in unemployment programs.

Relative pay concerns may also have relevance for the organization of production and firm boundaries. For example, they could influence whether workers of heterogeneous ability are organized within a firm or contract their labor through the external market. Consistent with this, Nickerson and Zenger (2008) argue that pay differences across firms serve as a hindrance to firm mergers. Similarly, firms may "specialize" in hiring workers of a given productivity level to avoid pay discrepancies. Relative pay concerns also have bearing on human resource policies—for example, they could help explain why about one-third of US firms require employees to sign nondisclosure contracts that forbid them from discussing their pay with their co-workers (Card et al. 2012).

In addition, our findings suggest that firms may have several potential tools at their disposal to manage morale in the presence of pay dispersion. For example, technologies that make it easier to quantify worker productivity could have aggregate output benefits not just through increased monitoring, but also through improved morale. Firms could also potentially alter the organizational structure of the workplace itself—through job titles, physical co-location of similar workers, or the construction of "teams"—to affect who a worker views as being in her reference group. Indeed, our experimental design leverages the insight that the organization of production can be manipulated to affect the reference group for relative pay comparisons.

While speculative, the above possibilities suggest a variety of ways through which relative pay concerns could affect pay structure, organizational arrangements, and other labor market outcomes. These possibilities are a promising direction for further research.

References

Adams, J. S.: 1963, Towards an understanding of inequity., *The Journal of Abnormal and Social Psychology* **67**(5), 422.

Akerlof, G.: 1982, Labor contracts as partial gift exchange, *The Quarterly Journal of Economics* **97**(4), 543–569.

Akerlof, G. A. and Yellen, J. L.: 1990, The fair wage-effort hypothesis and unemployment, *The Quarterly Journal of Economics* pp. 255–283.

Bartling, B. and von Siemens, F. A.: 2011, Wage inequality and team production: An experimental analysis, *Journal of Economic Psychology* **32**(1), 1–16.

Bewley, T. F.: 1999, *Why wages don't fall during a recession*, Harvard University Press.

Bracha, A., Gneezy, U. and Loewenstein, G.: 2015, Relative pay and labor supply, *Journal of Labor Economics* **33**(2), 297–315.

Brown, G. D., Gardner, J., Oswald, A. J. and Qian, J.: 2008, Does wage rank affect employees? well-being?, *Industrial Relations: A Journal of Economy and Society* **47**(3), 355–389.

Card, D., Mas, A., Moretti, E. and Saez, E.: 2012, Inequality at work: The effect of peer salaries on job satisfaction, *American Economic Review* **102**, 2981–3003.

Charness, G. and Kuhn, P.: 2007, Does pay inequality affect worker effort? experimental evidence, *Journal of Labor Economics* **25**(4), 693–723.

Charness, G. and Kuhn, P.: 2011, Lab labor: What can labor economists learn from the lab?, *Handbook of Labor Economics* **4**, 229–330.

Clark, A. E. and Oswald, A. J.: 1996, Satisfaction and comparison income, *Journal of Public Economics* **61**(3), 359–381.

Clark, A., Masclet, D. and Villeval, M.-C.: 2010, Effort and comparison income: Experimental and survey evidence, *Industrial and Labor Relations Review* **63**.

Cohn, A., Fehr, E., Herrmann, B. and Schneider, F.: 2012, Social comparison in the workplace: Evidence from a field experiment, *IZA Discussion Paper No. 5550* .

DellaVigna, S., List, J., Malmendier, U. and Rao, G.: 2015, Gift exchange at work, *Working Paper* .

Dreze, J. and Mukherjee, A.: 1989, Rural labor markets in india, *The Balance Between Industry and Agriculture in Economic Development, Vol, 3, Manpower and Transfers* .

Duesenberry, J. S.: 1949, *Income, saving, and the theory of consumer behavior*, Vol. 87, Harvard Univ Pr.

Easterlin, R. A.: 1974, Does economic growth improve the human lot? some empirical evidence, *Nations and Households in Economic Growth: Essays in Honor of Moses Abramowitz. P. A. David and M. W.Reder* .

Falk, A., Fehr, E. and Fischbacher, U.: 2008, Testing theories of fairness?intentions matter, *Games and Economic Behavior* **62**(1), 287–303.

Fang, H. and Moscarini, G.: 2005, Morale hazard, *Journal of Monetary Economics* **52**(4), 749–777.

Fehr, E., Goette, L. and Zehnder, C.: 2009, A behavioral account of the labor market: The role of fairness concerns, *Annual Review of Economics* .

Fehr, E., Kirchsteiger, G. and Riedl, A.: 1993, Does fairness prevent market clearing? an experimental investigation, *The Quarterly Journal of Economics* **108**(2), 437–459.

Frank, R.: 1984, Are workers paid their marginal products?, *American Economic Review* **74**, 549–571.

Frey, B. S. and Stutzer, A.: 2002, What can economists learn from happiness research?, *Journal of Economic literature* **40**(2), 402–435.

Gatcher, S. and Thoni, C.: 2010, Social comparison and performance: Experimental evidence on the fair wage–effort hypothesis, *Journal of Economic Behavior and Organization* .

Gneezy, U. and List, J. A.: 2006, Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments, *Econometrica* **74**(5), 1365–1384.

Guiteras, R. P. and Jack, B. K.: 2014, Incentives, selection and productivity in labor markets: Evidence from rural malawi, *NBER Working Paper* **19825**.

Hamermesh, D.: 2001, The changing distribution of job satisfaction, *Journal of Human Resources* **36**, 1–30.

Hamermesh, D. S.: 1975, Interdependence in the labour market, *Economica* **42**(168), 420–429.

Hicks, J.: 1932, *The Theory of Wages*.

Kahneman, D. and Tversky, A.: 1979, Prospect theory: An analysis of decision under risk, *Econometrica: Journal of the Econometric Society* pp. 263–291.

Kube, S., Maréchal, M. A. and Puppe, C.: 2013, Do wage cuts damage work morale? evidence from a natural field experiment, *Journal of the European Economic Association* **11**(4), 853–870.

Kuziemko, I., Buell, R., Reich, T. and Norton, M.: 2011, Last-place aversion: Evidence and redistributive implications", nber working paper no. 17234, *NBER Working Paper* .

Lazear, E. P.: 2000, Performance pay and productivity, *American Economic Review* pp. 1346–1361.

Levine, D. I.: 1993, What do wages buy?, *Administrative Science Quarterly* pp. 462–483.

Levitt, S. D. and List, J. A.: 2007, What do laboratory experiments measuring social preferences reveal about the real world?, *The Journal of Economic Perspectives* **21**(2), 153–174.

List, J. A.: 2009, Social preferences: Some thoughts from the field, *Annu. Rev. Econ.* **1**(1), 563–579.

Luttmer, E. F.: 2005, Neighbors as negatives: Relative earnings and well-being, *The Quarterly Journal of Economics* **120**(3), 963–1002.

MacLeod, W. B. and Parent, D.: 1999, Job characteristics, wages and the employment contract, *Federal Reserve Bank of Saint Louis Review* **81**, 13–28.

Marmot, M.: 2004, Status syndrome, *Significance* **1**(4), 150–154.

Marshall, A.: 1890, *Principles of Economics*.

Nickerson, J. A. and Zenger, T. R.: 2008, Envy, comparison costs, and the economic theory of the firm, *Strategic Management Journal* **29**(13), 1429–1449.

Pfeffer, J. and Langton, N.: 1993, The effect of wage dispersion on satisfaction, productivity, and working collaboratively: Evidence from college and university faculty, *Administrative Science Quarterly* pp. 382–407.

Rege, M. and Solli, I.: 2013, Lagging behind the joneses: The impact of relative earnings on job separation.

Veblen, T. and Almy, C.: 1899, *The theory of the Leisure Class: An Economic Study in the Evolution of Institutions*, Macmillan & Company, Limited.

Figures

Figure 1. Randomization Design

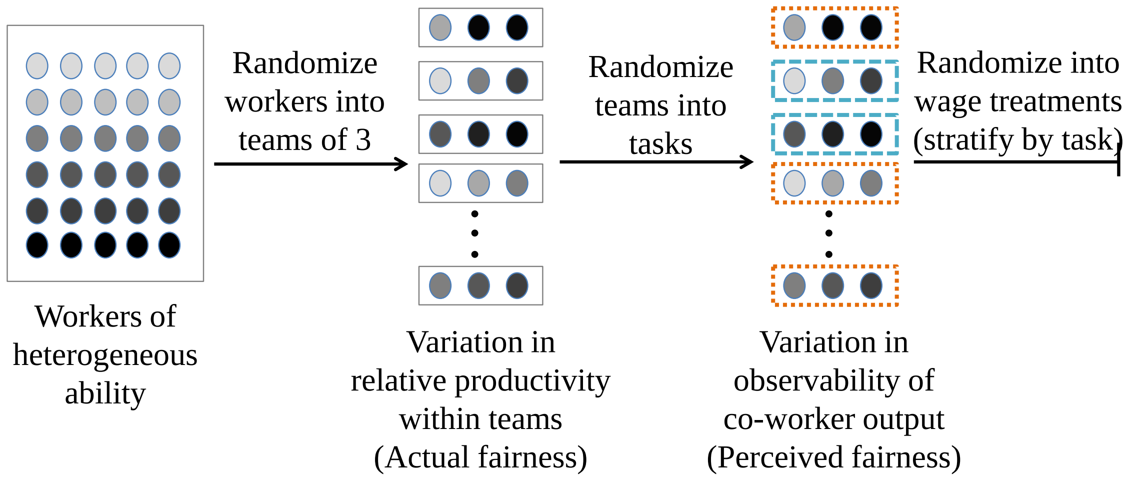


Figure 2. Time Line

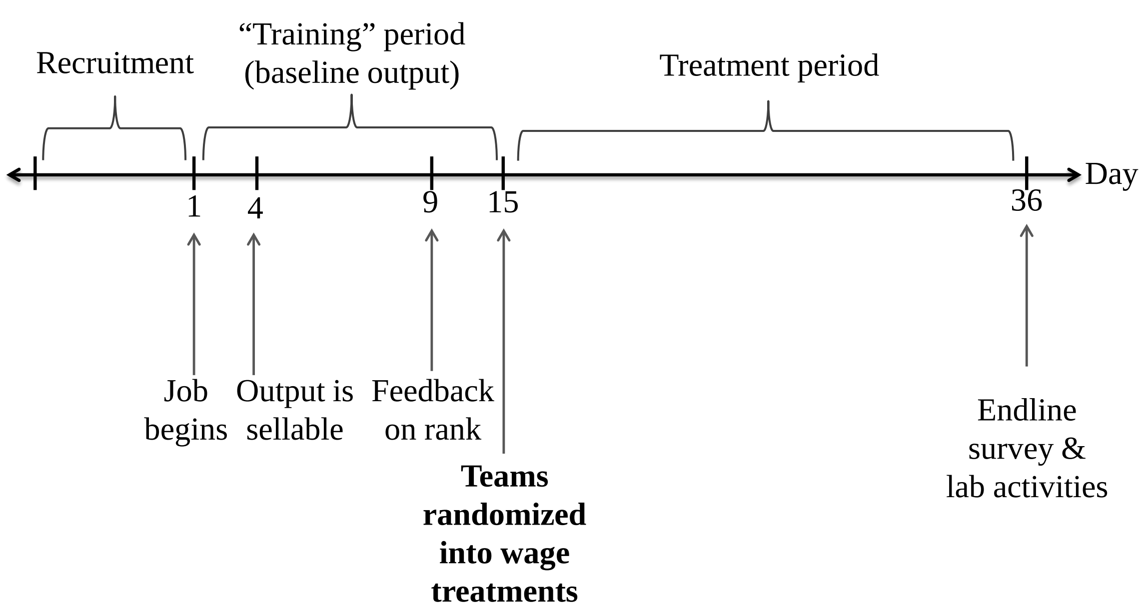
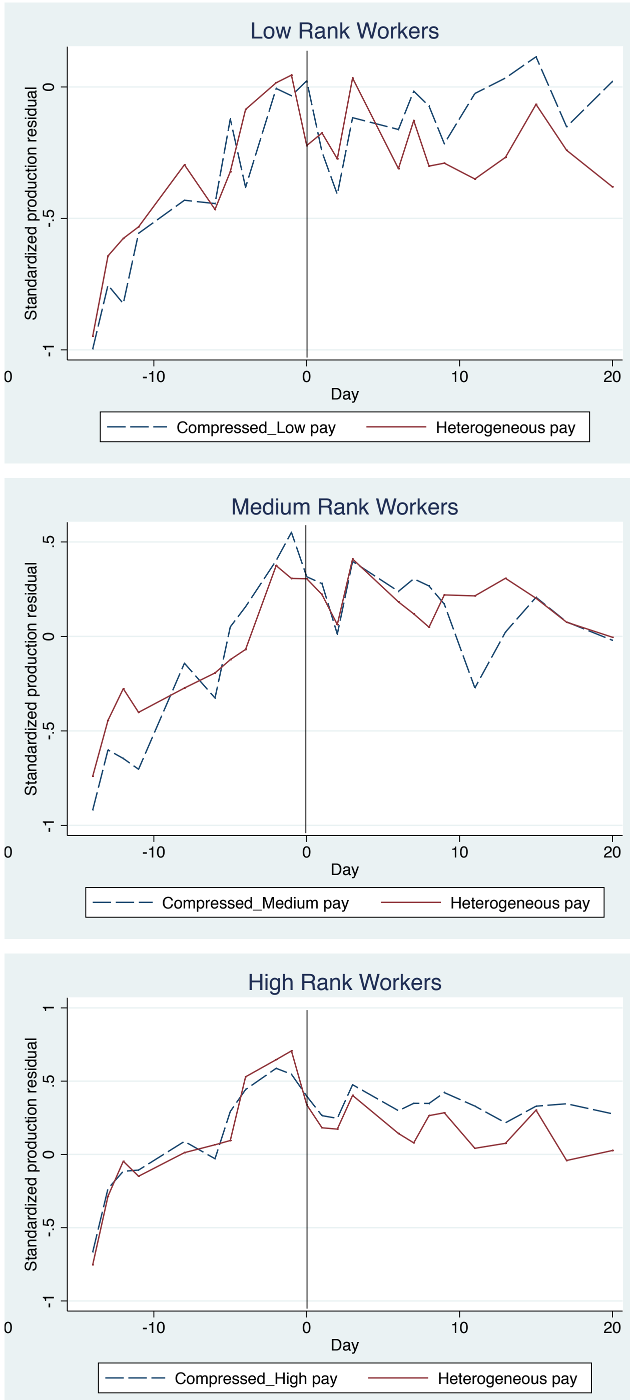


Figure 3. Effects of Heterogeneous Pay on Worker Output



Standardized production residual is the residual from a regression of standardized output on a dummy for festival days and dummies for each of the four treatment groups. The figures plot, for each day of the experiment, the average of the residuals for each group of workers (the relevant pairwise comparisons are shown of workers who earn the same absolute wage, but are in Compressed vs. Heterogeneous pay teams). Day=0 is the day wage treatments took effect (i.e. when workers were told their post-training wage).

FIGURE 4. Task Observability: Actual vs. Survey Correlations

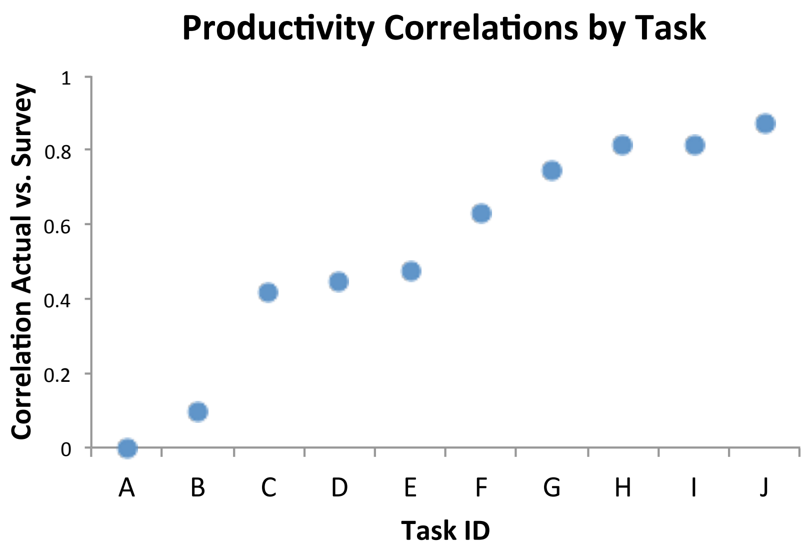


Figure plots the correlation between actual productivity rankings and perceived rankings by the workers (reported in endline surveys) for eight of the production tasks. Note that this data come from four pilot rounds where the research team did not inform workers of their production rankings. In our analysis, we split the tasks at the median level of observability (0.5 correlation).

Tables

Table 1. Treatments and Relevant Comparisons

Worker Type	Heterogeneous	Compressed_L	Compressed_M	Compressed_H
Low productivity	w_{Low}	w_{Low}	w_{Medium}	w_{High}
Medium productivity	w_{Medium}	w_{Low}	w_{Medium}	w_{High}
High productivity	w_{High}	w_{Low}	w_{Medium}	w_{High}

Table 2. Summary Statistics

Panel A: Demographic Characteristics	Mean
Own any land	0.54
Sharecrop any land	0.70
Land Owned (Acres)	0.68
Land Leased Out (Acres)	0.04
Own Land Cultivated (Acres)	0.67
Land Sharecropped In (Acres)	1.17
Female HH members	2.08
Male HH members	2.37
Female HH members engaged in labor force	0.81
Male HH members engaged in labor force	1.79
N	145

Panel B: Labor Market Experience	Mean
Received wage different from prevailing wage	0.72
Received wage different from other laborers in village	0.17
Ever worked on piece rates	0.71
N	313

Table 3. Effects of Pay Disparity

	Dependent variable: Output (standard dev.)				Dependent variable: Attendance			
	Full sample (1)	Full sample (2)	Non-paydays (3)	Paydays only (4)	Full sample (5)	Full sample (6)	Non-paydays (7)	Paydays only (8)
Post x Heterogeneous	-0.380***	-0.360***	-0.368***	-0.359*	-0.112*	-0.118**	-0.146**	-0.0264
	(0.134)	(0.137)	(0.135)	(0.205)	(0.059)	(0.057)	(0.057)	(0.083)
Post x Heterogeneous x Med wage	0.378**	0.385**	0.434**	0.162	0.0542	0.0584	0.0800	-0.0197
	(0.187)	(0.191)	(0.182)	(0.304)	(0.074)	(0.075)	(0.074)	(0.123)
Post x Heterogeneous x High wage	0.144	0.201	0.199	0.232	-0.0234	-0.0154	0.00985	-0.110
	(0.221)	(0.216)	(0.221)	(0.262)	(0.076)	(0.076)	(0.076)	(0.099)
Individual fixed effects?	No	Yes	Yes	Yes	No	Yes	Yes	Yes
F-test pvalue: (Post x Het) + (Post x Het x Med) = 0	0.989	0.864	0.621	0.462	0.193	0.223	0.158	0.636
F-test pvalue: (Post x Het) + (Post x Het x High) = 0	0.209	0.373	0.342	0.612	0.00979	0.00955	0.00983	0.0656
Post-treatment Compressed Mean	-0.000596	-0.000596	-0.0131	0.0445	0.939	0.939	0.931	0.968
R-squared	0.421	0.442	0.442	0.473	0.214	0.206	0.208	0.188
N	7691	7691	6105	1586	7691	7691	6105	1586

Notes: Difference in differences regressions. Post is an indicator that equals 1 if the day is after workers have been randomized into wage treatments, and 0 during the baseline training period. Regressions include day*round fixed effects, task-specific quadratic experience trends, and controls for neighboring teams. All coefficients are identified off comparisons of workers who earn the same absolute wage and have the same productivity rank within their team (see regression specification in text). Standard errors clustered by team.

Table 4. Effects Over Time

	Dependent variable: Output (standard dev.)			Dependent variable: Attendance		
	Before first payday in post period	Between first and second payday in post period	After second payday in post period	Before first payday in post period	Between first and second payday in post period	After second payday in post period
	(1)	(2)	(3)	(4)	(5)	(6)
Post x Heterogeneous	-0.105	-0.341**	-0.531**	0.0419	-0.0984	-0.262***
	(0.179)	(0.140)	(0.207)	(0.0855)	(0.0639)	(0.0882)
Post x Heterogeneous x Med wage	0.221	0.220	0.691**	-0.0794	0.00555	0.219*
	(0.243)	(0.204)	(0.291)	(0.102)	(0.0810)	(0.129)
Post x Heterogeneous x High wage	-0.0636	0.194	0.366	-0.153	-0.0370	0.109
	(0.265)	(0.287)	(0.245)	(0.112)	(0.102)	(0.102)
Individual fixed effects?	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.549	0.525	0.547	0.192	0.215	0.225
N	4,463	4,799	4,709	4,463	4,799	4,709

Notes: Post is an indicator that equals 1 if the day is after workers have been randomized into wage treatments, and 0 during the baseline training period. Regressions include day*round fixed effects, task-specific quadratic experience trends, and controls for neighboring teams. All coefficients are identified off comparisons of workers who earn the same absolute wage and have the same productivity rank within their team. Standard errors clustered by team.

Table 5. Perceived Justifications: Mediating Effects of Relative Productivity Differences

	Dependent variable: Output (standard dev.)				Dependent variable: Attendance			
Productivity difference measure	Production difference	Above mean difference	Above mean difference	Above mean difference	Production difference	Above mean difference	Above mean difference	Above mean difference
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post x Heterogeneous	-0.359**	-0.617***	-0.556***	-0.629***	-0.133**	-0.239***	-0.232***	-0.233***
	(0.162)	(0.198)	(0.203)	(0.222)	(0.0632)	(0.0679)	(0.0684)	(0.0728)
Post x Heterogeneous x Prod difference	0.222	0.633**	0.559*	0.621**	0.0911	0.275***	0.295***	0.278***
	(0.230)	(0.278)	(0.306)	(0.301)	(0.0734)	(0.0878)	(0.102)	(0.0917)
Post x Heterogeneous x Med wage	0.466**	0.800***	0.724***	0.809***	0.0987	0.222***	0.216***	0.215**
	(0.209)	(0.221)	(0.223)	(0.240)	(0.0813)	(0.0812)	(0.0824)	(0.0851)
Post x Heterogeneous x Med wage x Prod difference	-0.525*	-1.117***	-1.009***	-1.112***	-0.142	-0.375***	-0.390***	-0.374***
	(0.306)	(0.334)	(0.335)	(0.350)	(0.113)	(0.115)	(0.127)	(0.117)
Post x Heterogeneous x High wage	0.202	0.460*	0.396	0.480*	-0.000792	0.106	0.0940	0.102
	(0.216)	(0.245)	(0.250)	(0.265)	(0.0740)	(0.0738)	(0.0778)	(0.0794)
Individual fixed effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls for own baseline prodn x treatment x rank x post?	No	No	Yes	No	No	No	Yes	No
Dropping bottom 10% of low-rank workers?	No	No	No	Yes	No	No	No	Yes
R-squared	0.444	0.445	0.446	0.449	0.208	0.210	0.211	0.212
N	7,691	7,691	7,691	7,449	7,691	7,691	7,691	7,449

Notes: Regressions include day*round fixed effects, task-specific quadratic experience trends, and controls for neighboring teams. Prod difference is a measure of the baseline producitivty difference between a worker and his higher-ranked teammate (defined only for low and medium rank workers). Cols. 1 and 5 show interactions with the continuous productivity difference, and the remaining columns show interactions with a binary indicator for an above mean difference. Standard errors clustered by team.

Table 6. Perceived Justifications: Mediating Effects of Output Observability

	Dependent variable: Output (standard dev.)		Dependent variable: Attendance	
	Observability correlation (conitnous) (1)	High observability (indicator) (2)	Observability correlation (conitnous) (3)	High observability (indicator) (4)
Post x Heterogeneous	-0.704***	-0.487***	-0.139	-0.143**
	(0.245)	(0.184)	(0.095)	(0.072)
Post x Heterogeneous x Observability measure	0.861**	0.431*	0.090	0.101
	(0.392)	(0.247)	(0.137)	(0.079)
Post x Heterogeneous x Med wage	0.844***	0.470**	0.105	0.0789
	(0.281)	(0.219)	(0.105)	(0.083)
Post x Heterogeneous x Med wage x Observability measure	-1.072**	-0.317	-0.135	-0.086
	(0.499)	(0.298)	(0.198)	(0.113)
Post x Heterogeneous x High wage	0.399	0.105	0.005	-0.007
	(0.312)	(0.278)	(0.107)	(0.098)
Post x Heterogeneous x High wage x Observability measure	-0.562	0.192	-0.098	-0.054
	(0.471)	(0.303)	(0.154)	(0.100)
Post-treatment Control Mean	-0.0266	-0.0266	0.928	0.928
R-squared	0.193	0.191	0.200	0.201
Number of observations (worker-days)	7755	7755	7755	7755

Notes: Regressions include day*round fixed effects, task-specific quadratic experience trends, and controls for neighboring teams. In Cols. 1 and 3, the observability measure is the accuracy with which workers could assess co-worker output for a given production task (using a separate sample at baseline). The measure in Cols. 2 and 4 is a binary indicator for whether the accuracy rate for a production task is greater than 70%. Standard errors clustered by team.

TABLE 7. Effects on Team Cohesion: Endline Games

	Game: Tower building in teams		Game: Cooperative games in pairs		
Dependent variable	Tower height (1)	Tower height (2)	Number correct (3)	Number correct (4)	Above mean correct (5)
Heterogeneous team	-8.974**	-7.978**			
	(3.602)	(3.907)			
Compressed_High team		3.111			
		(5.553)			
Both workers from same team x Heterogeneous			-0.929**	-0.888*	-0.184***
			(0.464)	(0.465)	(0.064)
Both workers from same team			0.440	0.613**	0.0832*
			(0.287)	(0.290)	(0.045)
At least one Heterogeneous worker in pair			0.411	0.383	0.0796
			(0.345)	(0.334)	(0.050)
At least one low rank worker in pair				-0.820***	-0.0937**
				(0.269)	(0.038)
At least one medium rank worker in pair				-0.599**	-0.0924**
				(0.281)	(0.041)
Dependent variable mean	54.13	54.13	4.329	4.329	0.561
Observations	80	80	1,870	1,870	1,870
R-squared	0.265	0.269	0.199	0.207	0.194

Notes: This table shows results from cooperative team building games at endline. These games were only run in later rounds of the experiment. Cols. 1-2 shows results from a game where each production team was given materials to build as high a tower as possible (measured in cm). Regressions include round fixed effects. Observations are the number of teams. Cols. 3-5 show results from paired cooperative games with partners. The dependent variable is the number of items correct within each pair-game. Both workers from same team is a dummy that equals 1 if both members of the pair were on the same production team during the experiment. Regressions include round*game_station fixed effects and fixed effects of the order in which a game was played during the day. Observations are the number of pair-games. All standard errors are clustered by team.

Table 8. Effects by Caste Composition

	Dependent variable: Output (standard dev.)			Dependent variable: Attendance		
Caste Heterogeneity Measure	High rank worker is high caste	Low rank worker is high caste	High rank - high caste and low rank - low caste	High rank worker is high caste	Low rank worker is high caste	High rank - high caste and low rank - low caste
	(1)	(2)	(3)	(5)	(6)	(7)
Post x Heterogeneous	-0.248*	-0.487***	-0.231*	-0.0812	-0.146**	-0.0709
	(0.139)	(0.169)	(0.130)	(0.0612)	(0.0677)	(0.0578)
Post x Heterogeneous x Caste Measure	-0.514	0.344	-0.822*	-0.156	0.0404	-0.272**
	(0.360)	(0.255)	(0.421)	(0.114)	(0.0807)	(0.133)
Post x Heterogeneous x Med wage	0.283	0.429*	0.242	0.0420	0.0493	0.0357
	(0.194)	(0.225)	(0.190)	(0.0701)	(0.0921)	(0.0666)
Post x Heterogeneous x Med wage x Caste Measure	0.206	-0.211	0.198	0.0162	0.0895	0.0206
	(0.388)	(0.280)	(0.482)	(0.164)	(0.103)	(0.229)
Post x Heterogeneous x High wage	0.173	0.383*	0.0155	-0.0340	0.0456	-0.0865
	(0.191)	(0.216)	(0.233)	(0.0621)	(0.0669)	(0.0808)
Post x Heterogeneous x High wage x Caste Measure	0.231	-0.590*	1.103**	0.109	-0.199	0.396***
	(0.601)	(0.344)	(0.501)	(0.189)	(0.122)	(0.120)
Individual fixed effects?	Yes	Yes	Yes	Yes	Yes	Yes
F-test pvalue: (Post x Het) + (Post x Het x High) = 0	0.63	0.55	0.27	0.01	0.01	0.01
F-test pvalue: (Post x Het) + (Post x Het x Caste) = 0	0.03	0.46	0.01	0.03	0.07	0.01
F-test pvalue: (Post x Het x High) + (Post x Het x High x Caste)= 0	0.50	0.54	0.02	0.70	0.22	0.004
F-test pvalue: (Post x Het) + (Post x Het x High) + (Post x Het x Caste) + (Post x Het x High x Caste)= 0	0.47	0.23	0.86	0.23	0.01	0.53
R-squared	0.456	0.450	0.459	0.215	0.216	0.218
N	7,496	7,632	7,437	7,496	7,632	7,437

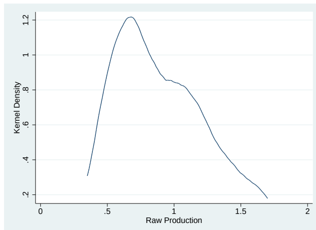
Notes : Regressions include day*round fixed effects, task-specific quadratic experience trends, and controls for neighboring teams. Prod difference is a measure of the baseline producitivty difference between a worker and his higher-ranked teammate (defined only for low and medium rank workers). Cols. 1 and 5 show interactions with the continuous productivity difference, and the remaining columns show interactions with a binary indicator for an above mean difference. Standard errors clustered by team. Differences in the number of observations between columns correspond to the fact that not all respondents were able to name their caste in the baseline survey.

Table 9. Cross Team Comparisons and Gift Exchange

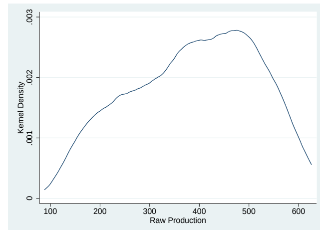
	Dependent variable: Output (standard dev.)			Dependent variable: Attendance		
	Full sample	First two days after wage change	First week after wage change	Full sample	First two days after wage change	First week after wage change
	(1)	(2)	(3)	(4)	(5)	(6)
Post x Compressed Medium Wage Team	-0.0872	-0.0704	-0.0369	0.00270	0.0529	0.0453
	(0.130)	(0.147)	(0.141)	(0.0373)	(0.0485)	(0.0372)
Post x Compressed High Wage Team	-0.0986	-0.134	-0.0741	0.0125	0.0382	0.0517
	(0.0998)	(0.114)	(0.112)	(0.0375)	(0.0455)	(0.0367)
Post x Heterogeneous Wage Team	-0.122	-0.0588	-0.0407	-0.0614	0.0147	0.00655
	(0.103)	(0.118)	(0.118)	(0.0372)	(0.0426)	(0.0362)
Individual fixed effects?	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.441	0.574	0.524	0.205	0.208	0.239
N	7,691	3,806	4,940	7,691	3,806	4,940

Notes : Post is an indicator that equals 1 if the day is after workers have been randomized into wage treatments, and 0 during the baseline training period. Regressions include day*round fixed effects, task-specific quadratic experience trends, and controls for neighboring teams. These regressions also include individual fixed effects and controls for post*rank. The omitted category is workers in the Compressed Low Wage Teams in the Post wage treatment period. Standard errors clustered by team.

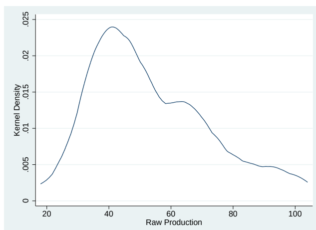
Appendix A. Supplemental Figures and Tables



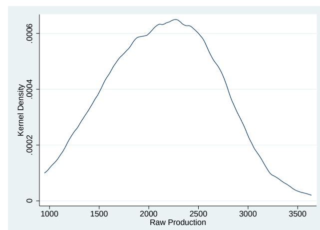
(A) Correlation = 0



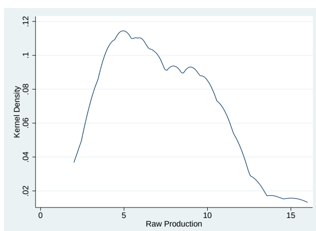
(B) Correlation = 0.09



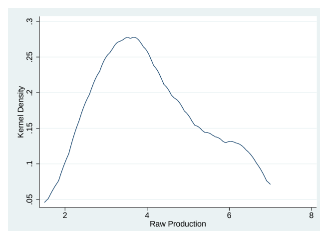
(C) Correlation = 0.42



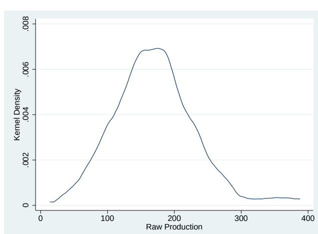
(D) Correlation = 0.45



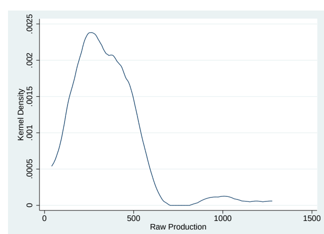
(E) Correlation = 0.47



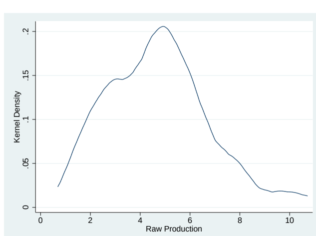
(F) Correlation = 0.63



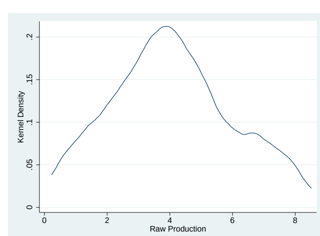
(G) Correlation = 0.74



(H) Correlation = 0.81



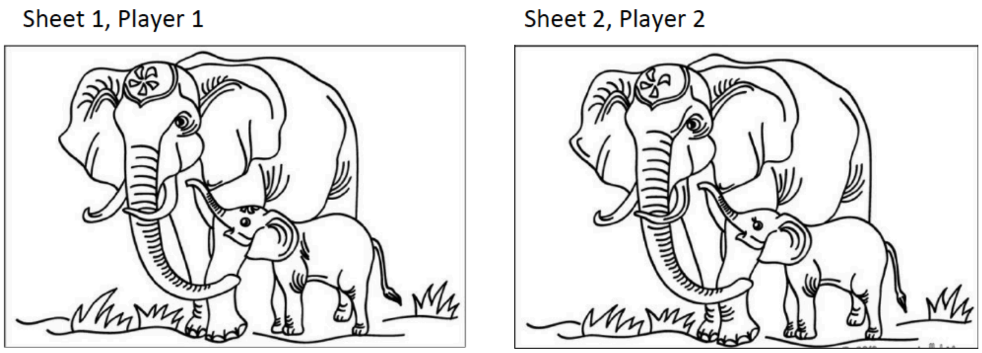
(I) Correlation = 0.81



(J) Correlation = 0.87

Figure 5. Density of Raw Production by Task, Ranked by Output Observability

Panel A: Spot the Difference - Example



Panel B: Symbol Matching - Example

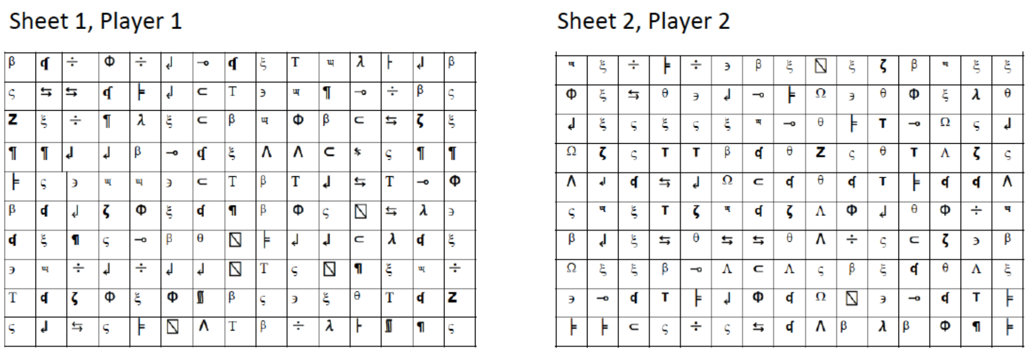


FIGURE 6. Cooperative Puzzle Games - Examples

Examples of the cooperative pair games. Each worker in a pair would receive one of the sheets. Workers had to compare their respective sheets, and circle items that were different (Spot the Difference) or matched (Symbol Matching) on both their sheets.

Table 10. Effects on Output Quality

Dependent variable	Output	Attendance	Quality rating	Qualty rating	Above mean rating	Above mean rating
Sample	Full sample	Full sample	Full sample	Condl on attendance	Full sample	Condl on attendance
	(1)	(2)	(3)	(4)	(5)	(6)
Post x Heterogeneous	-0.368**	-0.135***	-0.0888	0.0795	-0.0790	-0.0667
	(0.150)	(0.031)	(0.293)	(0.238)	(0.200)	(0.205)
Post x Heterogeneous x Med wage	0.510**	0.0576	-0.244	-0.364	-0.158	-0.178
	(0.214)	(0.044)	(0.372)	(0.297)	(0.243)	(0.249)
Post x Heterogeneous x High wage	0.373	0.0566	-0.438	-0.295	-0.236	-0.165
	(0.226)	(0.051)	(0.352)	(0.272)	(0.262)	(0.272)
Individual fixed effects?	Yes	Yes	Yes	Yes	Yes	Yes
Post-treatment Compressed Mean	-0.001	0.904	3.553	3.793	0.718	0.785
R-squared	0.519	0.283	0.340	0.531	0.368	0.437
N	3685	3685	3078	2816	3078	2816

Notes: Difference in differences regressions. Cols. 1-2 replicate the main results for the subset of rounds where quality rating data was collected. The dependent variable in Cols. 3-4 is the continuous quality rating on a scale of 1 (very poor) to 5 (very good). The dependent variable in Cols. 5-6 is an indicator for a quality rating of 4 or 5 (above the mean). Regressions include day*round fixed effects, task-specific quadratic experience trends, and controls for neighboring teams. Cols. 3-6 also include fixed effects for task*supervisor_id (the id of the supervisor who did the quality rating). Standard errors clustered by team.

Table 11. Sample Overlap Between Perceived Justification Measures

	Below mean production difference	Above mean production difference
Below mean observability	30.56	21.03
Above mean observability	24.60	23.81

Notes: Tabulations of overlap between the two sources of variation for perceived justifications: baseline productivity differences between a worker and his higher ranked peer, and the observability of co-worker output. The binary splits for each measure shown here are those used in the tables in the analysis. The table shows the percentage of observations in each cell.

Table 12. Robustness to Removal of Neighbor Controls

	Dependent variable: Output (standard dev.)		Dependent variable: Attendance	
	Observability correlation	Productivity differences	Observability correlation	Productivity differences
	(1)	(2)	(3)	(4)
Post x Heterogeneous	-0.495**	-0.312*	-0.120	-0.118**
	(0.225)	(0.175)	(0.0748)	(0.0565)
Post x Heterogeneous x Observability measure	0.891**		0.104	
	(0.407)		(0.108)	
Post x Heterogeneous x Prod difference		0.670***		0.137**
		(0.255)		(0.0607)
Post x Heterogeneous x Med wage	0.588**	0.470*	0.0874	0.0939
	(0.294)	(0.270)	(0.0998)	(0.0685)
Post x Heterogeneous x Med wage x Observability measure	-0.891*		-0.164	
	(0.533)		(0.145)	
Post x Heterogeneous x Med wage x Prod difference		-0.855*		-0.260**
		(0.463)		(0.128)
Post x Heterogeneous x High wage	-0.0163	0.0969	-0.0203	0.0469
	(0.316)	(0.218)	(0.0836)	(0.0620)
Post x Heterogeneous x High wage x Observability measure	-0.314		0.0284	
	(0.570)		(0.128)	
R-squared	0.214	0.203	0.174	0.178
Number of observations (worker-days)	7,755	7,755	7,755	7,755

Notes : Regressions include task fixed effects, day*round fixed effects. Standard errors clustered by team.

Table 13. Robustness to Changes in Worker Rankings

	Dependent variable: Output (standard dev.)			Dependent variable: Attendance		
	Full sample	Drop Fallers to Low Rank	Drop any Fallers	Full sample	Drop Fallers to Low Rank	Drop any Fallers
	(1)	(2)	(3)	(4)	(5)	(6)
Post x Heterogeneous	-0.360***	-0.357*	-0.362*	-0.118**	-0.0976	-0.0965
	(0.137)	(0.213)	(0.216)	(0.057)	(0.0722)	(0.0721)
Post x Heterogeneous x Med wage	0.385**	0.389	0.451	0.0584	0.0469	0.0566
	(0.191)	(0.247)	(0.281)	(0.075)	(0.0892)	(0.100)
Post x Heterogeneous x High wage	0.201	0.229	0.231	-0.0154	-0.0188	-0.0203
	(0.216)	(0.278)	(0.281)	(0.076)	(0.103)	(0.104)
Individual fixed effects?	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.442	0.481	0.500	0.206	0.217	0.227
N	7,691	6,487	5,921	7,691	6,487	5,921

Notes: Difference in differences regressions. Post is an indicator that equals 1 if the day is after workers have been randomized into wage treatments, and 0 during the baseline training period. Regressions include day*round fixed effects, task-specific quadratic experience trends, and controls for neighboring teams. All coefficients are identified off comparisons of workers who earn the same absolute wage and have the same productivity rank within their team (see regression specification in text). Standard errors clustered by team. "Drop Fallers to Low Rank" drops workers who were not the lowest rank before the ranking treatment, but who became the lowest rank during the ranking treatment. "Drop any Fallers" drops any worker who fell to any lower rank between the training period and the ranking treatment.

TABLE 14. Robustness: Effects on Team Cohesion - Conditional on Attendance

Dependent variable	Attendance	Attendance	Game: Tower building in teams: Tower height	Game: Tower building in teams: Tower height	Game: Cooperative games in pairs: Number correct	Game: Cooperative games in pairs: Above mean correct
	(1)	(2)	(3)	(4)	(5)	(6)
Heterogeneous team	-0.0045	-0.0157	-9.380**	-8.567*		
	(0.030)	(0.030)	(4.136)	(4.395)		
Compressed_High team		-0.0348		2.753		
		(0.042)		(6.412)		
Low rank worker		-0.0250				
		(0.031)				
Medium rank worker		-0.0375				
		(0.034)				
Both workers from same team x Heterogeneous					-0.645	-0.167**
					(0.486)	(0.063)
Both workers from same team					0.464**	0.0679*
					(0.231)	(0.038)
At least one Heterogeneous worker in pair					0.169	0.0546
					(0.299)	(0.046)
At least one low rank worker in pair					-0.696***	-0.0797**
					(0.247)	(0.037)
At least one medium rank worker in pair					-0.527**	-0.0867**
					(0.223)	(0.036)
Dependent variable mean	0.954	0.954	54.20	54.20	4.685	0.607
Observations	240	240	71	71	1706	1706
R-squared	0.0233	0.0331	0.267	0.270	0.218	0.205

Notes: This table replicates results for cooperative team building games, conditional on attendance. Cols 1-2 include the full sample of workers who participated in endline games (these games were only run in the later rounds of the experiment). Cols. 3-4 limit analysis to teams where all 3 team mambers were present on the day of endline games. Cols. 5-6 limit analysis to pair-games where both workers in an assigned pair were present the day of endline games. All standard errors are clustered by team.