THE MORALE EFFECTS OF PAY INEQUALITY

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

ABSTRACT. A long tradition in economics and psychology has advanced the notion that individuals care about not only their own pay, but also their pay relative to that of their co-workers. This has potentially broad implications for wage structure and the organization of labor markets, such as the prevalence of wage compression. We use a field experiment with full-time Indian manufacturing workers over a one-month period to test whether relative pay comparisons affect effort and labor supply. Workers perform individual production tasks, but are organized into distinct teams—defined by the type of product they produce. We randomize teams to receive either compressed wages (where all workers earn the same random daily wage) or heterogeneous wages (where each team member is paid a different wage according to his baseline productivity level). This enables effort comparisons across workers who receive the same absolute wage, but vary in the wages of their co-workers. We find that workers reduce output by 52% when their co-workers are paid more than themselves. They are also 13.5 percentage points less likely to come to work (on a base of 94% attendance) — giving up substantial earnings to avoid a workplace where they are paid less than their peers. These effects are concentrated among production tasks where it is more difficult to observe co-worker output. In addition, effort decreases are larger when the difference in baseline productivity between a worker and his higher-paid peers is small. These 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. This may help explain why piece rates—which generate earnings dispersion but are transparent—are not perceived as unfair, while flat wages are often compressed. It also has implications for the types of tasks in which wage compression may be more likely.

JEL CLASSIFICATION CODES: J3, O15, D03, E24

KEYWORDS: Reference dependence, peer comparisons, wage compression

---

Date: August 2015.

We thank James Andreoni, Dan Benjamin, Pascaline Dupas, Edward Glaeser, 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. Mohar Dey, Piyush Tank, and Deepak Saraswat at JPAL South Asia provided excellent research assistance. We gratefully acknowledge financial support from the National Science Foundation, the 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 payoffs 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 care about not only their own pay, but also their pay relative to that of their co-workers.[1] If relative pay enters into worker utility, then there are two potential sets of channels through which this could affect behavior: labor supply at a given absolute pay level (the compensating differential) and effort provision to the firm (morale).[2] When there is incomplete contracting, morale considerations may be particularly relevant for the firm's ability to elicit effort (Bewley 1999).[3]

If workers care about relative pay, this has potentially broad consequences for understanding labor markets. For example, Figure 1 documents that in a sample of Indian villages, over 70% of male workers are paid the same exact daily wage for work. This is consistent with a widely documented stylized fact about casual daily labor markets in developing countries: there tends to be one prevailing wage that applies to all workers within a labor market, regardless of worker ability (e.g., Dreze and Mukherjee 1989). More generally, wage compression—when wages vary less than the marginal product of labor—has been documented in both poor and rich countries (Frank 1984). If relative pay concerns affect worker utility, it could be profit maximizing for firms to compress wages (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). They may also affect organizational arrangements and firm boundaries—influencing how heterogeneous workers are sorted into firms, and whether labor is contracted within a firm or through the external market (Frank 1984, Nickerson and Zenger 2008).[4]

These implications rest on the presumption that workers do indeed care about relative pay. In this paper, we use a field experiment with manufacturing workers to test the validity of this view. We construct an experimental design to accomplish three goals: (i) For each worker, define a clear reference group of co-workers for pay comparisons; (ii) Be able to compare outcomes for workers whose absolute pay levels are the same, but who vary in

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

[2]For example, psychologist Stacey Adams (1963) proposed that when individuals perceive inequity, this induces a change in behavior that seeks to reduce this tension. Retaliatory effort reductions by workers when they feel unfairly treated could be interpreted in this light. Frank (1984) builds a model where relative pay status is a compensating differential.

[3]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.

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

their co-workers' (reference group) wages; (iii) Examine whether the justification for pay differences matters for fairness perceptions.

In the experiment, 273 workers are employed full-time for one month in seasonal manufacturing jobs in India.[5] 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.[6] Note that there is no joint production within teams—production is an individual activity, enabling us to measure each worker's individual output.[7] 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 daily wage according to his productivity rank within the team (as determined by baseline productivity levels)—so that more productive workers are paid relatively more than less productive ones. These relative pay differences are fairly modest: the discrepancy in wages between the highest and lowest paid worker in the *Heterogeneous* pay teams is 10% or less. In the *Compressed* pay condition, all team members are paid the same exact daily wage, where this wage level is randomly chosen. Each team is assigned to either the *Heterogeneous* or *Compressed* wage structure, and we then monitor output and other outcomes for the remainder of the one-month employment period.[8] This design enables comparisons of workers of similar productivity who are paid the same absolute wage, but differ in the distribution of their co-workers' (reference group) wages.[9]

In order to further understand the nature of reference dependence, we also randomly vary whether pay disparity seems justified in two ways. First, we induce variation in *perceived fairness*: whether a team engages in a production task for which it is easy to observe co-worker output.[10] Second, we alter *actual fairness*: the extent to which the difference in team members' wages overstates the difference in their relative productivity levels.

[5]The experiment is still running, and is in its final stages. The results reported in this draft are for the first set of rounds, comprised of 273 workers. The final sample will be comprised of 450 workers.

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

[7]In order to accomplish this, we pay to hire extra staff to measure each worker's output at the end of each day.

[8]This feature of the design shuts down dynamic incentive effects: one's performance today does not affect future wages. 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. Of course, in deciding on optimal pay structure, a firm would weigh various potential considerations, such as potential dynamic incentives.

[9]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 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.

[10]To quantify the observability of each task, we measure whether workers can accurately rank their output relative to that of their teammates, limiting analysis to teams where all team members are paid the same

On average, for a given absolute pay level, a worker's output declines by 0.557 standard deviations (approximately 52%) when both his co-workers are paid more than himself. Such workers are also 13.5 percentage points (14%) less likely to come to work any given day. Because workers are paid a daily wage based on attendance, this amounts to a substantial decline in earnings from the factory and implies that employees are willing to give up substantial earnings to avoid a workplace where they are paid less than their peers. We estimate that the attendance decrease accounts for 48% of the overall decline in worker output. These negative effects are persistent over the duration of the employment period, with some suggestive evidence that they actually become stronger in later weeks.

In contrast, we find little evidence that being paid more than one's co-workers affects output or attendance. Compared to their counterparts on the *Compressed* teams, the workers with the highest wage in each of the *Heterogeneous* teams have similar outcomes on average: the point estimates are small in magnitude and statistically insignificant. This is consistent with asymmetry around the reference point.[11] In addition, we also find no discernible impact on the behavior of *Heterogeneous* condition workers who receive the median wage on their team—with one person paid more and one person paid less than themselves.

We find an important role for the perceived fairness of pay differences. For the lowest paid workers, the negative effects on morale are concentrated in production tasks where it is difficult to observe co-worker output. For such tasks, the output decrease for the lowest paid workers is 0.658 standard deviations. In contrast, when co-worker output is highly observable, we cannot reject that there is no effect of being paid less than one's peers.

The results of our actual fairness manipulation paint a similar picture. In teams where the baseline productivity levels of workers are far apart—so that differences in productivity swamp differences in wages—we find no evidence for effort reductions among the lowest-paid workers. This is consistent with the task observability results: when it is very clear that one's co-workers are contributing higher output to the firm, there is no fairness violation. However, as long as a worker has some wiggle room to doubt the justification for the pay difference, output drops sharply.[12]

Again, in contrast, we find no role for perceived justifications—task observability or whether productivity differences swamp pay differences—in affecting the behavior of median

wage. We stratify the wage treatments by production task, ensuring variation in task observability within each wage treatment cell.

[11]Standard models of reference dependence predict asymmetry in the value function around the reference point: any utility gains from being above the reference point will be smaller in magnitude than the utility losses from being below the reference point (Kahneman and Tversky 1979).

[12]This pattern of results is broadly consistent with the predictions in Fang and Moscarini (2005). In their model, over-confidence makes workers inclined to over-estimate their relative productivity. Under this presumption, more clearly observing co-worker output or a bigger output difference would make it less likely that a worker would believe he is as productive as his more productive peers, mitigating perceptions of unfairness in relative pay.

or high earners on the *Heterogeneous* teams relative to their counterparts on the *Compressed* teams. Similarly, within *Compressed* teams (where all team members are paid the same wage), we find no evidence that workers behave any differently if their co-workers are substantially more or less productive than they are. These findings suggest that in our particular setting, workers do not necessarily feel entitled to higher pay commensurate with higher productivity. In other words, workers do not simply compare their own ratio of $^{pay}/_{productivity}$ relative to that of referent others. Rather, they appear to compare their pay in levels. When differences in pay levels trigger a potential fairness violation, this is mitigated if pay differences are clearly justified by relative productivity.

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 training 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. Of course, 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 incentives). Our findings suggest that perceived justifications and transparency can affect this calculus. In this paper, we seek to examine whether relative pay considerations matter in our setting; we do not make claims about optimal pay structure in any given setting.

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 workers 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 pay cuts matter more than absolute pay cuts for effort over a six-hour period.[13] In addition, our work relates to the broader literature on the effect of fairness preferences on effort provision under incomplete contracting, particularly gift-exchange (Akerlof 1982; Gneezy and List 2006).[14]

[13]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). Related laboratory experiments have examined the effects of rank (Brown et al. 2008, Clark et al. 2010, Kuziemko et al. 2011) and of wage justification (Bracha et al. (2015)). 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).

[14]These studies test for reference points that are determined by a worker's own past wage (or expected wage). A large number of studies find gift exchange in the lab (Fehr et al. 1993). Field evidence, however, is more limited; a small number of field experiments suggest that absolute wage increases don't create lasting

Our study advances the literature on relative pay comparisons—and gift exchange more broadly—in several ways. We document effects of relative pay comparisons on both effort (the output effect) and how much compensation workers give up to avoid unequal pay (the compensating differential)—the two distinct channels through which relative pay comparisons could affect wage structure and labor market outcomes. In addition, while previous work has focused on arbitrary wage changes, in our study wage differences reflect interpersonal differences in worker productivities; this better reflects why wage differences may arise in the labor market and is important given laboratory evidence that justifications matter for fairness violations (Falk et al. 2008). Relatedly, our study provides the first field evidence on whether justifications matter for morale effects. This has bearing on understanding why wage compression may arise in some settings and not in others—for example, based on the observability or quantifiability of output. It may also help explain why piece rates are generally not perceived as unfair. Finally, workers make decisions for a job from which they derive full-time earnings over one month. This helps ensure 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 at 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 strategies has the potential to affect wage compression, wage rigidity, 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 2 describes our empirical setting and experimental design, Section 3.1 presents our results, Section 4 discusses threats to validity, and Section 5 concludes.

effort increases, while absolute wage cuts cause effort reductions over a one-day period (Gneezy and List 2006; Kube et al. 2013). For excellent reviews of this literature, see Fehr et al. (2009), List (2009), and Charness and Kuhn (2011).

2. Experimental Design and Data

2.1. **Conceptual Framework.** We conceptualize relative pay concerns as reference- dependence in utility, where co-worker pay enters as an argument into the reference point. Let w_i denote worker i's own wage. In addition, denote the worker's reference point as w_R. We allow $w_R = f(\mathbf{w_{-i}}, \theta_\mathbf{i}, \theta_\mathbf{-i}, \mathbf{x})$, where $\mathbf{w_{-i}}$ is a vector of co-workers' wages, $\theta_\mathbf{i}$ is own productivity, $\theta_\mathbf{-i}$ is a vector of co-workers' productivity levels, and $\mathbf{x}$ contains other potential determinants of the reference point.[15] We do not take a strong ex-ante stance on the functional form of this reference point function, other than to assume that $\mathbf{w_{-i}}$ does indeed enter as an argument. As we describe below, our primary tests will be valid under most reasonable functional forms where this assumption is true. We will supplement these with secondary tests to enrich our understanding of the underlying reference point function.

In constructing tests, we begin with two core predictions of loss aversion models (Tversky and Kahneman 1991). First, for any given own absolute wage level w_i, there is a strict decrease in utility when $w_i < w_R$. Second, marginal utility is asymmetric around the reference point: for a given deviation ε from w_R, the utility losses from $w_i = w_R - \varepsilon$ will be larger in magnitude than any utility gains from $w_i = w_R + \varepsilon$.

The effects on utility from variation in w_R will potentially affect labor supply on the extensive margin through a worker's willingness to accept work at a given w_i. In addition, we follow the previous literature and assume that, in a setting where there is incomplete contracting on effort (providing workers with some latitude to choose effort without direct earnings consequences), workers will adjust their effort to offset utility effects of reference dependence (Charness and Kuhn 2007). This implies that workers will retaliate by decreasing effort when $w_i < w_R$ and that effort effects will be asymmetric around the reference point. We test these predictions by varying the composition of $\mathbf{w_{-i}}$, holding fixed w_i.

If the perceived justification for pay decisions matters for fairness violations (e.g., Falk et al. 2008), then the reference point may also depend on how pay differences relate to productivity differences. If so, utility effects will be more likely when the difference in own and co-workers' wages overstates the difference in own and co-workers' productivity. We explore this supplementary prediction by cross-cutting the variation in w_i and $\mathbf{w_{-i}}$ with variation in θ_i relative to $\theta_\mathbf{-i}$.

2.2. **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).

[15] For example, if there is also reference dependence in one's past wage history, as in some models of gift exchange, then $\mathbf{x}$ may contain own wages in previous periods.

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 w_H and w_L is less than 10%. 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.[16] 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 2.

[16] To 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. In these trials, all teammates were paid the same wage, so that wage was not a signal of productivity rank.

2.3. **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*. In addition, we predict that any such increases will be smaller in magnitude than effort decreases under Test 1.

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 in Tests 1 and 2 will be smaller when differences between a co-worker and his higher paid peers is large, and when these differences are observable.[17]

2.4. **Time Line, Recruitment and Survey Instruments.** We detail the implementation of our experiment in Figure 3. Before each round begins, we select three to four villages within 8km of the worksite from which to recruit. We then distribute information about the jobs to all labor households and hold a village meeting to allow potential employees a chance to learn more about the work. We then offer all interested males between the ages of 18 and 55 a chance to sign up for the work. From this list, we randomize who receives a job and ensure that we hire at most one individual from any given household. Given that we require exactly 30 workers in each worksite round, we also construct a wait list of randomly-selected names in case any of the initially-selected workers decide not to take the job. While we do run multiple rounds per physical worksite, we never recruit from the same village more than once.

Once the recruitment is over, we begin the 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.[18] We consider the training period to be over on day 10, at which time we inform workers about their relative production rankings within their teams. Individuals

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

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

are told their rankings in private, and we do not disclose the exact rankings of the other members of their teams. We inform individuals of their rankings so as not to confound our subsequent wage treatments with information revelation by the factory. Finally, 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, the workers are surveyed and participate in laboratory activities, described below.

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. In addition, we elicit information on their consumption and credit behavior during the time of employment. We also use a survey instrument to map out their social networks with other workers at the worksite.[19]

At the end of the contract period, we conduct a day of laboratory games with the workers. In these sessions, we ask workers to perform two types of activities. First, we randomly assign workers to teams of two and ask them to perform cooperative tasks. We aim to test whether pairs of workers who were assigned to the same *Heterogeneous* wage teams perform worse than individuals in *Compressed* wage teams or even pairs of workers who were on different teams for the duration of the contract period. Second, we ask workers to perform cooperative tasks in their production teams. Again, we hypothesize that the *Compressed* teams will perform better than the *Heterogeneous* teams even under identical team incentives.

In Panel A of Table 2, we briefly describe the workers employed at our factories.[20] 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

[19]Many of these survey instruments have not yet been entered, as the project is still in the field at this time.
[20]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.

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. While 72% of workers report ever receiving wages that differ from the village prevailing wage (which is likely due to non-agricultural work such as stone cutting or construction), only 17% of workers report ever receiving a wage different from that of other laborers in the village for the same task.

2.5. **Regression Specifications.**

Worker Level Regressions. 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". We begin by presenting a simplified regression specification in Equation 2.1 where we only include observations from so-called "relevant" individuals. We then present an augmented regression in Equation 2.2 where we identify the main treatment effects off of the relevant group, but use the other worker treatment types in the regression to estimate the fixed effects and other controls.

The differences-in-differences regression equation restricting the sample to only the relevant worker treatment types can be written:

$$
\begin{aligned}
(2.1)\quad y_{i,j,t} \;=\;& \beta_{Het}\tau_{G,j}1_{(t\geq 0)} + \beta_{M,Het}\tau_{G,j}1_{r(i)=M}1_{(t\geq 0)} + \beta_{H,Het}\tau_{G,j}1_{(r(i)=H)}1_{(t\geq 0)} \\
&+\delta_{Het}\tau_{G,j} + \delta_{M,Het}\tau_{G,j}1_{r(i)=M} + \delta_{H,Het}\tau_{G,j}1_{(r(i)=H)} \\
&+\eta_{k(j),t} + \eta_{w(j),t} + \eta_{r(i),t} + \varepsilon_{i,j,t}
\end{aligned}
$$

Each observation captures information for worker i on team j on day t of the experimental round. We rescale t such that $t = 0$ corresponds to the first day of the wage treatment in each of the worksite-rounds. We focus on two outcome variables $y_{i,j,t}$, attendance and production.[21] $attendance_{i,j,t}$ is a binary variable capturing whether worker i is present on day t. The output variable, $production_{i,j,t}$ 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.

While treatments are randomized at time $t = 0$, we estimate differences-in-differences regressions to economize on power. All key terms in the regression are interacted with the post-treatment indicator $1_{(t\geq 0)}$.

[21]When we have all of the survey data entered, we will also consider total worker-level earnings as a key outcome variable.

The four variables $\tau_{l,j}$ indicate the treatment status of team j, where $l \in \{G, CL, CM, CH\}$. G denotes the *Heterogeneous* wage treatment, and CL, CM, and CH denote *Compressed Low*, *Medium*, and *High*, respectively. It is also useful to define the vector of indicators $(k(j), w(j), r(i))$. Here, $k(j)$ indexes the task k produced by team j; $w(j)$ indexes the worksite-round w in which team j works; $r(i) \in (H, M, L)$ denotes the pre-period ranking of worker i.

Our main prediction is that $\beta_{Het} < 0$, that is Low rank workers in *Heterogeneous* teams will decrease production and attendance relative to the Low rank workers in *Compressed_L* teams. We also predict $\beta_{H,Het} > 0$ and $\beta_{Het} + \beta_{H,Het} \geq 0$, that the High-ranked workers in the Heterogeneous teams will produce weakly more than their counterparts in the *Compressed_H* teams. Recall that we do not have a strong prediction on $\beta_{M,Het}$.

We also include fixed effects for task $\times$ experience $\left(\eta_{k(j),t}\right)$, worksite-round $\times$ time $\left(\eta_{w(j),t}\right)$, and worker ranking $\times$ pre/post $\left(\eta_{r(i),t}\right)$ in our main regression specification.

Rather than discard the "irrelevant" observations, we instead specify the augmented regression:

$$
\begin{aligned}
(2.2)\quad y_{i,j,t} \quad = \quad & \beta_{Het}\tau_{G,j}1_{(t\geq 0)} + \beta_{M,Het}\tau_{G,j}1_{r(i)=M}1_{(t\geq 0)} + \beta_{H,Het}\tau_{G,j}1_{(r(i)=H)}1_{(t\geq 0)} \\
& + \delta_{Het}\tau_{G,j} + \delta_{M,Het}\tau_{G,j}1_{r(i)=M} + \delta_{H,Het}\tau_{G,j}1_{(r(i)=H)} \\
& + \theta^{post}_{Rel} relevant_{i,j}1_{(t\geq 0)} + \theta^{post}_{Rel,M}\left(\tau_{G,j} + \tau_{CM,j}\right)1_{(r(i)=M)}1_{(t\geq 0)} \\
& + \theta^{post}_{Rel,M}\left(\tau_{G,j} + \tau_{CH,j}\right)1_{(r(i)=H)}1_{(t\geq 0)} + \theta^{pre}_{Rel} relevant_{i,j} \\
& + \theta^{pre}_{Rel,M}\left(\tau_{G,j} + \tau_{CM,j}\right)1_{(r(i)=M)} + \theta^{pre}_{Rel,M}\left(\tau_{G,j} + \tau_{CH,j}\right)1_{(r(i)=H)} \\
& + \eta_{k(j),t} + \eta_{w(j),t} + \eta_{r(i),t} + \varepsilon_{i,j,t}
\end{aligned}
$$

The key regressors and fixed effects are identical to those in Equation 2.1, however, we also add a set of indicators (i.e., the θ terms) for the relevant comparison groups both pre- and post- wage change. We define $relevant_{i,j}$ to include all members of the relevant comparison set:

$$relevant_{i,j} = 1\left(\tau_G + \tau_{CL} * 1_{r(i)=L} + \tau_{CM} * 1_{r(i)=M} + \tau_{CH} * 1_{r(i)=H} = 1\right)$$

Note that the $\left(\tau_{G,j} + \tau_{CM,j}\right)1_{(r(i)=M)}$ terms denote the Medium rank workers in *Heterogeneous* or *Compressed_M* teams, and the $\left(\tau_{G,j} + \tau_{CH,j}\right)1_{(r(i)=H)}$ terms denote the High rank workers in the *Heterogeneous* or *Compressed_H* teams. We present the results of estimating Equation 2.2 in Section 3.1.

Team Level Regressions. Recall from Section 2.3 that if the negative response to being paid less than one's peers is larger in magnitude than the positive response to being paid more than one's peers, then we should expect for *Heterogeneous* teams to perform worse than the average over all *Compressed* teams, and perhaps even worse than the *Compressed_M* or *Compressed _L* teams alone. Recall that the total wages paid to the *Heterogeneous* teams are the same as the average wage across all three *Compressed* teams. We estimate the following team-level regression to explore whether the *Heterogeneous* teams under-perform the average production across all of the *Compressed* teams.

$$y_{j,t} = \beta_{Het}\tau_{G,j} * 1_{(t\geq 0)} + \eta_j + \eta_{k(j),t} + \eta_{w(j),t} + \varepsilon_{j,t} \quad (2.3)$$

Again, we are interested in outcomes $y_{i,j,t}$ of attendance and production. Here, team attendance is measured as the average attendance of team j on day t, $\left(attendance_{j,t} \in \left\{0, \frac{1}{3}, \frac{2}{3}, 1\right\}\right)$. Team-level production, $production_{j,t}$ is simply the total production of team j on day t, measured in the same standardized units as in Equation 2.1.[22] Again, η_j captures team-level fixed effects, and $\eta_{k(j),t}$ and $\eta_{w(j),t}$ capture task x experience and worksite-round x time fixed effects.

3. Results

In what follows, we present results from ten worksite rounds, employing a total of 273 workers. Because the experiment is still in the field, data collection is ongoing.

3.1. **Main Results.** Figure 4 shows graphically the main differences-in-differences results of pay heterogeneity on output among the relevant comparison groups, and Table 3 presents the estimation of Equation 2.2 on the full sample of workers in each round. Columns (1) and (2) measure the effects on standardized production, while Columns (3) and (4) measure effects on attendance.

The graphical and regression results support our main prediction, that Low rank workers decrease production and attendance when they are paid less than their team-mates, holding their absolute wage levels fixed ($\beta_{Het} < 0$). The output of Low rank workers declines by 0.557 standard deviations in response to the *Heterogeneous* treatment (approximately 52% of average pre-period production). Further, the Low rank worker is 13.5 percentage points less likely to come to work after the wage treatment (on a base of 94% attendance in the *Compressed* groups). Given that workers do not receive pay for days on which they are absent, this large decrease in attendance likely implies that workers are decreasing substantially their total earnings in response to receiving lower wages than their co-workers.[23]

Table 3 also presents evidence regarding how the Medium rank workers respond when they are paid the median team wage and how the High-rank workers respond when they earn the highest wage on the team. The results are consistent with no effect of *Heterogeneous* pay on production of the Medium and High types. In the specification in Column (1), the heterogeneous impacts on the Medium rank workers (i.e., $\beta_{M,Het}$) is large, positive and significant at the 10% level. The total treatment effects ($\beta_{Het} + \beta_{M,Het}$ and $\beta_{Het} + \beta_{H,Het}$)

[22]When we have data from more experimental rounds, we also plan to estimate: $y_{j,t} = \beta_{Het} Heterogenous_j * Post_t + \beta_{CM} Compressed\,Med_j * Post_t + \beta_{CM} Compressed\,High_j * Post_t + \eta_j + \eta_{k(j),t} + \eta_{w(j,t)} + \varepsilon_{j,t}$. Note that the Compressed Low team is the omitted category. We predict that $\beta_{Het} < \beta_{CM}$. A strong version of our prediction is that $\beta_{Het} < 0$.

[23]When our full survey data has been entered, we will be able to test this formally.

on the Medium and High rank workers are much smaller in magnitude than for the Low rank workers and are indistinguishable from zero. The differential impact in Column (1) of the *Heterogeneous* wage treatment on the productivity of Medium rank workers relative to Low rank workers is 0.441 standard deviations, for example. This large, positive differential effect largely offsets the overall decrease of 0.557 standard deviations experienced by the Low rank workers. Further, we find no evidence here that the High rank workers increase their production in response to receiving higher wages than their peers. 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.

One natural question is whether the large negative impacts on the Low rank workers persist or instead wear off over time. In Appendix Table 8, we separately estimate Equation 2.2 over the first and second halves of the post-treatment period.[24] We find large, negative, and statistically significant effects of the *Heterogenous* treatment on the Low rank workers in both samples. The reduction in production is 0.359 standard deviations in the early period and .0.737 standard deviations in the late period. Similarly, attendance falls by 0.8 percentage points in the early period and by 24.2 percentage points in the late period. If anything, we see that the results 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). If the effect on attendance is large, this poses problems for identifying the intensive margin effect on effort. If dis-advantageous peer wage comparisons affect some types of workers more than others, then running the regression in Equation 2.2 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.

First, we offer a back-of-the-envelope calculation to decompose these effects. In this calculation, we assume that output conditional on attendance in absence of the treatment is the same for *Heterogeneous* and *Compressed_L* workers. We calculate that the mean output conditional on attendance for Low rank workers in the *Compressed_L* team is 1.99. The effect of *Heterogeneous* wages for the Low rank workers on attendance is -0.135 percentage points. If the full treatment effect on production were coming through attendance, then we would predict an output decrease of -0.135*1.99 = -0.269 standard deviations. This corresponds to approximately 48% of the total effect on production.

Second, once we have completed the experiment with our full sample, we plan to estimate a two-step selection model as in Heckman (1979) to decompose the effort and attendance effects. In our current sample, we observe that attendance is 6pp higher on paydays than

[24]The full pre-treatment period is included in both regressions.

on non-paydays. Thus, we propose to use paydays interacted with treatment status as an excluded predictor of attendance in the two-step procedure.[25]

3.2. **Perceived and Actual Fairness.** Our design allows us to explore the circumstances under which workers react most strongly when faced with disadvantageous pay comparisons. We focus on unpacking two different dimensions of our design that map to the fairness, both perceived and actual, of the *Heterogeneous* pay treatment.

We begin with *perceived* fairness, and note that some of our tasks, by nature, are much more observable than others. This variation in observability, coupled with our randomized design allows us to ask if the effects of *Heterogeneous* pay are mitigated when output is more observable, and therefore, when team members can better justify the reason for the underlying pay differences.

We first need to define what constitutes an observable task. To do this, we use data from four pilot rounds (not included in the regression analysis here). On the last day of these pilot rounds, we asked workers to rank their co-workers by productivity. Given that in the pilot rounds, we never shared productivity rankings with the workers, we can measure how well the survey responses correspond to the actual productivity rankings. In Figure 5, we present the correlations between the actual and survey rankings of the workers on each of the eight pilot production tasks.[26] The correlations range from negative and close to zero to 0.87. For the following analysis, we define a task to be observable if the correlation is above 0.7. This cutoff also splits the tasks in two equally-sized groups.[27]

In Table 4 we estimate Equation 2.2 separately for the set of observable (Columns (1) and (3)) and unobservable tasks (Columns (2) and (4)). The treatment effects for the Low rank workers producing observable task are significant and very large – production falls by 0.66 standard deviations, and attendance falls by 11.5 percentage points. In contrast, the changes in production and attendance are much smaller in magnitude and are statistically insignificant for the non-observable tasks. Consistent with the results in Table 3, we again do not observe a significant treatment effect of *Heterogeneous* pay on the Medium or High rank workers in either the observable or unobservable tasks.

Next, we explore whether the treatment effects vary by the *actual* fairness of the *Heterogeneous* pay. Recall from our discussion of Figure 2 that randomizing workers to teams induces differences in the relative productivity differentials between workers within each team. If workers are less aggrieved when pay differences are better-justified, then we should observe smaller treatment effects when productivity differentials are large. In order to test

[25]Early estimates suggest effects of similar magnitudes to the back-of-the-envelope decomposition.

[26]At the end of the pilots, we added two additional tasks. These tasks are currently excluded in our analysis, though we plan to run separate rounds in the future to be able to measure the correlation between actual productivity and team member surveys.

[27]We also note that our results are unchanged if we shift the observability cutoff to either 0.6 or 0.8. Results available upon request.

this hypothesis, we use the pre-treatment production data to construct the size of the productivity differentials between each team member and the co-worker with the next-highest rank.[28] We then code the productivity differential as high if it is greater than 0.25 standard deviations.[29]

To test the prediction that negative reactions can be mitigated when pay differences are more fair (i.e, when then stem from larger productivity differences), we estimate Equation 2.2 separately for individuals experiencing high and low productivity differences with their next-ranked coworkers. Table 5 presents the results. Columns (1) and (3) present the results on production and attendance for low-difference workers, while Columns (2) and (4) present the results for high-difference workers. We find that the negative treatment effects are much less severe in magnitude when productivity differences are high. When productivity differences are not high, production decreases by 0.920 standard deviations and attendance drops by 31.0 percentage points in response to *Heterogeneous* pay. Yet again, we detect no differential effects on Medium or High rank workers. We can also show that workers in *Compressed* teams do not behave differently when they produce more than their peers.[30]

Taken together, the results suggest that the reactions of workers to pay inequality do depend on the underlying circumstances. When it is very clear that one's co-workers are contributing higher output to the firm, there is no fairness violation. However, as long as a worker has some wiggle room to doubt the justification for the pay difference, output drops sharply.

3.3. **Team Level Results.** All of our results thus far have focused on the so-called "relevant" comparisons between individuals who have the same rank and are paid the same wage, but whose peers have different wages. However, one might also be interested in consequences of *Heterogeneous* pay on team-level output. In Appendix Table 7, we present the results from estimation Equation 2.3.

Columns (1) and (3) compare the *Heterogeneous* teams to the pooled set of all *Compressed* teams, while Columns (2) and (4) allows for pairwise comparisons among all four team treatments. While the treatment-wise comparisons are underpowered, overall the *Heterogeneous* teams have production that is 0.703 standard deviations lower and report attendance rates that are 7.54% lower than the *Compressed* teams. The magnitudes of the comparisons with the *Compressed_L* teams are similar. These results echo our findings of asymmetric effects: that disadvantageous pay comparisons reduce output more than commensurate advantageous pay comparisons increase output.

[28]Thus, for Low and Medium-ranked individuals, we code the difference with the Medium and High-ranked worker on the same team respectively.

[29]This cutoff corresponds to the 50th percentile of productivity differences. We pick the median as our threshold, because we are interested in the case where productivity differences are most justified.

[30]Available upon request.

3.4. **Worker Perceptions.** Finally, we can use survey evidence from our study participants to explore how workers perceived their wages. Table 6 presents survey responses to questions about wage satisfaction, wage fairness and life satisfaction. We compare differences in responses between individuals in *Heterogeneous* vs. *Compressed* teams. Panel A presents these comparisons for Low and Medium rank workers, while Panel B presents these comparisons for High rank workers. We find that High rank workers in the *Heterogeneous* teams are much more likely to think that the wages are fair and to be satisfied with the wages. This, together with the fact that High rank workers in *Heterogeneous* teams are not more likely to increase their production and attendance compared with *Compressed* teams, suggests that High rank workers may simply rationalize that they deserve the higher wages and therefore do not feel a need to work harder in response.

4. threats to Validity and Discussion

Internal validity concerns. Could an explanation other than relative pay comparisons explain 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.

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.

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

5. 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 heterogonous 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, 422.

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

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

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

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

Bracha, A., U. Gneezy, and G. Loewenstein (2015): "Relative Pay and Labor Supply," *Journal of Labor Economics*, 33, 297–315.

Brown, G. D., J. Gardner, A. J. Oswald, and J. Qian (2008): "Does Wage Rank Affect Employees? Well-being?" *Industrial Relations: A Journal of Economy and Society*, 47, 355–389.

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

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

——— (2011): "Lab labor: What can labor economists learn from the lab?" *Handbook of Labor Economics*, 4, 229–330.

Clark, A., D. Masclet, and M.-C. Villeval (2010): "Effort and Comparison Income: Experimental and Survey Evidence," *Industrial and Labor Relations Review*, 63.

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

Cohn, A., E. Fehr, B. Herrmann, and F. Schneider (2012): "Social Comparison in the Workplace: Evidence from a Field Experiment," *IZA Discussion Paper No. 5550.*

Dreze, J. and A. Mukherjee (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., E. Fehr, and U. Fischbacher (2008): "Testing theories of fairness?Intentions matter," *Games and Economic Behavior*, 62, 287–303.

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

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

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

Frank, R. (1984): "Are Workers Paid Their Marginal Products?" *American Economic Review*, 74, 549–571.

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

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

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

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, 420–429.

Heckman, J. J. (1979): "Sample selection bias as a specification error," *Econometrica: Journal of the econometric society*, 153–161.

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

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

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

Kuziemko, I., R. Buell, T. Reich, and M. Norton (2011): "Last-place Aversion: Evidence and Redistributive Implications", NBER Working Paper No. 17234," *NBER Working Paper*.

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

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

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

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

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

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

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

Rege, M. and I. Solli (2013): "Lagging Behind the Joneses: The Impact of Relative Earnings on Job Separation," .

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

Figures

Figure 1. Village Agricultural Wages

Data collected by the authors from surveys of 184 laborers in 20 villages in Orissa, India. The figure shows the distribution of the daily wage for casual agricultural work minus the village mode.

Figure 2. Randomization Design

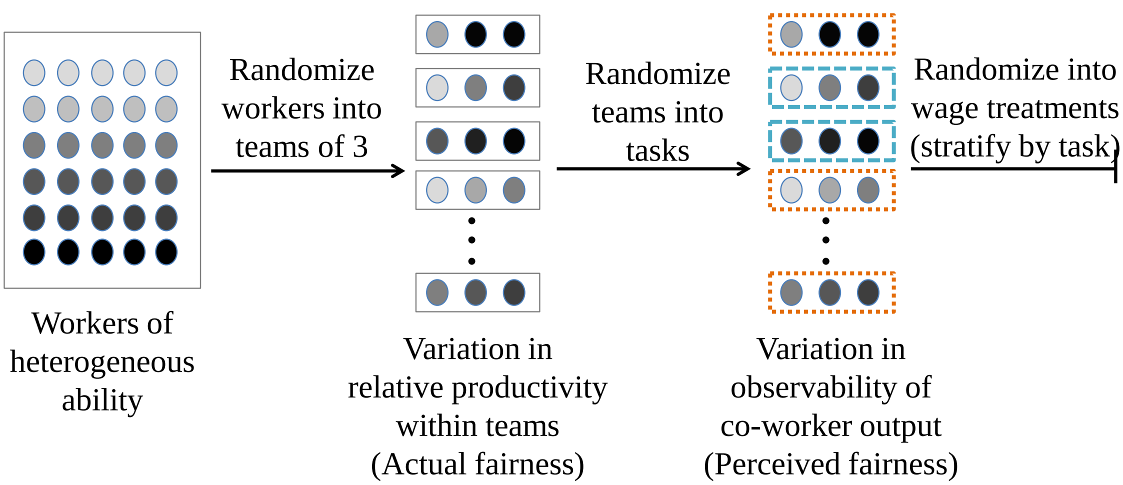


Figure 3. Time Line

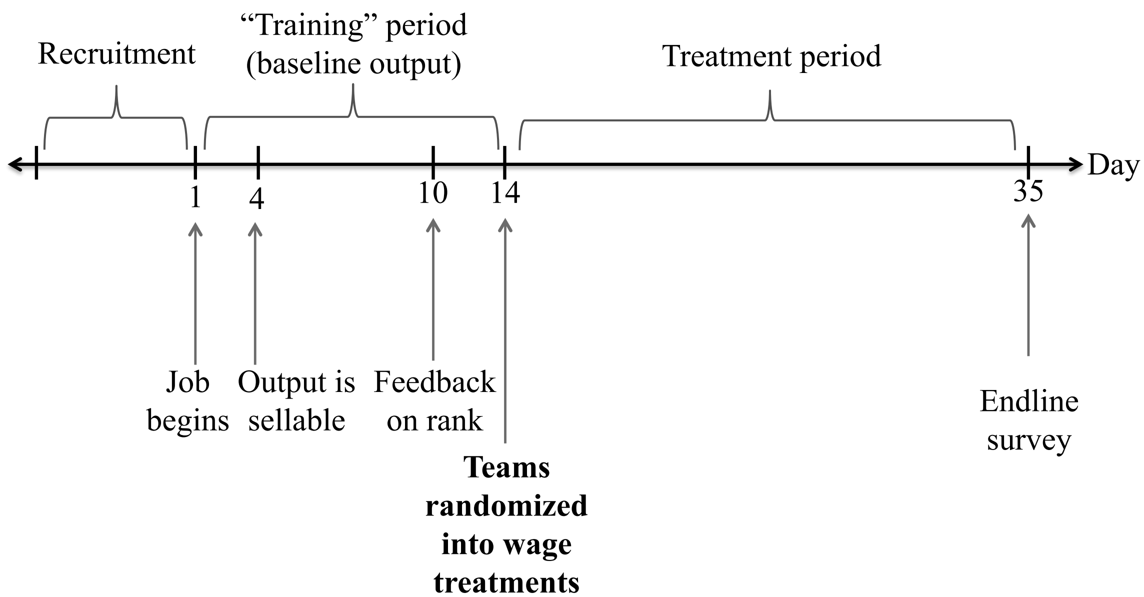
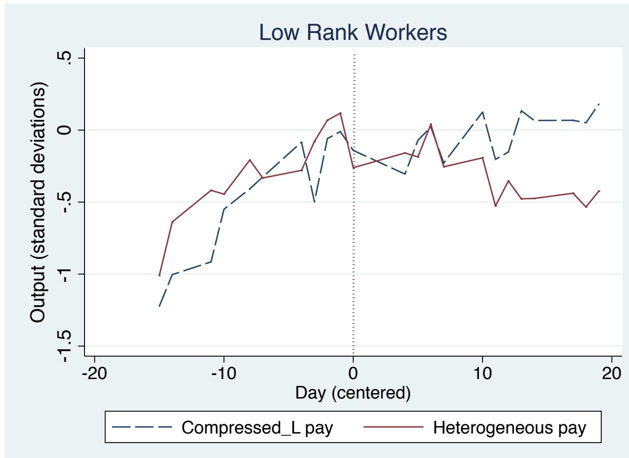
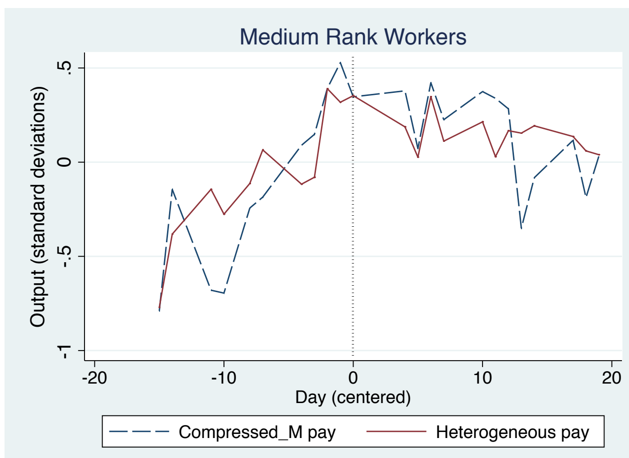


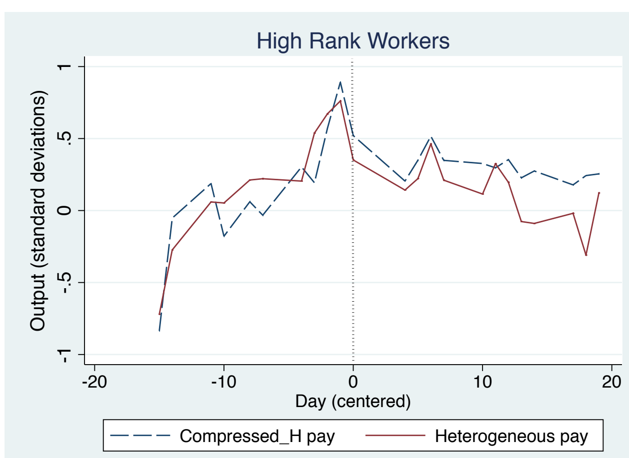
Figure 4. Effects of Heterogeneous Pay on Worker Output



(A) Low-Rank Workers



(B) Medium-Rank Workers



(C) High-Rank Workers

Figure plots production (standardized by task) in *Heterogeneous* vs. *Compressed* teams. Panel A compares Low-rank workers in the *Heterogeneous* vs. *Compressed_L* teams. Panel B compares Medium-rank workers in the *Heterogeneous* vs. *Compressed_M* teams. Panel C compares High-rank workers in the *Heterogeneous* vs. *Compressed_H* teams. Standardized production is demeaned by pre-period productivity by group. Day = 0 is the day when workers learned of their treatment assignments and when wages were changed.

Figure 5. Task Observability: Actual vs. Survey Correlations

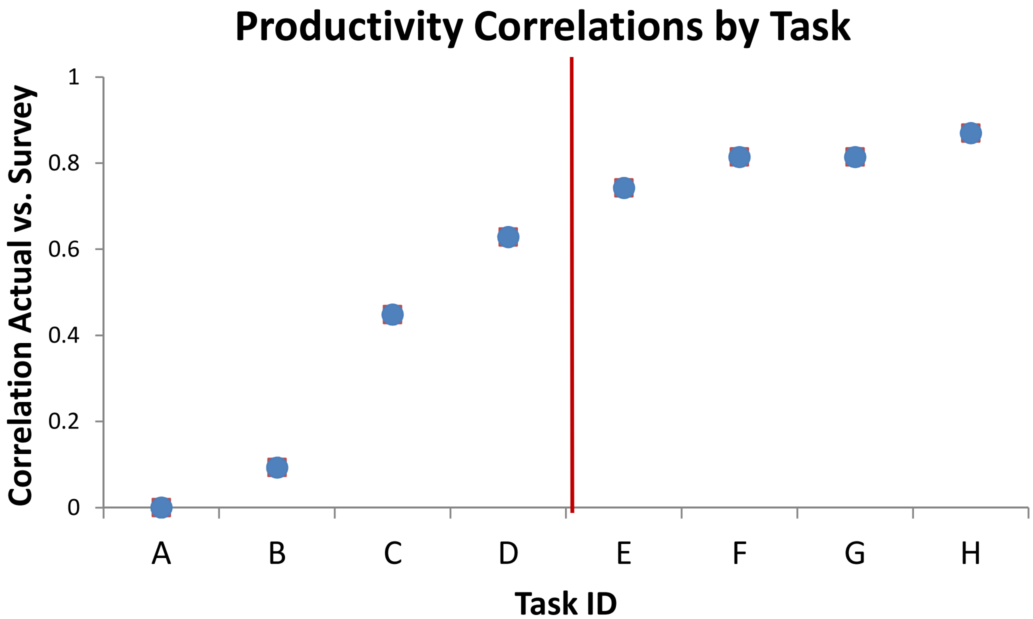


Figure plots the correlation between actual productivity rankings and perceived rankings by the workers (reported in endline surveys) for eight of the production tasks. Note that this data come from four pilot rounds where the research team did not inform workers of their production rankings. In our analysis, we split the tasks at the median level of observability (0.7 correlation).

Tables

Table 1. Treatments and Relevant Comparisons

Worker Type	Heterogeneous	Compressed_L	Compressed_M	Compressed_H
Low productivity	$\boxed{W_{Low}}$	$\boxed{W_{Low}}$	W_{Medium}	W_{High}
Medium productivity	$\boxed{W_{Medium}}$	W_{Low}	$\boxed{W_{Medium}}$	W_{High}
High productivity	$\boxed{W_{High}}$	W_{Low}	W_{Medium}	$\boxed{W_{High}}$

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. Standardized Production and Attendance

	(1) Standardized Production	(2) Standardized Production	(3) Attendance	(4) Attendance
Heterogeneous Team * Post Wage Change	-0.557***	-0.455***	-0.135	-0.141*
	(0.178)	(0.166)	(0.082)	(0.081)
Heterogeneous Team: M Rank * Post Wage Change	0.441*	0.370	0.130	0.124
	(0.235)	(0.228)	(0.096)	(0.097)
Heterogeneous Team: H Rank * Post Wage Change	0.257	0.189	-0.0440	-0.0425
	(0.309)	(0.301)	(0.113)	(0.112)
Task x Experience FE	Yes	Yes	Yes	Yes
Day x Round FE	Yes	Yes	Yes	Yes
Team FE	No	Yes	No	Yes
Post-treatment Control Mean	0.142	0.142	0.938	0.938
Post-treatment Control Std. Dev.	0.903	0.903	0.241	0.241
Post*Het+Post*Het*Med=0	0.521	0.617	0.925	0.755
Post*Het+Post*Het*High=0	0.228	0.262	0.0146	0.00923
R-squared	0.410	0.400	0.177	0.162
N	4343	4343	4343	4343

Standard errors clustered at the team level. Production is standardized (demeaned, and normalized by standard deviation) at the product-level, using production data from the pre-treatment period.

Table 4. Standardized Production and Attendance by Task Observability

	(1)	(2)	(3)	(4)
	Standardized Production		Attendance	
	Low Observability	High Observability	Low Observability	High Observability
Heterogeneous Team * Post Wage Change	-0.658	-0.052	-0.115	-0.025
	(0.252)***	(0.233)	(0.105)	(0.121)
Heterogeneous Team: M Rank * Post Wage Change	0.459	-0.027	0.100	-0.016
	(0.322)	(0.365)	(0.112)	(0.138)
Heterogeneous Team: H Rank * Post Wage Change	0.147	0.073	-0.094	-0.092
	(0.449)	(0.356)	(0.166)	(0.119)
Task x Experience FE	Yes	Yes	Yes	Yes
Day x Round FE	Yes	Yes	Yes	Yes
R-squared	0.446	0.446	0.262	0.262
N	1762	1768	1762	1768

Standard errors clustered at the team level. Production is standardized (demeaned, and normalized by standard deviation) at the product-level, using production data from the pre-treatment period. We define a task to be highly observable if the correlation between the perceived productivity of a worker relative his teammates has a correlation greater than 0.7 with the truth. Under this definition of observability, 4 tasks meet the criteria and the remaining 4 tasks have low observability.

Table 5. Standardized Production and Attendance by Relative Productivity Within Team

	(1)	(2)	(3)	(4)
	Standardized Production		Attendance	
	Low Difference	High Difference	Low Difference	High Difference
Heterogeneous Team * Post Wage Change	-0.920	-0.134	-0.310	0.083
	(0.303)***	(0.253)	(0.106)***	(0.098)
Heterogeneous Team: M Rank * Post Wage Change	1.282	-0.286	0.377	-0.146
	(0.388)***	(0.365)	(0.133)***	(0.136)
Task x Experience FE	Yes	Yes	Yes	Yes
Day x Round FE	Yes	Yes	Yes	Yes
R-squared	0.416	0.416	0.191	0.191
N	2759	1584	2759	1584

Standard errors clustered at the team level. Production is standardized (demeaned, and normalized by standard deviation) at the product-level, using production data from the pre-treatment period. Productivity differences for low- and medium-ranked workers are measured as the distance in pre-period productivity with the individual on the same team with a rank higher by one. We define the difference to be high if it is greaterr than the median of the distribution of productivity differences.

Table 6. Worker Perceptions of Fairness

Endline Survey Responses: Mean Differences

	Satisfied with wage (1)	Relative wage is fair (2)
Panel A: Low and Medium rank workers		
Heterogeneous team	0.033	-0.041
	(0.063)	(0.074)
Mean of workers in Compressed	0.892	0.892
Number of workers	120	120
Panel B: High rank workers		
Heterogeneous team	0.143	0.163
	(0.051)***	(0.054)***
Mean of workers in Compressed	0.857	0.837
Number of workers	63	63

The outcome variable in each column is the proportion of workers who agreed or responded positively to the questions: "How satisfied were you with your wage in your job here?" (Col. 1); and "Agree or Disagree: My wage was set fairly in relation to my other two teammates." (Col. 2). The sample in Panel A is all Low and Medium Ranked workers; the sample in Panel B is all High ranked workers. Each panel reports the mean difference in response for *Heterogeneous* teams relative to *Compressed* teams.

Appendix A. Supplemental Appendix

Table 7. Standardized Production and Attendance: Team Level Regressions

	(1) Standardized Production	(2) Standardized Production	(3) Attendance	(4) Attendance
Heterogeneous Team * Post Wage Change	-0.703*	-0.440	-0.0754**	-0.0425
	(0.367)	(0.409)	(0.033)	(0.032)
Compressed Medium Team * Post Wage Change		0.00225		0.0317
		(0.483)		(0.028)
Compressed High Team * Post Wage Change		0.756		0.0682**
		(0.471)		(0.027)
Task x Experience FE	Yes	Yes	Yes	Yes
Day x Round FE	Yes	Yes	Yes	Yes
R-squared	0.453	0.474	0.392	0.398
N	1483	1483	1483	1483

Standard errors clustered at the team level. Production is standardized (demeaned, and normalized by standard deviation) at the product-level, using production data from the pre-treatment period. Observations collapsed to the team level: sum(production), average(attendance). The omitted category in columns (1) and (3) is the set of all compressed teams (pooled), while in columns (2) and (4), the omitted category is Compressed Low teams.

Table 8. Standardized Production and Attendance Over Time

	(1)	(2)	(3)	(4)
	Standardized Production		Attendance	
	Early Period	Late Period	Early Period	Late Period
Heterogeneous Team * Post Wage Change	-0.359*	-0.737***	-0.00808	-0.242**
	(0.181)	(0.224)	(0.103)	(0.099)
Heterogeneous Team: M Rank * Post Wage Change	0.254	0.631**	0.00636	0.238**
	(0.262)	(0.258)	(0.120)	(0.110)
Heterogeneous Team: H Rank * Post Wage Change	0.175	0.379	-0.122	0.0373
	(0.331)	(0.336)	(0.143)	(0.124)
Task x Experience FE	Yes	Yes	Yes	Yes
Day x Round FE	Yes	Yes	Yes	Yes
R-squared	0.453	0.456	0.210	0.194
N	3041	3251	3041	3251

Standard errors clustered at the team level. Production is standardized (demeaned, and normalized by standard deviation) at the product-level, using production data from the pre-treatment period. All regressions include observations from the entire pre-treatment period. Columns (1) and (3) only include the first six work days after the wages were changed (treatment). Columns (2) and (4) include the last six work days of the treatment period.