Journal of the Economic Science Association (2024) 10:1–14
https://doi.org/10.1007/s40881-023-00144-4

ORIGINAL PAPER

Preference-based hiring decisions and incentives

Francesca Gioia[1] · **Giovanni Immordino**[2]

Received: 1 September 2022 / Revised: 13 June 2023 / Accepted: 4 August 2023 /
Published online: 28 September 2023

Abstract
This article replicates an experiment by Coffman et al. (Manag Sci 67(6):3551–3569, 2021) who separated taste-based and statistical discrimination by comparing employer choices in one of two hiring environments (treatments). Both treatments were characterized by the same ability distributions of workers in tasks on which men are found to outperform women on average, but only one allowed for gender-specific considerations. We found statistical discrimination against women when they are presented to employers not as women, but as people belonging to a low-performance group, but discrimination in their favor when their gender is revealed to potential employers. This discrimination in favor of women was observed in both male and female employers. It was greater when employers were women and disappeared when monetary incentives to employ more productive workers were higher for employers.

Keywords Gender · Labor discrimination · Preferences · Incentives

JEL Classification J16 · J24

1 Introduction

Studies of labor market discrimination typically refer to preference-based and statistical discrimination. When discrimination is preference-based, it is rooted in preferences and based on group dislike, by animus or prejudice, involving discrimination against individuals, who would be effectively willing to pay the price of their

✉ Giovanni Immordino
giovanni.immordino@unina.it

Francesca Gioia
francesca.gioia@unimi.it

1 University of Milan, Milan, Italy

2 University of Naples Federico II, CSEF, DISES, Via Cintia Monte Sant'Angelo, 81122 Naples, Italy

choices (Becker, 1957). By contrast, statistical discrimination relies on accurate or biased beliefs about the abilities or skills of the group an individual belongs to (Arrow, 2015; Phelps, 1972; Bordalo et al., 2016).[1] A lot of work has been done on this, both theoretically and empirically, to untangle the sources of discrimination.

To mitigate statistical discrimination, one might suggest providing ample specific information about individual candidates in order to avoid relying on average group information for important decisions such as hiring, promotions, and so on Guryan and Charles (2013). By contrast, where preference-based discrimination is concerned, decreasing subjectivity in judgment by providing information on the quality of a specific individual has no effect because it influences beliefs but not tastes (Bohren et al., 2019). Given that individuals engaging in preference-based discrimination seem willing to bear the costs involved in not selecting people from the group they dislike, increasing such costs could induce employers to forego their own preferences when making choices.

This paper replicates the carefully designed experiment by Coffman et al. (2021) who separated taste-based and statistical discrimination by comparing the choices of employers acting in one of two hiring environments (treatments), both characterized by the same ability distributions of workers in tasks on which men are found to outperform women on average, but only one allowing for gender-specific considerations. Moreover, we used this setting to vary the amount of employers' financial compensation (either low or high) in a within-subject design to study whether increasing employers' monetary costs eliminates preference-based gender discrimination in hiring.

Coffman et al. (2021) show that employers choosing between two workers with identical résumés are less willing to hire a worker from a group that performs worse on average, regardless of whether this group is defined by gender or by a non-stereotypical characteristic. Thus, discrimination against women is not specific to gender but statistical as it depends on beliefs about average group differences. However, the authors show that tastes do have a part to play: when the gender dimension is salient, females (males) are significantly more (less) likely to be hired than equally able males (females) when labeled as such than when labeled by means of their month of birth. Thus, while discrimination against women is statistical, preference-based considerations produce another type of discrimination that works in favor of women, disadvantaging equally able men (belonging to the better performing group). This sort of over-compensating behavior, inducing discrimination in favor of the *discriminated group*, has also been documented in other settings, for instance with immigrants (Alesina et al., 2018). Such preference-based discrimination in favor of females disappears when image concerns are reduced through the introduction of something that might excuse the hiring of a member of the non-preferred group, such as an apparent risk associated with hiring a female worker.

Our purpose in this paper is twofold. On the one hand, we replicated Coffman et al. (2021)'s experiment. We therefore selected a set of female workers born in

[1] Psychological biases and heuristics may lead to inaccurate beliefs or they may simply be due to a lack of information. A rational actor may lack the relevant information necessary to form correct beliefs.

https://doi.org/10.1007/s40881-023-00144-4 Published online by Cambridge University Press

even months and male workers born in odd months and studied how employers' hiring decisions change when the labels used to describe the *same* workers are framed in terms of gender or month of birth. On the other hand, we used the same design to study whether increasing the monetary cost of relying on tastes or other non-monetary motives makes individuals less likely to do so, thus eliminating or at least reducing gender discrimination.

We found that when employers' monetary incentives to employ more productive workers are as they were in Coffman et al. (2021), there is preference-based discrimination in favor of females whereby, controlling for beliefs about the average ability gap between the two groups, females are hired more often when defined as such than when they are labeled with their birth month. What we show that is new—thanks to our experimental design—is that preference-based discrimination, in terms of greater likelihood of hiring a female intentionally when labeled female instead of even-month worker, significantly reduces and is no longer present when employer's monetary incentives to employ more productive workers are higher. Results point in the same direction for the probability of hiring a male worker intentionally (i.e., the reduction in this probability for a male versus an odd-month worker is lower when monetary incentives to employ more productive workers are higher), but the effect is smaller and less precisely estimated. We also find that the decision to leave it to chance is not significantly different between the gender and month-of-birth treatment with low incentives. Yet, when monetary incentives to employ more productive workers are higher, employers in the gender treatment opt to leave the decision to chance significantly more frequently than in the month-of-birth treatment. Thus, our results suggest that increasing employers' financial compensation reduces the likelihood of them relying on preferences, especially in terms of hiring a female intentionally.

Moreover, our study shows that monetary incentives to employ more productive workers are a powerful tool to attenuate other broader motives in hiring decisions. Indeed, when we look at in-group preferences, namely at whether sharing the same identifier with an employee increases the likelihood of being hired (Tajfel et al., 1971; Chen and Li, 2009; Chen and Chen, 2011; Chen et al., 2014), we found that, when monetary incentives to employ more productive workers are low, in-group preferences are significantly stronger in the gender treatment than in the birth-month treatment. By contrast, such reliance on in-group status in the gender treatment is significantly reduced and no longer present with higher monetary incentives. In contrast with Coffman et al. (2021), we found no evidence of in-group preferences in the birth-month treatment: employers born in even months do not hire even-month employees more than odd-month employers do.

The main takeaway for policymakers and organizations interested in limiting discrimination within the workplace is that the remuneration of individuals making employment decisions (human resource offices or paid hiring committees) should be contingent on employees' performance. Further research should investigate whether monetary incentives to employ more productive workers also mitigate other (clearly identified) gender or race preference-based discrimination.

Concerning this last point, this article contributes to a recent current in economic literature which has found mixed results regarding the effect of financial incentives

https://doi.org/10.1007/s40881-023-00144-4 Published online by Cambridge University Press

on various sorts of discrimination. Anukriti (2018), for instance, studied an Indian policy that offers financial incentives to families designed to lower the fertility rate and the gender ratio at birth (increasing daughters over sons). She found that financial incentives have a modest effect on reducing the fertility but do increase the sex ratio at first birth. In two randomized vignette studies, Underhill (2019) varied the race of a hypothetical patient in need of a kidney transplant (black or white) and the financial incentives involved, showing that incentives encourage donation but introduce a significant bias favoring white patients. However, this "crowding in" of racial bias seems to be limited to medium-sized incentives, while large incentives may be so attractive as to override any bias effect. Hortaçsu et al. (2019) have estimated the effect of a policy in India that offers monetary incentives for inter-caste marriages. They found that a 10,000-rupee increase in the incentive raises exogamy between Scheduled Caste (historically disadvantaged caste) men and non-Scheduled Caste women by 4 percent in rural India.

The remainder of the paper is organized as follows: Section 2 describes our experimental design and the data. Section 3 presents the results and Sect. 4 concludes the paper.

2 Experimental design and data

Our hiring experiment follows Coffman et al. (2021)'s as regards treatments and employment settings but varies in the monetary incentives given to participants and the set of decisions they are required to make.

We conducted two preliminary studies to collect performance information for subjects to be used as the available "workers" for hire in the main hiring experiment. The first study asked just a few questions, including month of birth, separately to a female and a male pool, each made up of 150 people. Using the data from the first preliminary study, we then focused on two groups of workers: male workers born in odd months (79 subjects) and female workers born in even months (69 subjects). In a second preliminary study, these two groups of workers were invited to complete one easy and one hard three-minute math quiz, each consisting of 10 multiple-choice questions.[2] Workers received 10 cents for each question answered correctly in a randomly selected quiz.[3] On average, males answered 6.47 questions in the easy quiz correctly (ranging from 3 to 10) and females 5.67 (ranging from 0 to 9).

In the hiring experiment, participants were "employers" asked to make incentivized hiring decisions on the available workers.[4] At the time of the decision, employers received information about the easy-quiz performances of the applicants

[2] Overall, 105 workers (52 females and 53 males) accepted our invitation.

[3] They were made aware that their performances may have been shown to other participants in a follow-up experiment.

[4] The experiment was part of a bigger experiment including another preliminary survey. There was also a third part to the hiring experiment. Here we have detailed only the first two parts of the hiring experiment, i.e., those used to answer our research question. See Gioia and Immordino (2022) for details on the full experiment.

https://doi.org/10.1007/s40881-023-00144-4 Published online by Cambridge University Press

but not about their hard-quiz performance, which would determine their earnings in the experiment. On top of the fixed show-up fee, employers earned a bonus that depended on the hard-quiz performance of the selected worker in a randomly selected decision.

The treatment is the presence of a *gendered* label, instead of the neutral label of the control group, used to describe the same workers, thus keeping their performance constant. Thus, depending on treatment, the same available workers were labeled with their own gender (males and females) or with their own month of birth (categorized as even or odd months). We collected data on a total of 210 employers, 100 for the gender treatment and 110 for the month-of-birth treatment.

The experiment was divided into two parts. At the end, subjects completed a short questionnaire.

In the first part, we elicited participants' prior beliefs about the average ability gap (male/female gap and odd/even month gap)[5]. Then, employers were asked to make nine hiring decisions between groups of workers (i.e., female-even-month *vs* male-odd-month) and information on worker performance was provided in the form of a bar chart comparing the distribution of the two groups.[6] In each decision, employers could choose whether to hire a randomly selected worker from the female-even-month group, the male-odd-month group or leave the hiring decision to chance (in this case the computer randomly determined which group to hire from). The option to leave the hiring decision to chance was introduced by Coffman et al. (2021) with the aim of allowing for expressions of indifference; the availability of this option should increase the likelihood that choosing one of the other two groups reflects a strict preference. If one of the decisions from this part was selected for payment, the hired worker received an additional 25 cents as bonus payment and the employer received 10 cents for each question answered correctly—on the hard quiz—by the worker hired in the randomly selected decision. Finally, we elicited posterior beliefs, using the same questions asked to elicit prior beliefs, in order to see how a given set of beliefs affected behavior in the gender and month-of-birth treatments.

Therefore, in both treatments, employers received accurate information about the distribution of easy-quiz performance across worker groups (female and male workers in the gender treatment; odd- and even-month workers in the month-of-birth treatment) prior to making their hiring decisions and forming their beliefs on the group average performance difference.[7] The information set did not change since the

[5] The question used to elicit beliefs was the same used in Coffman et al. (2021): "If you compare the average score of a male (*odd-month* in the Month-of-birth treatment) worker to the average score of a female (*even-month*) worker from round 1 of the math questions, what do you think the difference in scores would be?". The same question was used for round 2 (hard quiz) and both questions were used also for the posterior beliefs.

[6] As in Coffman et al. (2021), all employers made the same nine hiring decisions. The distributions of the first eight decisions were formed on the basis of subsets of workers born during different date ranges and the order in which such decisions appear was randomized at the participant level. The last decision was the same for all participants and, for each group of workers, it contained the distributions of the full sample.

[7] Part 1 provides precise and comprehensive information about the performance on the easy quiz in our experiment. Its aim is to ensure that the perceived gender gap in performance aligns with the perceived birth-month gap in performance when comparing the two treatments. In this way, the birth-month treat-

https://doi.org/10.1007/s40881-023-00144-4 Published online by Cambridge University Press

set of workers was the same. However, employers in the gender treatment associated the performance information with male and female workers. By contrast, employers in the month-of-birth treatment, were never told workers' gender because the *same* workers were presented to them as two groups of workers, one born in odd months and the other in even months.

In the second part of the experiment, participants took two sets of nine hiring decisions between two specific workers, each belonging to a different group (male or female in the gender treatment; even- or odd-month in the birth-month treatment). For each hiring decision, employers knew the exact easy quiz performance of both applicants rather than knowing only the distributions of the two applicant groups as in Part 1, a design feature intended as a means of countering potential statistical discrimination. The displayed performances for the female-even-month and the male-odd-month workers for each decision were: 4-4, 5-4, 6-4, 7-4, 8-4, 6-6, 7-6, 8-6, 8-8.[8] Thus, individual performance was the same for both workers in three decisions and female-even-month workers did strictly better in the other six decisions. Considering only decisions in which the female-even-month workers weakly outperformed the male-odd-month workers enabled all decisions not to hire female workers to be classified as discrimination against women. Employers could choose whether to hire the male-odd-month worker, the female-even-month worker or leave the hiring decision to chance. If, for a given decision, employers chose to leave the decision to chance, the computer randomly determined which of the two workers (in that decision) to hire. If one of the decisions from this part was selected for payment, the worker hired received 25 cents.

Employers' payment depended on the hard quiz performance of the worker they decided to hire. In the first set of nine hiring decisions in Part 2 of the experiment, the employer earned 10 cents for each correct answer by the hired worker in the hard quiz. This monetary incentive to employ more productive workers was also used in Coffman et al. (2021). In the second set of nine hiring decisions, the monetary incentives to employ more productive workers were higher: the employer earned 50 cents for each hard quiz question answered correctly by the hired worker. Thus, while the structure of the experiment up to this point replicated that of Coffman et al. (2021), this set of decisions was an addition aimed at studying the role played by monetary incentives.

Table 1 shows the descriptive statistics for our variables separately by treatment and, for the dependent variables, by the magnitude of the monetary incentives. As in Coffman et al. (2021), we focused our attention on the three decisions in which employers chose between workers with the same easy quiz performance.

Footnote 7 (continued)

ment can serve as a suitable comparison for the gender treatment. To address our research question, we use Part 2 that entails choices between two specific workers.

[8] We made sure that, in the workers' sample, we had at least one worker with the easy quiz levels of performance chosen for the nine decisions in part 2 of the experiment. Thus, employers were choosing between real workers. If, for example, in the 4-4 decision they selected the female-even-month worker, to compute earnings, we randomly selected one out of all female-even-month workers with an easy quiz performance of 4.

https://doi.org/10.1007/s40881-023-00144-4 Published online by Cambridge University Press

Table 1 Descriptive statistics

	Mean (1)	Std. dev (2)	Min (3)	Max (4)
PANEL A: gender				
Low incentive				
Female-even-month and chance	0.478	0.275	0	1
Female-even-month	0.13	0.337	0	1
Chance	0.697	0.46	0	1
Male-odd-month	0.173	0.379	0	1
High incentive				
Female-even-month and chance	0.453	0.264	0	1
Female-even-month	0.097	0.296	0	1
Chance	0.713	0.453	0	1
Male-odd-month	0.19	0.393	0	1
Controls				
Posterior (easy gap)	2.69	2.86	− 5	10
Posterior (hard-easy gap)	− 0.19	2.261	− 12	8
No. of obs	300			
PANEL B: month of birth				
Low incentive				
Female-even-month and chance	0.332	0.286	0	1
Female-even-month	0.052	0.221	0	1
Chance	0.561	0.497	0	1
Male-odd-month	0.388	0.488	0	1
High incentive				
Female-even-month and chance	0.344	0.295	0	1
Female-even-month	0.067	0.250	0	1
Chance	0.555	0.498	0	1
Male-odd-month	0.379	0.486	0	1
Controls				
Posterior (easy gap)	3.827	2.434	− 3	10
Posterior (hard-easy gap)	− 0.282	1.671	− 6	6
No. of obs	330			

The dependent variables are *Female-even-month and chance* taking the value of 1 if a female-even-month worker is hired intentionally, 0.5 if chance is chosen and 0 if male-odd-month is chosen, *Female-even-month* taking the value of 1 if a female-even-month worker is hired intentionally and 0 otherwise, *Chance* taking the value of 1 if chance is chosen and 0 otherwise, *Male-odd-month* taking the value of 1 if a male-odd-month worker is hired intentionally and 0 otherwise

When making each choice, employers could decide whether to select a female-even-month worker, a male-odd-month worker or leave the hiring decision to chance. We classified employers' decisions across these three options using four methods. The first looks at the choice of a female-even-month worker in terms of the outcome of the choice (i.e., not only strict preference for female-even-month

https://doi.org/10.1007/s40881-023-00144-4 Published online by Cambridge University Press

workers but also the likelihood that these are randomly hired via the chance option). The corresponding dependent variable is *Female-even-month and chance*, taking the value of 1 if a female-even-month worker was hired intentionally, 0.5 if chance was chosen and 0 if male-odd-month was chosen, and it was the main dependent variable used in Coffman et al. (2021). The other three dependent variables, on the other hand, consider employers' preferences and are represented by strict preference for one of the two groups and indifference. Specifically, these are: *Female-even-month*, taking the value of 1 if a female-even-month worker was hired intentionally and 0 otherwise; *Chance*, taking the value of 1 if chance was chosen and 0 otherwise; *Male-odd-month*, taking the value of 1 if a male-odd-month worker was hired intentionally and 0 otherwise. The same classification was also used by Coffman et al. (2021) in the online appendix.

We saw that on average female workers were hired, intentionally and by chance, in 47.8% of the decisions when there were low monetary incentives to employ more productive workers and in 45.3% of the decisions when incentives were higher; both figures are significantly below the 50% benchmark (p value = 0.087 and 0.001, respectively).[9] However, this discrimination is not specific to gender. Indeed, even-month workers were hired in just 33% of low incentive decisions and 34% of higher incentive decisions; both figures are significantly below the 50% benchmark (p value = 0.000). Thus, as in Coffman et al. (2021), we found evidence of statistical discrimination because employers in both treatments preferred to hire workers associated with the higher-performing group.

On closer examination, we can see that in the gender treatment employers left the hiring decision to chance in about 70% of choices, while females were hired intentionally only with a probability of 13% with low monetary incentives to employ more productive workers and 9.7% with higher incentives. Males were hired intentionally in about 17% and 19% of decisions, respectively. In the month-of-birth treatment (Panel B) the probability of hiring an even-month worker intentionally was considerably lower (5.2% and 6.7% with low and high incentives, respectively), the probability of leaving the outcome to chance was about 56% and the probability of selecting an odd-month worker intentionally was higher (38.8% and 37.9%, respectively).

As in Coffman et al. (2021), we included employers' posterior beliefs as controls—after observing the distribution of abilities in the easy quiz for the two groups of workers—of the average performance gap in the easy quiz, *Posterior (easy gap)*, which on average favors male workers (2.69 in the gender treatment and 3.8 in the month-of-birth treatment) and a difference between the posterior beliefs of the average performance gap in the hard quiz and in the easy quiz, *Posterior (hard-easy gap)*, which is on average close to zero (− 0.19 and − 0.28 in the gender and birth-month treatment, respectively). Finally, as in Coffman et al. (2021), we found that employers initially believed that male workers outperform female workers by a significantly larger margin than odd-month workers outperform even-month workers0

[9] In the absence of discrimination (both belief-based and taste-based) in the observed decisions where workers have the same easy quiz performance, employers should hire female workers 50% of the time.

(p value < 0.01 in both quizzes). After taking the decisions in Part 1, posterior beliefs in both treatments point to a greater advantage of male-odd-month workers and the reported performance gap is slightly greater for employers in the month-of-birth treatment. By including posterior beliefs in our controls, we estimated gender-driven discrimination on top of stated beliefs.

3 Results

Table 2 presents the estimated results of a regression model including controls and decision fixed effects for each of our four dependent variables (columns 1–4), in the low incentive decision subset (Panel A), in the high incentive decision subset (Panel B) and across all decisions (Panel C).

When we used the same dependent variable in Coffman et al. (2021), we obtained similar results: when women were labeled with their gender instead of their month of birth preference-based discrimination emerged in favor of females who were hired more often when defined as such (Panel A); the effect is smaller when monetary incentives to employ more productive workers are higher (Panel B) but the difference is not significant at conventional levels (Panel C).

However, this variable may mask important information if monetary incentives to employ more productive workers affect employers' decision to hire females intentionally differently than when the decision is left to chance. Thus, in columns 2–4 we disaggregated our dependent variable. In column 2, where we looked at the probability of hiring females intentionally when these were labeled female rather than even-month workers, preference-based discrimination in favor of females significantly reduced and was no longer present when employer's incentives were higher. Thus, with higher employer incentives, women are no more likely to be hired intentionally in the gender treatment than in the month-of-birth treatment. The observed gender-based discrimination in favor of females may possibly be driven by image concerns, social desirability or other non-pecuniary motives, which may be mitigated with sufficient financial incentives.

Similar results emerge when we look at the probability of hiring a male worker intentionally (column 4) because the reduction in this probability for males *versus* odd-month workers was lower when monetary incentives to employ more productive workers were higher; however, the effect was smaller and less precisely estimated.[10] On the other hand, the decision to leave it to chance was not significantly different between the gender and the month-of-birth treatments with low incentives and was bigger and statistically significant when employer's incentives were higher (column

[10] The discrimination against men seen in the low incentives treatment flees to both *Chance* and *Male-odd-month* in the high incentives treatment (the − 0.048 coefficient gets distributed across the other two outcomes almost equally). This might lay behind the absence of statistically significant change in the "actively select a man" (*Male-odd-month*) outcome when monetary incentives to employ more productive workers are higher.

Table 2 Incentives and gender discrimination

	Female-even-month and chance	Female-even-month	Chance	Male-odd-month
	(1)	(2)	(3)	(4)
Panel A: low incentive				
Gender treatment	0.124***	0.073**	0.102	− 0.175***
	(0.032)	(0.033)	(0.062)	(0.054)
No. of obs	630	630	630	630
Adjusted R-squared	0.088	0.019	0.041	0.088
Panel B: high incentive				
Gender treatment	0.078**	0.017	0.121*	− 0.138**
	(0.034)	(0.031)	(0.062)	(0.056)
No. of obs	630	630	630	630
Adjusted R-squared	0.092	0.008	0.054	0.100
Panel C: all				
Gender treatment	0.120***	0.069**	0.100	− 0.170***
	(0.031)	(0.032)	(0.062)	(0.054)
High incentive	0.012	0.015	− 0.006	− 0.009
	(0.021)	(0.021)	(0.034)	(0.032)
Gender treatment × high incentive	− 0.037	− 0.048*	0.023	0.026
	(0.029)	(0.029)	(0.041)	(0.042)
No. of obs.	1260	1260	1260	1260
Adjusted R-squared	0.090	0.013	0.049	0.095
Controls	YES	YES	YES	YES
Decision FE	YES	YES	YES	YES

All estimates standard errors (reported in parentheses) are corrected for heteroscedasticity and clustered at the respondent level

*,**,***Significance at 10%, 5%, and 1% levels, respectively. The dependent variables are *Female-even-month and chance* taking the value of 1 if a female-even-month worker is hired intentionally, 0.5 if chance is chosen and 0 if male-odd-month is chosen, *Female-even-month* taking the value of 1 if a female-even-month worker is hired for intentionally and 0 otherwise, *Chance* taking the value of 1 if chance is chosen and 0 otherwise, *Male-odd-month* taking the value of 1 if a male-odd-month worker is hired for intentionally and 0 otherwise

3). However, the effect of higher incentives to employ more productive workers was not significantly different from zero (column 3, Panel C).

Thus, increasing employers' monetary incentives reduces their likelihood of explicitly choosing females when labeled as such, but this does not fully translate into a higher probability of choosing a male intentionally because it also increases the likelihood of leaving the decision to chance when workers are labeled with their gender rather than with month of birth.

Other preferences may play a part in employment decisions, for example preferences involving hiring an employee within the employer's same social/identity group. Since our treatment manipulation changes only the label associated with workers (such that all male workers are born in odd months and all female workers are born in even months and, depending on the treatment, the two categories are either labeled with their own gender or with the even/odd month of birth grouping) and not the information on the group-level performance of workers, we can compare in-group preferences between gender and birth-month treatment. In Table 3 we studied whether higher monetary incentives to employ more productive workers also help mitigate the effect of in-group preferences on employment decisions. The in-group preferences label was applied to decisions to hire employees of the same gender in the gender treatment (women hiring women or men hiring men) and the same month of birth in the month-of-birth treatment (people born in even months hiring even-month-born people or people born in odd months hiring odd-month-born people).

We found that when monetary incentives to employ more productive workers were low (column 1), in-group preferences played a statistically significant role in the gender treatment: the coefficient of the interaction between *Gender treatment* and *In-group* was positive and statistically significant, suggesting that female employers hire female employees when labeled with their own gender with a significantly higher likelihood than male employers. Specifically, female-even-month workers were 7% points (p value = 0.109) more likely to be hired by a male rather than an odd-month worker and about 18% points (0.070 + 0.106 = 0.176, p value = 0.044) more likely to be hired by a female rather than an even-month worker with a statistically significant difference of about 11% points between in- and out-group for the gender treatment.

However, in-group preferences no longer play a part when employer's incentives are high (column 2) and the reduction in the effect of in-group preferences for the gender treatment is statistically significant (column 3). Similar results hold when *Female-even-month* was taken as dependent variable while in-group preferences never played a statistically significant role on *Chance* and *Male-odd-month*.

Therefore, similarly to Coffman et al. (2021), we found evidence of in-group preferences in the gender treatment. However, in contrast with them, we did not find evidence of in-group preferences in the birth-month treatment[11]: employers born in even months did not hire even-month employees more than odd-month employers did.[12]

[11] The lack of replication in the birth-month treatment does not constitute significant cause for concern. Indeed, this was not the main result in Coffman et al. (2021) and, as highlighted by Camerer et al. (2016), interaction effects are less likely to be replicable than main or simple effects (see also the Reproducibility Project: Psychology (RPP, 2015)).

[12] This may be explained by the fact that in our sample more people actually identify with their gender (and hence perceive an "in group" and an "out group" along that dimension) than with the evenness or oddness of their month of birth.

https://doi.org/10.1007/s40881-023-00144-4 Published online by Cambridge University Press

Table 3 Incentives, gender discrimination and in-group preferences

	Female-even-month and chance			Female-even-month	Chance	Male-odd-month
	(1)	(2)	(3)	(4)	(5)	(6)
	Low	High	All			
Gender treatment	0.070	0.085*	0.067	0.031	0.071	− 0.103
	(0.044)	(0.045)	(0.043)	(0.038)	(0.086)	(0.076)
In-group	− 0.006	0.054	− 0.004	0.023	− 0.054	0.031
	(0.044)	(0.050)	(0.044)	(0.030)	(0.087)	(0.082)
Gender treatment × in-group	0.106*	− 0.021	0.101*	0.070	0.063	− 0.133
	(0.061)	(0.067)	(0.062)	(0.063)	(0.123)	(0.106)
High incentive			− 0.014	− 0.011	− 0.006	0.017
			(0.023)	(0.014)	(0.042)	(0.042)
Gender treatment × high incentive			0.021	0.011	0.020	− 0.031
			(0.036)	(0.030)	(0.050)	(0.055)
In-group × high incentive			0.057	0.058	− 0.001	− 0.057
			(0.043)	(0.045)	(0.070)	(0.065)
Gender treatment × in-group × high incentive			− 0.118**	− 0.121**	0.006	0.115
			(0.059)	(0.059)	(0.084)	(0.083)
Controls	YES	YES	YES	YES	YES	YES
Decision FE	YES	YES	YES	YES	YES	YES
No. of obs.	630	630	1260	1260	1260	1260
Adjusted R-squared	0.100	0.095	0.097	0.023	0.048	0.097

All estimates standard errors (reported in parentheses) are corrected for heteroscedasticity and clustered at the respondent level. *, **, and *** indicate significance at 10%, 5%, and 1% levels, respectively. The dependent variables are *Female-even-month and chance* in columns 1–3; *Female-even-month* in column 4; *Chance* in column 5 and *Male-odd-month* in column 6. Column 1 considers only decisions made with low incentives; column 2 considers only decisions made with high incentives; columns 3–6 consider all decisions

https://doi.org/10.1007/s40881-023-00144-4 Published online by Cambridge University Press

4 Conclusion

We replicated Coffman et al. (2021)'s analysis and found statistical discrimination against women when they belonged to a low-performance group, but discrimination in their favor when the female identity of employees was revealed to potential employers. We found evidence of discrimination for both male and female employers that is larger for female employers and disappears when employers' financial compensation for making employment decisions was higher.

With the data at hand, we were not able to identify the mechanism driving our discrimination result in favor of women displaced by high monetary employer incentives. One possible explanation is that employers make gender discriminatory choices in favor of women because they have image concerns or social desirability biases. This possible mechanism is supported by Coffman et al. (2021) who found that this discriminatory behavior disappears when employers' intentions are concealed through the introduction of something that might excuse the hiring of a member of the non-preferred group, such as an apparent risk associated with hiring a female worker. However, other mechanisms may also be at work. Indeed, women participants would arguably feel less social pressure to dispel concerns about a possible bias against women and our data show that women are more likely to prefer women applicants. This would suggest that attaching a gender label to a group may also trigger social-identity motives that translate into in-group preferences for women in particular. When monetary incentives to employ more productive workers are higher for employers, the utility deriving from better self-image, socially adequate behavior or shared social identity is traded off for an alternative source of utility represented by higher potential earnings, thus alleviating discrimination.

Future research could test whether the effect of incentives is monotonic or if there exists a maximizing point beyond which the increased cost of higher compensation (for example for hiring committees within the firm) is greater than the marginal benefit in terms of reduction of discrimination and differential quality of the hired candidate.

Funding Open access funding provided by Università degli Studi di Napoli Federico II within the CRUI-CARE Agreement.

Data availability The data that support the findings of this study are available on request from the corresponding author.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Alesina, A., Carlana, M., La Ferrara, E., & Pinotti, P. (2018). Revealing stereotypes: Evidence from immigrants in schools. Technical report, National Bureau of Economic Research.

Anukriti, S. (2018). Financial incentives and the fertility-sex ratio trade-off. *American Economic Journal: Applied Economics, 10*(2), 27–57.

Arrow, K. J. (2015). *The theory of discrimination*. Princeton University Press.

Becker, G. S. (1957). *The economics of discrimination*. University of Chicago press.

Bohren, J. A., Imas, A., & Rosenberg, M. (2019). The dynamics of discrimination: Theory and evidence. *American Economic Review, 109*(10), 3395–3436.

Bordalo, P., Coffman, K., Gennaioli, N., & Shleifer, A. (2016). Stereotypes. *The Quarterly Journal of Economics, 131*(4), 1753–1794.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., et al. (2016). Evaluating replicability of laboratory experiments in economics. *Science, 351*(6280), 1433–1436.

Chen, R., & Chen, Y. (2011). The potential of social identity for equilibrium selection. *American Economic Review, 101*(6), 2562–89.

Chen, Y., & Li, S. X. (2009). Group identity and social preferences. *American Economic Review, 99*(1), 431–57.

Chen, Y., Li, S. X., Liu, T. X., & Shih, M. (2014). Which hat to wear? Impact of natural identities on coordination and cooperation. *Games and Economic Behavior, 84*, 58–86.

Coffman, K. B., Exley, C. L., & Niederle, M. (2021). The role of beliefs in driving gender discrimination. *Management Science, 67*(6), 3551–3569.

Gioia, F., & Immordino, G. (2023). Gender-science implicit association and employment decisions. CSEF Working Paper no. 681.

Guryan, J., & Charles, K. K. (2013). Taste-based or statistical discrimination: The economics of discrimination returns to its roots. *The Economic Journal, 123*(572), F417–F432.

Hortaçsu, A., Hwang, S. I. M., & Mathur, D. (2019). Monetary incentives on inter-caste marriages in India: Theory and evidence. *Journal of Development Economics, 141*, 102371.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *The American Economic Review, 62*(4), 659–661.

RPP. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology, 1*(2), 149–178.

Underhill, K. (2019). Price and prejudice: An empirical test of financial incentives, altruism, and racial bias. *The Journal of Legal Studies, 48*(2), 245–274.