Term paper on *The Morale Effects of Pay Inequality* - Based on Breza, Kaur and Shamdasani (2018)

Mingzhao Dai*

March 22, 2025

1 Introduction

There is a vast literature on the effects of pay inequality, especially on how the pay inequality is embedded into the social preference function and thus affects the behavior of employees (e.g., output, attendance, quit rate, etc.). Some papers, Dube et al. (2019) for example, have done in-depth research on this by conducting lab or field experiments to quantify and study the extent of morale effects towards pay inequality. Few studies, however, have specifically discussed the incentives and behavioral mechanism behind these morale effects, especially on an experiment setting before Breza et al. (2018)[1]. The perceived justifications, as one of the behavioral mechanism behind the morale effects, takes into consideration the worker's perceptions of other colleagues' wages, productivity, and how the worker understands the wage and productivity difference. The discussed paper examines the role of perceived justifications in morale effects of pay inequality and aims to fill the gap by conducting a comprehensive experiment in India, where the morale effects of pay inequality are particularly pronounced. The results of the experiment show that the perceived justifications acts as a significant factor of the morale effects when the worker faces pay inequality.

*Freie Universität Berlin and Berlin School of Economics

[1]For simplicity, I will refer to Breza et al. (2018) as "**the discussed paper**" throughout this term paper.

This term paper aims to provide a comprehensive review of the discussed paper, a critical review in detail, and a discussion on the connection to our social preference course. The term paper is structured as follows: The rest of section 1 will introduce the research questions of the paper, and why they are important. The section 2 will briefly provide a theoretical framework of the analysis, in order to differentiate various sources of the morale effects with respect to different perceived pay inequality levels. The section 3 will explain detailedly the design of the experiment, especially how the crucial randomizations are conducted. The section 4 will present the regression model the paper uses and show the results which could answer all of our research questions. The section 5 and 6 follows with critical reviews and connections to our social preference course.

1.1 Research Questions

Although the intuition and motivation of the discussed paper has been stated clearly in the introduction part above, I will still formally list all the research questions the discussed paper explores, which correspond to all the results found in the discussed paper. The research questions are as follows:

- **RQ1:** Do workers really know their colleague's wage?
- **RQ2:** Does pay inequality have negative effect on the morale behavior (output / attendance)?
- **RQ3:** How does perceived justification affect the morale effect?
- **RQ4:** How is the pooled unit effect (regardless of productivity rank)?
- **RQ5:** How is the absolute wage effect?
- **RQ6:** How is the effect from the channel of social cohesion?

The **RQ1** serves as a pre-check of other research questions. Only if the workers are aware of the pay inequality, the results of the experiment will become valid. The **RQ2** and **RQ3**

are the main questions this paper aims to explore. The **RQ4** to **RQ6** correspond to the extended analysis of the experiment results, which provide more aspects to understand the data got from the experiment and enhance the validity of the results of two main questions. Some of the research questions may appear vague at this stage. However, it will be clear after we go through the theoretical framework and experimental design in the following sections.

2 Theoretical Framework

The authors assume that a typical worker's payoff function simply follows the reduced form below:

$$V\left(w_i, w_R, e_i; \theta_i\right) = w_i - c\left(e_i; \theta_i\right) + M\left(w_i, w_R\right) e_i, \tag{1}$$

where w_i is the own wage of worker i, w_R is the wage of the reference worker, e_i is the effort level of worker i, θ_i is the worker's type, stated as the productivity rank of the worker i in the discussed paper, $c\left(e_i; \theta_i\right)$ is the cost of effort e_i with a productivity rank θ_i, and $M\left(w_i, w_R\right)$ is the morale effect of pay inequality.

The morale effect is assumed to be a function of the own wage of worker i and the wage difference between worker i and the reference worker. The construction of the morale effect function follows the main idea of Fehr and Schmidt (1999), as follows:

$$M\left(w_i, w_R\right) = \alpha f\left(w_i - w_R | w_i < w_R\right) + \beta f\left(w_i - w_R | w_i > w_R\right) + g(w_i). \tag{2}$$

The morale effect function shows that positive and negative differences between the own wage and the reference wage have different levels of effect on the morale of the worker. The absolute wage effect, $g(w_i)$, also has impact on the worker's morale to some certain extent, regradless of the reference wage.

It is important to notice that the reference wage w_R plays a crucial role in the equations

above. The general formation of the reference wage of a typical worker i is given by:

$$w_R = r\left(w_{-i}, \theta_i, \theta_{-i}\right). \tag{3}$$

As shown above, the reference wage of worker i is a function of the wages of all other workers, the productivity rank of worker i, and the productivity ranks of all other workers. The reference wage is the key to the morale effect of pay inequality, as it determines the level of pay inequality perceived by worker i. Since different workers have different perceptions of the productivity ranking, they may have different formations of reference wage specifically[2] within the general framework of equation 3. For example, 1) when the worker i is unaware of or neglects the difference in productivity ranks, she forms the reference wage as the function of only the wages of all other workers, i.e., $w_R = r\left(w_{-i}\right)$. This means, the worker only cares about the pay inequality in wages, but does not consider the potentially underlying reasons of the pay inequality, which might be the difference in productivity. 2) when the worker i is aware of the difference in productivity rank, and able to perfectly observe the productivity rank of all workers, she forms the reference wage as the function of the wages of all other workers and the productivity ranks of all workers, i.e., $w_R = r\left(w_{-i}, \theta_i, \theta_{-i}\right)$, as the equation 3 presents. 3) when the worker is aware of the productivity difference, but unable to observe the productivity, she forms the reference wage as $w_R = r(w_{-i}, \hat{\theta}_i, \hat{\theta}_{-i})$, where $\hat{\theta}_i$ and $\hat{\theta}_{-i}$ are the beliefs, i.e. the perceived productivity of worker i and other workers.

The three examples above shows that the reference wage is determined by the observability of wages, observability of the productivity and how the worker evaluate the productivity difference. I will show in the following sections how the authors examines these in the experiment, which corresponds to the research questions **RQ1** and **RQ3**.

[2]which are not directly observable to the authors and to be tested in the experiment

3 Experimental Design

In order to answer the research questions, the authors design an experiment following the intuition from the theoretical framework. In this section, I will explain the design of the experiment in detail, including the experiment background and context, the sample construction design, and the experiment procedure. Afterwards, I additionally discuss the information structure and underlying mechanisms of this experiment.

3.1 Background and Context

- **Location**. The experiment was conducted in India, where researchers found the morale effects of pay inequality are particularly strong from previous survey studies. This is crucial for the experiment to have enough variation in the morale effects.

- **Subjects**. The authors organized rural Indian workers who are seeking seasonal manufacturing jobs to work in local factories. Basically, these workers do agricultural work during around half of the year, and try to find short-term jobs in local factories during another half when the agricultural work is of off-season. This ensures that 1) the work is not so complicated that requires work experience for years; 2) the working contract is short enough to be completely covered by the experiment duration. It should be noticed that all the subjects are male, which may limit the generalizability and cause potential bias (I will discuss this in the critical review section).

- **Job description**. The subjects are asked to work on a low-skill non-cooperative manufacturing task, for example, producing candle wicks. The task is simple enough to be completed by all subjects, but the productivity of the subjects may vary. This is crucial for the experiment to have comparable productivity variations among the subjects and to quantify the productivity ranks.

- **Duration**. The experiment / short-term job lasts for one month during the non-

agricultural season.

- **Sample size**. The experiment consists of 14 rounds, with 30 workers in each rounds. No subject participates in more than one round.

- **Pay structure**. The workers are paid a flat daily wage according to the productivity. The flat daily wage is higher than the local prevailing wage $\bar{w}$.

3.2 Sample Construction

In the sample construction part, I will introduce three major randomizations the authors use to get a well established dataset in order to study the treatment and the effect of perceived justification. The three randomizations are as follows.

3.2.1 Randomization 1: reference group formation

Prior to the main experiment phase, there is a training period. At the beginning of the training period, all the workers are randomly divided into different **units**. Each unit has 3 workers, which means each worker has 2 actual random colleagues as the reference. The production tasks and wage schemes are different among units. At the end of the training period, all recruited workers are labeled with a productivity rank according to their performance in the training phase. The productivity rank is among "low productivity", "medium productivity" and "high productivity", according . In this way, the reference group of a typical worker is formed by the two colleagues in the same unit, with whom she can compare the wages and productivity.

3.2.2 Randomization 2: perceived justifications

The workers may have different perceptions regarding the productivity of themselves and their colleagues. Based on these perceptions, the workers may form perceived justifications

for the pay inequality, thus reducing or exaggerating their morale effects. Typically, there are two types of perceived justifications among the workers:

- **Discrete productivity difference**. Since the productivity ranks are discrete, the workers may have different morale effects even when they face the same pay inequality with the same productivity rank of themselves. This is because the actual productivity level is continuous, while the discrete productivity rank may not be able to fully capture the actual productivity difference. This causes an inconsistency between the productivity difference and pay difference. Here is an example.

  Example 1. *Suppose in an unit, worker A's productivity rank is "low productivity", worker B's productivity rank is "medium productivity", and worker C's productivity rank is "high productivity". In scenario 1, the actual productivity levels of the three workers are 8, 9 and 10, respectively. In scenario 2, the actual productivity levels of the three workers are 2, 9 and 10, respectively. The difference of A and C, in particular, is shown in the table below.*

	Worker A	Worker C	Morale effect
Scenario 1	8	10	High
Scenario 2	2	10	Low

  The table shows that, even though worker A is ranked as "low productivity" in both scenarios, he may have different morale effects towards the pay inequality with worker C. When the difference of actual productivity levels is as large as 8, the pay inequality turns more understandable to worker A, and thus the morale effect is low. In contrast, when the difference of actual productivity levels is as small as 2, the negative morale effect becomes higher.

  We can see, however, if we have a continuous productivity ranking, the same effect will not present because the worker will get the rank the same as his productivity level.

Therefore, it is important to test the effect of perceived justifications caused by productivity difference when we have a discrete ranking. The authors achieved this in the discussed paper by randomly assigning the workers to different units (Randomization 1), so that there will exist different "scenarios" in different units. The authors can thus make comparisons and estimate the (average) effects of this type of perceived justifications.

- **Observability of productivity**. As we discussed in section 2, the observability of productivity is one of the source of perceived justifications as it affects the formation of the reference wage. To test how the observability affects the morale effects, the authors randomly assign different tasks to different units (Randomization 2). Each type of task has a certain level of observability of productivity. Therefore, the authors can estimate the effect of observability of productivity since all the tasks are randomly distributed, and this distribution is not correlated with other variables such as unit characteristics and the wage scheme I am going to introduce in the next part.

  Let me introduce an example[3] to illustrate how different tasks have different observability of productivity. If one works in a coffee shop, her productivity is less observable than if she works in a candle wick factory. This is because the work in the coffee shop (e.g., making latte art) is not as standardized as the work in the candle wick factory (producing the same product in a fixed working period). Also, when working in the coffee shop, the frequency of the customers statistically follows a poisson process and is not deterministic along the time, so it is difficult to quantify the worker's productivity. Of course, the tasks in the discussed paper are more standardized and the difference among tasks (in terms of observability of productivity) is not as large as the example I provided. However, the example can still help us understand the concept of observability of productivity.

[3]The tasks in this example are not from the tasks of the experiment in the discussed paper.

3.2.3 Randomization 3: wage treatment

Again, different units are assigned with different wage schemes out of 4 types (Pay disparity, Compressed low, Compressed medium, Compressed high). The Pay disparity wage scheme is considered the treatment group and the other three are control groups. The scheme designs are shown in the table below.

	Pay disparity	Compressed low	Compressed medium	Compressed high
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}

The wages are discrete and $\bar{w} < w_{low} < w_{medium} < w_{high}$.

Therefore, we can just compare the subjects with the same stars shown above, in order to get the treatment effect of pay inequality.

So far, we have done all three randomizations, and the sample is ready to be used in the experiment. Before proceeding to the experiment procedure, there is a virtual summary of the sample construction design below.

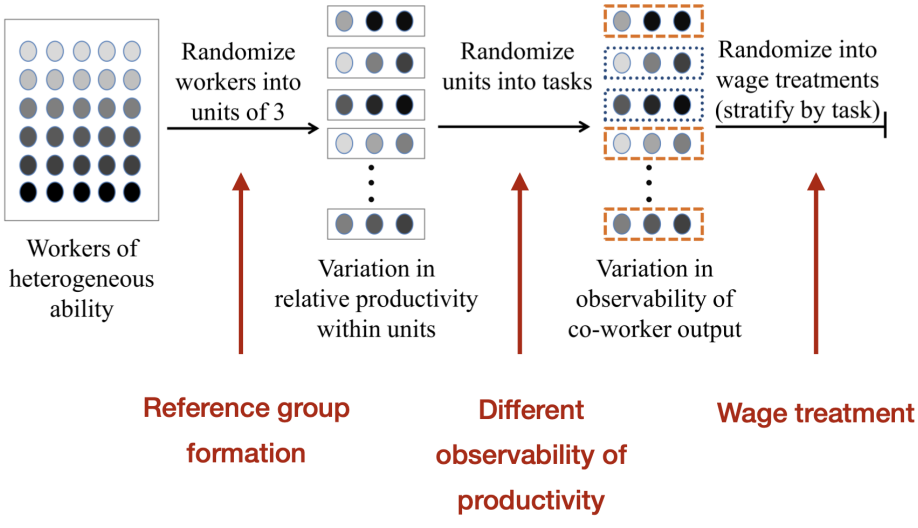

Figure 1: Sample construction design

3.3 Experiment Procedure

The timeline of this experiment is in the order as follows:

- **Pilot experiment**. Prior to the experiment we mainly discuss, the authors conducted several pilot experiments to estimate the observability of productivity of the tasks which are going to be assigned to the workers in the main experiment. The pilot experiments are conducted on different subjects. The subjects are asked to rank their own productivity compared to their colleagues' productivity. The authors used the accuracy of the subjects' own evaluation as the proxy of the task's observability of productivity. That means, higher accuracy is driven by better knowledge of the colleagues' productivity, which implies higher observability of productivity.

- **Recruitment**. When the main experiment starts, the authors recruited workers and randomize them in the way that section 3.2 introduces except wage scheme treatment (Randomization 3) only.

- **Training period (Day 1 - Day 14)**. The workers get trained to produce the products of the assigned task in this period. In this period, all workers are paid the same flat wage $\bar{w}$ regardless of their output. Typically, they get familiar with the task and their output becomes stable after Day 4. Afterwards, their output starts to be recorded in order to evaluate their productivity rank. They privately receive their productivity rank relative to their colleagues' at the end of the training period, on Day 14. They are also randomized into different wage scheme groups (Randomization 3) and receive their salary level (private information) before the next stage starts.

- **Treatment period (Day 15 - Day 34)**. In this period, Workers produce the products and get paid according to the wage scheme they are assigned to.

- **Endline test and survey (Day 35)**. At the end of the experiment, the workers are asked to do a group cohesion test and a survey.

Specifically, it is important to emphasize what information a typical worker has and does not have when he makes morale decisions.

Information structure. *A typical worker privately knows his productivity rank θ_i and wage level w_i (after Day 14). His knowledge on θ_{-i} depends on the observability of productivity of the task, while the knowledge on w_{-i} is also limited and determined by colleagues' self-disclosure and rumors. He has no knowledge of whether he and other people are in the treatment group (Pay disparity), i.e., he does not know $PayDisp_i$ for all i.*

3.4 Underlying Mechanisms

The authors point out that there may be two potential channels through which the morale effects of pay inequality work. The first channel is driven by individual preference. That is, the worker may feel unfair directly according to their preference, and thus reduce their morale. This channel is not connected to any certain person/colleague. The second channel, however, is driven by social cohesion. The worker may feel unfair because of one or two certain colleagues' wages, and thus become less cooperative and reluctant to work in the same unit with those colleagues. This channel can be tested and identified by the endline cohesion test and survey.

4 Regression Models and Results

4.1 Regression Model: Treatment effect

The authors use DID regression model to estimate the treatment effect of pay inequality on the morale effects. The regression model is as follows:

$$
\begin{aligned}
y_{it} = & \alpha_1 \left[\text{Post}_t \times \text{PayDisp}_i \times \text{Low}_i\right] \\
& + \alpha_2 \left[\text{Post}_t \times \text{PayDisp}_i \times \text{Med}_i\right] \\
& + \alpha_3 \left[\text{Post}_t \times \text{PayDisp}_i \times \text{High}_i\right] \\
& + \alpha_4 \left[\text{Post}_t \times \text{Low}_i\right] + \alpha_5 \left[\text{Post}_t \times \text{Med}_i\right] + \alpha_6 \left[\text{Post}_t \times \text{High}_i\right] \\
& + \bar{\text{Irrel}}'_{it}\theta + \bar{\text{Neigh}}_{it}\gamma + \lambda_{\mathbf{i}} + \tau_{\mathbf{t}} + \eta_1 x_{kt} + \eta_2 x_{kt}^2 + \varepsilon_{it},
\end{aligned} \tag{4}
$$

where the Post_t is the post-treatment period indicator, PayDisp_i is the treatment indicator, Low_i, Med_i and High_i are the productivity rank indicators, $\bar{\text{Irrel}}_{it}$ is the vector of irrelevant variables, $\bar{\text{Neigh}}_{it}$ is the vector of neighborhood variables, $\lambda_{\mathbf{i}}$ is the fixed effect of worker i, $\tau_{\mathbf{t}}$ is the fixed effect of time t, x_{kt} is the control variable, and ε_{it} is the error term.

We care about α_1, α_2 and α_3 in the regression model, which are the treatment effects of pay inequality on the morale effects of low, medium and high productivity workers, respectively (**RQ2**). For example, $\alpha_1 = \mathbf{E}\left[y(\text{PayDisp} = 1) - y(\text{PayDisp} = 0) \mid \text{Post} = 1, \text{Low} = 1\right]$ estimates the average treatment effect (ATE) of pay inequality on the morale effects of low productivity workers in the post-treatment period.

The authors also estimate the effects of perceived justifications (**RQ3**) by interacting the main indicators $[\text{Post}_t \times \text{PayDisp}_i \times \text{Rank}_i]$ with the perceived justification indicators Observ_k and ProdDiff_i. Additionally, the pooled effect of Pay disparity regardless of the rank (**RQ4**) can be estimated by the following regression model:

$$
y_{it} = \alpha \left[\text{Post}_t \times \text{PayDisp}_i\right] + \bar{\text{Irrel}}'_{it}\theta + \bar{\text{Neigh}}'_{it}\gamma + \lambda_i + \tau_t + \eta_1 x_{kt} + \eta_2 x_{kt}^2 + \varepsilon_{it}. \tag{5}
$$

The absolute wage effect (**RQ5**) can be estimated simply by running regression on the own wage level w_i. The social cohesion effect (**RQ6**) and the wage observability **RQ1** can be estimated by the endline cohesion test and survey.

4.2 Results to the Research Questions

The results of the experiment are as follows:

- **RQ1:** Do workers really know their colleague's wage?

  Answer: Yes, mostly. In the control groups, 95.8% of subjects know at least one colleague's wage, 90.9% know both two colleagues' wages. In the treatment group, they are 87.1% and 74.2%.

- **RQ2:** Does pay inequality have negative effect on the morale behavior (output / attendance)?

  Answer: Yes, the effects of pay inequality on morale behavior are as follows[4]:

	Low productivity	Medium productivity	High productivity
Output	-*	-	-
Attendance	-*	-*	-*

- **RQ3:** How does perceived justification affect the morale effect?

  Answer: Negative effects on the morale from both discrete productivity difference and observability of productivity. That means, when the perceived justification is significant, the morale effects are reduced. Especially, when productivity difference is very big, the morale effect almost disappears.

- **RQ4:** How is the pooled unit effect (regardless of productivity rank)?

  Answer: The pooled unit effect of pay inequality is negative. That means, when the pay inequality presents in a big system, the overall efficiency decreases.

- **RQ5:** How is the absolute wage effect?

  Answer: There is no significant effect of the absolute wage on the morale. This is perfect for this study, as workers focus more on relative wage than absolute wage.

[4]The significant results are shown with a star*.

- **RQ6:** How is the effect from the channel of social cohesion?

  Answer: The social cohesion channel is active. Pay inequality harms unit cohesion.

5 Critical Reviews

5.1 Strengths

- **Multi-dimensional experimental design**. The experiment in the discussed paper does not only include the main experiment which answers the **RQ2**, but also includes the pilot experiments, the training period, the endline test and survey, the randomizations of sample construction, and the perceived justification analysis. This multi-dimensional design is the main reason why this paper can answer as many as six research questions in detail. It ensures the validity of the results, the robustness of the conclusions, and also shows the power of experimental economics.

  The well-organized experiment design also makes it possible to conduct such a complicated and comprehensive experiment within 35 days in India. The dataset got from this experiment contains very rich information, which could be further explored and used to answer some other research questions such as the dynamics of the effect of pay inequality on morale (since the experiment lasts for 35 days and all the daily output data are recorded). This is a very good example of how to design a comprehensive experiment with limited resources.

- **Exploration in perceived justifications**. With the topic of social preference, this discussed paper extends the ordinary analysis of the effect of pay inequality on morale, productivity, quit rate and etc. (like the paper Dube et al. (2019) introduced in Section 1) to the exploration of the role of the worker's belief, perceptions and justifications when the pay inequality presents. This is a very important extension, as the social preference theory itself focuses on how interactions between individuals affect the be-

haviors, and the perceived justifications are formed based on interactions and networks. How the worker perceives the wage disparity, the productivity difference and how he understands, justifies these differences are crucial to understand the morale effects of pay inequality and the social preference behind the worker's behavior. This discussed paper provides a good example of how to explore the perceived justifications in an experiment setting, and how to test the effects of these justifications on the morale effects. Also, the discussed paper contributes to the related literature by providing a new perspective to understand the social preference in the context of labor economics.

- **Policy implications**. This discussed paper provides a clear policy implication that pay inequality has negative effects on the morale, and the perceived justifications can reduce the negative effects. This is important for the policy makers to design the wage schemes and the working environment in order to increase the efficiency and the morale of the workers. The results of this paper can be used to design the wage schemes in the factories, to design the training programs for the workers, and to design the working environment in order to increase the efficiency and the morale of the workers. This is a good example of how to use the experimental results to provide policy implications.

5.2 Weaknesses

- **Concerns about the pilot experiment**. The pilot experiment is conducted on different subjects from the main experiment. This may cause potential bias and limit the generalizability of the results. The reason is, the subjects in the pilot experiment may have different characteristics from the subjects in the main experiment, which may affect the results of the main experiment. For example, some workers are more self-concentrated and don't really notice how is everything going on with their colleagues. This may cause an underestimation of the observability of productivity for that task. Sometimes, however, the worker in the pilot experiment may be too confident and overestimate his own productivity, thus making the observability of productivity

higher than it actually is. This could make our analysis on how the observability of productivity affects the morale effects less reliable.

In my opinion, the authors could have included the question about the evaluation of other colleagues' productivity at the endline survey in the main experiment as well. In this way, the authors could have compared the results of the pilot experiment and the main experiment, and thus could have tested the generalizability of the results by checking that the observability measurement has not changed much within the main experiment sample. Also, the authors should have done full randomizations in the pilot experiment and repeated the pilot experiment for enough times, and taken the average of the observability measurement to avoid bias. These methods could have made the results more reliable and robust.

- **The wage scheme setting**. One seemingly trivial but important issue is about the wage setting. In the discussed paper, all the wage level (w_{Low}, w_{Medium}, w_{High}) are higher than the local prevailing wage $\bar{w}$, even the lowest one. This could cause a problem considering the background and household conditions of the workers. Firstly, if the lowest wage level w_{Low} is already higher than the local prevailing wage $\bar{w}$, the workers may not feel the pay inequality as a big issue. This is because they are paid extra money that cannot be found in somewhere else. The following result would be that workers care less about the pay inequality. This could lead to an underestimation of the morale effects of pay inequality. Secondly, if the wage level is too high, the workers may not feel the necessity to work hard. This could cause the morale effects of pay inequality to be overestimated. Thirdly, considering that all of the workers are with seasonal contracts and they don't have agricultural work to earn money, it could be crucial to notice that the household may rely heavily on the wage income from this job. This may put pressure on the workers and prevent them from working inactively, being absent or even quitting the job. Because if they do so, they may face unemployment and the household may face financial difficulties. This could cause the

morale effects of pay inequality to be underestimated.

Therefore, the wage level should be set at a reasonable level, for example, let w_{Low} be set close to the local prevailing wage $\bar{w}$, in order to get a valid and reliable result. Also, the highest wage w_{High} should not be too high to weaken the worker's motivation and the gaps among three wage levels should be within a reasonable interval. The authors should have provided more information about the local prevailing wage and the household conditions of the workers, and should have explained why they set the wage level as they did. Ideally, there should be more theoretical analysis on the equilibrium of the wage system in the local labor market, to better understand the worker's incentives.

- **All workers are male**. This is a big issue. Specifically, there are two explanations why this issue is so serious here.

  Firstly, it is very straightforward that if all the subjects in the sample are male, the results will lose generalizability. The results of the discussed paper cannot, for example, simply apply with a sample in Sweden. This is because in those countries where people care more about gender diversity, there are many female workers working in the factory and the characteristics of the sample changes. It is still unclear whether female workers care more or less about the pay inequality, especially with a seasonal contract in a local factory, so we cannot have a quick intuition that how the results will change. Therefore, there is not that much to imply according to this paper about the insights outside India, and even outside the experiment regions.

  Secondly, we should notice that the pay inequality already exists between male and female workers in most of the societies around the world. This indicates an existing endogeneity in this regression model if we include female workers. On the one hand, if we do the reference group randomization, and give the male and female workers the same wage level, they may have different preference towards the pay inequality,

and different perceptions about the wages and productivity. This is due to the lower outside option of the female workers. For example, let the wages be like $\bar{w}_F < \bar{w}_M < w_{low} < w_{medium} < w_{high}$ where the local prevailing wage of female workers is lower than the male workers'. The female worker who is assigned the lowest wage w_{low}, then, may have less morale effects towards pay inequality than male workers as the female wage jump from the outside option is bigger than the male one. On the other hand, if the female worker faces the pay inequality, she may have one additional understanding of this disparity - gender inequality. She may think the pay inequality is not caused by productivity difference. Instead, it is due to gender difference. Also, it can be imagined that when the female worker earns the highest wage, the male colleague may have more envy about the difference between him and the female colleague. This is not tested and included in the experimental design in the discussed paper, of course, and should be considered if we include female subjects in the sample.

Therefore, in reality cases where female workers are also included, things become much more complicated and the experiment should be re-designed more carefully as well as the regression strategy. The current discussed paper is not able to provide a full picture of the morale effects of pay inequality in reality, and the results should be interpreted with caution.

6 Connections to Social Preference Theory

The discussed paper, although with a topic of labor economics, shows a strong connection to the social preference theory. The theoretical part follows the idea of Fehr and Schmidt (1999), that is, the worker has a social preference on fairness, and dislikes the pay inequality. As a result, the morale effects are driven by the social preference. The model in the discussed paper, however, has different parameters compared to Fehr and Schmidt (1999) in terms of the parameter signs. In the discussed paper, the worker always prefers his own wage

over other people's wage, which means the parameter $\beta > 0$ in the model. In Fehr and Schmidt (1999), however, the agent is set to potentially dislike, in a general setting, the pay inequality even when he is the one who gets the higher wage, which corresponds to the negative parameter β in the discussed paper. Unlike Bolton and Ockenfels (2000), the model in the discussed paper does not directly include the relative share of the total wage into the preference function. However, the same thing actually happens in the model, as the different wage levels already imply the different shares of the total wage (although discrete). The only difference is among control groups, where workers have the same wage share, but different morale effects. This is what Bolton and Ockenfels (2000) is silent about.

Bosman and Van Winden (2002) points out that emotions plays an crucial role in social preference, and emotional hazard is identifed as a source of effciency costs. The discussed paper shares the similar idea. Since the workers within the same unit work in the same environment, and they also eat lunch together, chat together, and have some other social activities together, the pay inequality may cause negative emotions among the workers. Also, the emotions affect the cohesion inside a unit, which serves as one of the channels of the morale effects of pay inequality. The social cohesion channel is active, and the pay inequality harms unit cohesion in the discussed paper, thus reducing the overall effciency. This is consistent with the idea of Bosman and Van Winden (2002).

The discussed paper provides another example of what Bénabou and Tirole (2006) states. Bénabou and Tirole demonstrate that external incentives, such as rewards or punishments, can lead to an "overjustification effect", where the introduction of extrinsic motivators undermines intrinsic prosocial behavior, potentially resulting in a net decrease in such actions. In the discussed paper, offering high wages (corresponding to the rewards in Bénabou and Tirole (2006)) to the workers as a motivator of working hard (corresponding to the prosocial behavior in Bénabou and Tirole (2006)) may reversely reduce the total contribution, due to the existence of pay inequality. Therefore, the analysis in the discussed paper also provides a good explanation of why the overjustification effect happens in the labor market setting,

especially for explaining why the α_3 in the model, i.e. the morale effect of high productivity workers, is also negative.

The discussed paper is an excellent example of presenting the existence of heterogeneous fairness norms. Cappelen et al. (2007) investigates how individuals apply different fairness principles—such as strict egalitarianism, liberal egalitarianism, and libertarianism—when distributing resources. Through experimental methods, Cappelen et al. demonstrate that people hold diverse fairness ideals and that these norms significantly influence their distributive choices. In the discussed paper, the treatment group (salary based on productivity) serves as a typical example of libertarianism, while the control groups (flat wage) serve as a typical example of strict egalitarianism. The results show that the morale effects of pay inequality are negative, which means the workers in the fairness norms of the treatment group turns out to be closer to strict egalitarianism instead of libertarianism. This could be explained as another channel of the morale effects of pay inequality, and it proves that the heterogeneous fairness norms do influence the workers' behaviors and choices.

There are two other papers which has a similar topic with the discussed paper. Dube et al. (2019), another paper I discussed in the class, studies how pay inequality affects the worker's quit behavior in a US retailer. The research question somehow overlaps with the discussed paper, i.e. exploring how pay inequality affects worker's production behavior (morale, output, quit rate, etc.). However, Dube et al. (2019) does not include the perceived justifications and the social cohesion channel in the analysis, and the method (RDD) and dataset used in that paper is more like empirical instead of experimental. The research questions are also not as comprehensive as the ones in the discussed paper. This comparison shows that the discussed paper extends the analysis of social preference to a deeper level by investigating the perceived justifications - the perceptions and incentives behind the worker's morale effects. It does not simply apply the Fehr and Schmidt (1999) model to the social preference function, but also consider how the worker forms the reference wage (equation 3) given different perceptions and understanding of the wages, observability of productivity and

the productivity difference. This is the main contribution of the discussed paper compared to similar research like Dube et al. (2019).

Another paper DeCelles and Norton (2016) talks about how another inequality - the cabin class difference - triggers passenger's air rage. This paper has a similar topic with the discussed paper, but shows further evidence of how emotions affect individual's behavior, which is a supplement to Bosman and Van Winden (2002). Particularly, this paper shows evidence that air rage is more likely to be due to the unfair emotions instead of the physical discomfort. This could have analogies with the discussed paper where workers may have reduced morale probably due to the unfair emotions caused by the pay inequality rather than the real wage difference. This could be tested, I think, by changing the gap between the wage levels, and see how the morale effects change. If the morale effect does not change much as we increase the wage gap, the unfair emotions may be the main reason of the morale effects.

References

Roland Bénabou and Jean Tirole. Incentives and prosocial behavior. *American economic review*, 96(5):1652–1678, 2006.

Gary E Bolton and Axel Ockenfels. Erc: A theory of equity, reciprocity, and competition. *American economic review*, 91(1):166–193, 2000.

Ronald Bosman and Frans Van Winden. Emotional hazard in a power-to-take experiment. *The Economic Journal*, 112(476):147–169, 2002.

Emily Breza, Supreet Kaur, and Yogita Shamdasani. The morale effects of pay inequality. *The Quarterly Journal of Economics*, 133(2):611–663, 2018.

Alexander W Cappelen, Astri Drange Hole, Erik Ø Sørensen, and Bertil Tungodden. The pluralism of fairness ideals: An experimental approach. *American Economic Review*, 97 (3):818–827, 2007.

Katherine A DeCelles and Michael I Norton. Physical and situational inequality on airplanes predicts air rage. *Proceedings of the National Academy of Sciences*, 113(20):5588–5591, 2016.

Arindrajit Dube, Laura Giuliano, and Jonathan Leonard. Fairness and frictions: The impact of unequal raises on quit behavior. *American Economic Review*, 109(2):620–663, 2019.

Ernst Fehr and Klaus M Schmidt. A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, 114(3):817–868, 1999.