China Economic Review 77 (2023) 101908

Contents lists available at ScienceDirect

China Economic Review

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

# Do men and women discriminate against women for the same reason? Evidence from China

Jianfei Shen [a], Qiqi Wang [b,*]

[a] Shandong University of Finance and Economics, 7366 Erhuandong Road, Jinan, Shandong 250000, China
[b] Xi'an University of Finance and Economics, 360 Changning Avenue, Xi'an, Shaanxi 710100, China

A R T I C L E  I N F O

*JEL classification:*
C90
J31
J71

*Keywords:*
Discrimination
Gender
Labor market
Experiment

A B S T R A C T

In this study, we examine whether men and women form gender discrimination for the same reason. To do that, we build an experimental Chinese labor market in which employers evaluate the productivity of workers who perform a real-effort task. Before evaluation, the employer observes the worker's personal and group information. The personal information contains gender identity and a signal of productivity. The group information reveals the productivity distributions of some other male and female workers who do not differ in average productivity. However, it shows more male workers at the very top productivity levels and more female workers at the very bottom productivity levels in one treatment than in the other. According to the belief-based theory, there will be a greater degree of discrimination against female workers in the former. We find that, however, only male employers' evaluations are well predicted by this approach. Female employers behave oppositely: their degree of gender discrimination is smaller in the treatment emphasizing men's advantage in the tails of the productivity distributions. To explain female employers' evaluations, we adopt the preference-based approach. Our findings suggest that employers of different genders can have different motivations for gender discrimination, and thus call attention to the theoretical foundation of gender discrimination and policy measures aimed at reducing gender discrimination.

## 1. Introduction

Gender discrimination may be driven by three forces. According to the preference-based literature, gender discrimination is caused by a preference against rewarding or interacting with women (Becker, 1957). The statistical approach argues that people judge individual women based on gender differences in abilities (Arrow, 1972; Arrow, 1973; Phelps, 1972). The inaccurate belief theory, which has recently been developed, proposes that people judge women based on systematic biases in beliefs (Bohren, Imas, & Rosenberg, 2019; Bordalo, Coffman, Gennaioli, & Shleifer, 2016).

In most of the exising studies, men and women are not separately discussed, and the underlying assumption is that they form gender discrimination for the same reason. However, men and women diverge in a wide range of preferences (Croson & Gneezy, 2009; Niederle & Vesterlund, 2007), despite that they are alike in most cognitive abilities (Hyde, 2005; Rudman & Glick, 2008). Taking such divergence as point of departure, our research question is: If gender discrimination exists, what are the underlying motives for men and women, respectively? To explore this question, we employ a series of laboratory experiments in the setting of the Chinese labor market

* Corresponding author.
*E-mail addresses:* jianfeishen@sdufe.edu.cn (J. Shen), qiqiwang@xaufe.edu.cn (Q. Wang).

https://doi.org/10.1016/j.chieco.2022.101908
Received 10 October 2021; Received in revised form 31 October 2022; Accepted 15 December 2022
Available online 21 December 2022

as a worker's gender is commonly observable in job applications in China. We find that the sources of male and female employers' gender discrimination could be fundamentally different and, more importantly, context dependent. Our findings demonstrate that it is critical to consider the interaction between employer characteristics and embedding environments in order to learn about the precise reason for gender discrimination.

The experiment proceeds as follows. Participants play roles as workers and employers. The workers are asked to solve a real-effort task, and the employers are incentivized to evaluate worker productivity based on individual characteristics and other information. Employer evaluations determine workers' earnings, and gender discrimination is measured by the evaluation differentials between male and female workers with identical productivity.

Before employers evaluate workers, we provide each employer with workers' individual and group information. The individual information contains a signal of productivity and gender identity, and the group information reveals the productivity distributions of some other male and female workers. The employers are partitioned into two treatments. In both treatments, the group information – on average – shows no gender difference in ability. However, it differs in the tails: there are more men than women at the very highest performance levels, and more women than men at the very lowest performance levels in one treatment than in the other treatment.

The purpose of varying group information between treatments is to test the inaccurate belief theory by Bordalo et al. (2016). The theory states that even if there are no gender differences in average productivity, discrimination can still arise. The reason is that when the tails of the productivity distributions become highly representative and are overweighted in employers' assessment of the average productivity, a belief that women are aggregately less productive than men will be wrongly formed. For example, women have made great progress in school performance and today are comparable to men in average school grades and overall school performance, including mathematic grades (Goldin, Katz, & Kuziemko, 2006; Hyde, Lindberg, Linn, Ellis, & Williams, 2008). However, the belief that women are worse than men at mathematics still persists (Ecceles, Jacobs, & Harold, 1990; Carrell, Page, & West, 2010; Guiso, Monte, Sapienza, & Zingales, 2008). Such belief is explained by the fact that in some exams men appear more frequently than women at the very highest performance levels, leading one to mistakenly use the extreme performers as the representative types for each gender and overstate the size of the average differences (Bordalo et al., 2016).

According to the above mechanism of inaccurate belief, there will be a greater degree of discrimination against female workers in the treatment where the group information shows more men at the very top and more women at the very bottom. We find that, however, only male employers' evaluations follow such a pattern. Female employers' behavior goes to the opposite direction: their degree of discrimination against female workers is actually smaller in the treatment emphasizing men's advantage in the tails.

To explain female employers' behavior, we adopt the preference-based approach in addition to the belief-based one. That is, female employers associate the underrepresentation of women at the top performance levels with real life experiences, like glass ceiling and lack of women at the elite positions in workplaces. Such associations lead female employers to feel a threat to self-esteem and/or develop sympathy toward female peers. As a response, they activate a positive preference toward other women to boost self-esteem and/or to protect female peers. This positive preference translates to higher evaluations for female workers, and its effect is large enough to not only neutralize the treatment effect on belief, but also mitigate the original gender discrimination of female employers.

Studying the motivation of discrimination is challenging as different motivations – preferences, statistical discrimination, and inaccurate beliefs – generate similar patterns of observable behavior (Fang & Moro, 2011). Our current paper further finds that these different motivations of discrimination are interactive. That is, while male employers' behavior is explained by the belief-based approach, female employers' behavior cannot be completely explained by their beliefs as we disturb the information structure they face. Since one's gender has little or no bearing on cognitive skills (Hyde, 2005; Rudman & Glick, 2008), it follows that some other factors must vary with information. Therefore, our findings suggest that employer gender is key to disentangle different motivations for gender discrimination.

Based on these findings, our study contributes to the literature that investigates the reasons behind gender discrimination by separately addressing the motives of males and females. The existing works mostly focus on identifying an overall, single motive for discriminatory behavior. If analyzing employer behavior at the overall level, however, we would have failed to identify the exact reason of gender discrimination which depends on the interaction between employer gender and information environment. By considering employer gender, we observe that male and female employers respond differently to the variation of information. This observation highlights the complexity of gender discrimination: the reasons for men and women's gender discrimination are not only different but also context dependent. It is therefore crucial for future works, both theoretic and empiric, to pay attention to the interaction between the gender of the discriminating agents and the environments the agents are embedded in.

Empirically, our study has three implications for the reduction of gender discrimination. First, although women's status has improved remarkably in the 21st century, the gender hierarchy is still significant worldwide and has been documented in a large scope of contexts, including the labor market context (Goldin & Rouse, 2000; Kuhn & Shen, 2013; Neumark, Bank, & van Nort, 1996; Zhang, Jin, Li, & Wang, 2021), the bargaining context (Ayres & Siegelman, 1995; Castillo, Petrie, Torero, & Vesterlund, 2013), and the academics context (Card, Vigna, Funk, & Iriberri, 2020; Dupas, Modestino, Niederle & Wolfers, 2021; Mengel, Sauermann & Zolitz, 2019; Sarsons, Gerxhani, Reuben & Schram, 2021; Wu, 2020). Our study provides an explanation for why gender discrimination persists even though considerable resources have been devoted to solve this problem. Many public policy measures deployed to reduce gender discrimination do not distinguish males and females for the target population. We demonstrate, however, one measure that attenuates discrimination against females for one gender actually exacerbates discrimination for the other gender. This is a caveat for policy interventions in the future, that it is crucial to implement different strategies for each gender. Second, several works have studied how females treat female peers. The evidence, however, is mixed. While some works demonstrate that females are nice to female peers (Bapna & Ganco, 2021; De Paola & Scoppa, 2015; Kunze & Miller, 2017), some argue for the opposite (Bagues & Esteve-Volart, 2010; Broder, 1993; Kunze & Miller, 2017). Our study offers an explanation for why the evidence has been mixed. That is, the context matters. In the current work, females become nice to female peers when observing that males occupy the majority of top-level positions. Finally, our study proposes a solution for policy interventions via information. In a male dominant environment, the bias against women would be mitigated when information puts less emphasis on men's advantage over women at elite positions. In a female dominant environment, however, information triggering positive preferences toward female peers would be useful.

The rest of this paper is organized as follows. Section 2 reviews the literature related to this work. Section 3 introduces the experimental design and procedures. Section 4 reports and discusses the results, and Section 5 concludes.

## 2. Related literature

The present study is part of growing research on discrimination motivations. Below we first summarize the theories of discrimination, and then review the empiric works in the gender context with emphasis on our contribution. In the end of the section, we also review the empiric works exploring the motives of racial/ethnic/minority discrimination.

In terms of the underlying causes for discrimination, three theories have been suggested. The early literature focuses on the role of preferences (tastes). Becker (1957) proposes that employers with prejudice against a certain group of workers perceive the costs of hiring workers from this target group to be higher than the actual wages. Assuming that workers from the target group and other workers are identical in productivity, the former must accept lower wages in order to be hired. As a result, non-prejudiced employers hire workers from the target group with lower wages and prejudiced employers hire other workers with higher wages. This preference-based approach predicts that discrimination is not profitable, and prejudiced employers will not survive in competitive markets. By contrast, Arrow (1972, 1973) and Phelps (1972) argue that discrimination can arise even in the absence of prejudice. Statistical discrimination theory states that non-prejudiced, profit-maximizing employers use group statistics to judge workers with similar individual characteristics, leading to a wage gap based on group differences. This statistical approach relies on actual group differences, and implies that discrimination will be reduced as group differences narrow down. However, the disappearance of group differences in education and the implementation of anti-discrimination policies suggest that the explanation for the remaining gap lies elsewhere. In this respect, a strand of literature has emerged. In particular, Bordalo et al. (2016) present a model of inaccurate beliefs based on cognitive biases. In their model, beliefs are rooted in the differences between tails of productivity distributions rather than averages. Specifically, the upper and the lower tails become highly representative for each group, and are overweighted in the decision maker's assessment of the group averages. This inaccurate belief approach improves the discrimination theory and predicts discrimination for non-prejudiced, profit-maximizing employers against workers who have group averages similar to others but are disadvantaged in the tails.

Based on the theories, some empiric works have examined the motivations for gender discrimination. Perhaps the works that are mostly related to ours are Bohren et al. (2019) and Coffman, Exley, and Niederle (2021). These works use experimental methodology to disentangle different motivations by manipulating one particular factor while holding other factors constant between treatments. Specifically, Bohren et al. (2019) test how users of an online Q&A forum evaluate female and male posts. The treatment variable is the evaluation history of the user account. The authors find that for new user accounts that have no evaluation history, female posts receive fewer positive votes than male posts. For advanced accounts that have accumulated a history of evaluations, female posts receive more positive votes than male posts. These findings indicate that gender discrimination is driven by inaccurate belief. Coffman et al. (2021) focus on isolating the effect of gender preferences. In their experiment, employers predict the performance of workers in a real-effort task. Before prediction, employers repeatedly observe a group of male and female workers' performance in another similar task, so that beliefs are held constant across employers. For the prediction, there are two treatments: one does not show worker gender and one shows worker gender. Since beliefs are held constant, any difference in prediction between the two treatments can be attributed to gender tastes. The authors then find positive preferences favoring females.

The key difference between the above two works and ours is that their findings do not depend on the gender of the discriminating agents, while in our study employer gender is key for analyzing the specific motive for gender discrimination. More importantly, we show that for each gender, the specific reason for discriminating against females depends on information environment. Methodologywise, Bohren et al. (2019) manipulate user individual information on the Q&A forum, while we consider the labor market context and display different types of group information to employers. Coffman et al. (2021) hold group information constant but vary the availability of worker gender, while we hold the availability of worker gender constant but vary group information across treatments. Furthermore, the workers' performance distributions shown to employers in Coffman et al. (2021) reflect gender differences in both averages and extreme performance, making it difficult to separate statistical discrimination and inaccurate belief. In our study, the workers' productivity distributions contain only gender differences in extreme performance, allowing us to attribute gender discrimination to employers' inaccurate beliefs.

Beside the gender context, several works have explored discrimination motives in other contexts. For preference-based discrimination, Charles and Guryan (2008) construct racial prejudice indices using them US General Social Survey and find that prejudice explains one-quarter of the racial wage gap in the US labor market. Doleac and Stein (2013) vary the skin color of the hand holding iPod Nano for sale on local websites, finding an offer gap between iPods held by white hands and iPods held by black hands. The offer gap decreases as level of market competition increases, providing evidence to preference-based discrimination. For statistical discrimination, List (2004) send buyers and sellers to trade sportscard in a real market, and find differences in offers made to majority and minority groups. In complementary experiments, the data suggests that statistical discrimination is the main force driving the majority-minority gap. Nunley, Owens, and Stephen Howard (2011) act as white named and black named sellers and sell identical products on eBay. Because price differences between white sellers and black sellers disappear when seller information becomes more transparent through feedback system, the data is indicative of statistical discrimination. In Israel, Zussman (2013) conduct a correspondence study on the Israeli used car market to study discrimination against Arabic buyers and sellers. The results indicate that Arabic buyers receive fewer seller responses, and Arabic sellers receive fewer buyer responses. The following surveys reveal that discrimination is explained by beliefs about Arabic buyers and sellers. For inaccurate belief, Fershtman and Gneezy (2001) study ethnic discrimination in the Israeli Jewish society using abstract games. They find that people's biases toward Ashkenazic Israelis are systematic, belief-based, and inaccurate.

## 3. The experiment

The experimental framework adopts the design of Mobius and Rosenblat (2006) and Mobius, Rosenblat, and Wang (2016). It simulates a Chinese labor market where "employers" evaluate the productivity of "workers" who are employed in completing a real-effort task. The experimental labor market consists of two stages. The first stage is a baseline experiment that provides preliminary evidence for the existence of gender discrimination and further information for the second stage. The second stage is the main experiment that separates males and females' motivations for gender discrimination. In both stages, the participants play the worker's role and the employer's role. Below we introduce the two stages and the connection between them. The overall sequence of the experiment is shown in Fig. 1.

### 3.1. *The first stage experiment*

In the first stage experiment, all participants start the experiment as a worker who is incentivized to complete a real-effort task. The task involves solving a series of character puzzles that contain two quadratic 7 by 6 arrays of Latin alphabet characters. Fig. 2 shows an example of the puzzle. The two arrays are identical except for two random positions where the characters differ. To solve the puzzle, one has to find these two locations.[1]

Before starting the experiment, workers are informed that they need to solve puzzles in the following three periods. In the *warm-up period*, they solve two non-payoff-relevant puzzles to get familiar with the task. Then, in the *practice period*, workers solve one single puzzle. The length of time each worker completes this puzzle is recorded by the experimenter as his/her *practice performance*. Finally, workers solve puzzles in a five-minute *task period* at a piece rate of 40 credits per puzzle.[2] Workers are instructed that their performance in the practice period and other personal characteristics such as gender will be shown on a personal resume that is observed by employers who evaluate their performance in the task period. The workers also know that they will be rewarded at a rate of 40 credits per average employer evaluation.

The worker's role is followed by the employer's role. The participants who were previously a worker switch to an employer who evaluates worker performance.[3] The employer receives a resume displaying a worker's practice performance and gender. Depending on resume treatment, there are other worker characteristics such as ethnicity, urban/rural status (Huji), and province of origin on the resume.[4]

After receiving worker resume, the employer predicts the number of puzzles an individual worker would solve correctly in the task period. Each employer evaluates 10 other randomly selected workers. The employer receives a payoff of 150 credits for predicting a worker's task performance. However, the payoff is reduced by 10 credits per miscalibrated puzzle. Before the evaluation, employers are informed that each average employer evaluation rewards the evaluated worker with 40 credits.

In labor market terms, we interpret the worker's performance in the task period as the worker's labor market productivity. Practice performance, gender, and other characteristics are noisy signals of productivity. Gender discrimination is defined by employers' different evaluations for male and female workers with identical productivity.

Two important findings are obtained from the first stage experiment. First, we find that although men and women have similar average productivity, they differ in the tails of the productivity distributions. Second, discrimination against female workers at the overall level is found. These findings are indispensable for the design of the second stage experiment.

### 3.2. *The second stage experiment*

The second stage experiment is identical to the first stage experiment except that before evaluation, employers receive the worker's group information consisting of the productivity of 50 male and 50 female workers from the task period of the first stage experiment.[5] The group information is displayed to employers in the form of productivity distributions. An example of group information is given by Table 1. When employers get to the step of observing group information, the computer screen shows three columns. The first column lists the productivity levels in terms of numbers of puzzles solved, starting at zero and ending at forty-nine.[6] The second and the third

[1] The participants in our experiments are supposed to have the ability to identify basic alphabets as they are university students who have studied English for at least six years from grades 7–12.

[2] The same puzzles appear to the workers in the same sequence so that the measured abilities of solving puzzles are comparable.

[3] The design of this dual roles provides participants with complete information about the nature of the puzzle-solving task.

[4] One can study discrimination in different contexts through these resume treatments. For example, ethnic discrimination is reported in Mobius et al. (2016).

[5] To focus on gender discrimination, the resumes in the second stage experiment show only practice performance and gender.

[6] The largest number of puzzles solved in the task period of the first stage experiment is forty-nine.

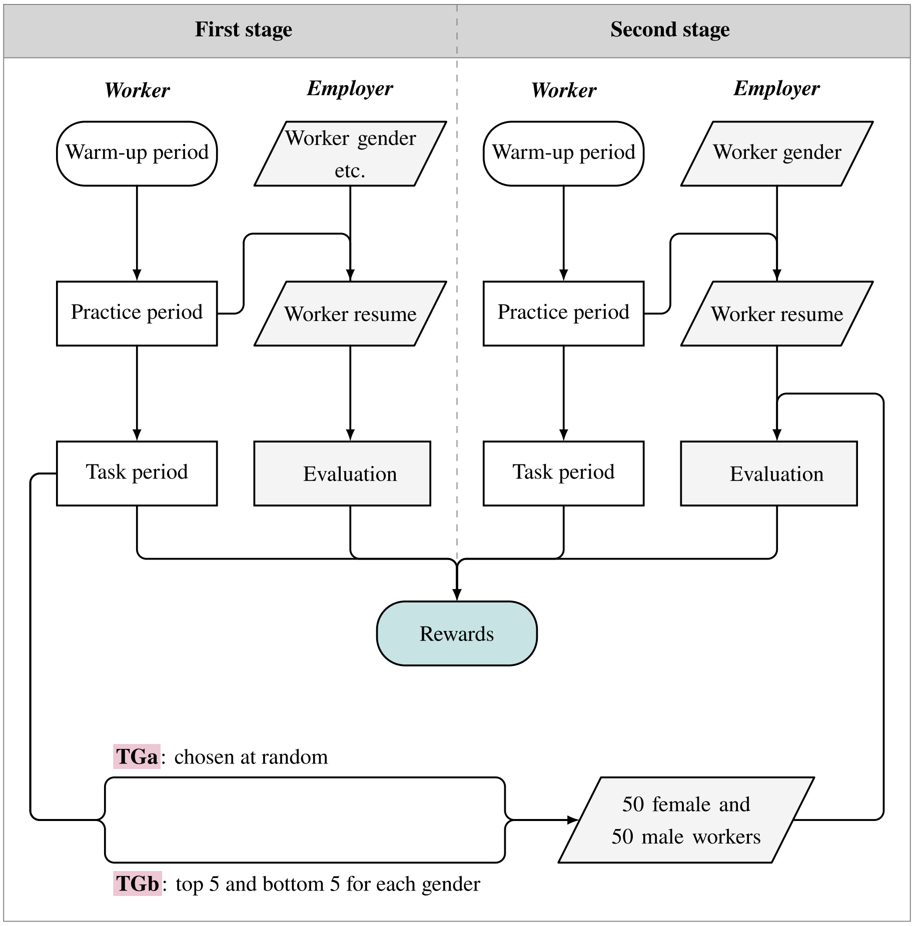


**Fig. 1.** The sequence of the experiment.

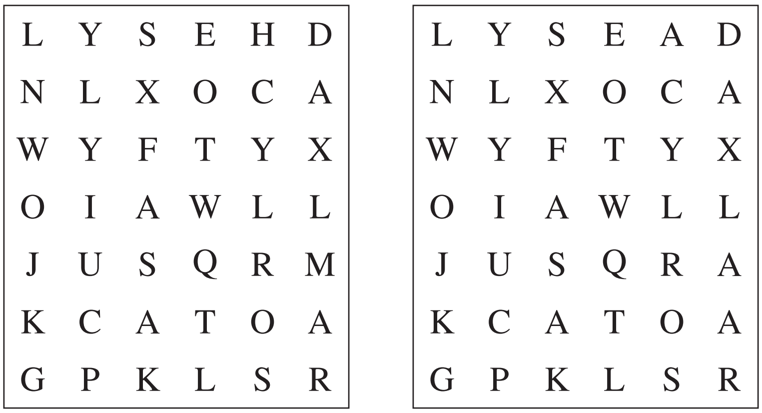


**Fig. 2.** The character puzzle.

columns show the numbers of males and females at each corresponding productivity level, respectively. The first row of Table 1 for instance, means that at the productivity level of zero puzzle, there are one male and one female.

The purpose of providing group information to employers is to shape their beliefs about average productivity at the gender level. There are several features in the experimental procedure guaranteeing that this purpose is achieved. First, displaying group information in the form of productivity distributions makes it easy for employers to observe the number of workers at each productivity level. Second, contrasting males with females is straightforward as they are listed in adjacent columns at each productivity level. Third,

**Table 1**
An example of the group information.

| Productivity | Number of male workers | Number of female workers |
|---|---|---|
| 0 | 1 | 1 |
| … | … | … |
| 49 | 1 | 0 |

there is no time limit for observing group information, so that only when employers are ready to proceed to the next screen, they click on a button to proceed.

Although in the task period of the first stage experiment there is no gender difference in productivity on average, there are some slight differences in the tails of the distributions. To look at the details, we present Fig. 3 showing the productivity distributions of the task period, and Table 2 showing the specific productivity of the top five and the bottom five performers by gender. As the data shows, at the very highest performance levels there are more men while at the very lowest performance levels there are more women.

Based on the features of the productivity distributions of the first stage experiment, we design two treatments, TGa and TGb, that differ in the tails of the productivity distributions. For treatment TGa, all observed workers are randomly drawn from the worker pool of the first stage experiment. For treatment TGb, the top five and the bottom five male performers are always selected, and the rest 40 male performers are randomly drawn from the worker pool. The 50 female performers are chosen in a similar way. This process of selecting workers insures that relative to that in the TGa treatment, the group information in the TGb treatment displays more men than women at the highest performance levels, and more women than men at the lowest performance levels. The employers are randomly assigned to either treatment TGa or treatment TGb, and in both treatments they are informed of the selection process of the observed workers.

The above two treatments allow us to test the inaccurate belief mechanism by Bordalo et al. (2016). The theory proposes that even there are no differences in averages, discrimination could still arise as people overweight the low probability but highly representative tails. In the current context, it predicts discrimination against women for employers who observe productivity distributions that reveal no gender differences in average productivity but emphasize men's advantage in the tails. Because the two treatments differ in the probabilities that employers observe the tails in which men are advantaged, the degrees of discrimination against females will be different. Specifically, because the tails are always presented to employers in the TGb treatment, we expect to observe a higher degree of discrimination against female workers in this treatment than in the TGa treatment.

### 3.3. Data

The participants were undergraduate students from two Chinese universities. Each participant was given a unique user identification number and a password to log in. We received participants' demographic information from the university administration in advance and were able to link each identification number to its corresponding demographic information. All experimental instructions were in Chinese. Table 3 displays the numbers of workers and employers.[7]

The participants earned different amounts of credits depending on their performance. The credits of the first stage experiment were converted into cash at a rate of 1 RMB per 100 credits. Considering rising price levels and living standards, the conversion rate increased to 1.5 RMB per 100 credits in the second stage experiment. On average, the participants earned 22 RMB in the first stage experiment and 35 RMB in the second stage experiment.

## 4. Results

In this part, we report and discuss the experimental results. First, we investigate the performance of the workers in the practice period and the task period. Second, we assess gender discrimination at the overall level. Third, we separately measure gender discrimination of male and female employers. Finally, we discuss the motivations for male and female employers' gender discrimination.

### 4.1. Worker performance

We present the means of the participants' performance as worker in Table 4. In the first stage experiment, the average practice performance of males and females are 20.8 and 20.1 seconds, respectively. The average performance in the task period are 15.3 puzzles for males and 15.8 puzzles for females. In the second stage experiment, males on average spend 16.2 seconds while females on average spend 15.5 seconds on solving the practice puzzle. In the task period, males and females on average solve 20.3 and 21.5 puzzles, respectively. In both stages of experiments, females appear to outperform males in practice performance and productivity, although the differences are not statistically significant.

To look for factors that could possibly affect productivity, we construct the following productivity regression:

[7] A few employers did not finish the evaluation task, including: six males and nine females in the first stage experiment, two males and two females in the TGa treatment, and one female in the TGb treatment.

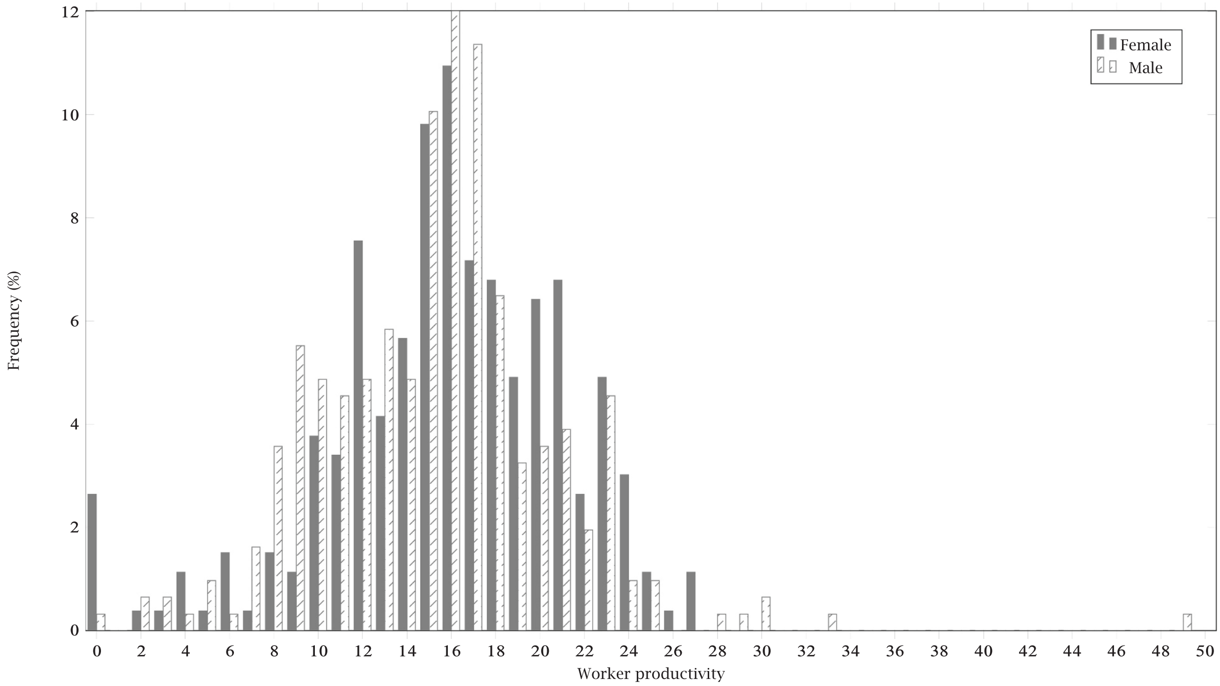


**Fig. 3.** The distributions of productivity in the task period of the first stage experiment.

**Table 2**
The top five and the bottom five performers in the first stage experiment.

| | Number of puzzles solved in the task period | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Top five performers | | | | | Bottom five performers | | | | |
| Male | 49 | 33 | 30 | 30 | 29 | 0 | 2 | 2 | 3 | 3 |
| Female | 27 | 27 | 27 | 26 | 25 | 0 | 0 | 0 | 0 | 0 |

**Table 3**
Numbers of workers and employers.

| | Number of workers | | | Number of employers | |
|---|---|---|---|---|---|
| | Male | Female | | Male | Female |
| First stage | 308 | 265 | | 302 | 256 |
| Second stage | 61 | 113 | TGa | 25 | 52 |
| | | | TGb | 34 | 58 |

$$Productivity_j = \beta_0 + \beta_1 Female_j + \gamma Practice_j + \varepsilon_j, \tag{1}$$

where *Productivity* is the number of puzzles that worker *j* solves in the task period, *Female* is the worker's gender (1 if female and 0 if male), and *Practice* is number of seconds that worker *j* spends on solving the single practice puzzle.

The coefficient $\beta_1$ in (1) measures the productivity differential between males and females in numbers of puzzles. To find equal productivity, we expect not to reject the null hypothesis that $\beta_1 = 0$. The estimation results are shown in Table 5. In all specifications, we do not reject the null hypothesis that $\beta_1 = 0$, meaning that there is no gender difference in average productivity among the workers.

## 4.2. Average productivity revealed by group information

Before evaluating worker productivity, each employer in the second stage experiment observes group information showing the productivity distributions of 50 males and 50 females from the task period of the first stage experiment. If group information shows no gender difference in average productivity, we can exclude the possibility of statistical discrimination and just consider gender preferences or inaccurate beliefs. It is therefore necessary to examine the average productivity of males and females that are revealed by

**Table 4**
Summary statistics of worker performance.

| | Workers | | | |
|---|---|---|---|---|
| | All<br>(1) | Male<br>(2) | Female<br>(3) | Difference<br>(4) |
| First stage experiment | | | | |
| Practice performance | 20.5 | 20.8 | 20.1 | 0.7 |
| | (12.7) | (12.1) | (13.5) | (1.1) |
| Productivity | 15.6 | 15.3 | 15.8 | −0.5 |
| | (5.4) | (5.4) | (5.4) | (0.5) |
| N | 573 | 308 | 265 | |
| Second stage experiment | | | | |
| Practice performance | 15.7 | 16.2 | 15.5 | 0.7 |
| | (6.4) | (6.2) | (6.5) | (1.0) |
| Productivity | 21.1 | 20.3 | 21.5 | −1.2 |
| | (4.9) | (4.9) | (5.0) | (0.8) |
| N | 174 | 61 | 113 | |

*Notes*: (i). The unit of practice performance is number of seconds, and the unit of productivity is number of puzzles. (ii). The standard deviations are in parentheses in columns (1)–(3). (iii). The standard errors are in parentheses in column (4). (iv). N is the number of workers.

**Table 5**
The determinants of worker productivity.

| Dependent variable | *Productivity* | | | | |
|---|---|---|---|---|---|
| Experiment | First stage | | | Second stage | |
| | (1) | (2) | (3) | (4) | (5) |
| *Female* ($\beta_1$) | 0.536 | 0.457 | 0.365 | 1.126 | 0.977 |
| | (0.452) | (0.434) | (0.421) | (0.782) | (0.719) |
| *Practice* | | −0.119*** | −0.100*** | | −0.311*** |
| | | (0.017) | (0.017) | | (0.054) |
| Other controls | No | No | Yes | No | No |
| $R^2$ | 0.003 | 0.081 | 0.144 | 0.014 | 0.175 |
| N | 573 | 573 | 573 | 174 | 174 |

*Notes*: (i). The standard errors are in parentheses. (ii) Hereafter, significance levels of 10%, 5% and 1% are denoted by *, ** and ***, respectively. (iii). Other controls are the worker's ethnicity (1 if minority and 0 if Han), Huji (1 if rural and 0 if urban), and province of origin (1 if western province and 0 if eastern province). (iv). N is the number of workers.

group information, in order to learn the exact reasons of gender discrimination. To do that, for the group information that is sent to a particular employer, we test the following hypothesis at the significance level of 5%:

$$H_0 : \text{Observed male average productivity} = \text{Observed female average productivity}.$$

The above hypothesis states that for a particular employer, the average productivity of the 50 observed male workers equals that of the 50 observed female workers. If it is rejected, we further examine which gender has a higher observed average. We conduct this hypothesis testing for each employer, and show the numbers of employers in each category in Table 6. In the TGa treatment, the group

**Table 6**
Average productivity revealed by group information.

| | All employers<br>(1) | Male employers<br>(2) | Female employers<br>(3) |
|---|---|---|---|
| Average productivity revealed by group information | | | |
| Treatment TGa | 77 | 25 | 52 |
| Male average > Female average | 0 | 0 | 0 |
| Male average < Female average | 5 | 1 | 4 |
| Male average = Female average | 72 | 24 | 48 |
| Treatment TGb | 92 | 34 | 58 |
| Male average > Female average | 0 | 0 | 0 |
| Male average < Female average | 0 | 0 | 0 |
| Male average = Female average | 92 | 34 | 58 |

information reveals that male average equals female average for 72 out of 77 employers. For the remaining five employers, one male and four females, the group information reveals that male average is lower than female average. In the TGb treatment, the group information reveals equal averages for all employers. Based on these results, we conclude that group information in both treatments does not reveal any gender differences in average productivity favoring male workers. Therefore, gender discrimination against female workers that we look for in subsequent analysis, is not likely to be a product of observed differences in group averages.

### 4.3. Gender discrimination at the overall level

We start examining gender discrimination by comparing employers' mean evaluations for male and female workers. In Table 7, columns (1), (4), and (7) present the means of all employers. In the first stage experiment, employers assign 18.2 puzzles to male workers and 17.8 puzzles to female workers. This evaluation gap of 0.4 puzzle in favor of the male workers is significant at the 10% level. In the TGa treatment, the mean evaluations are 16.2 and 16.0 puzzles for male and female workers, respectively. In the TGb treatment, the average evaluation is 18.2 puzzles for male workers and 18.3 puzzles for female workers. The evaluation gaps of 0.2 puzzle in the TGa treatment and −0.1 puzzle in the TGb treatment are not statistically significant.

To study gender discrimination more rigorously, we construct an evaluation regression:

$$Evaluation_{ij} = \overline{\beta}_0 + \overline{\beta}_1 Female_j + \overline{\gamma} Practice_j + E_i + \overline{\varepsilon}_{ij}, \tag{2}$$

where $Evaluation_{ij}$ is employer $i$'s evaluation for worker $j$. Because each employer evaluates 10 workers, we add the term $E_i$ for each employer that controls for unobservable individual characteristics, and report the estimation results for both fixed effect and random effect models.

Regression (2) is a mirror of regression (1). For perfectly non-biased employers, the coefficient $\overline{\beta}_1$ should be identical to its counterpart in regression (1). In particular, we expect not to reject the null hypothesis that $\overline{\beta}_1 = 0$ because we have verified that there is no gender difference in average productivity. Otherwise, gender discrimination is found as employers assign different evaluations to male and female workers.

In Panel A of Table 8, columns (1) and (2) report the estimation results for the first stage experiment, columns (3) and (4) report the results for the TGa treatment, and columns (5) and (6) report the results for the TGb treatment. The estimates of $\overline{\beta}_1$ suggest that there are gender gaps ranging from 0.320 to 0.556 puzzle in favor of the male workers. These gender gaps are statistically significant at the 5% level in the first stage experiment, the 10% level in the TGa treatment, and the 5% level in the TGb treatment. As shown in Panel B, the estimates of $\overline{\beta}_1$ translate to predicted gender differentials of 1.8%-3.0%.

### 4.4. Gender discrimination by employer gender

We turn to decompose the overall discrimination into discrimination by male and female employers. The mean evaluations of male and female employers are reported in columns Table 7. For female employers in the first stage experiment, the evaluation gap of 0.737 puzzle favoring male workers is significant at the 1% level. None of the other cases are statistically significant.

Next, we modify regression (2) by adding an interaction term of worker gender and employer gender to separate the evaluations of

**Table 7**
Summary statistics of employer evaluations.

| Experiment | First stage | | | Second stage | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Treatment | | | | TGa | | | TGb | | |
| Employers | All (1) | Male (2) | Female (3) | All (4) | Male (5) | Female (6) | All (7) | Male (8) | Female (9) |
| Workers | | | | | | | | | |
| All | 17.9 | 17.1 | 18.9 | 16.1 | 18.4 | 15.0 | 18.2 | 16.0 | 19.5 |
| | (6.9) | (5.5) | (8.1) | (7.0) | (6.3) | (7.1) | (5.3) | (5.8) | (4.5) |
| N | 5486 | 2977 | 2509 | 770 | 250 | 520 | 920 | 340 | 580 |
| Male | 18.2 | 17.2 | 19.3 | 16.2 | 18.0 | 15.3 | 18.2 | 15.9 | 19.6 |
| | (7.2) | (5.6) | (8.5) | (7.0) | (6.3) | (7.2) | (5.6) | (5.7) | (5.1) |
| N | 2967 | 1654 | 1313 | 385 | 125 | 260 | 460 | 170 | 290 |
| Female | 17.8 | 17.2 | 18.6 | 16.0 | 18.7 | 14.7 | 18.3 | 16.1 | 19.5 |
| | (7.6) | (6.8) | (8.5) | (7.4) | (6.8) | (7.4) | (5.6) | (6.4) | (4.8) |
| N | 2519 | 1323 | 1196 | 385 | 125 | 260 | 460 | 170 | 290 |
| Difference | 0.4* | 0.0 | 0.7*** | 0.2 | −0.7 | 0.6 | −0.1 | −0.2 | 0.1 |
| | (0.2) | (0.2) | (0.3) | (0.4) | (0.7) | (0.5) | (04) | (0.6) | (0.5) |

*Notes*: (i). The standard deviations are in parentheses. (ii). For the differences, the standard errors are in parentheses. (iii). N is the number of evaluations.

**Table 8**
The determinants of employer evaluations.

| Panel A. Regression results | | | | | | |
|---|---|---|---|---|---|---|
| Dependent variable | *Evaluation* | | | | | |
| Experiment | First stage | | Second stage | | | |
| Treatment | | | TGa | | TGb | |
| Model | FE | RE | FE | RE | FE | RE |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| *Female* ($\widetilde{\beta}_1$) | −0.349** | −0.320** | −0.446* | −0.446* | −0.556** | −0.555** |
| | (0.157) | (0.156) | (0.269) | (0.269) | (0.248) | (0.248) |
| *PracticeTime* | −0.093*** | −0.092*** | −0.405*** | −0.405*** | −0.625*** | −0.623*** |
| | (0.005) | (0.005) | (0.024) | (0.024) | (0.023) | (0.023) |
| Other controls | Yes | Yes | No | No | No | No |
| $R^2$ | 0.020 | 0.022 | 0.095 | 0.095 | 0.216 | 0.216 |
| N | 5486 | 5486 | 770 | 770 | 920 | 920 |
| Hausman *p*-value | 0.685 | | 0.923 | | 0.656 | |
| Panel B. Predicted gender gap | | | | | | |
| $\Delta\widehat{Evaluation}\%$ | −2.0 | −1.8 | −2.7 | −2.8 | −3.0 | −3.0 |

*Notes*: (i). The standard errors are in parentheses. (ii). N is the number of evaluations. (iii). The predicted gender gaps for the first stage experiment are calculated between a pair of male and female workers whose practice performance is 20.4 seconds, Han and rural identities are observed by employers, and provinces of origin are not observed by employers. For the second stage experiment, the predicted gender gaps are calculated between a pair of male and female workers with the average practice performance of 15.7 seconds. (iv). Other controls are ethnicity, Huji, and province of origin. Regarding ethnicity, there are three groups: Han workers whose ethnicity is observed by employers, minority workers whose ethnicity is observed by employers, and workers whose ethnicity is not observed by employers. In the regression, we use the last group as the reference group and include dummies for the first two groups. The Huji and the province of origin variables follow similar procedures.

male and female employers:

$$Evaluation_{ij} = \widetilde{\beta}_0 + \widetilde{\beta}_1 Female_j + \widetilde{\beta}_2 EmployerFemale_i + \widetilde{\beta}_3 Female_j \cdot EmployerFemale_i + \widetilde{\gamma} Practice_j + E_i + \widetilde{\varepsilon}_{ij}, \tag{3}$$

where *EmployerFemale* is the employer's gender identity (1 if female and 0 if male) and $Female_j \cdot EmployerFemale_i$ is the interaction term of worker gender and employer gender. We report the estimation results for both fixed effect and random effect models.

The coefficient $\widetilde{\beta}_1$ in regression (3) is male employers' gender discrimination. The coefficient $\widetilde{\beta}_3$ measures the discrimination differential between male and female employers. The sum of these two coefficients, $(\widetilde{\beta}_1 + \widetilde{\beta}_3)$, is female employers' gender discrimination. If male and female employers assign perfectly non-biased evaluations to workers, the coefficients $\widetilde{\beta}_1$ and $(\widetilde{\beta}_1 + \widetilde{\beta}_3)$ should be identical to $\beta_1$ in regression (1). In other words, for non-discriminating male and female employers, we expect not to reject the null hypotheses that $\widetilde{\beta}_1 = 0$ and $(\widetilde{\beta}_1 + \widetilde{\beta}_3) = 0$, as we have verified that there is no gender difference in average productivity in regression (1).

In Panel A of Table 9, columns (1) and (2) report the estimation results for the first stage experiment, columns (3) and (4) report the results for the TGa treatment, and columns (5) and (6) report the results for the TGb treatment. In each case, the fixed effect and the random effect estimates are similar. In all cases, the *p*-values of the Hausman Test suggest that the random effect models are the preferable ones. We thus interpret the random effect estimates.

In column (2), the estimate of $\widetilde{\beta}_1$ is insignificant, suggesting that male employers in the first stage experiment do not discriminate against female workers. Likewise, male employers in the TGa treatment do not discriminate against female workers as $\widetilde{\beta}_1$ in column (4) is insignificant. However, the estimate of $\widetilde{\beta}_1$ in column (6) is statistically significant at the 5% level, suggesting that when male employers in the TGb treatment evaluate workers, there is an evaluation gap of 0.806 puzzle against female workers. According to Panel B, this evaluation gap 0.806 puzzle translates to a 4.9% gender gap.

For female employers, the estimate of $(\widetilde{\beta}_1 + \widetilde{\beta}_3)$ in column (2) is statistically significant at the 1% level, implying that female employers discriminate against female workers in the first stage experiment. According to Panel B, the evaluation gap of 0.665 puzzle converts to a 3.5% gender gap. The estimate of $(\widetilde{\beta}_1 + \widetilde{\beta}_3)$ is also significant in column (4), meaning that female employers in the TGa treatment discriminate against female workers as well. The evaluation gap of 0.868 puzzle predicts a gender gap of 5.6%. On the other hand, the estimate of $(\widetilde{\beta}_1 + \widetilde{\beta}_3)$ in column (6) is not statistically significant. This suggests that female employers do not discriminate against female workers in the TGb treatment.

### *4.5. Male and female employers' motivations for gender discrimination*

In this section, we discuss male and female employers' motivations for gender discrimination based on the regression results in Table 9. According to the *p*-values of the Hausman Test in Panel A, we adopt the random effect estimates for the discussion. For the

**Table 9**
The determinants of male and female employers' evaluations.

| Panel A. Regression results | | | | | | |
|---|---|---|---|---|---|---|
| Dependent variable | *Evaluation* | | | | | |
| Experiment | First stage | | Second stage | | | |
| Treatment | | | TGa | | TGb | |
| Model | FE | RE | FE | RE | FE | RE |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| *Female* ($\widetilde{\beta}_1$) | −0.038 | −0.026 | 0.433 | 0.433 | −0.810** | −0.806** |
| | (0.214) | (0.213) | (0.471) | (0.471) | (0.408) | (0.408) |
| *EmployerFemale* ($\widetilde{\beta}_2$) | | 2.054*** | | −2.602 | | 3.050*** |
| | | (0.601) | | (1.686) | | (1.116) |
| *Female* · *EmployerFemale*($\widetilde{\beta}_3$) | −0.666** | −0.639** | −1.301** | −1.301** | 0.401 | 0.398 |
| | (0.314) | (0.312) | (0.573) | (0.572) | (0.512) | (0.512) |
| *PracticeTime* | −0.092*** | −0.092*** | −0.405*** | −0.405*** | −0.626*** | −0.623*** |
| | (0.005) | (0.005) | (0.024) | (0.024) | (0.023) | (0.023) |
| Other controls | Yes | Yes | No | No | No | No |
| $R^2$ | 0.017 | 0.032 | 0.110 | 0.132 | 0.222 | 0.265 |
| N | 5486 | 5486 | 770 | 770 | 920 | 920 |
| Females' discrimination ($\widetilde{\beta}_1 + \widetilde{\beta}_3$) | −0.705*** | −0.665*** | −0.868*** | −0.868*** | −0.409 | −0.408 |
| | (0.229) | (0.228) | (0.327) | (0.326) | (0.311) | (0.311) |
| Hausman *p*-value | 0.636 | | 0.992 | | 0.572 | |
| Panel B. Predicted gender gap | | | | | | |
| $\Delta\widehat{Evaluation}\%$, male employers | −0.2 | −0.2 | 2.7 | 2.4 | −4.4 | −4.9 |
| $\Delta\widehat{Evaluation}\%$, female employers | −4.0 | −3.5 | −5.4 | −5.6 | −2.2 | −2.1 |
| Panel C. Comparison of coefficients | | | | | | |
| | $\widetilde{\beta}_1^{TGb} - \widetilde{\beta}_1^{TGa}$ (Treatment effect on belief) | | | $\widetilde{\beta}_3^{TGb} - \widetilde{\beta}_3^{TGa}$ (Treatment effect on preference) | | |
| Model | FE | RE | | FE | RE | |
| Size | −1.243 | −1.239 | | 1.702 | 1.699 | |
| *P*-value | 0.075 | 0.024 | | 0.069 | 0.016 | |

*Notes*: (i). The standard errors are in parentheses. (ii). N is the number of evaluations. (iii). The predicted gender gaps for the first stage experiment are calculated between a pair of male and female workers whose practice performance is 20.4 seconds, Han and rural identities are observed by employers, and provinces of origin are not observed by employers. For the second stage experiment, the predicted gender gaps are calculated between a pair of male and female workers with the average practice performance of 15.7 seconds. (iv). Other controls are ethnicity, Huji, and province of origin. Regarding ethnicity, there are three groups: Han workers whose ethnicity is observed by employers, minority workers whose ethnicity is observed by employers, and workers whose ethnicity is not observed by employers. In the regression, we use the last group as the reference group and include dummies for the first two groups. The Huji and the province of origin variables follow similar procedures.

male employers, the gender gap changes from 0.433 puzzle favoring females in the TGa treatment to 0.806 puzzle disfavoring females in the TGb treatment. As shown in Panel C, these two estimates, $\widetilde{\beta}_1^{TGa}$ and $\widetilde{\beta}_1^{TGb}$, are significantly different from each other.

To interpret the change of −1.239 puzzles from treatment TGa to treatment TGb, we use the belief-based approach. Specifically, Bordalo et al. (2016) propose that if some types occur more frequently in the target group than in the reference group, these representative types are overweighted in the decision maker's assessment of the target group, and stereotypes favoring the target group are formed. In the current context, men appearing at the top performance levels become representative men and are overweighted toward a higher average by male employers, while women appearing at the bottom performance levels become representative women and are overweighted toward a lower average by employers. As a consequence, a wrong belief about gender differences in productivity is formed against female workers.

Since men and women do not differ in most cognitive abilities (Hyde, 2005), we assume that the treatment effect on belief is of the same magnitude for female employers. That is, female employers' belief also changes by −1.239 puzzles. However, the overall change of female employers' evaluations, $(\widetilde{\beta}_1^{TGb} + \widetilde{\beta}_3^{TGb}) - (\widetilde{\beta}_1^{TGa} + \widetilde{\beta}_3^{TGa})$, is actually 0.460 puzzle favoring female workers. This means that there must be some non-belief-based factors that have positively affected female employers' evaluations.

We add the preference-based approach to explain female employers' behavior. The reason is the following. The group information in the TGb treatment emphasizing that women are missing at the top performance levels relative to men could be associated with negative real life experiences such as glass ceilings or lack of women at elite positions in workplaces. Under such associations, female employers become sympathetic and develop a positive preference toward their own group. This preference factor leads female employers to evaluate female peers more positively than males, and is large enough to counter the belief-based bias. Another preference-based explanation is that females treat the disadvantaged group better in general.[8] Thus when females in the TGb treatment observe that there is a group, the female workers in this case, is disadvantaged, they assign to this disadvantaged group higher evaluations. Such altruism motivation is mitigated in the TGa treatment as the group information reveals similarities in both averages and tails.

We can further derive the size of the treatment effect on female employers' preference. Recall that for the male employers, the treatment effect on belief is $(\widetilde{\beta}_1^{TGb} - \widetilde{\beta}_1^{TGa})$. For the female employers, the change in discrimination from the TGa treatment to the TGb treatment, $(\widetilde{\beta}_1^{TGb} + \widetilde{\beta}_3^{TGb}) - (\widetilde{\beta}_1^{TGa} + \widetilde{\beta}_3^{TGa})$, is the treatment effect on both belief and preference. The difference between these two differences, $(\widetilde{\beta}_3^{TGb} - \widetilde{\beta}_3^{TGa})$, isolates the treatment effect on preference. According to Panel C, the treatment effect on preference is 1.699 puzzles and statistically significant.

### 4.6. *Female employers' gender discrimination in the first stage experiment and the TGa treatment*

A surprising result is that in the first stage experiment and the TGa treatment, female employers significantly discriminate against female workers. A possible explanation for such phenomenon is the psychological scar of women. Scanlon (2018) contends that in a society in which people are discriminated on the basis of characteristics, such as gender or skin color, objectionable humiliating differences in status will emerge and persist. Even worse, the pressure of status inequality can lead to those who are discriminated against to accept their inferiority. That is, females begin to accept the oppressive belief that casts women as inferior, incapable, or a burden on society. Therefore, even when the message revealing that women have the same ability as men is provided to women, it is not strong enough to change their accepted beliefs that women are inferior to men. For men, this psychological scar has no effect, as they have always enjoyed more favorable assumptions from others.

## 5. Conclusion

We use experimental methodology in a labor market setting to explore the different motivations for gender discrimination by men and women. Employers in our experimental labor market observe workers' personal and group information and evaluate worker productivity. The group information in both treatments reveals no gender difference in average productivity, but in one treatment shows more men at the top productivity levels and more women at the bottom than in the other. By finding that male and female employers behave differently between the treatments, we conclude that the motivations for gender discrimination depend on employer gender and context.

This research offers new policy implications for addressing gender discrimination. According to social psychology research, cognitive learning is an effective strategy to reduce discrimination (Pettigrew, 1997; Pettigrew & Tropp, 2006; Pettigrew, Tropp, Wagner, & Christ, 2011). That is, if individuals are educated, persuaded, or informed that women and men are intellectually equal, gender discrimination would be reduced. However, our study implies that a more systematic attempt, which involves more complex public polices, is needed to break down discrimination against women. For males, or for environments in which males are the majority, cognitive learning should not make men's occupancy of the top-level positions prominent, but instead emphasize the progress of women and the disappearance of gender differences in education and many other dimensions. For females, or for environments in which females are the majority, it is essential to recognize that gender discrimination is groundless and inhumane. Moreover, measures that trigger women's positive preferences toward women should be encouraged.

Our experimental framework can be extended to address other labor market issues. For example, the character puzzle is a gender neutral real-effort task. In the real world, some tasks are male-typed or female-typed. For the female-typed tasks, one might observe that females appear more frequently at the top positions while males are lagging behind. How do women respond to such environment? Do men develop positive preferences toward men? These questions are of our interest but left for future research.

## Data availability

Data will be made available on request.

## References

Arrow, Kenneth (1972). Models of Job Discrimination. In A. Pascal (Ed.), *Racial Discrimination in Economic Life*. Lexington, MA: Lexington Books.

Arrow, Kenneth (1973). Discrimination in Labor Markets. In O. Ashenfelter, & A. Rees (Eds.), *The Theory of Discrimination* (pp. 3–33). Princeton, NJ: Princeton University Press.

Ayres, Ian, & Siegelman, Peter (1995). Race and Gender Discrimination in Bargaining for a New Car. *American Economic Review, 85*(3), 304–321.

Bagues, Manuel, & Esteve-Volart, Berta (2010). Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment. *The Review of Economic Studies, 77*(4), 1301–1328.

Bapna, Sofia, & Ganco, Martin (2021). Gender Gaps in Equity Crowdfunding: Evidence from a Randomized Field Experiment. *Management Science, 67*(5), 2679–2710.

Becker, Gary (1957). *The Economics of Discrimination*. Chicago, IL: The University of Chicago Press.

Bohren, Aislinn J., Imas, Alex, & Rosenberg, Michael (2019). The Dynamics of Discrimination: Theory and Evidence. *American Economic Review, 109*(10), 3395–3436.

[8] We thank one referee for pointing out this hypothesis.

Bordalo, Pedro, Coffman, Katherine, Gennaioli, Nicola, & Shleifer, Andrei (2016). Stereotypes. *Quarterly Journal of Economics, 131*(4), 1753–1794.
Broder, Ivy E. (1993). Review of NSF Economics Proposals: Gender and Institutional Patterns. *American Economic Review, 83*(4), 964–970.
Card, David, Vigna, Stefano Della, Funk, Patricia, & Iriberri, Nagore (2020). Are Referees and Editors in Economics Gender Neutral? *Quarterly Journal of Economics, 135*(1), 269–327.
Carrell, Scott, Page, Marianne, & West, James (2010). Sex and Science: How Professor Gender Perpetuates the Gender Gap. *Quarterly Journal of Economics, 125*(3), 1101–1144.
Castillo, Marco, Petrie, Ragan, Torero, Maximo, & Vesterlund, Lise (2013). Gender differences in Bargaining Outcomes: A Field Experiment on Discrimination. *Journal of Public Economics, 99*, 35–48.
Charles, Kerwin Kofi, & Guryan, Jonathan (2008). Prejudice and Wages: An Empirical Assessment of Becker's The Economics of Discrimination. *Journal of Political Economy, 116*(5), 773–809.
Coffman, Katherine, Exley, Christine, & Niederle, Muriel (2021). The Role of Beliefs in Driving Gender Discrimination. *Management Science, 67*(6), 3551–3569.
Croson, Rachel, & Gneezy, Uri (2009). Gender Differences in Preferences. *Journal of Economic Literature, 47*(2), 448–474.
De Paola, Maria, & Scoppa, Vincenzo (2015). Gender Discrimination and Evaluators' Gender: Evidence from Italian Academia. *Economica, 82*(325), 162–188.
Doleac, Jennifer, & Stein, Luke C. D. (2013). The Visible Hand: Race and Online Market Outcomes. *Economic Journal, 123*(572), F469–F492.
Dupas, Pascaline, Modestino, Alicia Sasser, Niederle, Muriel, & Wolfers, Justin (2021). *Gender and the Dynamics of Economics Seminars*. No: National Bureau of Economic Research, Article w28494.
Ecceles, Jacquelynne, Jacobs, Janis, & Harold, Rena (1990). Gender Role Stereotypes, Expectancy Effects, and Parents' Socialization of Gender Differences. *Journal of Social Issues, 46*(2), 183–201.
Fang, Hanming, & Moro, Andrrea (2011). Theories of Statistical Discrimination and Affirmative Action: A Survey. In Jess Benhabib, Matthew O. Jackson, & Alberto Bisin (Eds.), *Handbook of Social Economics*. Amsterdam: North-Holland.
Fershtman, Chaim, & Gneezy, Uri (2001). Discrimination in a Segmented Society: An Experimental Approach. *Quarterly Journal of Economics, 116*(1), 351–377.
Goldin, Claudia, Katz, Lawrence, & Kuziemko, Ilyana (2006). The Homecoming of American College Women: The Reversal of the College Gender Gap. *Journal of Economic Perspectives, 20*(4), 133–156.
Goldin, Claudia, & Rouse, Cecilia (2000). Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians. *American Economic Review, 90*(4), 715–741.
Guiso, Luigi, Monte, Ferdinando, Sapienza, Paola, & Zingales, Luigi (2008). Culture, Gender, and Math. *Science, 320*(5880), 1164–1165.
Hyde, Janet, Lindberg, Sara, Linn, Marcia, Ellis, Amy, & Williams, Caroline (2008). Gender Similarities Characterize Math Performance. *Science, 321*(5888), 494–495.
Hyde, Janet Shibley (2005). The Gender Similarities Hypothesis. *American Psychologist, 60*(6), 581–592.
Kuhn, Peter, & Shen, Kailing (2013). Gender Discrimination in Job Ads: Evidence from China. *Quarterly Journal of Economics, 128*(1), 287–336.
Kunze, Astrid, & Miller, Amalia R. (2017). Women Helping Women? Evidence from Private Sector Data on Workplace Hierarchies. *Review of Economics and Statistics, 99*(5), 769–775.
List, John (2004). The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field. *Quarterly Journal of Economics, 119*(1), 49–89.
Mengel, Friederike, Sauermann, Jan, & Zölitz, Ulf (2019). Gender Bias in Teaching Evaluations. *Journal of the European Economic Association, 17*(2), 535–566.
Mobius, Markus, & Rosenblat, Tanya (2006). Why Beauty Matters? *American Economic Review, 96*(1), 222–235.
Mobius, Markus, Rosenblat, Tanya, & Wang, Qiqi (2016). Ethnic Discrimination: Evidence from China. *European Economic Review, 90*, 165–177.
Neumark, David, Bank, Roy, & van Nort, Kyle (1996). Sex Discrimination in Restaurant Hiring: An Audit Study. *QuarterlyJournal of Economics, 111*(3), 915–941.
Niederle, Muriel, & Vesterlund, Lise (2007). Do Women Shy Away from Competition? Do Men Compete Too Much? *Quarterly Journal of Economics, 122*(3), 1067–1101.
Nunley, John M., Owens, Mark F., & Stephen Howard, R. (2011). The effects of information and competition on racial discrimination: Evidence from a field experiment. *Journal of Economic Behavior & Organization, 80*(3), 670–679.
Pettigrew, Thomas (1997). Generalized Intergroup Contact Effects on Prejudice. *Personality and Social Psychology Bulletin, 23*(2), 173–185.
Pettigrew, Thomas, & Tropp, Linda (2006). A Meta-Analytic Test of Intergroup Contact Theory. *Journal of Personality and Social Psychology, 90*(5), 751–783.
Pettigrew, Thomas, Tropp, Linda, Wagner, Ulrich, & Christ, Oliver (2011). Recent Advances in Intergroup Contact Theory. *International Journal of Intercultural Relations, 35*(3), 271–280.
Phelps, Edmund (1972). The Statistical Theory of Racism and Sexism. *American Economic Review, 62*(4), 659–661.
Rudman, Laurie A., & Glick, Peter (2008). *The Social Psychology of Gender: How Power and Intimacy Shape Gender Relations*. New York: Guilford Press.
Sarsons, Heather, Gerxhani, Klarita, Reuben, Ernesto, & Schram, Arthur (2021). Gender differences in recognition for group work. *Journal of Political Economy, 129*(1), 101–147.
Scanlon, Thomas M. (2018). *Why Does Inequality Matter*. New York: Oxford University Press.
Wu, Alice (2020). Gender Bias among Professionals: An Identity-based Interpretation. *Review of Economics and Statistics, 102*(5), 867–880.
Zhang, Jian, Jin, Songqing, Li, Tao, & Wang, Haigang (2021). Gender Discrimination in China: Experimental Evidence from the Job Market for College Graduates. *Journal of Comparative Economics, 49*(3), 819–835.
Zussman, Asaf (2013). Ethnic Discrimination: Lessons from the Israeli Online Market for Used Cars. *Economic Journal, 123*(572), 433–468.